SYNTHESE LIBRARY 347

Mauricio Suárez
Editor

Probabilities, Causes and Propensities in Physics

Springer

Probabilities, Causes and Propensities in Physics

SYNTHESE LIBRARY

STUDIES IN EPISTEMOLOGY,
LOGIC, METHODOLOGY, AND PHILOSOPHY OF SCIENCE

VOLUME 347

For further volumes:
http://www.springer.com/series/6607

Probabilities, Causes and Propensities in Physics

Edited by

Mauricio Suárez
Complutense University of Madrid, Madrid, Spain

Springer

Editor
Prof. Mauricio Suárez
Universidad Complutense de Madrid, Faculty Filosofía
Departomento Lógica y Filosofía de la Ciencia
Planta Sótano, Edificio B
28040 Madrid
Spain
msuarez@filos.ucm.es

ISBN 978-1-4020-9903-8 e-ISBN 978-1-4020-9904-5
DOI 10.1007/978-1-4020-9904-5
Springer Dordrecht Heidelberg London New York

Library of Congress Control Number: 2010936456

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

Some of the papers collected in this volume were first presented in draft outline at a workshop that I organised at Complutense University in Madrid in October 2006. This was the second workshop organised within the Classical and Causal Concepts in Science network of philosophers of physics in Spain. I want to thank the leaders of the other two legs of the network, Carl Hoefer and Henrik Zinkernagel, for our collaboration and sustaining partnership over the years. Thanks also to the members of the Complutense research group MECISR for logistical and editorial help. Financial support is acknowledged from the Spanish Ministries of Education and Science (research projects HUM2005-07187-C03-01 and HUM2006-27975-E/FISO), and Science and Innovation (research projects FFI2008-06418-C03-01 and PR2008-0079). I also would like to thank the Editors of Synthese Library – Vincent Hendricks and John Symons -, two anonymous referees, and Margherita Benzi, Anjan Chakravartty, Roman Frigg, Mathias Frisch, Meir Hemmo, Carl Hoefer, Colin Howson, Federico Laudisa, Huw Price, Iñaki San Pedro, Ian Thompson for refereeing and consulting work as well as their encouragement. Ingrid van Laarhoven was the friendly, efficient, and patient first port of call at Springer. I finished working on the manuscript while I was visiting Harvard University during 2009 and I want to thank the Department of Philosophy, particularly Hilary Putnam, for sponsorship.

Cambridge, Massachusetts — Mauricio Suárez

Contents

Contributors

Guido Bacciagaluppi Department of Philosophy, University of Aberdeen, Aberdeen, AB24 3UB, UK, g.bacciagaluppi@abdn.ac.uk

Sorin Bangu Department of Philosophy, University of Illinois at Urbana-Champaign, IL 61801, USA, sib10@uiuc.edu

Joseph Berkovitz Centre for Time, Department of Philosophy, University of Sydney, Sydney, NSW, Australia; IHPST, Victoria College, University of Toronto, 91 Charles St. West, Toronto, ON, Canada M5S 1K7, joseph.berkovitz@utoronto.ca

Mauro Dorato Department of Philosophy, University of Rome 3, Viale Ostiense 234, 00144 Rome, Italy, dorato@uniroma3.it

Roman Frigg Department of Philosophy, Logic and Scientific Method, London School of Economics, London WC2A 2AE, UK, r.p.frigg@lse.ac.uk

Balazs Gyenis Department of History and Philosophy of Science, University of Pittsburgh, Pittsburgh, PA, USA, gyepi@pitt.edu

Federico Laudisa Department of Human Sciences, University of Milan-Bicocca, Piazza dell'Ateneo Nuovo 1, 20126 Milan, Italy, federico.laudisa@unimib.it

Nicholas Maxwell Department of Science and Technology Studies, University College London, London WC1E 6BT, UK, nicholas.maxwell@ucl.ac.uk

Miklós Rédei Department of Philosophy, Logic and Scientific Method, London School of Economics, London WC2A 2AE, UK, m.redei@lse.ac.uk

Iñaki San Pedro Department of Logic and Philosophy of Science, Complutense University of Madrid, Madrid 28040, Spain, inaki.sanpedro@filos.ucm.es

Mauricio Suárez Department of Logic and Philosophy of Science, Complutense University of Madrid, 28040 Madrid, Spain, msuarez@filos.ucm.es

Ian J. Thompson Current address: Lawrence Livermore Laboratory, L-414, Livermore, CA 94551, USA, IJT@ianthompson.org

Chapter 1
Four Theses on Probabilities, Causes, Propensities

Mauricio Suárez

1.1 Overview of the Book

The present volume collects eleven essays by philosophers of science and physics on three inter-related themes: probability, causality and propensities. The discussion centres on modern physics and, in particular, on the pre-eminently probabilistic branches of physics in our time, quantum and statistical mechanics. In spite of the technical nature of most of the papers, this is a collective effort in the philosophical foundations of physics, and of science more generally. In other words, it is essentially a book on the foundations of science rather than its application, and its main aims are conceptual, philosophical and methodological. In this introduction I provide a summary and a philosophical defence of some of the claims made in the book. The introduction is not meant to back all of the specific claims made by the different authors (nor can it be understood as endorsement, since some of the authors disagree with, or at least qualify, some of the claims I have made in my own work). Instead it is meant to underscore the importance of the topics on which the authors focus their analytical gaze, and their detailed development of these ideas.

The book is divided into three sections each devoted to one of the main themes. Thus the first part contains three essays devoted to probability in science; the second part contains four on the nature of causality particularly in quantum mechanics; and the final part contains some essays on propensities again mainly in quantum mechanics. In spite of the diversity of aims and interests, there are some common themes running throughout the book. In particular there is agreement in general on the following four joint themes or theses (N.B. not all authors would agree with all four): (i) An emphasis on taking probabilities in physics to be objective features of the world as opposed to degrees of belief; (ii) A correlated emphasis on the importance of transition probabilities – i.e. probabilities for objective changes of physical

M. Suárez (✉)
Department of Logic and Philosophy of Science, Complutense University of Madrid,
28040 Madrid, Spain
e-mail: msuarez@filos.ucm.es

M. Suárez (ed.), *Probabilities, Causes and Propensities in Physics*, Synthese Library
347, DOI 10.1007/978-1-4020-9904-5_1,

state – over merely conditional probabilities; (iii) An additional reluctance to interpret all objective probabilities in any one of the traditional ways (actual or virtual frequencies, single case or long-term propensities); and finally (iv) A general tendency to identify various causal commitments and presuppositions in foundational physics – including in several cases the causal relation between underlying dispositional properties, or propensities, and their empirical manifestations in terms of probability distributions.

The first three sections of this introduction review the contents of each of the parts of the book, always with an eye on these four interrelated philosophical themes. Then in Sections 1.5, 1.6, and 1.7 I develop my own philosophical understanding of these four theses, relating them to previous discussions in the literature, particularly the literature on probabilistic causation, causal inference, and dispositional properties. Section 1.8 draws some conclusions and provides some pointers for future work.

1.2 Probabilities

The first part of the book contains papers by Guido Bacciagaluppi on transition probabilities; Sorin Bangu on the principle of indifference; and Roman Frigg on the typicality approach to equilibrium. All these papers concern the nature of probability as it appears in science, mainly in physics. I next provide a brief summary of their main results, with an eye on the particular themes that run through the book.

1.2.1 Transition Probabilities and Time-Symmetry

In Chapter 2: 'Probability and Time Symmetry in Classical Markov Processes' Guido Bacciagaluppi argues that time-symmetric transition probabilities can also be employed to represent typical examples of time-directed phenomena. Therefore transition probabilities, even if representing the chances of possible changes of physical states, can neither entail nor ground an objective distinction between past and future. To a first approximation, this implies that defenders of tensed theories of time and other philosophers inclined to deny the reality of becoming need not fear the concept of transition probability: it is not an essentially time-directed concept although it may of course be used to represent processes that are fundamentally directed in time. (Later on in Section 1.7 of this introductory essay it is argued that Bacciagaluppi's thesis may have interesting implications regarding the nature of the propensities that might underlie transition probabilities).

Bacciagaluppi follows the usual definition of transition probabilities in terms of Markov stochastic processes. Roughly a process is Markov if the probability of any state at any given time is dependent only on the immediately preceding state; all previous states are statistically irrelevant. For a stochastic process this entails roughly:

$$P_{j+1/j}(S(t_{j+1})/S(t_j)\&S(t_{j-1})\&\ldots\&S(t_1)) = P_{j+1/j}(S(t_{j+1})/S(t_j)) \quad \text{(MP)}$$

where $S(t_j)$ is the state of the system at time t_j, and so on.

This equation is a simplified version of Bacciagaluppi's (equation 2.3), where I have made explicit the dynamical properties of states, identifying them by means of time index variables. I have then kept states in the variable range of the probability function – as opposed to placing them in the subscript.[1] And I have represented a static probability, when in a stochastic process each probability more generally carries a time index too – which determines the values of the probability at that stage of the process. Equation (MP) hence expresses a kind of statistical independence: the state at any given time is statistically independent of any previous state, conditional on the state just prior to it. In the language of contemporary theories of causal inference, the state at time t_j, $S(t_j)$ screens off the later state $S(t_j+1)$ from any previous states $S(t_{j-1}),\ldots,S(t_1)$. In this simplified terminology the concept of transition probability can be expressed concisely:

$$P_{j+1/j}(S(t_{j+1})/S(t_j)) = P_{j\&(j+1)}(S(t_{j+1})\&S(t_j))/P_j(S(t_j)) \quad \text{(FTP)}$$

The equation expresses the *transition probability* that a system will physically undergo a change *from* state $S(t_j)$ at time t_j *to* state $S(t_{j+1})$ at a later time t_{j+1}. We may refer to this as a *forwards transition probability* (FTP) since it expresses the transition probability $P_{j+1/j}$ *from an earlier to a later time* of a change of state $S(t_j)$ into a state $S(t_{j+1})$.[2] (FTP) may be contrasted with the expression for the backwards transition probability (BTP), i.e. the probability of the same change of state *but from the later to the earlier time*[3]:

$$P_{j/j+1}(S(t_{j+1})/S(t_j)) = P_{(j+1)\&j}(S(t_{j+1})\&S(t_j))/P_{j+1}(S(t_j)) \quad \text{(BTP)}$$

[1] Bacciagaluppi's terminology employs the technical notion of an n-fold joint distribution, which is standard in the literature on stochastic processes (see e.g. Doob, 1953). According to this terminology, states 1 to n appear in the subscript of the probability function, and time indexes in its variable range. We then consider the n-fold joint probability distributions that the n states define over the time indexes. This terminology is more convenient for the derivation of technical results but it strikes me as less intuitive, at least for the purposes of this chapter.

[2] These notions are again expressed in my own terminology. The notation of n-fold distributions has, undoubtedly, an advantage at this point since it allows us to distinguish the concept of *symmetry* of the transition probability from the concept of *detailed balance* (see Bacciagaluppi's Section 3, where it is also claimed that under standard conditions these concepts are equivalent as statements of time-symmetry). But the distinction plays no role in this introductory essay which focuses instead on conceptual issues regarding objective probability.

[3] So, importantly, a backwards transition probability is *not* the forwards transition probability of the time-inverse of the state change: $\text{Prob}_{j/j+1}(S(t_{j+1})/S(t_j)) \neq \text{Prob}_{j+1/j}(S(t_j)/S(t_{j+1}))$, with $t_{j+1} > t_j$. The latter is rather a different transition probability altogether, belonging to an entirely different Markov process.

Forwards and backwards transition probabilities need not be equal, and typically they are not. A stochastic process that is fundamentally time-asymmetric would normally establish different forwards and backwards probabilities for the same change of state. For instance a process directed 'forwards' in time would set one or zero backwards transition probabilities, while setting forwards transition probabilities between zero and one *for the very same change of state*. A process directed 'backwards' in time would do conversely. If the forwards and backwards probabilities for all changes of state are equal, then the process is time-symmetric in a robust sense. More specifically, if all processes are time-symmetric then a consideration of the probabilities defined for the world-dynamics (i.e. the probabilities for all the changes throughout history of all the states of all systems in the world) would leave the direction (the 'arrow') of time completely undetermined. There would be no way to pick out a particular direction of time from any transition probabilities. Although such ideal and abstract world dynamics is not helpful in modelling any particular stochastic process, it does show that there is nothing in the concept of transition probability per se that contradicts time-symmetric fundamental laws. In other words, it is possible to define genuine transition probabilities in worlds endowed with fundamentally time-symmetric laws.

In the main section of his paper (Section 2.4), Bacciagaluppi considers and rejects three different arguments that may be raised against this conclusion. These arguments purport to show that transition probabilities do in fact conflict with time-symmetric laws and, therefore, require a direction of time. Roughly they go as follows. First, there is the argument that ergodicity on its own defines an arrow of time because it entails that most systems will tend towards equilibrium. In our case this should mean that the stochastic process will tend to equilibrate in time, i.e. that it will tend to define identical and hence symmetrical probabilities for all state transitions in the limit (or to put it another way its single time n-fold distribution $p_n(t)$ becomes time-invariant in the limit). This seems to require asymmetry at some point in the process before equilibrium is reached. Second, there is the idea that, at least for some common processes, backwards transition probabilities fail to be time translation invariant. Consider decay processes where the probability of decay from an excited to a ground state in unit time is finite. Finally, there is the thought that backwards transition probabilities are not invariant across experiments with varying initial distributions, i.e. experiments where the initial time series data differs.

In all these cases transition probabilities seem to conflict with time symmetric laws because a fundamental distinction seems to emerge between forwards and backwards transition probabilities. Yet since we have just argued that the concept of transition probability itself cannot be used to introduce any fundamental time-asymmetry, it follows that these arguments must employ additional assumptions. It is to be expected that these assumptions are responsible for the conflict with time-symmetry and Bacciagaluppi argues convincingly that they reduce to the same mistaken presupposition in all three cases, namely: that the calculation of transition probabilities is to be worked out on samples that are *not* in equilibrium. In such cases the inference from the frequencies in the sample to the transition probabilities will yield an apparent time-asymmetry. However, once the samples have been 'cleansed' in order to generate 'unbiased' ones, the apparent time-asymmetry disappears. There is an interesting philosophical insight buried in this argument, which I shall take up briefly later in Section 1.5 of this essay.

1.2.2 The Principle of Indifference

In the third chapter, Sorin Bangu reconsiders the role of the *principle of indifference* in the ascription of probabilities with a particular emphasis on its use in physics. Keynes first stated it as follows[4]: 'The *principle of indifference* asserts that if there is no known reason for predicating of our subject one rather than another of several alternatives, then relatively to such knowledge the assertions of each of these alternatives have an equal probability'. There are a number of well known arguments against the principle, many of them taking the form of counterexamples, or paradoxes. Typically these counterexamples show that the application of the principle leads to several inconsistent probability ascriptions to the same event. The so-called Bertrand paradoxes stand out: on the basis of geometrical considerations, and under several assumptions of continuity and smoothness of the probability density, they show that the principle of indifference leads to inconsistent probability ascriptions. A very simple version due to Van Fraassen is often discussed.[5] Consider a factory that produces cubes of length l up to 2 cm. What is the probability that the next cube produced has an edge $\leq$1 cm? A straightforward application of the principle of indifference yields probability = ½. But, we could have formulated the question in several different ways. For instance, what is the probability that the next cube has sides with an area $\leq$1 cm^2? The principle now yields the answer ¼. And how about the probability that the next cube has volume $\leq$1 cm^3? The answer provided by the principle is now 1/8. These are all inconsistent with each other since they ascribe different probabilities to the occurrence of the very same event.

More generally the principle of indifference employs a problematic inference from our epistemic situation of relative ignorance regarding the outcome space of a stochastic process to a definite probability ascription over the various outcomes. The inference is problematic in just the way any inference from ignorance to truth is problematic.[6] But in addition there is a sense, which I discuss in Section 1.7.2, in which the principle may invite an illegitimate inference from a merely epistemic fact about our knowledge (or lack thereof) to an objective fact about the physical world – specifically its dispositional properties.

Bangu agrees that there is at least a priori no reason to support the principle, and he does not attempt to provide new arguments to support it. His aim is rather to contest two other arguments against the principle, a classic argument by Hans

---

[4] In the *Treatise on Probability* (Keynes, 1921) which traces it back to Bernouille's application of the principle of sufficient reason. For discussion see also Gillies (2000a, Chapter 3).

[5] Van Fraassen (1989, 303–304).

[6] See Strevens (1998, 231) for further discussion.

Reichenbach (1971/1949), and a more recent one by Donald Gillies (2000a). These arguments attempt to show that the principle is not an a priori truth, and is moreover redundant even as a contingent truth about the correct adscription of probability values in specific situations. In other words the principle is not even a necessary tool or condition for the practice of probabilistic inference. Or, to invoke Reichenbach's own terminology,[7] the principle of indifference can be neither *validated* a priori nor *vindicated* a posteriori.

Reichenbach's argument appears to aim for a stronger conclusion than Gillies'. Reichenbach proposes a proof that the principle of indifference grounds no sound inferences at all to the probabilities of physical events *that can not be established by other empirical means*. In other words, the principle does no outstanding work at all in practical inference. By contrast, Gillies accepts that the principle does some heuristic work – in suggesting new hypotheses or physical theories entailing probability values for various outcomes. However, although it may be heuristically useful in generating new physical theories or hypotheses, it has no standing as a logical principle. Employing Reichenbachian terminology once again, we may say that, according to Gillies, the principle has an inferential function in the context of discovery, while lacking it in the context of justification. By contrast, Reichenbach appears to claim that the principle has no inferential function in any context whatever.

Nevertheless both arguments share the aim to show that the principle is redundant in the ascription and justification of probabilities: any work the principle could appear to do in providing probability values for outcomes, in any context, is work that can be done by other methods. More generally both Reichenbach and Gillies aim to provide alternative means for the justification of probabilistic hypotheses and stochastic laws, which would eliminate any need for the principle, or render it otiose for this purpose. We may thus refer to their arguments as 'eliminativist'.

Bangu finds both 'eliminativist' arguments defective. He first shows that Reichenbach's argument is either circular or unsound: either the principle of indifference is itself assumed in the proof or it remains thoroughly unjustified. Reichenbach's argument is a development of yet another argument found in Poincaré, and goes roughly as follows.[8] Consider a roulette wheel, evenly divided into red and black intervals, corresponding to red and black numbers. In the absence of any further information, an application of the principle of indifference entails that the probability of obtaining a red or a black outcome should be the same and equal to $\frac{1}{2}$. The question is whether there is a distinct procedure that would enable us to derive the same result but without invoking the principle at all. Poincaré and Reichenbach reason as follows. Consider that the outcome of the game is determined by where the wheel stops, and may be represented by a variable θ ranging between 0 and 2π. Let then $\mathrm{d}(\theta)$ be the probability distribution over θ. The probability of

[7] As applied to the rather different problem of induction – see Reichenbach (1951, Chapter 14) and Salmon (1991) for a critical discussion.

[8] Reichenbach (1949); Poincaré (1902). For a summary and review see Strevens (1998, 236–238).

obtaining a red number is given by the sum over the probabilities that θ falls in a particular red square. Now assuming that the intervals alternate rapidly in θ, *and* that the function d(θ) is smooth over the intervals (even though not necessarily constant), then the probability of red and black is equal. This reasoning appears to provide us with a procedure that enables us to derive the correct ½ probability values for red and black from the physical symmetry of the roulette wheel without apparently invoking the principle of indifference. However as Bangu points out, the argument depends upon the function d(θ) is smooth. And the only reason for this is that the symmetry of the wheel requires that d(θ) is uniform, i.e. that it is the same for every discrete value of θ. But this is just another statement of the principle of indifference: we ascribe equal probability to all possible outcomes because there is no reason to anticipate one rather than another result. Unfortunately what this means is that the smoothness of d (θ) depends upon the principle of indifference itself, so the procedure described by Reichenbach and Poincaré does not actually do away with the principle in practice. Hence a vindication remains a possibility, and the principle of indifference has not been eliminated.

Bangu then discusses Gillies' argument and he claims that it does not hold water either. He points out that the kinds of methods that Gillies invokes as replacement for the principle of indifference for the justification of probabilistic hypotheses are subject to precisely the same kind of objections that show the principle itself to be untenable. Gillies claims, following Jaynes, that the principle of indifference provides us with a heuristics for seeking new statistical theories and hypotheses.[9] But he also claims that the principle is dispensible as a method for justifying statistical hypotheses, which may always be justified by means of a more appropriate methodology. In particular Gillies defends a 'methodological falsificationist' approach to the testing of statistical hypotheses, partly inspired by Popper and partly by the classical statisticians Fisher, Neymann and Pearson. In this account, a *falsifying rule for probability statements* (FRPS) is formulated, which enables us to construe probabilistic statements as falsifiable 'in practice', even though from a strictly deductive point of view, such statements are in principle unfalsifiable.[10] A statistical hypothesis H is then methodologically falsified by a sample of data points $\{e_1, e_2, e_3, \ldots, e_n\}$ if there is a test statistic X whose value lies below the statistical significance level, which is typically fixed at 5%.[11]

Howson and Urbach have argued that the falsifying rule requires a decision regarding the outcome space of the test statistic X. And whether or not the data points may be said to falsify the hypothesis H may well depend on this decision. In particular they claim that a decision is required to determine the 'stopping rule'

---

[9] Gillies (2000a, 47–49), where several examples from physics are provided, such as the viscosity of gases and Bose Einstein statistics.

[10] See Gillies (2000a, 147).

[11] A test statistic for an experiment is a random variable X, whose value can be calculated as a function of the data sampled, $X\{e_1, e_2, e_3, \ldots, e_n\}$, and that can be taken to represent the outcome of the experiment. Note that the same experiment may yield different values for the test statistic, depending on the data sampled.

describing the conditions under which the experiment is terminated or finalised. For instance in assessing the hypothesis that a particular coin is fair, we must repeat the experiment a number of times and different rules may be applied to the termination point. As a result the outcome space (the space of all possible sequences of outcomes) is affected.[12] Bangu goes further in claiming that the decision regarding the outcome space is akin to the decision that the principle of indifference promotes in order to ascribe equal probability to outcomes evidentially on a par. In both cases the decision involves fixing the outcome space. According to Bangu this compromises Gillies' argument for the dispensability of the principle of indifference. The type of methodology that we would be attempting to replace the principle with is thoroughly infused with just the sort of difficulty that led us to abandon the principle in the first place.[13] Thus, Bangu concludes that there is not yet a good argument against the vindication of the principle of indifference in practice.

1.2.3 Typicality in Statistical Mechanics

In the fourth and last chapter in the probability section of the book, 'Why Typicality does not Explain the Approach to Equilibrium?', Roman Frigg critically evaluates attempts in the philosophy of statistical mechanics to provide typicality-based explanations of thermodynamic irreversibility. Consider a classical system consisting of n particles, each endowed with three degrees of freedom, and governed by Hamiltonian dynamics. Its state may be represented in a constrained $6n$–1 dimensional energy hypersurface Γ_E of the corresponding $6n$-dimensional phase space Γ. Each macroscopic state (defined by sets of macroscopic properties) M_i will define disjoint and exhaustive subregions Γ_{Mi} of Γ_E. The second law of thermodynamics is then supposed to entail that the evolution of the entropy of the macrostate of any (freely evolving) system mirrors the increase of thermodynamic entropy over time, reaching a maximum value at equilibrium. Suppose the initial state of the system is $x(t_0)$, and the final state is $x(t)$. Then let Γ_{Past}, Γ_{Equi} be the past and the equilibrium macrostates of the system, so $x(t_0) \in \Gamma_{Past}$, and $x(t) \in \Gamma_{Equi}$. It seems to follow from the second law that any system whose initial macrostate is Γ_{Past} will eventually wind up in Γ_{Equi}.

[12] Howson and Urbach (1993, 210–212). In their example we may choose either to terminate the experiment as soon as 6 heads occur, or rather after 20 trials regardless of the outcome. The size of the outcome space is then predetermined in the latter case ($= 2^{20}$) but not so in the former. Even if the outcome spaces happened to have the same size in both cases (because say the 6th head happens to occur on the 20th trial), it would still be the case that the stopping rule could affect the result of the application of the falsifying rule, falsifying it in the former but not the latter case.

[13] Note that Gillies disagrees that a falsificationist methodology is in any way threatened by Howson and Urbach's argument. See particularly the discussion in his interesting review of their book (Gillies, 1990, 90–97). Howson and Urbach respond in the 2nd edition of their book (214–215). This debate turns on whether or not the stopping rule is relevant to the performance of the experiment, and therefore relevant to the evaluation of the application of the falsifying rule. It is surprising that this debate does not yet appear to have been linked to the question of the nature of the probabilities involved, and in particular whether they are subjective or objective probabilities.

Why is this so? And more particularly: is there an explanation for this fact in statistical mechanics? [14] We may refer to any approach that aims to provide an explanation by invoking the notion of 'typical state', as a 'typicality explanation' (of the approach to equilibrium). This type of approach relies on the thought that the equilibrium macrostate Γ_{Equi} is the largest among all the regions Γ_{Mi} under some standard natural measure, such as the Lebesgue measure μ.[15] Frigg discusses three different typicality approaches and his sober conclusion is that none are actually viable. As is often the case in a philosophical dispute much hinges on the initial formulation of the problem. Frigg first outlines a standard formulation which he helpfully refers to as 'gloss', and which he goes on to dispute (in Section 1.4 of his paper). This formulation is however sometimes adopted by other authors as a fact, namely 'the fact that equilibrium microstates are typical with respect to Γ_{E} and the Lebesgue measure μ' (p. 5). Indeed the three approaches discussed by Frigg in some way link this 'gloss' to the dominance of the equilibrium macrostate.

The first approach appeals to the brute fact of typicality itself. In other words it aims to explain the approach to equilibrium as a result of the typicality of equilibrium states. Frigg rightly points out that there is no reason to suppose that atypical states need evolve into typical states just because the former are atypical and the latter are not. And this is true even if the atypical states make up a measure zero set. The evolution of the states depends rather on the specific dynamical laws that operate, and cannot be settled just by looking at the measures (relative sizes in the case of the Lebesgue measure) of different regions of phase space.

The second approach consequently focuses on dynamics. Boltzmann's original ergodic theorem is an attempt at a dynamic explanation. (Roughly the ergodic theorem states that the dynamics of the state is such that any trajectory sooner or later visits every point in Γ_{E}. In other words regardless of the initial microstate a system will eventually take every other microstate compatible with the macroscopic constraints.)[16] There are however well known problems with Boltzmann's original ergodic theorem, and improved ergodic explanations of the approach to equilibrium have also been criticised.[17] And in any case, the solution seems to be rejected by those who advocate the typicality explanation anyway. Another reading of the second (dynamical) approach regards chaotic dynamics as the key to the explanation of the approach to equilibrium. Frigg in turn distinguishes two versions of a chaotic explanation. The first is based upon the sensitive dependence on initial conditions characteristic of chaotic behaviour, and only requires chaos locally in a particular subset of the phase space. Sensitivity to initial conditions has been argued to ground

---

[14] Should there be one? The presumption that there should is of course tantamount to the view that thermodynamics should be in some sense reduced to statistical mechanics. It is controversial whether such attempts have been successful. Moreover it is unclear that they should be in order to ground thermodynamic irreversibility. See for instance Sklar (1993, Chapter 9). Such interesting questions are beyond the purview of this essay or this book.

[15] A measure usually defined over the semi-closed intervals of the real line (see Halmos, 1974, 65ff.)

[16] Sklar (1993, 159–160).

[17] For a thorough critique see Earman and Rédei (1996).

a typicality explanation of equilibrium, in the sense that the trajectories that will exhibit random walk behaviour are ‘typical’. More specifically, the region of the phase space that contains the initial states of trajectories that exhibit this type of random walk behaviour has a Lebesgue measure arbitrarily close to 1. Frigg refers to this condition as the Typicality Past Hypothesis (TPH) but rejects the idea that all those trajectories that satisfy this condition actually carry typical initial conditions into the equilibrium region. He claims that there is an important set of such trajectories belonging to KAM systems that do not do so. So this typicality explanation also seems to fail for reasons not dissimilar to the ergodic explanation. The second version of the dynamical explanation is more promising according to Frigg. This focuses on the notion of global chaos, where the entire phase space exhibits chaotic features and not just isolated subsets of the phase space. Frigg discusses several attempts to make the notion of global chaos more precise and ground the explanation of the approach to equilibrium. The most promising are still prey to some of the objections that were raised against ergodic approach.

Frigg discusses yet a third approach, due to Lebowitz and Goldstein, which focuses on the internal structure of the micro regions Γ_{Mi} rather than the entire phase space. The important feature, according to Frigg, is the property of each state in Γ_{Mi} of being ‘entropy-increasing’. This is a relational property of states and dynamical trajectories: a state is entropy increasing if it lies on a trajectory that takes lower entropy states into higher entropy states. A system is then defined as ‘globally entropy increasing’ roughly if every subset of its phase space is densely populated by such entropy increasing states. One would then hope that global entropy increasing systems are all necessarily equilibrium approaching. However this is unfortunately not the case, and any attempt to work out a fit between these two notions still requires us to make assumptions regarding the typicality of entropy increasing states within the phase space regions in accordance to the standard Lebesgue measure.

Frigg’s conclusion is that any proper explanation of the approach to equilibrium will necessarily involve dynamics; merely grounding it upon the typicality of the corresponding states within the phase space won’t ever be sufficient. It does not matter whether entropy increasing states are typical in this sense – what matters is rather the details of the dynamical laws that evolve low entropy into higher entropy states. Without a reference to the dynamical transformation of the states, such explanations appear empty or vacuous. (see Section 1.7 for a discussion of the dynamics of propensity states).

1.3 Causes

The second part contains essays by Federico Laudisa on the nature of causation in modern physics, Joseph Berkovitz on the more specific issue of backwards in time causality in quantum mechanics, Balazs Gyenis and Miklós Rédei on the causal completeness of probabilistic models, and a joint paper of mine with Iñaki San Pedro on causal inference in the context of EPR experiments.

1.3.1 From Metaphysics to Physics

In Chapter 5, Federico Laudisa takes up the issue of causation in quantum mechanics, particularly in connection with the EPR correlations. Laudisa first rejects the idea that causality is anathema to quantum mechanics in general. He then endorses a form of causal pluralism that leads him to the view that many questions regarding causality in quantum mechanics may receive different answers in different frameworks, or depending on interpretation. (In fact he later makes it known that he subscribes to a stronger claim vis a vis the EPR experiment, namely: that such issues have no determinate answers independently of the details of the models of the correlations provided within each interpretation). The rest of the paper is a review of the main difficulties that emerge in the attempt to provide causal accounts, mainly with reference to the EPR correlations within some of the different models and interpretations of quantum mechanics. In particular Laudisa focuses on the GRW and Bohm's theories.

One feature of Laudisa's analysis is his assumption that performing a measurement and obtaining an outcome is essentially the same event. The causal connections that he has in mind are between measurement-and-outcome events. (It is arguable that this rules out a propensity interpretation of the quantum state, something that I shall discuss in due course). Laudisa thinks that the superluminal nature of any putative connection in this case yields a 'weak' form of causality, which seems to violate intuitions regarding the necessary temporal priority of causes. Hence after reviewing some of the literature that disputes that there is necessarily a conflict between a causal reading of the EPR correlations and special relativity, Laudisa raises the question: is it possible to provide a causal understanding of the connection that does not require backwards in time causation? The key to a proper analysis, according to Laudisa, lies in a better ontological account of the theory in the first place.

This leads Laudisa to address two different interpretations, the GRW theory (Section 5.4) and Bohmian mechanics (Section 5.5). The GRW interpretation is well known for its postulate of spontaneous collapses of the wavefunction. These spontaneous localisation events occur sufficiently often for the detection of macroscopic superpositions not to be possible in practice. One outstanding problem with the account is related to its relativistic extension since the localisation events seem to privilege a particular hypersurface and might select a frame. Laudisa distinguishes two different proposals for its ontology, the 'matter density' and the 'flash' ontology.[18] The former assumes that a continuous field on 3-dimensional space represents the matter density in each point of space at each instant. The latter by contrast assumes a discrete ontology, in which matter is made up of discrete points ('flashes') in spacetime such that to each of these flashes there correspond one of the spontaneous collapses of the wavefunction. One advantage of the flash ontology is that it has been shown to be Lorentz-invariant, while prescribing the relevant probability

[18] See Tumulka (2007) for the distinction and a development of the 'flash' ontology.

distributions for all observables. This avoids any conflict between GRW and the temporal priority of causes over effects thesis.

Laudisa then considers the non-relativistic alternative to select a preferred foliation of spacetime. He finds that while this assumption is unjustified for orthodox quantum mechanics, it is unavoidable in the case of Bohmian mechanics. In this context, as is well known, whatever mutual causal influence there is between the quantum potential or wavefunction in configuration space and the particles inhabiting 3-dimensional space, is both simultaneous and epistemically inaccessible in the sense that only the consequences of the causal interaction (the positions of the particles) are detectable by measurement apparati, but not the causal interaction itself. (It is worth noting that a propensity interpretation of the state in orthodox quantum mechanics would share this feature).

1.3.2 Causal Loops in Retro-Causal Models

In Chapter 6, Joseph Berkovitz carefully considers a number of retro-causal models of the Einstein-Podolsky-Rosen correlations. These are models that postulate the existence of causes acting backwards in time. A traditional objection against such causes in general states that they may generate loops in time which give rise to inconsistent effects. In the simplest case, suppose e causes c, but that c precedes e and is moreover an inhibitor of e, i.e. c is a cause of $\neg e$. Now suppose the causing is deterministic in both instances: it then follows that e if and only if $\neg e$. The most straightforward way to avoid such inconsistency would be a total ban on retrocausality. But there might be other less sanguine ways to keep such inconsistencies at bay, similar to those often used to keep at bay the inconsistencies generated by 'bilking'.[19] Berkovitz focuses on the particular conditions that obtain in an EPR experiment, with an eye to investigating ways in which causal loops maybe evaded even if the postulated causal structure contains causes that act back in time in at least some frames of reference. In the end Berkovitz's assessment is sober: even where such models may be postulated and do not entail inconsistency, there are problems regarding their predictive or explanatory power; and the problems are sufficiently severe to make the models dubious or at least unnecessary.

Berkovitz applies retrocausality to a specific experimental setting that he calls experiment X. This is an EPR experiment where the right hand side measurement takes place before the left hand side setting in the laboratory rest frame. Let us

[19] In the case of the famous 'bilking' argument (Black, 1956), the assumption is simply that an event c is the positive cause of an event e that lies in its past. The issue is then how to prevent the bilking of c after e has occurred. For if we prevent c from happening after e has already occurred, then this would generate the inconsistency that both 'c is the cause of e' and 'c is not the cause of e' are simultaneously true. Much will depend on whether 'bilking' is actually physically possible in the particular circumstances that give rise to c and e. Similarly for the type of inconsistency that causal loops may generate: much will hinge on the particular circumstances that bring about the EPR correlations.

denote by l, r the settings on the left and right hand sides; and by L and R the measurement outcome events on the left and right hand sides respectively. Suppose further that the right hand side outcome, R, is a deterministic cause of the left hand side setting l. Since we have assumed that R occurs before l in the rest frame of the laboratory, the causal connection between R and l is hence forwards in time in that frame. However, in a retrocausal model we additionally require either that (i) l retro-causes the complete state at the source, or (ii) both R and L jointly cause the complete state at the source.

We may then go on to appropriately distinguish two different kinds of retrocausal models: deterministic and indeterministic. In agreement with the standard understanding of these terms, a deterministic cause invariably brings about its effects in the appropriate circumstances. An indeterministic cause by contrast, determines the probabilities of its effects between zero and one – so it brings about its effects but only with certain probabilities. For instance in a typical retrocausal model of experiment X, the measurement setting on the left, l, may be a partial but deterministic cause of the complete state at the source, which in turn is a partial but indeterministic cause of the outcome events. (This seems to be what Berkovitz has in mind with his 'DS model'). By contrast, if the setting l only prescribes the probabilities for the complete state at the source, the model is indeterministic. In either case, there is a causal influence from settings or outcomes back towards the complete state at the source at the time of emission.

More specifically retrocausal models are typically assumed to violate the condition known as λ-independence[20]:

$$\rho(\lambda/\psi \ \& \ \mathrm{l} \ \& \ r) = \rho(\lambda/\psi)$$

where λ is the complete (hidden variable) state of the pair at the source, ψ is the quantum mechanical state, and l and r are the settings of the measurement apparatuses on the left and right side of the experiment respectively. In other words, in these models the hidden state at the source is statistically dependent upon the quantum state and the left and right settings. However, recall that in a typical EPR experiment the setting events take place in the rest frame of the laboratory *after* the emission event at the source and thus after the hidden state is determined. If the statistical dependence expressed by λ-independence reflects a violation of direct causal influence it follows that posterior events causally influence antecedent ones.[21]

[20] 'Hidden autonomy' is Van Fraassen's (1982) terminology.

[21] But does statistical dependency reflect causal dependency? Arguably the relationship is more complex and subtle. First, it is well known that statistical dependencies may mask hidden factors or hidden common causes. And second, the relation of conditional probability $P(x/y)$ need not indicate that the conditioned upon event y is a direct cause of the event x. This requires a further assumption (see Section 1.6 in this essay). I will follow Berkovitz here and assume for the sake of argument that causal dependencies can be read off statistical relations. In the second part of this chapter, I argue that conditional probabilities are not generally a reasonable way to read propensities.

Berkovitz carefully analyses different kinds of retrocausal models of experiment X and concludes that these models entail the existence of causal loops. The issue is then how to interpret such loops and their consequences, and in particular whether they imply inconsistent predictions. Berkovitz concludes that the causal loops within some deterministic models entail inconsistent predictions, while those within indeterministic models are unable to determine the distributions over complete states or measurement outcomes (unless supplemented with the appropriate statistical rules).[22] So in the deterministic case, retrocausality possesses the potential to generate contradictions, while in the indeterministic case it is unable to generate any meaningful predictions at all. Either way these are important arguments against retrocausal models of the EPR correlations in general.

1.3.3 Causal Completeness of Probability Theories

In Chapter 7 Balazs Gyenis and Miklós Rédei provide a review and reassessment of recent work regarding the notion of causal completeness for probability spaces. They provide very precise formal definitions of some of the most important terms in this literature. For instance, they define the concept of generalised Reichenbachian common cause (in Section 1.3) and the notion of causal completeness that follows from it (Section 1.4). They then review some of the main results on causal completeness derived within the so-called 'Budapest school'.[23]

The basic formal notion is that of a general probability measure space (£, Φ), where £ is an orthocomplemented lattice and Φ is a generalized probability measure or *state*, a σ-additive map Φ: £ $\rightarrow$ [0, 1] where $\Phi(0) = 0$ and $\Phi(1) = 1$. (Roughly: the elements of the lattice $\{A, B\}$, or variables, correspond to one-dimensional observables while the measure Φ defines the probabilities over the values of these variables ascribed by a quantum mechanical state). We may then define a correlation as follows: Corr_Φ (A, B) is the measure of correlation between compatible variables A and B in the state Φ.

A generalised version of Reichenbach's *criterion* of the common cause [24] may then be formally characterised as follows[25]:

C_k is a Reichenbachian common cause of the correlation Corr_Φ $(A_I, B_J) > 0$ between A_I and B_J if $\Phi(C_k) \neq 0$ for all k ε K and the following conditions hold:

1. $\mathrm{Corr}_\Phi(A_I, C_k) > 0$.
2. $\mathrm{Corr}_\Phi(B_j, C_k) > 0$.
3. $\mathrm{Corr}_\Phi(A_I, B_J/C_k) = 0$ for all k ε K.

---

[22] Throughout his paper Berkovitz assumes a single-case propensity interpretation of probabilities. But he shows that analogous results stand if the probabilities are understood as frequencies.

[23] The name 'Budapest school' was introduced by Jeremy Butterfield (2007, 807).

[24] For the distinction between the 'criterion' and the 'postulate' of common cause see Suárez (2007b).

[25] See Gyenis and Rédei's Definition 3.1.

Gyenis and Rédei then show that these conditions reduce to the usual Reichenbach characterisation of common causes in the limiting case of two-valued variables. The intuitive idea is indeed the same, namely screening off: conditionalising upon the common cause renders its effects statistically independent. (The first two conditions assert that the common cause is statistically relevant to each effect taken separately).

The question of causal completeness of probability spaces is then in a nutshell the following: given any correlated variables A_I, B_J ε £, can we expand the probability space (£,Φ) so as to find a common cause variable C_K, satisfying the relations above, which is included in the space? Gyenis and Rédei formalise the notion of *causal completeness* as follows: A probability space (£,Φ) is causally complete with respect to a causal independence relation R and correlation function Corr_Φ if for any two compatible variables A_I, B_J in £ there exists a generalized Reichenbachian common cause C_K of size $K \geq 2$ in £ of the correlation.[26] The causal independence relation R minimally requires logical independence – but it must impose additional conditions.[27]

Under these conditions Gyenis and Rédei review a number of important results on causal completeness; the most important seems to be 'proposition 8', which states that 'every atomless general probability space is causally event-complete'. This means that there are statistical theories that are causally complete: i.e. they contain the Reichenbachian common causes of their correlations. Gyenis and Rédei point out that it follows from this result that one may not refute Reichenbach's common cause principle by appealing to the thought that statistical theories are *generally* causally incomplete.[28]

1.3.4 Robustness and the Markov Condition

Chapter 8 is my own discussion (jointly with Iñaki San Pedro) of the relationship between the robustness condition once defended by Michael Redhead for the quantum correlations and the Causal Markov condition (CMC) that has been much discussed recently in the causal inference literature. We argue for a tight connection between these two conditions, namely: robustness follows from the CMC together with a number of additional assumptions. First we take Richard Healey's (1992) distinction between two forms of robustness, each appropriate for the assumption

---

[26] See Gyenis and Rédei's definition 4.1. A common cause variable C_K has size 2 if it has two values. For instance an indicator function (on-off) can be represented as a size two variable $(C, \neg C)$.

[27] Gyenis and Rédei leave open what this further conditions may be, which seems wise since their aim is to describe formal models applicable to any physical set ups. In causal modelling one would of course like to know more about this relation, and in particular the physical conditions that must obtain for A, B to be causally independent in the prescribed sense.

[28] The reasoning is convincing but one wonders to what extent the arguments against Reichenbach's Principle depend on the claim of (formal) incompleteness. For discussion see San Pedro (2007, Chapter 3).

of total or partial causes. (Healey reserves the term 'robustness' for the first condition only, while using 'internal robustness' for the second condition.) We then show that each notion of robustness follows from CMC and the assumption of either total or partial causes under the only further assumption that there exists one independent disturbing cause acting on the putative cause of the cause-effect link (in other words, that a form of intervention is possible). This entails that from the standpoint of an interventionist account of causality there is no real difference between applying robustness or the CMC. And the latter condition is more general since it does not require interventions (or disturbing causes). So it may be safely assumed in all future discussions regarding the status of causality in quantum mechanics. The robustness literature is thus shown to be superseded, and we recommend philosophers of science and causal methodologists alike to focus on the status of the CMC in quantum mechanics instead.

This argument so far supports the programme of the causal Markov condition theorists, such as Jim Woodward and Dan Hausman. However, in the second half of the chapter we go on to disagree with Hausman (1999) and Hausman and Woodward (1999) over the status of causation in quantum mechanics. It has traditionally been supposed that quantum mechanics provides a striking refutation of the principle of common cause and other standard methods of causal inference. This would arguably compromise the validity of CMC – at least in indeterministic contexts. Hausman and Woodward have claimed that the CMC is not false in quantum mechanics, but rather *inapplicable*. That is, they maintain that the conditions that would allow us to apply CMC are not met in this setting, and it is impossible to tell whether CMC obtains or is violated. We argue that on the contrary there is in principle no reason why the CMC cannot be applied. Not only that, but the application of CMC does not support the traditional judgement regarding causation in quantum mechanics. On the contrary our assessment is that whether or not CMC is violated depends very sensitively upon both the detailed statistics modelled, and the interpretation of quantum mechanics applied. As an example we discuss the status of causality in EPR in the context of the model of Bohmian mechanics. Steel (2005) has argued that in this context the CMC fails; we argue that to the contrary it arguably obtains, provided enough attention is paid to the details of the model itself. More generally, our paper is a call to apply the CMC to quantum mechanics in order to figure out causal structures, but to do so judiciously – and this, we claim, requires a healthy dose of methodological pragmatism. Philosophers ought to start by looking at the diverse range of models available first within a number of different interpretations and then draw their judgements on the basis of a consideration of their details.

1.4 Propensities

The third and final part of the book contains three essays on propensities, mainly in the quantum domain. Mauro Dorato reassesses the role of dispositions in quantum mechanics, Nicholas Maxwell reviews the latest stage of his 'propensiton' theory, and Ian Thompson provides a philosophical analysis of nested dispositions in physics.

1.4.1 Dispositions in the Ontology of Quantum Mechanics

In Chapter 9 Mauro Dorato considers the role of dispositions in quantum mechanics. In particular the most substantial part of the paper defends a role for dispositions within the so called Ghirardi-Rimini-Weber (GRW) interpretation. Dorato defends the view that the probabilities for collapse ascribed by these theories can be given an objective reading – in particular, they are interpretable as propensities. He suggests two different ways for doing this. First, he aims to show that dispositional readings of the spontaneous collapses postulated by these theories are not only possible but natural. Second, he argues against alternative non-dispositional interpretations of collapse probabilities, particularly the Lewis-style best system analysis account.

On the first issue, Dorato argues that dispositions are natural on both the original mass density localisation proposals of Ghirardi-Rimini-Weber (1986) and the most recent proposal attributed to Tumulka (2007), the so-called 'flash ontology' proposal. (The supposed advantage of the latter is the existence of a relativistic extension). Secondly, Dorato argues against Frigg and Hoefer's (2007) attempt to read quantum probabilities in the GRW interpretation in a Humean way, in accordance with the best system analysis. Dorato's main claim seems to be that the quantum probabilities are conditional probabilities and therefore relations between sets of events or properties at the quantum level. A Humean reading of such probabilities would then incur a fallacy of omission – since it fails to explain what such conditional probabilities are *conditional upon*.

In the final section of the paper Dorato argues against my own selective propensity interpretation (Suárez, 2004, 2007a), which he appropriately links to some aspects of Bohr's response to the measurement problem. As I understand it Dorato is charging the selective propensity interpretation with a possible fallacy in its description of the actualisation of dispositional properties. Such actualisations may or not be physical processes. If they are physical processes, then the selective propensity account is incomplete since it does not describe them. If on the other hand such actualisations are not physical processes then the application of propensities remains mysterious (and its explanatory power is compromised): we are back to the old 'dormitive virtue' objection to dispositions in general.

The selective propensity account indeed remains silent on the physical processes that underlie the actualisation of propensities. It takes the standard propensity view that dispositions are displayed in probability distributions, each in its proper context of application.[29] But it does not aim to explain the mechanisms – if any – that connect dispositions and probabilities. Such mechanisms would appeal either to categorical properties in which case dispositions are ultimately reduced, or to further dispositional properties. Either option seems viable from a dispositionalist point of view, but neither seems called for since the very existence of such a mechanism seems a remnant from categorical property-speech. Consequently I disagree with the need to provide a categorical basis for the dispositions which Dorato and I do agree are applicable to Bohmian mechanics. (We agree on the applicability of

[29] Mellor (1971).

dispositions, but the agreement seems to end there – I take such dispositions may well be ultimately irreducible while Dorato thinks they must be reducible to the only categorical property available in Bohmian mechanics, i.e. position).[30]

1.4.2 The Propensiton Theory Revisited

Chapter 10 contains Nicholas Maxwell's latest defence of his 'propensiton' version of quantum theory, which he has been developing for more than three decades now (see Maxwell, 1972 for the earliest defence). Maxwell argues that the propensition quantum theory (PQT) has testable consequences that could in principle distinguish it empirically from the orthodox quantum theory (OQT). So the PQT is not merely an interpretation of quantum theory: it is an alternative theory in its own right. Its main merit, according to Maxwell, is to combine indeterminism – understood as the idea that there are essentially stochastic or probabilistic processes out there in the world which generate certain outcomes with certain probabilities – and realism – the view that at the quantum level nature too is determinate: properties have values all the time independent of whether or not subjected to measurement.

Maxwell is right that indeterminism and realism are not necessarily in contradiction. Some of the extant alternative interpretations of quantum mechanics – such as the Ghirardi-Rimini-Weber (GRW) collapse interpretation, and the Quantum State Diffusion (QSD) theory – are already living proof.[31] And Maxwell is right to claim that his propensiton theory (PQT) was formulated before these theories came onto the market. The PQT is distinct from either of these more established alternatives on several counts. The most important difference is that Maxwell postulates the existence of distinct entities – propensitons – which live in physical 3-d space and whose states are described by the quantum wavefunction. It is the physical interaction between such entities that 'fires' the spontaneous collapse of the wavefunction.

The theory has several virtues, not the least of which is to have anticipated collapse interpretations, and Maxwell canvasses and studies them well. Like any other

[30] I introduce irreducible dispositions into Bohmian mechanics in Suárez (2007a, Section 7.2). However, I was not the first person to suggest such a reading. Pagonis and Clifton (1995) are an antecedent (although to my mind they mistakenly understand dispositions relationally, and identify them with aspects of Bohmian contextuality). An attempt closer to my own ideas is due to Martin Thomson-Jones (Thomson-Jones, unpublished). We both defend irreducible dispositions with probabilistic manifestations for Bohmian mechanics but unlike Thomson-Jones I restrict the applicability claim to the causal or maximal interpretation. Thomson-Jones' unpublished manuscript is dated after the submission date of the final version of my paper. However, I was in the audience both in Bristol (2000) and Barcelona (2003) where preliminary versions of Thomson-Jones' paper were presented. Although I don't recall the details of these talks I am sure I was influenced by them, as well as many friendly chats with Martin over the years – for which I am very grateful.

[31] It is not surprising that such theories have already received interpretations in terms of dispositions – see Frigg and Hoefer (2007) and Suárez (2007a, Section 7.1).

version or interpretation of quantum theory the propensiton theory also has its own difficulties and challenges.[32]

1.4.3 Derivative Dispositions

In the last chapter of the book Ian Thompson faces up to a fundamental question for dispositionalism, namely the nested exercise of dispositions in physics. The manifestation properties for dispositions need not be categorical. Rather dispositions will often be manifested in further dispositional properties. Thompson cites potential energy and force as characteristically nested dispositions. (Potential energy is the disposition to generate a force, while force is the disposition to accelerate a mass). These are, in his terminology, *derivative* dispositions. It is interesting to apply the idea to the dynamical evolution of quantum systems (Section 11.4.3). Suppose a system in an initial state $\Psi(t_0)$ is evolved by a Hamiltonian $\hat{H}$ to a new state $\Psi(t_1)$. Thompson suggests that the Hamiltonian be a disposition to evolve the state, while the states be themselves dispositional properties, namely propensities to produce measurement outcomes with the various probabilities $p_\lambda = |< \mu_\lambda | \psi(t) >|^2$. The Hamiltonian represents a 'dynamical' or diachronic disposition that generates further 'static' or synchronic dispositional properties, or propensities, on measurement.[33] We may then refer to the latter as *derivative* dispositions.

The full range of derivative dispositions generates a 'grid' of dispositions that we may refer to as *multiple generative levels*. Thompson introduces a number of additional distinctions and terminology to supplement this idea. The terminology is essentially causal because Thompson assumes that the action of primary dispositions over the inferior levels down the grid is causal in nature. (Thus he would say the Hamiltonian disposition *causes* the successive sets of static propensities). The thesis that dispositions and their manifestations are causally related is not new.[34] It suggests that there is a particular time or instant at which the disposition fires to generate its manifestation. And this introduces questions regarding the nature of the 'firing' event, and whether it is grounded upon further dispositional properties. We do not enter these difficulties here. The point Thompson's essay makes admirably is the more basic one that the manifestation properties of dispositions may be dispositional too.

---

[32] See Thompson (1988) and Suárez (2007a, Section 4) for discussion.

[33] The idea strongly recalls the distinction between dynamical and value states within the modal interpretation of quantum mechanics. See Van Fraassen (1991, Chapter 9).

[34] David Lewis (1997, 149 ff.) introduced the idea of causal bases for dispositions. Bird (2010) discusses objections to the idea that stimulus conditions cause dispositions to manifest themselves. For the purposes of this introduction I have ignored stimuli and concentrated on the disposition – manifestation relation itself (e.g. in the discussion in Sections 5–6).

1.5 Transition Versus Conditional Probabilities

Most of the authors in this volume discuss, often approvingly, the idea that the properties dealt with in fundamental physics and, particularly in quantum mechanics, may be essentially dispositional, or propensities. Objective physical propensities or chances are sometimes represented as forwards in time conditional probabilities. In this section, I provide a brief argument for the view that the best representation is instead by means of transition probabilities, and that both representations are distinct.

1.5.1 Transition Probability: Take One

Consider the equation for a forwards transition probability discussed in Section one:

$$P_{j+1/j}(S(t_{j+1})/S(t_j)) = P_{j\&(j+1)}(S(t_{j+1})\ \&\ S(t_j))/P_j(S(t_j)) \quad \text{(FTP)}$$

This equation does not express a well-defined conditional probability. The probability functions are different in each side of the equality since the time sub-indexes are different. Rather the formula enables us to calculate the probability for a physical transition from the state $S(t_j)$ to the state $S(t_{j+1})$ by working out the probability of the earlier state at the time of its occurrence and then the joint probability of both states at the conjunction of *both* distinct times. I discuss more precisely the meaning of this expression shortly. For now let us just note that the expression of a transition probability crucially differs from the similar expression for the conditional probability of successive states at time t_j:

$$P_j(S(t_{j+1})/S(t_j)) = P_j(S(t_{j+1})\ \&\ S(t_j))/P_j(S(t_j)) \quad (\text{CP}_j)$$

It also differs from the conditional probability of such states but calculated at the later time t_{j+1} [35]:

$$P_{j+1}(S(t_{j+1})/S(t_j)) = P_{j+1}(S(t_{j+1})\ \&\ S(t_j))/P_{j+1}(S(t_j)) \quad (\text{CP}_{j+1})$$

Thus, a transition probability is at least *prima facie* distinct from the corresponding conditional probability regardless of the time that it is calculated at. The formal difference between the expressions reflects a deep physical distinction.

[35] On the assumption of a fixed past and an open future (CP_{j+1}) does not express anything informative since $P_{j+1}(S(t_j)) = 1$ and $P_{j+1}(S(t_{j+1})/S(t_j)) = P_{j+1}(S(t_{j+1}))$ for any states $S(t_j)$, $S(t_{j+1})$. But Bacciagaluppi is interested in the meaning that these expressions, and the corresponding concepts, may have in the absence of any assumptions regarding becoming or any other asymmetry in time. So he is right in considering them as distinct possibilities. The only reason I ignore (CP_{j+1}) in what follows is that all the considerations in the text above against reading (CP_j) as a transition probability apply just as well to it.

1.5.2 Transition Probability: Take Two

As a matter of fact FTP does not express a conditional probability at all since a transition probability is neither conceptually identical nor reducible to a conditional probability. We would be better advised to write down transition probabilities as follows:

$$P_{j\text{»}j+1}(S(t_j)\text{ » }S(t_{j+1})) = P_{j\&(j+1)}(S(t_{j+1})\ \&\ S(t_j))/P_j(S(t_j)) \quad \text{(TP)}$$

A new symbol '»' has been introduced to represent the actual physical transition from state $S(t_j)$ at t_j to state $S(t_{j+1})$ at t_{j+1}. The symbol characterises what is distinct about a *transition*, namely the actual dynamical *change* or transformation, of the state. Consequently one must distinguish carefully the probability of a state to state transition from the conditional probability of one of the states conditional on the other. $P(S(t_j)\text{ » }S(t_{j+1}))$ expresses the probability of a transition, while $P(S(t_{j+1})/S(t_j))$ expresses the probability of the later state conditional on the earlier one. Conditional probability does not require nor entail a dynamical process that physically transforms the prior into the present state; it simply expresses statistical dependencies between different states regardless of what goes on 'in between'. (Conditional probability is compatible with such a process – the point is that it neither requires it nor does it ascribe it a probability). In other words TP and CP$_j$ are not equivalent in the fundamental sense that they do not express the probability of the *same event*. TP expresses the probability of a dynamical change of state and it presupposes that such events exist and moreover that they may be meaningfully represented in the sigma field that constitutes the domain of the probability function. CP$_j$ by contrast expresses a conditional probability of the state at a certain time given the state at another time, and it is perfectly legitimately well defined on a sigma field where only states are represented. It does not require *changes* or physical transitions from one state to another to be represented in the domain of the probability function; in fact it does not require such changes or transitions to be events at all.

The advantage of starting out with TP as a definition of transition probability is that it becomes immediately clear that a good amount of substantial argument would be needed to show that transition probabilities conceptually reduce to conditional probabilities of *either* the CP$_j$ or CP$_{j+1}$ types.[36] In particular, the argument required is not simply formal, but would imply a difficult to justify restriction of the sigma fields over which these functions are defined.

[36] A different further question is whether these probabilities (in particular TP and CP$_j$, whenever they are both well defined) should coincide numerically for the initial and final states of any state transition. A study of the conditions under which they coincide is beyond the reach of this essay – but it seems to me to be an interesting and promising research project.

1.5.3 Transitions are Not Conditionalisation Processes

Transition probabilities TP are also distinct from Bayesian conditionalisation events, which are often taken to express the rule for rational change of subjective degree of beliefs:

$$P_{j+1}(S(t_{j+1})) = P_j(S(t_{j+1})/S(t_j)) = P_j(S(t_{j+1})\ \&\ S(t_j))/P_j(S(t_j)) \quad \text{(Cond)}$$

Conditionalisation is often invoked by Bayesians as a mechanism for the updating of rational degrees of belief in theories, laws, or other general hypotheses. It is rather unclear what it could possibly mean in the context of state-transitions. It could start to make sense if we could speak of a learning process whereby some agent first learns that state $S(t_j)$ occurs, and then wants to update her estimate of the probability of $S(t_{j+1})$ in light of this new knowledge. However, the corresponding change in degrees of belief would take place at time t_{j+1}, the time at which the state changes to become the new state $S(t_{j+1})$. So by the time we are supposed to update, the new state already has objective probability one. Why would anyone want to use conditionalisation in order to update her degree of belief in a state that has already occurred? Why, more generally, conditionalise on the basis of information that is already old? Whatever Cond means it is certainly formally distinct from the expression of a transition probability TP – the latter neither requires nor entails any updating rule for the probabilities at any given time.[37]

1.5.4 Biased and Unbiased Samples

The key to transition probability is the expansion of the sigma field of a probability function in order to include a representation of physical transitions or state-changes. An interesting question is whether this compromises the notion that an unbiased data sample must be in equilibrium since we know that samples out of equilibrium may generate qualitative time asymmetries between forwards and backwards transition frequencies (c.f. the discussion of Bacciagaluppi's argument in Section 1.2.1). There is reason to think that it does if there is reason to think that physical transitions or changes of state necessarily take place forwards in time. If so, the only events that are additionally represented in the sigma field of a transition probability are physical changes of state *forwards* ($S(t_j) \gg S(t_{j+1})$), but not physical changes of state *backwards* ($S(t_{j+1}) \gg S(t_j)$). As a result the data samples can at best contain the former type of events but not the latter. Not surprisingly, forwards transition frequencies over these data samples will appear to be time invariant but not so backwards transition frequencies.[38] In the view defended in this essay propensities are represented by forward looking transition probabilities. So in this view it is automatic that

[37] For a different argument to a similar anti-Bayesian conclusion see Guerra (2009, Chapter 8).

[38] See Arntzenius (1995, esp. Section 2) for a detailed example and discussion.

forwards transition frequencies measure the relative outcomes of genuine dynamical changes, while backwards transition frequencies are merely relative ratios of states calculated by means of the forwards transition probabilities and initial conditions.[39]

I conclude that the ontological primacy of forwards over backwards transition probabilities can only be denied if either (i) genuine physical changes of state occur backwards as well as forward in time, or (ii) no genuine physical changes of states occur ever at all. The former option entails denying that propensities, or objective transition probabilities, are time oriented. The latter option entails denying that such things as propensities, or transition probabilities, exist at all – in either direction in time. Both entail a major shift in our ordinary ontology.

1.6 Propensity as Probability

Transition probabilities are thus probabilities of genuinely physical changes of state. They somehow reflect the tendencies or propensities that systems possess to exert such changes. How should we represent them? I will adopt the view that quantum propensities are displayed in probability distributions, namely the usual transition probabilities provided by Born's rule. In this section, I elaborate on this notion of propensity, in particular with reference to some of the key texts and positions in the more general literature. I first distinguish it from the more widely known *propensity interpretation of probability*. I then discuss some historical precedents for the sort of view that I discuss here. Finally, I address the principal objection against the propensity interpretation in recent years, namely 'Humphrey's paradox'.[40]

1.6.1 Long Run Versus Single Case Propensities

The philosophy of probability literature appropriately distinguishes two types of propensity interpretations: long run and single case.[41] The difference between these two types lies in the object that is identified as the propensity. Long run interpretations of propensity identify propensity with the dispositional property of a chance set up to generate frequencies in sequences of outcome trials, while single case

[39] Penrose (1989, 355–359) defends an apparently similar view regarding the quantum mechanical algorithm for computing transition probabilities (the Born rule) in general. He claims that the algorithm can err if applied to compute backwards state-transitions: 'The rules [. . .] cannot be used for such reversed-time questions' (ibid, p. 359). The representation of transition probabilities proposed here makes it clear why this should be the case.

[40] The view of propensities that I shall be defending here is very much my own (see Suárez, 2004, 2007a), and none of the contributors in the book has explicitly committed to it. However I believe that this view, or a similar one, is required for the coherence of many pronouncements made in the book, particularly in the third part. If so, we may take this or a similar view to be implicit in the book, and its defence in this section to provide support for it.

[41] Gillies (2000a, 124–126); Fetzer (1981, Chapter 5).

interpretations identify it with the tendency to generate a particular outcome in a given trial. There are at least two long run interpretations: those which accept infinite virtual sequences and those which accept only long yet finite sequences. All long run interpretations have the following in common: a chance set up (an arrangement of distinct parts capable of generating a sequence of stochastic outcomes of some trial) may possess a propensity for some type of outcome if and only if the limiting frequency of such a trial outcome is well defined in each (long but finite, or virtual and infinite) sequence. Hence long run propensity interpretations agree with frequency interpretations in requiring sequences for the ascription of probabilities. The difference is that a long run propensity interpretation will not focus on the properties of the sequences (frequencies) but rather on the properties of the chance set ups that generate those sequences. In other words, a long run propensity interpretation does not identify probability with frequency, but with the tendency to generate the frequency.

Similarly, a single case propensity interpretation will not identify probability with any trial outcome but with whatever dispositional property generates a particular trial outcome. So a probability in this case is a tendency that is exerted in every trial; no frequency in any finite – however long – sequence of such trials may need to agree with the particular probability. The only frequencies that, on a single case propensity interpretation, need to agree with the probabilities are those pertaining to the virtual infinite sequences that would be generated if it were possible to repeat the same experiment an infinite number of times. Yet, unlike the long run propensity interpretation, the single case interpretation does not identify propensity with the tendency to generate any frequency, whether finite or infinite. Rather, it associates propensity with the tendency to generate *each* particular outcome in the sequence.[42]

When authors in the book discuss propensities they almost invariably have in mind a single-case interpretation. There are, however, a number of interesting differences among different single case interpretations and it is worth to review them quickly.[43] Gillies divides propensity interpretations into two types depending on what is regarded as an appropriate chance set up – i.e. the set of conditions that must obtain at a given time for the appropriate tendencies to be instantiated. Humphreys by contrast divides single case propensity interpretations into three additional types differing in their account of dynamics for propensities – i.e. their time evolution over a period and their effect on different events at successive stages.

Let me consider Gillies' taxonomy first, which divides all propensity interpretations into *repeated conditions* and *state of the universe* interpretations. The chance set up may be a simple enough arrangement that could be specified by means of just a few free variables or parameters. (The toss of a coin is an example). If so, a chance

[42] Long run propensities as tendencies to generate infinite sequences seems to be what the early Popper defended in his classic (1959), and as tendencies to generate long but finite sequences have been defended by Gillies (2000a, Chapter 7). Single case propensities are defended by Fetzer (1981, Chapter 5) and Miller (1994).

[43] I essentially follow the exposition in Gillies (2000a) and Humphreys (2004) and introduce further considerations along the way.

set up is defined by just a few conditions that are repeatable and hence allows for the same sort of trial to be repeatedly carried out. A single case interpretation of this sort implicitly requires all propensities to be conditional on such a set of repeatable conditions. Alternatively, a chance set up may include the complete hypersurface corresponding to a particular time *t*. If so, a chance set up is defined by the whole state of the universe at *t*. This type of single case interpretation too requires all propensities to be conditional – albeit conditional on a complete hypersurface.[44] In either view, there are no absolute propensities $\Pr(A_{t''})$ for any event or proposition *A* at any time t''. Any seemingly absolute propensity is really a conditional propensity, $\Pr(A_{t''}/S_{t'})$ with $t' < t''$, where *S* is either the full state of the universe at t', or the particular set of conditions required by an appropriate chance set up at t'.

On the assumption that all propensities are conditional, Paul Humphreys provides a different taxonomy based on the dynamical evolution of conditional propensities.[45] A *coproduction interpretation* assumes that the conditional propensity is fixed once and for all at the initial time *t* whether by a particular set of relevant conditions at *t* or by the *t* hypersurface or time slice. Thus all propensities carry an implicit time index which need not coincide with the time index of either conditioned or conditioning event. For example $\Pr_t(A_{t''}/S_{t'})$ is the propensity at *t* for *A* at t'' given *S* at t'. Under the natural assumption that $t < t' < t''$ a coproduction interpretation assumes that the conditional propensity of $A_{t''}$ given $S_{t'}$ is already fixed at the original time *t* given the background conditions at that time. A *temporal evolution interpretation* by contrast assumes that propensities evolve continuously in time so the propensity of $A_{t''}$ at *t* need not be identical to that at t'. The conditional propensity of $A_{t''}$ given $S_{t'}$ must then be evaluated at t' : $\Pr(A_{t''}/S_{t'})$ as the temporal update of the original propensity $\Pr_t(A_{t''}/S_{t'})$. Finally, a *renormalisation interpretation* assumes that updating is necessary even though there is no continuous temporal evolution of the propensity. (The difference between the renormalisation and the temporal evolution interpretations is that the former does not presuppose continuous evolution so updating in intermediate stages is not required. In the temporal evolution interpretation, by contrast, an updating at t'' of a propensity first defined at *t* necessarily requires an intermediate updating at t').[46]

The two taxonomies are orthogonal and, in principle, any of the six combinations is logically possible. Humphreys and Gillies in effect argue that as long as applied to single case propensities all fifteen of them are ruled out by Humphrey's paradox. In what follows I review the notorious paradox. For now I just note that all propensity

[44] The different interpretations are then classified as follows: Fetzer (1981) defends a single case repeated conditions interpretation, while Miller (1994) defends a single case state of the universe interpretation. Gillies (2000a, 130–136) argues that these interpretations succumb to Humphrey's paradox, and defends instead a long run repeated conditions interpretation.

[45] Humphreys (2004).

[46] Humphreys actually lists a fourth case, the *causal interpretation* (Humphreys, 2004, 673). However, the causal interpretation is not really on a par with the other three since it is not *per se* a dynamical interpretation of the evolution of propensities. In fact it does not seem to exclude any of the other three dynamical interpretations, being rather compatible with any of them.

interpretations so far analysed have one thing in common: they presuppose that there are no genuine absolute propensities and that all propensities are implicitly or explicitly conditional. Later on I shall argue that there is nothing in the dynamical interpretations per se that implies that this should be the case; and that there are alternative ways of understanding both *relevant conditions* and *state of the universe* interpretations.

1.6.2 Humphrey's Paradox

'Humphrey's Paradox' (HP) was first described by Wesley Salmon (1979, 213–214) and James Fetzer (1981, 283) who was also responsible for naming it. Most commentators describe it not so much as a 'paradox' as a powerful argument against the propensity interpretation of probability.[47] The key idea underlying this argument is roughly that propensities partake of the asymmetry of causation in a way that probabilities do not. But if propensities are causally asymmetric and probabilities are not, then they can not be the same kind of thing. Hence a wholesale propensity *interpretation* of (the classical – Kolmogorov – calculus of) probability is out of the question.

Let me use a simple everyday example to try to make the rough idea a bit more precise. Some of my friends have remarked on my propensity to travel to North America in the spring. On the basis of the relative frequency in the last 10 years, we may estimate the probability corresponding to this propensity roughly at $P(\text{NA/S}) = 0.9$ (where NA is my travelling to North America, and S stands for the – northern hemisphere – spring). We can then apply Bayes theorem in order to find out the value of the inverse probability of spring conditional on my travelling to North America: $P(\text{S/NA}) = P(\text{NA/S}) \times P(\text{S})/P(\text{NA})$. Dividing the year in four seasons and applying some estimates for the priors, we obtain $P(\text{S/NA}) = 0.56$. Let us suppose that there is a set of causal facts $\{F\}$ underlying my friends' propensity adscription along the lines of the intended implication, namely that $\{F\}$ are features unique to the spring season that attract me to North America, and cause me to travel there. We can suppose that $\{F\}$ includes (in addition to facts regarding the seasonal weather in spring in both continents) some facts about my psychology, habits and values, my work schedule, my family and financial situation, etc. Whatever these causal facts $\{F\}$ are, they fail to justify any propensity corresponding to the inverse probability. For whatever it is that causes me to travel does not also cause spring. In these terms, the inverse probability $P(\text{S/NA})$ does not seem to have any possible causal underpinning. The relevant causal facts relate to the conditioning event S, while the effects of interest relate to the conditioned event NA. But the inverse probability has inverted conditioned and conditioning events. And it is implausible to suppose that there are other facts $\{F'\}$ about North America – or about my travelling there – that

[47] Fetzer (1981, 283–286); Gillies (2000a, 2000b); McCurdy (1996); Miller (1994); Milne (1986).

cause or bring about spring with a 0.56 chance. (Certainly those very causal facts which underlie my propensity to travel there in the spring do not probabilistically cause it to be spring when I travel; so $\{F'\} \neq \{F\}$; and it is hard to see what other facts could be cited).[48]

On the basis of examples like this, many commentators have asserted that Humphreys' Paradox shows that very many well defined conditional probabilities are not propensities. This seems to rule out the propensity interpretation of probability in general since there is nothing about P(S/NA) that makes it in any way suspect as a well defined probability (certainly not as long as P(NA/S) is well defined). Notice that there are two assumptions underlying this use of the example. The first (*Assumption* 1) is that the propensity interpretation applies to conditional probabilities.[49] The second (*Assumption 2*) is that a propensity interpretation applies only when the conditioning event is a cause or partial cause of the conditioned event. This assumption trades on a supposedly intimate link between propensity and causation whereby the former inherits the asymmetry characteristic of the latter.

Paul Humphrey's own version of HP is not explicitly built on either of these two assumptions. But the assumptions are brought in implicitly. This is perhaps clearest in the discussion of the notorious example involving the transmission and reflection of a photon from a half-silvered mirror.[50] A source emits photons spontaneously; a few of these photons reach the mirror; among these a few are actually transmitted. Now let us consider the propensity for a single photon to be emitted at the source at time t_1; to hit the mirror at time t_2; and to be transmitted at time t_3. And let us consider the complete state of the source and mirror at time t_1; i.e. at the time of emission of the photon at the source. Humphreys invites us to consider the following assignment of propensities at time t_1:

(i) $\mathrm{Pr}_{t1}(T_{t3}/I_{t2}B_{t1}) = p > 0$
(ii) $1 > \mathrm{Pr}_{t1}(I_{t2}/B_{t1}) = q > 0$
(iii) $\mathrm{Pr}_{t1}(T_{t3}/\neg I_{t2}B_{t1}) = 0$

where B_{t1} represent the background conditions at t_1; I_{t2} the incidence of the photon upon the mirror at time t_2; and T_{t3} the transmission event of the photon. According to Humphreys these three propensity ascriptions are entailed by the physical and experimental circumstances. They do not follow from the formal features of the calculus of probability because the arguments in the propensity functions designate physical events and do not necessarily pick out subsets of a measure theoretic outcome

[48] For the convenience of the story, I am assuming that the relata of causation are facts along the lines of Mellor (1995). But the argument does not hinge on this assumption.

[49] This need not rule out absolute propensities, although some commentators – notably Gillies (2000a, 131–132) – go further and claim that all propensities are implicitly if not explicitly conditional. In this view a propensity interpretation of probability is always of (and only of) conditional probability.

[50] Humphreys (1985, 561).

space.[51] Indeed once the formal framework for the representation is chosen the content of ascriptions (i), (ii) and (iii) is not formal but empirical. However, it does not follow from the physical and experimental circumstances that the propensities involved are conditional nor does it follow that they must be formally represented in a way akin to conditional probabilities. This is a point that I shall take on later – and reveals that *Assumption 1* is built into the discussion of the example.

Humphreys invites us next to consider the following principle of conditional independence for propensities[52]:

Conditional Independence (CI): $\mathrm{Pr}_{t1}(I_{t2}/T_{t3}\ \&\ B_{t1}) = \mathrm{Pr}_{t1}\ (I_{t2}/\neg\ T_{t3}\ \&\ B_{t1}) = \mathrm{Pr}_{t1}(I_{t2}/B_{t1})$.

Together with the ascription of propensities above, this principle contradicts the (Kolgomorov) axioms of classical probability. The contradiction with the fourth axiom, in the form of Bayes Theorem for conditional probability is particularly easy to show.[53] So, at least one among these assumptions must go. Some responses to HP have focused on trying to show that principle CI is false when applied *to this particular example*.[54] But in retort Humphreys produced yet another example that conclusively obeys CI.[55] Other authors endorsed the HP argument as a definitive reason to abandon the propensity interpretation altogether.[56] Humphreys himself concluded that the axioms of classical probability can not represent propensities accurately. But instead of abandoning propensities, he recommends abandoning the classical (Kolmogorov) calculus of probability as a representation of chance or objective probability.

1.6.3 Conditional Propensities

The CI principle and its use in the derivation of Humphreys' Paradox require some careful analysis. Strictly speaking CI merely states that the propensity of the photon impinging on the mirror at t_2 is independent of the (later) event of transmission at t_3, and depends only on the background conditions at t_1. But Humphreys seems to think that the actual principle of conditional independence is more general, and CI as formally expressed above is merely a consequence of such a general principle. For he writes that the CI principle 'claims that any event that is in the future of I_{t2} leaves the propensity of I_{t2} unchanged. [...] This principle reflects the idea that

[51] I have adopted Humphreys' suggested terminology and refer to propensities as Pr(–) and probability functions as either Prob(–) or simply $P(-)$.

[52] Humphreys (1985, 561; 2004, 669).

[53] Humphreys (1985, 562).

[54] McCurdy (1996).

[55] See Humphreys (2004). My objections below to CI are very different in nature and cannot be answered by means of new examples.

[56] Milne (1986).

there exists a non-zero propensity at t_1 for I_{t2} to occur, and that this propensity value is unaffected by anything that occurs later than I_{t2}.' (Humphreys, 2004, 670).

Conditional independence in general, unlike CI in particular, applies to *any event* later than t_2, and not just to T_{t3} in particular. So the expression above is not a definition of conditional independence in general, but rather the application of conditional independence to the particular example. The main intuition is presumably that the propensities of the photon at t_i can be altered only by events at times $t < t_i$. But the only reason to suppose this is the temporally asymmetric nature of the 'altering' relation – so *Assumption 2* is involved after all. More generally the intuition seems to be that a system's propensities at $\{x, y, z, t\}$ can only be altered by events in $\{x, y, z, t\}$'s past light cone. If so, CI presupposes the view that propensities are time asymmetric in just the way causation is asymmetric in relativity theory under the 'causal' interpretation: no cause may lie outside the past light cone of its effects. So, a version of *Assumption 2* is again built into the application of a general principle of conditional independence to the photon example.

How plausible is this relativistic version of *Assumption 2*? There are many good arguments against the 'causal' interpretation of special relativity.[57] And even in a non-relativistic setting, *Assumption 2* is inconclusive since backwards in time causation in a fixed frame has not been decisively ruled out.[58]

Humphreys claims that CI holds in the *co-production interpretation* of propensities,[59] presumably because in this interpretation all propensities are fixed at the initial time t_1. But if this grounds independence at all, it is the very general claim that all propensities *at time later than t_1* (including therefore but not only the propensity for T_{t3}) are independent of the propensity for I_{t2}. This claim goes well beyond the general conditional independence that we have considered so far – which included only events in the future of t_2. The *co*-production interpretation on its own grounds CI but it also grounds other similar independence conditions that we would not want to have to assert in this case. The only apparent way to extract precisely CI out of the co-production interpretation is by adding *Assumption 2* or a similar causal principle. The co-production interpretation, in conjunction with *Assumption 2*, then entails that I_{t2} is conditionally independent with respect to those events outside of its proper past light cone. In particular it follows that I_{t2} is conditionally independent of T_{t3}, as stated in CI. So, CI requires *Assumption 2* after all, even in the co-production interpretation.[60]

---

[57] See Maudlin (1995), particularly Chapter 5.

[58] In fact many of the arguments against backwards in time causation turn out to depend on the fine grained space-time structure of the putatively refuting examples. Others, such as the bilking argument, attend to agency only, but seem inconclusive. See Black (1956) and Dummett (1964) for two classic sources and discussion.

[59] See for instance the table in Humphreys (2004, 677).

[60] We may wonder about the status of conditional independence in other interpretations of propensities. CI holds in the temporal evolution interpretation – since the propensity of I_{t2} is updated at time t_2. So $\Pr_{t2}(I_{t2}/T_{t3}B_{t1}) = \Pr_{t2}(I_{t2}/\neg\, T_{t3}B_{t1}) = \Pr_{t2}(I_{t2}/B_{t1}) = 1$ or 0. But it fails in the renormalisation interpretation since $\Pr_{t3}(I_{t2}/T_{t3}B_{t1}) \neq \Pr_{t3}(I_{t2}/\neg T_{t3}B_{t1})$ in general. However,

Humphreys argues against the *co-production interpretation* anyway, on the basis that it is not a genuine single case propensity interpretation. He claims that it does not classify conditional propensities as real conditional chances in an ontological sense, but only in the measure theoretic sense.[61] I suppose that he must have in mind the view that at time t_1 all of the probabilities are fixed for all the propensities afterwards. So barring the very ascription conditions at time t_1 and events prior to this, all other events are included in the outcome space and must be represented in the sigma field that defines the probability function.

But if this is a reason to reject the *co-production interpretation*, it is also a reason more generally to reject the representation of propensities as conditional probabilities. Let us accept like Humphreys that a 'conditional propensity' is a *sui generis* ontological relation between two events (or event types) *a* and *b*. This relation is entirely independent of any formal representation in measure theory (given the typical underdetermination of mathematics by physics it is in fact natural to suppose that the same propensities may be represented by means of very many different measure functions). Why are we then obliged to represent them by means of the standard representation for conditional probabilities? Why are we obliged to provide a measure theoretic representation at all?

1.7 Propensity as Dispositional Property

As long as propensity is understood as an *interpretation of probability*, we have no choice. Probability is routinely represented in measure theoretic terms, and there are even some good representation theorems.[62] But why suppose that propensity *interprets* probability? [63] Once the idea has been given up that propensity is a particular kind of probability, or an interpretation of the term 'probability', it becomes possible to suppose that the relation between these two terms is something different; for example, something akin to theoretical explanation.

1.7.1 Propensities Display Probabilities

Propensities and objective probabilities are distinct notions and it is the job of a propensity theory to establish how they are conceptually related. The two theories

---

Humphreys (2004, 673) finds that a similar principle holds in the renormalisation interpretation, namely the *fixity* principle. (The fixity principle states that: $\Pr_{t1}(I_{t2}/T_{t3}) = 0$ or 1, which holds in the renormalisation interpretation since $\Pr_{t3}(I_{t2}/T_{t3}) = 0$ or 1). In all cases, I contend, Assumption 2 is implicit in the derivation of CI.

[61] Humphreys (2004, 675).

[62] As good as they come – typically not up to uniqueness. In particular, and rather to the point, the fourth Kolmogorov axiom is sometimes disputed – see, e.g. Hajek (2004).

[63] Why suppose that objective probability, or chance, requires any interpretation at all? After all many theoretical concepts bring their own interpretation and/or require no interpretation. Elliott Sober for one has recently argued for a no-theory theory of probability in Sober (2005).

that have come closer to taking this insight to heart are due to Hugh Mellor (1971) and James Fetzer (1988). In this view propensities are dispositional properties that are displayed in probability distributions but may not be identified with them. Instead of providing semantics for probabilities in the model-theoretic sense, propensities may be said to *explain* probabilities since they explain how a certain distribution rather than another one comes about in specific circumstances. But if we accept this understanding of propensities as dispositional properties, there seems to be no reason why the relations between such properties need be represented as conditional probabilities. Consider first the relation between the possession and manifestation conditions of a propensity, such as those involved in the fragility (F) of a glass and its breaking (B). Supposedly this is a deterministic disposition under certain conditions C; we may assume that it displays the conditional probability $P(B/F \ \& \ C) = 1$. Every fragile glass that is hit under specific conditions (certain strength, etc) will break. But why represent this propensity as a conditional probability? Under different conditions D, the same propensity gets displayed in a probability of breakage less than one: $P(B/F \ \& \ D) = x \leq 1$. So, in general, it makes sense to formally distinguish propensities from the probability distributions that display them.

There are at least three alternatives to the conditional probability representation. First, we may represent the displays of propensities always as absolute probabilities in the restricted probability outcome space. Thus, instead of writing $P(B/F \ \& \ C)$ and $P(B/F \ \& \ D)$ we may always write $P_{F\&C}(B)$ and $P_{F\&D}(B)$, defining these probability functions on the smaller space. Since the functions are different, their values may correspondingly differ too. The advantage of this representation is that every probabilistic display of a propensity ascription is then relative to a set of circumstances or manifestation conditions. The disadvantage is that it does not allow us to ascribe probabilities to the propensities themselves since $\{F, F'$ etc.$\}$ are not represented in the sigma field that defines the probability.

An alternative is to come up with a distinct representing symbol for what is, after all, a distinct relation. There are at least two different ways of doing this. We may first consider transitions of state, and put to use the notation that we devised to this effect in the previous section. Thus, $F \gg_c B$ denotes the transition from the dispositional 'state' F to the manifestation 'state' B under circumstances c. In the case of propensities the manifestation property is itself a probability distribution, and we may write $F \gg_c P(B_i) = p_i$ where B_i are the different possible values of the manifestation property B. In this representation the outcome space includes *both* property possession and property manifestation events as part of the propensity and manifestation 'states'. So, we can define probability distributions over propensities, manifestation properties, and their transitions. I shall for the most part employ this notation in my discussion below.[64]

---

[64] I therefore assume that the actualisation of a propensity is tantamount to a state transition from the propensity to the manifestation property. This is necessarily the case whenever the new manifestation property is incompatible with the original propensity. Otherwise it is a contingent matter of fact whether the actualisation process entails a transition, but this seems plausible in most ordinary cases. Thus the smithereens of a broken glass are rarely themselves fragile. And even if they

Yet, note that another alternative would allow us to represent the relations among different propensities, whereby the possession of some propensity may causally affect another set of propensities. This is obvious in the case of logical entailment among properties, which may be modelled as deterministic causation. (A typical macroscopic case is colour under a dispositionalist reading; so for instance being white ipso facto entails being coloured, etc). But in addition there may be genuine 'productive' causation among dispositional properties.[65] Both may be understood under a very general causal relation and represented by some appropriate symbol such as '$\hookrightarrow$'. [66] We may then write 'A causes B' as '$A \hookrightarrow B$'. We saw in Section 1.4 that there is a debate in the literature about whether propensities cause their manifestations. If it is the case that the manifestation relation is causal, then we can write $P_c(F \hookrightarrow B_i) = p_i$ instead of $F \gg_c P(B_i) = p_i$ without loss of generality. However, in line with my previous discussion I shall not assume that the manifestation relation is itself a causal relation, but shall instead employ the 'neutral' notational system for transitions of state in general. From now propensities and their probabilistic distributions shall be denoted as $F \gg P(B_i) = p_i$ where I shall drop the c subscript for convenience.

1.7.2 Absolute Propensities

We are now able to represent changes of propensity state as follows. Suppose that S_1 is the full state of the system expressing all its properties, whether dispositional or not, at time t_1 and S_2 is the full state at time t_2. Then $S_1 \gg P(S_2) = p$ expresses the fact that the transition probability for a change of state S_1 into S_2 is p. This notation makes it unnecessary to represent a transition probability as a conditional probability $P(S_2/S_1) = p$. As we saw in Section 1.5 the conditional probability notation for transition probability is not only unnecessary but undesirable.

There are a number of advantages to this new notional system for propensities. Let me just comment on two of them since they relate to issues that were already mentioned in this essay. First, I address the distinction between different long run propensity theories that were reviewed in Section 1.6. Second, I address some difficulties related to the principle of indifference that were briefly mentioned in the summary of Bangu's paper in Section 1.2.2.

Firstly, in Section 1.6 Gillies' distinction between *repeated conditions* and *state of the universe* interpretations was reviewed. Let us continue to refer to the propensity as F. In the standard propensity interpretation of probability this propensity is

---

were, the property of fragility would no longer be a property of the original entity. So it is arguable that the evolution of the system as described is best represented by means of a state transition anyway.

[65] For the distinction between 'productive' and 'dependence' or counterfactual causality, see Hall (2004).

[66] The symbol employed by Cartwright (1983, Chapter 3) for this relation.

identified with the corresponding conditional probability: $F = \Pr_t\ (A_i/S_{t'}) = p_i$ where the $\{A_i\}$ are the values of a given quantity to be measured at time t, the $\{p_i\}$ represent their probabilities, and $S_{t'}$ is either the (hypersurface) state of the universe at t' or the set of repeated conditions at t' (with $t' < t$). However, in the account defended here, propensities must be reformulated as dispositional properties that display absolute probabilities. In accordance with our notation, we must write $F \gg S_{t'} P_t(A_i) = p_i$ when under the circumstances S at t', the propensity A manifests itself as a probability distribution over the values of A at t. We leave open whether $S_{t'}$ represents the state of the universe at t' or the set of repeated conditions at t'. In either case propensities are dispositional properties that ensue – or evolve into – probability distributions. The conditional probability representation is altogether unnecessary.

Secondly, in the discussion of Bangu's chapter in the first section, a source of difficulties associated with the principle of indifference was mentioned. In particular, I voiced the concern that the principle may invite an illegitimate inference from a merely epistemic fact about our knowledge (or lack thereof) to an objective fact about the physical world – and in particular about its dispositional properties. I can make the claim more precise now. Under a conditional propensity account such as Humphreys, the principle of indifference leads from facts about our lack of knowledge regarding the outcome of a particular experiment to an incorrect ascription of objective properties in the world. For instance, under total lack of knowledge regarding the outcomes of an experiment A performed under repeated conditions S we would be advised by indifference to ascribe equal probability to all such outcomes and the corresponding propensity would be given by Pr $(A_i/S) = p_i$, with $\Sigma p_i = 1$ and $p_i = p_j$ for any i, j. It seems clear that no knowledge (or lack thereof) of any finite sequence can justify such an ascription of a propensity. So, under this construal of propensities, the principle of indifference leads to an incorrect ascription of objective facts about the physical world, namely its propensities.[67]

Now, interestingly, the problem disappears as soon as a dispositional account of propensities is embraced, with a concomitant representation in terms of the notation that we have developed. We must then write $F \gg_S P(A_i) = p_i$ for the manifestation of F as A under circumstances c. It is then perfectly possible to apply the principle of indifference in order to fix p_i in the absence of any knowledge regarding the outcomes. We obtain that $\Sigma p_i = 1$ and $p_i = p_j$ for any i, j, as in the previous case. However, we now make no statement whatever regarding the propensity F that underlies this distribution. The principle of indifference applies only to the probability distribution that displays F but not F itself. Thus, we no longer commit the fallacy of going from lack of knowledge to objective facts.[68]

---

67 The problem is most acute for long run propensity theories. Gillies (2000a) attempts to solve the problem by appealing to the notion of a falsifying rule for hypotheses. But this is a controversial solution as discussed in Section 1.2.2.

68 There is an interesting question here for the 'causal' notation alternative mentioned earlier. In that case we would write $P(F \hookrightarrow A_i) = p_i$ with $\Sigma p_i = 1$, and $p_i = p_j$ for any i, j. Here the

1.7.3 Humphreys' Paradox Revisited

Let us now bring the discussion to bear on Humphreys' Paradox. Recall that the inappropriate symmetry of conditional probability lies at the heart of HP. We are discussing the view that propensities are dispositional properties that are manifested as probability distributions under the appropriate circumstances. Suppose that under circumstances c propensity F is displayed as the probability distribution P over the values of some manifestation property B. I have argued that this is appropriately expressed as $F \gg_c P(B_i)$. One of the relata of the manifestation relation is a probability distribution – in agreement with the thought that propensities manifest themselves in probability distributions.

It should be obvious that symmetry fails on this representation. It does not follow from $F \gg_c P(B_i)$ that $B \gg_c P(F_i)$; it does not even follow that B has any manifestation properties at all! [69] The 'inverse' manifestation relation is not generally well defined. Moreover, Bayes Theorem has no application in these cases since all the probabilities are absolute and not conditional. So even restricting ourselves to the probability distributions that display the propensities, the 'inverse' probabilities need not be well defined either.

One possible objection is that there is always an equivalent representation in terms of conditional probabilities. However, I do not think that a conditional probability representation of the above manifestation and, more specifically, causal relations is possible without significant loss of meaning. As we already saw in Section 1.5, transition probability, which is possibly the most favourable case for the equivalence claim, is best understood as a change of propensity state and not as the outcome of conditioning.

In this account, the reasoning underlying Humphreys Paradox goes wrong at the very start. The representation of photon state transitions as 'conditional propensities' (i.e. conditional probabilities) is incorrect. Instead, these processes should be represented properly as involving probabilities for manifestation or causal relations between propensities. It is the photon incidence upon the mirror that manifests itself in its transmission (or partially causes it together with the background conditions at t_1). The incidence of the photon is a manifestation of its ejection or at least partially caused by it. Etc. The first three conditions should then be re-expressed accordingly:

---

application of the principle of indifference would lead us to infer objective facts. However, these facts do not regard the distribution of propensities but refer exclusively to the causal efficacy of propensities in generating distributions. It is an open question to what extent such an inference is prohibited by the sort of arguments routinely employed against the principle of indifference. Bertrand style paradoxes, for instance, are prima facie inapplicable given the apparent absence of any causal relations in those geometrical examples. This is an interesting topic for further work.

[69] Note that failure of symmetry is the case in the 'causal' notation too. Thus it does not follow from $P_c(F \hookrightarrow B)$ that $P_c(B \hookrightarrow F)$. It does not, in fact, follow that F has any causes at all, never mind that B is one of them.

(i) $I_{t2}B_{t1}»P_{t1}(T_{t3}) = p > 0$.
(ii) $B_{t1}»P_{t1}\ (I_{t2}) = q$, where $1 > q > 0$.
(iii) $\neg I_{t2}\ B_{t1}»P_{t1}\ (T_{t3}) = 0$.

These equations represent the probabilities displayed by propensities and their relations. Since these probabilities are absolute, Bayes Theorem has no significant application. It is impossible to derive from these conditions a violation of Bayes Theorem whether in conjunction with a conditional independence principle such as CI – or any other of the principles discussed such as the zero influence or the fixity principle.[70]

1.8 Causal and Dispositional Presuppositions in Physics

The overall outlook of the book is decidedly in favour of dynamical, causal, or dispositional presuppositions underlying the practice of probabilistic modeling in science. The authors find that probabilistic modeling often carries an implicit or explicit commitment to such notions. When it does not implicitly or explicitly carry such a commitment, it often needs to be supplemented with some inferential rules that can be grounded only upon such notions. Thus, transition probabilities express dynamical processes; the selection of probabilistic hypotheses often requires information regarding the physical properties of the systems described; and the explanation of equilibrium in statistical mechanics requires essential reference to the dynamical character of statistical laws. Causal hypotheses and causal reasoning are required to understand statistical inference in quantum correlation phenomena; such causal hypotheses may imply some temporal orientation on pain of causal paradoxes or loops. On the other hand, a proper analysis of these questions requires philosophers to come to grips and apply the latest techniques in the field of causal inference, including the latest versions of the principle of common cause and the causal Markov condition. Finally, quantum systems are likely endowed with dispositional properties that get displayed under the appropriate circumstances as the characteristic probability distributions provided by Born's rule.

In this introductory chapter I argued that these diverse presuppositions are interlinked in a variety of interesting ways. For instance, transition probabilities must be understood as the probabilities of dynamical changes of state, and often express a system's dispositional properties. The manifestation of propensities may be understood as a kind of causal relation between the possession conditions and the manifestation outcomes. Statistical inference from frequencies to probabilities

[70] The same conclusion follows in accordance to the 'causal' notation. Humphreys conditions would be formalised as follows: (i) $P_{t1}(I_{t2}B_{t1} \hookrightarrow T_{t3}) = \mathrm{p}$; (ii) $P_{t1}(B_{t1} \hookrightarrow I_{t2}) = q$, where $1 > q > 0$; (iii) $P_{t1}(\neg I_{t2}B_{t1} \hookrightarrow T_{t3}) = 0$. Since Bayes Theorem has no application, no contradiction can ensue.

in quantum mechanics often requires causal hypotheses which are extremely sensitive to the particular interpretation of quantum mechanics employed in deriving those models. Etc. Every single one of these connections opens up a host of interesting philosophical problems and issues. The book demonstrates that work in the foundations of physics calls for deep and sustained philosophical reflection on such issues.

References

Arntzenius, F. (1995), Indeterminism and the direction of time, *Topoi* 14, 67–81.
Bird, A. (2010), Causation and the manifestation of powers, In A. Marmodoro (ed.), *Powers: Their Manifestation and Grounding,* London: Routledge, pp. 160–168.
Black, M. (1956), Why cannot an effect precede a cause?, *Analysis* 16(3), 49–58.
Butterfield, J. (2007), Stochastic Einstein locality revisited, *British Journal for the Philosophy of Science* 58, 805–867.
Cartwright, N. (1983), *How the Laws of Physics Lie*, Oxford: Oxford University Press.
Doob, J. L. (1953), *Stochastic Processes*. New York, NY: Wiley.
Dummett, M. (1964), Bringing about the past, *The Philosophical Review* 73(3), 338–359.
Earman, J. and Rédei, M. (1996), Why Ergodic theory does not explain the success of equilibrium statistical mechanics, *British Journal for the Philosophy of Science* 47, 63–78.
Fetzer, J. (1981), *Scientific Knowledge: Causality, Explanation and Corroboration.* Dordrecht: Reidel.
Fetzer, J. (1988), Probabilistic metaphysics, In J. Fetzer (ed.), *Probability and Causality*, Dordrecht: Reidel, pp. 109–131.
Frigg, R. and Hoefer, C. (2007), Probability in GRW theory, *Studies in History and Philosophy of Modern Physics* 38, 371–389.
Gillies, D. (1990), Bayesianism versus falsificationism, *Ratio* III, 82–98.
Gillies, D. (2000a), *Philosophical Theories of Probability.* New York, NY: Routledge.
Gillies, D. (2000b), Varieties of propensity, *British Journal for the Philosophy of Science* 51, 807–835.
Guerra, I. (2009), *Quantum Conditional Probability: Implications for Conceptual Change in Science*, PhD Thesis, Complutense University of Madrid.
Hajek, A. (2004), What conditional probability could not be, *Synthese* 137(3), 273–323.
Hall, N. (2004), Two concepts of causation, In J. Collins, N. Hall, and L. Paul (eds.), *Causation and Counterfactuals*, Cambridge, MA: MIT, pp. 181–204.
Hausman, D. (1999), *Causal Asymmetries*. Cambridge: Cambridge University Press.
Hausman, D. and Woodward, J. (1999), Independence, invariance and the causal Markov condition, *British Journal for the Philosophy of Science* 50, 521–583.
Halmos, P. (1974/1950), *Measure Theory*, New York, NY: Springer.
Healey, R. (1992), Causation, robustness, and EPR, *Philosophy of Science 59*, 282–292.
Howson, C. and P. Urbach (1993), *Scientific Reasoning: The Bayesian Approach*, (2nd Edition,) La Salle, IL: Open Court.
Humphreys, P. (2004), Some considerations on conditional chances, *British Journal for the Philosophy of Science* 55, 667–680.
Humphreys, P. (1985), Why propensities can not be probabilities, *The Philosophical Review* 94(4) 557–570.
Keynes, J. M. (1921/1963), *Treatise on Probability*. London: MacMillan.
Lewis, D. (1997), Finkish dispositions, *The Philosophical Quarterly* 47(187), 143–158.
Maudlin, T. (1995), *Quantum Non-locality and Relativity*, Malden, MA: Blackwell.

McCurdy, C. (1996), Humphrey's paradox and the interpretation of inverse conditional probabilities, *Synthese* 108(1), 105–125.

Mellor, H. (1971), *The Matter of Chance*. Cambridge: Cambridge University Press.

Mellor, H. (1995), *The Facts of Causation*. London: Routledge.

Miller, D. (1994), *Critical Rationalism: A Restatement and Defence*. Chicago, IL: Open Court.

Milne, P. (1986), Can there be a realist single-case interpretation of probability?, *Erkenntnis* 25(2), 129–132.

Pagonis, C. and Clifton, R. (1995), Unremarkable contextualism: Dispositions in the Bohm theory, *Foundations of Physics* 25(2) 281–296.

Penrose, R. (1989), *The Emperor's New Mind*. Oxford: Oxford University Press.

Poincaré, H. (1902), *La Science et L'Hypothèse*. Flammarion. Translated as *Science and Hypothesis*. 1952. New York, NY: Dover.

Popper, K. (1959), The propensity interpretation of probability, *British Journal for the Philosophy of Science* 10, 25–42.

Reichenbach, H. (1951), *The Rise of Scientific Philosophy*. Los Angeles, CA: University of California Press.

Reichenbach, H. (1971/1949), *Theory of Probability*, Berkeley, CA: University of California Press.

Salmon, W. (1979), Propensities: A discussion review, *Erkenntnis* 14, 183–216.

Salmon, W. (1991), Reichenbach's vindication of induction, *Erkenntnis* 35(1/3), 99–122.

San Pedro, I. (2007), *Reichenbach's Common Cause Principle and Quantum Correlations*, PhD thesis, University of the Basque Country.

Sklar, L. (1993), *Physics and Chance: Philosophical Issues in the Foundations of Statistical Mechanics*. Cambridge: Cambridge University Press.

Sober, E. (2005), Evolutionary theory and the evolution of macroprobabilities, In E. Eells and J. Fetzer (eds.), *Probability in Science*. La Salle, IL: Open Court.

Steel, D. (2005), Indeterminism and the causal Markov condition, *British Journal for the Philosophy of Science* 56, 3–26.

Strevens, M. (1998), Inferring probabilities from symmetries, *Nous* 32(2), 231–246.

Suárez, M. (2004), Quantum selections, propensities and the problem of measurement, *British Journal for the Philosophy of Science* 55(2), 219–255.

Suárez, M. (2007a), Quantum propensities, *Studies in the History and Philosophy of Modern Physics* 38, 418–438.

Suárez, M. (2007b), Causal inference in quantum mechanics: A reassessment, In F. Russo and J. Williamson (eds.), *Causality and Probability in the Sciences*, London: College, pp. 65–106.

Thompson, I. (1988), Real dispositions in the physical world, *British Journal for the Philosophy of Science* 39, 67–79.

Tumulka, R. (2007), Two unromantic pictures of quantum theory, *Journal of Physics A* 40, 3245–3273.

Van Fraassen, B. (1982), The charybdis of realism: Epistemological implications of Bell's inequality, *Synthese* 52, 25–38. Reprinted with corrections in Cushing and McMullin (eds.), pp. 97–113.

Van Fraassen, B. (1989), *Laws and Symmetry*. Oxford: Oxford University Press.

Van Fraassen, B. (1991), *Quantum Mechanics*. Oxford: Oxford University Press.

Part I
Probabilities

Chapter 2
Probability and Time Symmetry in Classical Markov Processes

Guido Bacciagaluppi

2.1 Introduction

The problem of the arrow of time in physics is that certain phenomena appear systematically to take place much more frequently than their time reversals, and this despite the fact that the fundamental laws are mostly believed to be fully time-symmetric, at least as long as they are deterministic. The two common general strategies for addressing this problem use, respectively, time-asymmetric laws or time-symmetric laws with special initial or boundary conditions.

It is less clear that such a problem exists also if one assumes indeterministic laws, since, intuitively, probabilities may be thought of as intrinsically time-directed. However, one should distinguish sharply between issues in the interpretation of probability, where these intuitions are strongest ('open future' versus 'fixed past'), and issues of formalism, which are the only ones involved in the description of the phenomena (can time-directed behaviour be described by formally time-symmetric laws?).

In this paper we propose to investigate, in the simple abstract setting of discrete Markov processes (more precisely, Markov processes with discrete state space and continuous time), whether and in what sense time-directed behaviour might indeed be compatible with time-symmetric probabilistic laws. We shall argue that time-symmetric stochastic processes, in a classical setting, are indeed quite capable of describing time-directed behaviour (or, when otherwise, that the remaining time asymmetry is quite benign). Thus, we suggest that a move to indeterministic laws is not likely to change the terms of the debate on the arrow of time. There will still be two fundamental alternatives for describing time-directed behaviour: adopting time-asymmetric laws, or adopting time-symmetric laws and suitable boundary conditions.[1]

G. Bacciagaluppi (✉)
Department of Philosophy, University of Aberdeen, Aberdeen, AB24 3UB, UK
e-mail: g.bacciagaluppi@abdn.ac.uk

[1] Note that Markov processes are indeed sometimes used in the context of thermodynamics to explain the thermodynamic arrow in terms of a 'probabilistic arrow of time'. Uffink (2007,

M. Suárez (ed.), *Probabilities, Causes and Propensities in Physics*, Synthese Library 347, DOI 10.1007/978-1-4020-9904-5_2,

On the basis of these results we then argue that considering the arrow of time in a probabilistic setting fails to justify a qualitative distinction in status between the future and the past. Of course, investigating notions of time symmetry or asymmetry at the level of the formalism can yield no normative conclusion about the interpretation of probability. However, we take it that it can provide useful guidelines for choosing or constructing a good interpretation, and in this sense we suggest that the common interpretation of probabilities as time-directed is unjustified.

Our results apply to classical probabilities. In a separate paper (Bacciagaluppi, 2007), we discuss the case of quantum probabilities as they appear in no-collapse approaches to quantum mechanics, specifically in the context of the decoherent histories formalism of quantum mechanics. The conclusions drawn in the two papers are quite different. Whereas in the classical case we shall argue against drawing such distinctions, in the quantum case we find that, albeit in a restricted sense, a qualitative distinction between forwards and backwards probabilities can be justified.

The structure of this paper is as follows: after reviewing some elementary theory in Section 2.2, we shall discuss notions of time symmetry for discrete Markov processes in Section 2.3. Then, in Section 2.4, we shall review reasons given for a time-asymmetric treatment of probabilities (Section 2.4.1); argue that, contrary to appearances, the relevant examples can very well be treated using processes that are time-symmetric or only harmlessly time-asymmetric (Section 2.4.2); and, finally, draw lessons for the interpretation of probability (Section 2.4.3).

2.2 A Few Essentials About Markov Processes

A stochastic process is defined to be a family of random variables, indexed by t, from a probability space Ω to a (common) state space S, which for the purposes of this paper we shall assume to be discrete (and sometimes finite):

$$X(t,.): \quad \Omega \to S\,. \tag{2.1}$$

It is, however, simpler to discuss a stochastic process in terms of joint distributions at finitely many times. Indeed, a classic theorem by Kolmogorov (1931) states that a stochastic process can be reconstructed from the collection of its *finite-dimensional distributions*, the n-fold joint distributions for all n:

$$p_{i_1 i_2 \ldots i_n}(t_1, t_2, \ldots, t_n)\,. \tag{2.2}$$

We shall also assume that the process is Markov, i.e. for any $t_1 < t_2 < \ldots < t_j < t_{j+1} < \ldots < t_n$,

Section 7) has independently criticised such attempts in a way that is very close to the ideas expressed in this paper.

$$p_{i_{j+1}\ldots i_n|i_1\ldots i_j}(t_{j+1},\ldots,t_n|t_1,\ldots,t_j) = p_{i_{j+1}\ldots i_n|i_j}(t_{j+1},\ldots,t_n|t_j)\,, \tag{2.3}$$

i.e.

$$\frac{p_{i_1\ldots i_n}(t_1,\ldots,t_n)}{p_{i_1\ldots i_j}(t_1,\ldots,t_j)} = \frac{p_{i_j\ldots i_n}(t_1,\ldots,t_n)}{p_{i_j}(t_n)}\,. \tag{2.4}$$

The finite-dimensional distributions of a Markov process can be reconstructed from its two-dimensional distributions,

$$p_{ij}(t,s)\,, \tag{2.5}$$

as is easily shown by induction. It should also be noted that the Markov condition is only apparently time-directed. Indeed, (2.4) is equivalent to

$$\frac{p_{i_1\ldots i_n}(t_1,\ldots,t_n)}{p_{i_j\ldots i_n}(t_j,\ldots,t_n)} = \frac{p_{i_1\ldots i_j}(t_1,\ldots,t_j)}{p_{i_j}(t_j)}\,, \tag{2.6}$$

i.e.

$$p_{i_1\ldots i_{j-1}|i_j\ldots i_n}(t_1,\ldots,t_{j-1}|t_j,\ldots,t_n) = p_{i_1\ldots i_{j-1}|i_j}(t_1,\ldots,t_{j-1}|t_j)\,, \tag{2.7}$$

so that the Markov condition is itself still perfectly time-symmetric.

Now we can introduce (two-time) transition probabilities. That is, for $t > s$ we define:

$$p_{i|j}(t|s) := \frac{p_{ij}(t,s)}{p_j(s)} \tag{2.8}$$

(*forwards* transition probabilities), and

$$p_{i|j}(s|t) := \frac{p_{ij}(s,t)}{p_j(t)} = \frac{p_{ji}(t,s)}{p_j(t)} \tag{2.9}$$

(*backwards* transition probabilities).

Using the forwards transition probabilities we can express the time evolution of the single-time distributions as

$$p_i(t) = \sum_j p_{i|j}(t|s)p_j(s)\,, \tag{2.10}$$

which we can also write in more compact form as

$$\mathbf{p}(t) = P(t|s)\mathbf{p}(s)\,. \tag{2.11}$$

$P(t|s)$ is called the transition matrix, mapping the probability vector $\mathbf{p}(s)$ into $\mathbf{p}(t)$. The matrix $P(t|s)$ is a so-called stochastic matrix, i.e. all elements of $P(t|s)$ are between 0 and 1, and each column of $P(t|s)$ sums to 1.

Similarly, we have the time-reversed analogues of (2.10) and (2.11):

$$p_i(s) = \sum_j p_{i|j}(s|t)p_j(t) , \tag{2.12}$$

and

$$\mathbf{p}(s) = P(s|t)\mathbf{p}(t) . \tag{2.13}$$

Note that the backwards transition matrix $P(s|t)$ is not in general the inverse matrix $P(t|s)^{-1}$, as can be seen easily by noting that the former is always well-defined, via (2.9), but the latter is not: e.g. if for given t and s,

$$P(t|s) = \begin{pmatrix} 1-\varepsilon & \alpha \\ \varepsilon & 1-\alpha \end{pmatrix} , \tag{2.14}$$

invertibility rules out the case $\alpha = 1-\varepsilon$.

The intuitive reason for this discrepancy is that, given (2.8) and (2.9), $\mathbf{p}(s)$ is not in general specifiable independently of both $P(t|s)$ and $P(s|t)$. Therefore, the condition that for all s and t,

$$\mathbf{p}(s) = P(s|t)P(t|s)\mathbf{p}(s) , \tag{2.15}$$

does not imply

$$P(s|t)P(t|s) = \mathbf{1} , \tag{2.16}$$

because $\mathbf{p}(s)$ in (2.15) is not arbitrary.

Now, let us take two possibly different initial distributions and evolve them both in time using the same (forwards) transition probabilities. It is then elementary to show that

$$\begin{aligned} \sum_i |p_i(t) - q_i(t)| &= \sum_i \Big| \sum_j p_{i|j}(t|s)p_j(s) - \sum_j p_{i|j}(t|s)q_j(s) \Big| \\ &\leq \sum_i \sum_j |p_{i|j}(t|s)||p_j(s) - q_j(s)| \\ &= \sum_j |p_j(s) - q_j(s)| . \end{aligned} \tag{2.17}$$

It follows that $\sum_i |p_i(t) - q_i(t)|$ converges to some positive number, not necessarily zero. Under suitable conditions, in particular if there are 'enough' transitions, one can hope to strengthen this result to

$$\lim_{t\to\infty} \sum_j |p_j(t) - q_j(t)| = 0 \, , \tag{2.18}$$

i.e. any two distributions would converge asymptotically. Under appropriate conditions, there would even be convergence of any initial distribution towards a unique (time-independent) limit distribution.

'Limit theorems', or 'ergodic theorems' for discrete Markov processes describe precisely the asymptotic properties of processes with a given set of (forwards) transition probabilities, in particular the circumstances under which such processes converge to a limit (uniquely or non-uniquely), and the relevant notion and corresponding speed of convergence. Analogous results hold, of course, if one fixes the set of backwards transition probabilities.[2]

Let us define state j to be a *consequent* of state i, if for all times s with $p_i(s) \neq 0$ there is a $t > s$ such that $p_{j|i}(t|s) \neq 0$. A state i is *transient* iff there is a state j that is a consequent of i, but such that i is not a consequent of j. The relation of consequence defines equivalence classes on the non-transient states (so-called ergodic classes).

In the case of finitely many states a sufficient condition for the existence of an (invariant) *limit distribution* for $t \to \infty$ is that the (forwards) transition probabilities are time-translation invariant – synonyms: if the (forwards) transition probabilities are stationary, or if the process is (forwards) *homogeneous*. The limit distribution decomposes into a convex combination of the limit distributions on each ergodic class, while the probability of any transient state converges to zero (see e.g. Doob, 1953, Chapter VI). In the next section and the appendix, we shall need to refer to the case of discrete time, where the result is slightly weaker, since in some ergodic classes one may have cyclic behaviour rather than convergence (see e.g. Doob, 1953, Chapter V).

Returning to the case of continuous time, if one has denumerably many states, homogeneity is not sufficient for the existence of limit distributions, and additional conditions can be used. On the other hand, homogeneity is not a necessary condition either for the existence of limit distributions, and alternative sufficient conditions are known. As an example, take a two-state process that has equal probabilities for jumping from 0 to 1 as from 1 to 0 in any given time interval, and such that in unit time these probabilities are always larger than a given δ. Then one can easily see that the process will converge exponentially fast towards the invariant distribution $p_0(t) = p_1(t) = 1/2$, whether or not the transition probabilities are time translation invariant. Similarly, there are conditions that ensure asymptotic convergence when the process has no invariant limit distribution (see e.g. Hajnal, 1958).

If the single-time distribution $p_i(t)$ of a process is invariant, it is itself equal to the limit distribution of the process, and we shall say that the process is in *equilibrium*.

[2] For a good introduction to the complex theme of ergodic theory in the deterministic case, see Uffink (2007, Section 6).

(We shall occasionally also refer to an invariant distribution as an equilibrium distribution.) Note that if a process is in equilibrium, it has no transient states. Finally, a process that is both homogeneous and in equilibrium is said to be *stationary*.

2.3 Definitions of Time Symmetry

The framework we have introduced above is quite austere, and we must realise that, at least for the purpose of investigating time symmetry, it has its limitations. For instance, we do not have enough structure to define the time reverse of a state (there is no analogue of inversion of momenta in Newtonian mechanics, for instance). More importantly, we are not going to be able to identify and abstract from *systematic* components of the process, in particular components that may appear time-asymmetric but might in fact be generated by some time-symmetric law (think of a diffusion process taking place in a Newtonian gravitational field). Nevertheless, the insights we shall gain will be enough to discuss how typical examples of time-directed behaviour can be described in terms of time-symmetric processes, and to provide clues as to the time-symmetric or time-asymmetric status of the probabilities with respect to their interpretation.

It is natural to consider transition probabilities as what defines the dynamics of a system described by a Markov process. This in turn suggests to consider the following condition as a possible condition for a time-symmetric process: that forwards and backwards transition probabilities coincide, i.e. (for all i, j, t and s)

$$p_{i|j}(t|s) = p_{i|j}(s|t) \tag{2.19}$$

or (for all t, s)

$$P(t|s) = P(s|t) \,. \tag{2.20}$$

This is by analogy to the condition, familiar from the deterministic case, that the backwards equations of motion have the same form as the forwards equations.

In the literature on Markov processes, however, the usual condition of time symmetry is the so-called condition of *detailed balance*[3]:

$$p_{i|j}(t|s)p_j(s) = p_{j|i}(t|s)p_i(s) \,. \tag{2.21}$$

The meaning of detailed balance can be readily seen using the notion of *probability current*, i.e. the net probability flow from a state j to a state i between s and t:

$$j_{ij}(t,s) := p_{i|j}(t|s)p_j(s) - p_{j|i}(t|s)p_i(s) \,. \tag{2.22}$$

[3] My thanks to Werner Ehm for discussions about this notion.

Detailed balance simply means that there are no probability currents.

Our main purpose in this section will be to see that the two conditions (2.19) and (2.21) are equivalent, at least under certain conditions. Note that (2.21) is often formulated under the additional presupposition that the process is stationary, but we shall *not* make this assumption.

Symmetry of the transition probabilities obviously involves both forwards and backwards transition probabilities, while detailed balance explicitly involves only the forwards transition probabilities. On the other hand,

$$\begin{aligned} j_{ij}(t,s) &= p_{ij}(t,s) - p_{ji}(t,s) \\ &= p_{ij}(t,s) - p_{ij}(s,t)\ , \end{aligned} \tag{2.23}$$

therefore detailed balance is equivalent to symmetry of the two-time distributions,

$$p_{ij}(t,s) = p_{ij}(s,t)\ , \tag{2.24}$$

which is clearly a time symmetry condition.

Now, (2.24) and hence detailed balance are easily seen to be a sufficient condition for both equilibrium and the symmetry of transition probabilities (2.19). Indeed, performing a sum over i in (2.24) yields invariance of the single-time distributions:

$$p_j(s) = p_j(t)\ , \tag{2.25}$$

i.e. equilibrium. But from (2.24) and (2.25) we obtain:

$$p_{i|j}(t|s) = \frac{p_{ij}(t,s)}{p_j(s)} = \frac{p_{ij}(s,t)}{p_j(s)} = \frac{p_{ij}(s,t)}{p_j(t)} = p_{i|j}(s|t)\ , \tag{2.26}$$

i.e. (2.19), as long as either side is well-defined.

Notice that, conversely, (2.19) and equilibrium together imply (2.24) and therefore detailed balance. Indeed,

$$\begin{aligned} p_{ij}(t,s) &= p_{i|j}(t|s)p_j(s) \\ &= p_{i|j}(s|t)p_j(s) \\ &= p_{i|j}(s|t)p_j(t) = p_{ij}(s,t)\ . \end{aligned} \tag{2.27}$$

Instead, equilibrium on its own does not imply detailed balance (and therefore not symmetry of transition probabilities either). Indeed, take a three-state system with

$$P(t|s) = \begin{pmatrix} 1/3 & 1/6 & 1/2 \\ 1/2 & 1/3 & 1/6 \\ 1/6 & 1/2 & 1/3 \end{pmatrix}^{t-s}\ . \tag{2.28}$$

We have in particular that

$$\begin{aligned} p_{i|i}(t+1|t) &= 1/3\ , \\ p_{i+1|i}(t+1|t) &= 1/2\ , \\ p_{i-1|i}(t+1|t) &= 1/6 \end{aligned} \tag{2.29}$$

(where $i+1$ and $i-1$ are to be read as addition mod 3). The equilibrium distribution for this process is $p_i(t) = 1/3$, but there is clearly a non-zero current $0 \to 1 \to 2 \to 0$, and detailed balance fails.

This example is generic in the sense that the only way to have currents in equilibrium, whether for finite or denumerable state space, is to have a circular current, i.e. a current along a closed chain of states with at least three members,[4]

$$i \to j \to k \to i\ . \tag{2.30}$$

Therefore equilibrium and zero circular currents together are equivalent to detailed balance. In the special case of a two-state system, there are no three-element chains, and equilibrium is in fact equivalent to detailed balance.

Simple examples suggest that, under suitable conditions, symmetry of the transition probabilities (2.19) might in fact imply equilibrium and therefore (by (2.27)) be equivalent to detailed balance. Take a homogeneous two-state process with (forwards) transition matrix

$$P(t|s) = \begin{pmatrix} 1-\alpha & \varepsilon \\ \alpha & 1-\varepsilon \end{pmatrix}^{t-s}\ . \tag{2.31}$$

If we take $\alpha \neq 0$ and ε arbitrary, this is a toy model of decay (with non-zero probability α of decay in unit time), with or without re-excitation (depending on whether $\varepsilon \neq 0$ of $\varepsilon = 0$).

Imposing (2.19) in this example leads to

$$p_0(t) = \frac{\alpha}{\alpha+\varepsilon}\ , \qquad p_1(t) = \frac{\varepsilon}{\alpha+\varepsilon} \tag{2.32}$$

for all t, i.e. the single-time distribution is fully constrained to be the equilibrium distribution of the process (and the process is stationary).

[4] In the case of denumerable state space, assume there are non-zero currents in equilibrium but no circular currents. Let us say that, between s and t, state 0 gains probability ε from states $1, \ldots, i_1$ (distinct from 0). Obviously, $\sum_{i=1}^{i_1} p_i(s) \geq \varepsilon$. In the same time interval, the states $1, \ldots, i_1$ must gain probability at least ε from some states $i_1+1, \ldots, i_2$ (all distinct from $0, \ldots, i_1$), and $\sum_{i=i_1+1}^{i_2} p_i(s) \geq \varepsilon$. Therefore $\sum_{i=1}^{i_2} p_i(s) \geq 2\varepsilon$. Repeat the argument until $\sum_{i=1}^{i_n} p_i(s) \geq n\varepsilon > 1$, which is impossible.

Indeed, for arbitrary t and s define α_{t-s} and ε_{t-s} such that

$$P(t|s) = \begin{pmatrix} 1-\alpha_{t-s} & \varepsilon_{t-s} \\ \alpha_{t-s} & 1-\varepsilon_{t-s} \end{pmatrix} . \tag{2.33}$$

Then, from

$$p_{0|1}(t|s) = p_{0|1}(s|t) = \alpha_{t-s} \tag{2.34}$$

and

$$p_{1|0}(t|s) = p_{1|0}(s|t) = \varepsilon_{t-s} , \tag{2.35}$$

one obtains

$$\varepsilon_{t-s} p_0(s) = \alpha_{t-s} p_1(t) , \tag{2.36}$$

$$\varepsilon_{t-s} p_0(t) = \alpha_{t-s} p_1(s) . \tag{2.37}$$

Thus, since there are only two states,

$$\varepsilon_{t-s} p_0(s) = \alpha_{t-s}(1-p_0(t)) , \tag{2.38}$$

$$\varepsilon_{t-s} p_0(t) = \alpha_{t-s}(1-p_0(s)) , \tag{2.39}$$

whence

$$p_0(s) = p_0(t) = \frac{\alpha_{t-s}}{\alpha_{t-s}+\varepsilon_{t-s}} . \tag{2.40}$$

Therefore, $p_0(t)$ is constant, since t and s are arbitrary. Finally, substituting $s = t-1$ in (2.40), we have

$$p_0(t) = \frac{\alpha}{\alpha+\varepsilon} , \tag{2.41}$$

and the claim follows.

We now ask for conditions under which symmetry of the transition probabilities strictly implies equilibrium and thus becomes equivalent to detailed balance.

Let us first specialise to homogeneous Markov processes, i.e. the transition probabilities are time-translation invariant. Then equilibrium follows very easily. (Incidentally, note that a forwards or backwards homogeneous process satisfying (2.19) will be both forwards and backward homogeneous.) Indeed, for all t, s,

$$\mathbf{p}(t+s) = P(t+s|t+s/2)P(t+s/2|t)\mathbf{p}(t) . \tag{2.42}$$

By translation invariance,

$$\mathbf{p}(t+s) = P(t+s/2|t)P(t+s/2|t)\mathbf{p}(t) , \tag{2.43}$$

and by symmetry

$$\mathbf{p}(t+s) = P(t|t+s/2)P(t+s/2|t)\mathbf{p}(t)\,, \tag{2.44}$$

but by definition also

$$\mathbf{p}(t) = P(t|t+s/2)P(t+s/2|t)\mathbf{p}(t)\,. \tag{2.45}$$

Therefore,

$$\mathbf{p}(t+s) = \mathbf{p}(t) \tag{2.46}$$

for all t, s, i.e. the process is in equilibrium.

If we relax the assumption that the process is homogeneous, it is still a theorem that (2.19) implies equilibrium, at least under the further assumptions that (a) the state space has finite size n, and (b) for all i, j and s the transition probabilities $p_{i|j}(t|s)$ are continuous in t. (The appendix provides an elementary derivation of this result from the ergodic theorem for discrete time.) Thus, under the appropriate conditions, the two definitions of time symmetry (2.19) and (2.21) are indeed equivalent.

2.4 Probability and Time Symmetry

2.4.1 Arguments for Asymmetry

Imagine a world in which fundamental laws are probabilistic. Imagine further that this world contains an arrow of time, that is, typical examples of time-directed behaviour, and that this behaviour is investigated by observers who can set up experiments under controlled initial conditions (but not final ones). That is, like ourselves, observers in this world are subject to some macroscopic arrow of time that may or may not be related to the time-directed behaviour under scrutiny. Finally, let this be a classical world; in particular, assume that gaining knowledge of the state i of a system at a certain time (in particular with regard to alternative initial conditions) can be done in principle without disturbing the system, so that we can still consider it as governed by the same stochastic process.

It will be tempting to interpret the probabilistic laws in this world as intrinsically time-directed. Such laws will specify objective probabilities for events in the future given events in the present (if the laws are Markovian), while probabilities for past events will be regarded as merely epistemic. The underlying intuition is that, under indeterminism, the future is genuinely ‘open’, while the past, while perhaps unknown, is ‘fixed’.

Formally, however, there is a very good argument for saying that in a classical stochastic process there is no distinction between future and past: a classical stochastic process is defined as a probability measure over a space of trajectories, so the formal definition is completely time-symmetric. Transition probabilities towards the future can be obtained by conditionalising on the past; but, equally, transition probabilities towards the past can be obtained by conditionalising on the future. Individual trajectories may exhibit time asymmetry, and there may be a quantitative asymmetry between forwards and backwards transition probabilities, but at least as long as the latter are not all 0 or 1, quantitative differences fall short of justifying a notion of fixed past.

On the other hand, at least in a world as the one sketched above, there are ways of arguing for *qualitative* formal differences between forwards and backwards transitions probabilities that could suggest also a different interpretational status for the two kinds of probabilities:

(A) In a probabilistic setting one has good ergodic behaviour, in particular, if time translation invariance of the transition probabilities holds (assuming finiteness of the state space or other suitable conditions), one will have a tendency for a stochastic process to equilibrate in time, regardless of the initial distribution. Such an arrow of time would thus appear to be very deeply seated in the use of probabilistic concepts. A related argument is that in the homogeneous case (and, as we have mentioned, more generally) the symmetry of transition probabilities implies equilibrium, and thus rules out not only any equilibration process but any time development of the probabilities whatsoever (Sober, 1993).

(B) Another interesting argument for asymmetry between forwards and backwards probabilities runs along the following lines. Take the simple model of exponential decay (2.31), with probability α of decay from the excited state 1 to the ground state 0 in unit time, and starting with all 'atoms' excited, i.e. $p_1(0) = 1$. We have:

$$p_{0|1}(t+1|t) = \alpha \ , \tag{2.47}$$

for all t, but:

$$p_{0|1}(t|t+1) = \begin{cases} \to \alpha \ \text{for}\, t \to \infty \ , \\ \to 0 \ \ \text{for}\, t \to 0 \ . \end{cases} \tag{2.48}$$

In this example, the forwards transition probabilities are time translation invariant, but the backwards transition probabilities are not. This difference has been used to argue that forwards transition probabilities are indeed law-like, while backwards transition probabilities are epistemic (Arntzenius, 1995).

(C) Finally, backwards transition probabilities are not invariant across experiments when one varies the initial distribution. One can thus argue that if the initial distribution of the process is an epistemic distribution over contingent initial states, then the backwards transition probabilities cannot be law-like, or not entirely law-like, because they depend on the epistemic initial distribution. A related argument is that, in general, at most one set of transition probabilities can be law-like, otherwise also the single-time probabilities will be, so that it appears that initial conditions cannot be freely chosen (Watanabe, 1965, Section 5).

These arguments infer from typical time-directed behaviour to formal qualitative differences in the transition probabilities. It is this type of inference that we shall question below. Without a qualitative difference in the formalism, however, we take it that there is no reason to deny the same interpretational status to both sets of transition probabilities alike.

2.4.2 Time-Directed Behaviour and Time-Symmetric Probabilities

The situation of convergence to equilibrium – indeed, the simple example of decay – can be used to exemplify at once all three purported differences between forwards and backwards transition probabilities and, at least at first sight, seems thus to be totally intractable in terms of symmetric processes. Indeed, (A) we have seen that time symmetry of transition probabilities implies equilibrium of the process ((2.32) above). (B) We have also seen the lack of time translation invariance for the backwards transition probabilities ((2.47) and (2.48) above). Finally, (C) if we start with all 'atoms' in the ground state, i.e. $p_0(0) = 1$, we obtain:

$$p_{0|1}(t+1|t) = \alpha \, , \tag{2.49}$$

for all t, but:

$$p_{0|1}(t|t+1) = \begin{cases} \to \alpha \ \text{for}\ t \to \infty \, , \\ \to 1 \ \ \text{for}\ t \to 0 \, . \end{cases} \tag{2.50}$$

Thus, a different choice of initial condition will indeed lead to different backwards transition probabilities.

The question we wish to raise is: can we indeed infer that there are such differences in the transition probabilities from time asymmetries of the phenomena, i.e. from the time-directed behaviour of *samples*?

Obviously, one must distinguish between the transition *probabilities* of the process and the transition *frequencies* in any actual sample. Observed behaviour, in particular time-directed behaviour, will always be defined in terms of frequencies, and in order to conclude from frequencies to probabilities, we have to ensure that the sample is unbiased. Indeed, suppose that we bias the sample by performing a post-selection of the final ensemble. Then in general we shall influence the *forwards* transition frequencies, in particular destroying their time translation invariance.

If we assume that the process has a limit distribution for $t \to \infty$, a simple criterion to make sure that the final ensemble is sufficiently unbiased is to check whether the distribution of the sample is at least approximately time-independent, i.e. whether or not the sample has been prevented from equilibrating or has subsequently departed from equilibrium for any reason (a statistical fluctuation, a final cause, or an uncooperative lab assistant sneakily post-selecting the ensemble). Only then will the observed transition frequencies be taken as evidence for any law-like forwards transition probabilities.

Estimating backwards transition probabilities should proceed analogously. If we assume that the process has a limit distribution for $t \to -\infty$, then we cannot accept a sample as unbiased unless the initial distribution of the sample is in fact a limit distribution of the process. And if we assume that there is no limit distribution for $t \to -\infty$, then we are begging the question, because we have introduced a qualitative difference between forwards and backwards transition probabilities by hand.

Thus, while time-symmetric transition probabilities imply invariant equilibrium, a sample appropriate for estimating both forwards and backwards transition probabilities will be in equilibrium anyway. But now, the above criticisms all rely implicitly or explicitly on considering samples other than in equilibrium. Indeed, (A) uses convergence towards equilibrium (or the possibility of time-dependent distributions), so cannot be applied if the sample is in equilibrium already; (B) also requires the use of non-equilibrium ensembles because, trivially, forwards homogeneity and equilibrium imply backwards homogeneity; finally, (C) relies on considering alternative initial conditions, some of which will be non-equilibrium distributions.[5] The idea that convergence to equilibrium could be formally described using a time-symmetric stochastic process, plus a constraint on the initial distribution of the specific sample, is thus perfectly viable.

A case apart is provided by samples exhibiting what appear to be transient states. In the example, this is when we observe decay from the excited state to the ground state but no re-excitation, which is a case of particularly marked time-directed behaviour. At first sight, one might think that our argument above applies even in this case. Indeed, in order to have the forwards transition frequencies match the forwards transition probabilities, the sample must be totally decayed at the final time. By analogy, in order for the backwards transition frequencies to match the backwards transition probabilities, the sample must be totally decayed at the initial time (invariant distribution). But then, the samples exhibiting transience of the excited state are always biased for the purpose of estimating the backwards transition probabilities. There are two problems, however. Firstly, in a sample that is appropriate for estimating the transition probabilities in one direction of time, the transition frequencies in the opposite direction are partially ill-defined: thus, there are no samples appropriate for estimating both sets of transition probabilities (if such there be). Secondly and crucially, a non-zero initial frequency for excited states forces the backwards transition frequencies to be non-zero when the corresponding transition probabilities (assuming symmetry) should be zero, and thus is clearly not an allowable constraint.

A better way of treating samples with transient states will be to maintain that there is in fact a small but non-zero probability of re-excitation, which is a move analogous to standard reasoning in the deterministic case. (The fact that

---

[5] Conditionalising on two different equilibrium distributions (if there are several ergodic classes) will not yield different backwards transition frequencies, because the transition frequencies are fixed separately in each ergodic class.

Julius Caesar was alive and is now dead is not conclusive evidence against the time symmetry of classical mechanics.)

Recapitulating the above, we have seen that we can describe convergence to equilibrium using the transition probabilities of a stochastic process in equilibrium plus an assumption about special initial conditions (with an additional assumption in the case of apparently transient states). Therefore, the qualitative formal distinctions between forwards and backwards transition probabilities used as premises in the criticisms considered above are unwarranted.

We have not shown, however, that convergence to equilibrium can always be described using time-symmetric transition probabilities, because, other than in the two-state case, equilibrium is a necessary but not a sufficient condition for time symmetry. Indeed, there are also examples in which *circular currents* are called for: the transition matrices (2.28) above are stationary, so any initial distribution will converge to equilibrium, but in equilibrium there is a circular current. Intuitively, the 'atom' has a ground state 0 and two excited states 1 and 2, and state 2 decays to 0 directly with much larger probability than via the intermediate state 1. Thus, the transition probabilities fail to be time-symmetric.[6]

The import of these asymmetric cases can perhaps be minimised. The asymmetry appears to be more benign than in the criticisms considered above (e.g. if the forwards transition probabilities are time translation invariant, so are the backwards transition probabilities). Indeed, it does not appear that this asymmetry could justify a *qualitative* distinction between forwards and backwards transition probabilities. Furthermore, as briefly mentioned at the beginning of Section 2.3, the framework we have adopted allows us to describe these currents, but lacks any further structure that might explain them as determined perhaps by some underlying laws allowing a fuller analysis as regards time symmetry. Given such structure, the currents might turn out to be time-symmetric after all, in the sense that they would swap direction under time reversal of the underlying law.

A related example is provided by the inhomogeneous processes used in Nelson's (1966) approach to quantum mechanics. Without going into details, Nelson's approach is somewhat similar to the pilot-wave theory of de Broglie and Bohm, in that it takes quantum systems (in standard non-relativistic quantum mechanics) to be systems of point particles described in configuration space. Whereas de Broglie and Bohm take the velocity of the particles to be deterministically determined by the wave function of the system, Nelson postulates a stochastic process (a diffusion process) on the configuration space, and tries to impose conditions that would ensure that the process is determined in a certain way by the amplitude and phase of a complex function satisfying the Schrödinger equation. Whether or not Nelson's conditions achieve this, the process on configuration space definable through the wave function has as its current velocity the same velocity that arises in pilot-wave theory, which indeed changes sign with the time reversal of the Schrödinger equation. Thus, both time translation invariance and time symmetry, which are not apparent at

[6] My thanks to Iain Martel for making this point in conversation.

the level of the probabilities, are restored by the additional structure provided by the Schrödinger equation. Note that Nelson's approach can be adapted to the discrete case (Guerra and Marra, 1984). In this case the systematic component of the process is a probability current in the sense of (2.22), which again swaps sign under time reversal of the Schrödinger equation.[7]

While our above considerations apply only to processes that admit an invariant limit distribution, Nelson's processes generally have only an asymptotic distribution (also called equivariant), given by the usual quantum distribution $|\psi(\mathbf{x}, t)|^2$ (similarly in Guerra and Marra's approach). We thus see that our considerations can be generalised to interesting cases of asymptotic convergence. That is, one can describe asymptotic convergence using a process that is time-symmetric – in the sense that the only time asymmetry is given by a current that swaps sign under time reversal – plus special initial conditions.[8]

2.4.3 Interpretation of Probability

We have tried to characterise the time symmetry of a Markov process in terms of forwards and backwards transition probabilities. To characterise similarly the *interpretation* of probabilities means that forwards and backwards transition probabilities would have the *same* or a *different* status. In particular, one could say that the idea of an (objectively) 'open future' and 'fixed past' means that forwards transition probabilities are law-like *chances*, while backwards transition probabilities are merely *epistemic*.

To say that both forwards and backwards transition probabilities are law-like seems less intuitive, since the two sets of probabilities determine the possible single-time distributions of the process (even uniquely), so the latter would also have to be taken as law-like. But law-likeness of probability distributions does not mean that relative frequencies have to always match the given probabilities. As long as an ensemble is finite, a law-like probability is compatible with infinitely many actual distributions, and it makes sense to consider *constraints* on, for instance, initial distributions or final distributions alongside with the laws. Indeed, the situation is quite analogous to that in the deterministic case. Deterministic laws determine the

---

[7] A more detailed introduction to Nelson's approach, including an explicit discussion of time symmetry and the status of the transition probabilities, is given in Bacciagaluppi (2005). As Nelson's approach relates to de Broglie and Bohm's pilot-wave theory, so Guerra and Marra's discrete case relates to the stochastic versions of pilot-wave theory, known as 'beable' theories, defined by Bell (1984).

[8] Observation in these cases, however, is definitely not classical. If one includes observers in the description (by adding some appropriate quantum mechanical interaction), when they gain knowledge about the state of the process, thus narrowing their epistemic distribution over the states, they effectively modify the wave function of the system, thus effectively modifying also the transition probabilities of the process, *both* forwards and backwards. (Note that convergence behaviour would thus be altered if monitored.)

time development of a system given, for instance, some initial condition; but which trajectory a system will actually follow is a contingent matter. Similarly, stochastic laws (whether symmetric or not) can be said to determine, in an appropriate sense, the time development of a system; but a stochastic process is a probability measure over a space of trajectories, and which trajectory the system will actually follow is a contingent matter. If we have a finite ensemble of systems, it is still a contingent matter which trajectories they will follow, regardless of whether the laws are deterministic or stochastic. (And, in fact, if the stochastic laws are assumed to be fundamental, then there is ultimately only one system – the universe – and only one trajectory.) Thus, at least as long as we are not dealing with literally infinite ensembles, we can make the same distinction between law-like time development and contingent initial or final states, or distributions over states, in the case of both deterministic and stochastic laws, and this even if we assume that both forwards and backwards transition probabilities are law-like, despite the ensuing law-likeness of single-time distributions.[9]

We can imagine a stochastic world in which observed transition frequencies typically show not merely a quantitative but a qualitative difference between forwards and backwards transition frequencies, as in the examples in Section 2.4.1. However, our analysis in Section 2.4.2 shows that arguments from observed frequencies fail to establish an asymmetry between the corresponding probabilities: although ensembles that are not in equilibrium lead to distorted frequencies, neither the preponderance of non-equilibrium ensembles in such a world nor any conclusions drawn on the basis of these frequencies can be arguments against time-symmetric transition chances (and this despite the fact that equilibrium is a necessary condition for (2.19)). The only serious source of time asymmetry at the level of the formalism and therefore potential motivation for a time-asymmetric interpretation would seem to be the presence in some cases of circular currents, which indeed yield quantitatively asymmetric transition probabilities. However, circular currents yield no *qualitative* difference that could justify a different status for forwards and backwards transition probabilities. In particular, if the only difference between past and future is the presence of a current in one direction or another along a closed chain of states, it is difficult to see which of the two directions should correspond to an *open* ‘future’ as opposed to a *fixed* ‘past’. Thus, the possibility of an asymmetry in terms of circular currents does not seem to be of the kind that would justify a time-asymmetric interpretation of probability.

---

[9] The notion of a constraint is of course more intuitive when one is talking about a subsystem on which one performs experiments (as in thermodynamics or statistical mechanics when compressing a gas into a small volume), but it is meant to apply generally. As emphasised by the anonymous referee, in the case of a stochastic theory such constraints will not only be ‘special’ in some sense but they will be improbable in the sense specified by the process itself. The further question of whether and how the contingent trajectories (or distributions) should be explained thus acquires a new twist as compared to the deterministic case.

At least in the case of processes with an invariant limit distribution, our analysis suggests that both forwards and backwards transition probabilities can be considered law-like. Therefore, whatever approach to the foundations of probabilities one might take, a *time-symmetric interpretation of probabilities* appears to be a natural option in the context of classical Markov processes.

Acknowledgements The first version of this paper was written while I was an Alexander-von-Humboldt Fellow at the Institut für Grenzgebiete der Psychologie und Psychohygiene (IGPP), Freiburg i. Br. I wish to thank in particular Werner Ehm at IGPP, Iain Martel, then at the University of Konstanz, and David Miller at the Centre for Time, University of Sydney, for useful discussions and suggestions, as well as Mauricio Suárez for the kind invitation to contribute to this volume and an anonymous referee for interesting comments.

Appendix

We now prove that symmetry of the transition probabilities (2.19), together with the further assumptions that the state space is finite and that the transition probabilities are continuous, implies equilibrium of the process.

We proceed by induction on the size n of the state space. The case $n = 1$ is trivial. Assume that the result has been proved for all sizes $1 \leq m < n$. We now prove it for n by *reductio*.

Assume that the single-time distribution is not invariant, i.e.

$$\exists s \exists t \geq s, \quad \mathbf{p}(t) = P(t|s)\mathbf{p}(s) \neq \mathbf{p}(s) \,. \tag{2.51}$$

For the rest of the proof we now fix such an s.

Since we assume (2.19), i.e. $P(t|s) = P(s|t)$, we also have

$$\mathbf{p}(s) = P(t|s)\mathbf{p}(t) \,, \tag{2.52}$$

and therefore

$$P(t|s)^2\mathbf{p}(s) = \mathbf{p}(s) \qquad \text{and} \qquad P(t|s)^2\mathbf{p}(t) = \mathbf{p}(t) \,. \tag{2.53}$$

Now fix a time $t \geq s$ and consider the matrix $P := P(t|s)^2$. This is an $n \times n$ stochastic matrix that we can consider as the transition matrix of a homogeneous Markov process with discrete time. By (2.53), $\mathbf{p}(t)$ and $\mathbf{p}(s)$ are both invariant distributions for this Markov process, and by (2.51) they are different.

By the ergodic theorem for discrete-time Markov processes, existence of at least two different invariant distributions implies that there are at least two ergodic classes. Therefore (whether or not there are any transient states), P must have a block diagonal form

$$P = \begin{pmatrix} P' & \mathbf{0} \\ \mathbf{0} & P'' \end{pmatrix} \,, \tag{2.54}$$

where P' is an $m \times m$ matrix and P'' an $(n-m) \times (n-m)$ matrix, for some $0 < m < n$.

For fixed s, $P = P(t|s)^2$ depends on t, and so *a priori* could m; but in fact $m(t)$ is independent of t. Indeed, assume there is an $m \neq m(t)$ such that for all $\varepsilon > 0$ there is a t' with $|t - t'| < \varepsilon$ and $m(t') = m$. The matrix elements of $P = P(t|s)^2$, in particular the ones off the diagonal blocks, are continuous functions of the transition probabilities. Therefore, by the continuity of the transition probabilities, $P(t|s)^2$ must also have zeros off the same diagonal blocks, i.e. $m = m(t)$, contrary to assumption. Therefore, for each $m \neq m(t)$ there is an $\varepsilon(m) > 0$ such that for all t' with $|t - t'| < \varepsilon(m)$ we have $m(t') \neq m$. Taking the smallest of these finitely many $\varepsilon(m) > 0$, call it ε_0, it follows that $m(t') = m(t)$ for all t' in the open ε_0-neighbourhood around t. However, again by the continuity of the matrix elements, this ε_0-neighbourhood is also closed, and therefore it is the entire real line. Since t was arbitrary, $P(t|s)^2$ has the form (2.54) with the same m for all $t \geq s$.

We now focus on the matrix $P(t|s)$ itself rather than on $P(t|s)^2$. Assume that for some $t \geq s$ it has some element $p_{k|l}(t|s)$ outside of the $m \times m$ and $(n-m) \times (n-m)$ diagonal blocks. In order for $P(t|s)^2$ to have the given block diagonal form, several other elements of $P(t|s)$ have to be zero, in particular all elements in the k-th column of $P(t|s)$ that lie inside the corresponding diagonal block.

Since $P(t|s)$ is a stochastic matrix and every column sums to 1, it follows that already those elements of the k-th column that lie outside the diagonal blocks sum to 1, and therefore the sum of all elements in the diagonal blocks of $P(t|s)$, call it $d(t)$, is at most $n-1$, i.e.

$$d(t) = \sum_{i,j \leq m} p_{i|j}(t|s) + \sum_{i,j \geq m+1} p_{i|j}(t|s) \leq n - 1 \,, \tag{2.55}$$

for any $t \geq s$ such that $P(t|s)$ has some element outside of the diagonal blocks. Let t_0 be the infimum of such t. By continuity, we have also

$$d(t_0) \leq n - 1 \,. \tag{2.56}$$

Now, if $t_0 \neq s$, then for all $t < t_0$ we have that $d(t) = n$, but then by continuity $d(t_0) = n$, contradicting (2.56). If instead $t_0 = s$, since $P(s|s) = \mathbf{1}$, we again have $d(t_0) = n$, contradicting (2.56). For all $t \geq s$, thus, $P(t|s)$ has the same block diagonal form as $P(t|s)^2$ with fixed m.

But then, our original Markov process decomposes into two sub-processes, with state spaces of size m and $n-m$, respectively. If $\mathbf{p}(t) \neq \mathbf{p}(s)$ (assumption (2.51)), then the same must be true for at least one of the two sub-processes, but, by the inductive assumption, this is impossible. Therefore, (2.51) is false and

$$\forall s \forall t \geq s, \; \mathbf{p}(t) = \mathbf{p}(s) \,, \tag{2.57}$$

QED.

References

Arntzenius, F. (1995), Indeterminism and the direction of time, *Topoi* 14, 67–81.

Bacciagaluppi, G. (2005), A conceptual introduction to Nelson's mechanics. In R. Buccheri, M. Saniga and E. Avshalom (eds.), *Endophysics, Time, Quantum and the Subjective*, Singapore: World Scientific, pp. 367–388.

Bacciagaluppi, G. (2007), Probability, arrow of time and decoherence, *Studies in History and Philosophy of Modern Physics* 38, 439–456.

Bell, J. S. (1984), Beables for quantum field theory, CERN-TH. 4035/84. Reprinted in J. S. Bell, *Speakable and Unspeakable in Quantum Mechanics*, Cambridge: Cambridge University Press, pp. 173–180.

Doob, J. L. (1953), *Stochastic Processes*, New York, NY: Wiley.

Guerra, F. and Marra, R. (1984), Discrete stochastic variational principles and quantum mechanics, *Physical Review* D 29, 1647–1655.

Hajnal, J. (1958), Weak ergodicity in non-homogeneous Markov chains, *Proceedings of the Cambridge Philosophical Society* 54, 233–246.

Kolmogorov, A. (1931), Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung, *Mathematische Annalen* 104, 415–458.

Nelson, E. (1966), Derivation of the Schrödinger equation from Newtonian mechanics, *Physical Review* 150, 1079–1085.

Sober, E. (1993), Temporally oriented laws, *Synthese* 94, 171–189.

Uffink, J. (2007), Compendium of the foundations of classical statistical physics. In J. Butterfield and J. Earman (eds.), *Handbook of the Philosophy of Physics*, Part B, Amsterdam: North-Holland, pp. 923–1074.

Watanabe, S. (1965), Conditional probability in physics, *Progress of Theoretical Physics Supplement*, Extra issue (1965), 135–167.

Chapter 3
Probability Assignments and the Principle of Indifference. An Examination of Two Eliminative Strategies

Sorin Bangu

3.1 Introduction

A central and controversial component of the 'classical' conception of probability, the Principle of Indifference (PI) claims, roughly, that equi-possibility entails equi-probability.[1] A more precise version can be formulated as follows[2]:

> Given a null state of background information, equal regions of the space of possible outcomes should be assigned equal probabilities.

The principle plays an important role not only in physics (in the foundations of statistical mechanics[3]), but also in everyday probabilistic inferences (e.g., in predicting the outcomes of various games of chance). Yet many philosophers and scientists have also pointed out that PI is subject to two serious objections. The first one stresses the well-known inconsistencies (aka Bertrand (1889) paradoxes) associated with the application of the principle.[4] The second problem was signalled by Hans Reichenbach (1949/[1971]) and does not focus primarily on the role of PI in generating paradoxes. Reichenbach observed that the principle seems to license the inference of the frequency of occurrence of some physical phenomena on the basis of our epistemic state (of ignorance). Hence, the acceptance of the principle would be tantamount to the acceptance of the idea that the occurrences of events

S. Bangu (✉)
Department of Philosophy, University of Illinois at Urbana-Champaign, IL 61801, USA
e-mail: sib10@uiuc.edu

[1] See Hacking (1975, Chapter 14) for more on the traditional objection that the principle is a tautology, as equi-possible just means equi-probable.

[2] Virtually every treatise on probabilities mentions the principle. The present formulation is Howson and Urbach's (2006, 266).

[3] Standard textbooks on statistical mechanics start with what they call the 'fundamental assumption' (Kittel and Kroemer, 1980, 29), or 'the fundamental postulate of equal apriori probabilities' (Reif, 1965, 54): 'an isolated system in equilibrium is equally likely to be in any of its accessible states'.

[4] van Fraassen, for instance, notes that 'the great failure of symmetry thinking' is revealed in those situations 'where indifference disintegrated into paradox.' (1989, 293).

M. Suárez (ed.), *Probabilities, Causes and Propensities in Physics*, Synthese Library 347, DOI 10.1007/978-1-4020-9904-5_3, © Springer Science+Business Media B.V. 2011

in the physical world somehow ‘follow the directives of human ignorance’ (1971, 354). Finding this worrying, he attempted to show that PI is in fact *not necessary* in probabilistic reasoning. (See especially §§ 68–71 of his (1971)).

Reichenbach’s approach is an illustration of what I’ll call here an eliminative strategy. As noted, the guiding idea of such a strategy is to show that the inferences to the correct observed frequencies made by using the seemingly apriori principle can also be completed without employing it. Naturally, this type of strategy is appealing to those who fear that the acceptance of PI collides head-on with one of the fundamental principles of our modern scientific worldview: namely, that the natural order should not be expected to be sensitive to the standards of human reason – more generally, that apriori knowledge is highly problematic. Similar concerns have recently been voiced by Michael Strevens (1998). After noting the systematic success of employing PI in deriving probabilities (‘[By using PI] we infer the *correct physical* probability for the event in question.’ – 1998, 231, emphasis in original), Strevens points out that endorsing the principle (at least in the traditional form) would amount to endorsing a mysterious connection between our epistemic condition and the world. In a Reichenbachian spirit, he remarks that we should be reluctant to accept such a connection, because we believe that ‘[T]he nature of the world is independent of our epistemic deficits. The fact that we do not know anything about A does not constrain the way things are with A.’ (1998, 231).

My goal in this paper is to analyze in detail two attempts to implement the eliminative strategy.[5] (I discuss only two proposals since, to the best of my knowledge, no others have been advanced so far.) Although I agree that eliminativism is a very promising type of strategy,[6] I maintain that both attempts to be examined here are fraught with problems. The paper is divided in two sections, each aiming to highlight the difficulties faced by each of those attempts. In the first section I argue that Reichenbach’s implementation of the eliminative strategy fails. More precisely, one of the premises of his eliminative argument either simply assumes PI, or is lacking any justification altogether. I examine the same example taken by Reichenbach (following Poincaré) as paradigmatic for showing the effectiveness of the principle: the employment of PI in predicting the correct probabilities in the game of roulette. Since there is presumably nothing special about roulette (in the sense that the eliminative strategy can be adapted to other cases in which PI seems effective),[7] it is natural to expect that the failure to eliminate PI in the roulette case will have similar consequences for other attempts.

---

[5] With one exception, none of the analyses of PI listed below pays attention to these strategies; instead, they focus exclusively on the relation between PI and the (Bertrand-type) paradoxes. See Norton (2008), Shackel (2007), Mikkelson (2004), Bartha and Johns (2001), Gillies (2000), Castell (1998), Marinoff (1994). The notable exception is Strevens (1998), and I’ll discuss it shortly.

[6] A bonus of this approach is that the threat of paradox vanishes as well. Another promising line of objections to PI is of course based on its role in deriving the Bertrand type paradoxes. While I’ll be saying something about this role in Section 2, the focus here is not on the paradoxes.

[7] Reichenbach himself noted that there is nothing philosophically special about the games of chance (1971, 358)

In the second section I turn to another, more recent attempt to eliminate PI, Donald Gillies's heuristic approach (Gillies, 2000). Despite the fact that Gillies allows PI a certain role in our probabilistic inferences (namely, to help us conjure probabilistic hypotheses), I construe his view as an attempt to dispense with PI as well. My reason for construing it this way is motivated by Gillies' emphasis on the incapacity of the principle to *justify* those hypotheses, and thus to yield a substantial epistemic benefit. After I present Gillies' position, I raise doubts with regard to its cogency. I point out that the alternative method of justification/rejection of probabilistic hypotheses endorsed by Gillies – in essence, the method of statistical relevance tests – is subject to the same kinds of difficulties as the method of a priori justification involving the use of PI.

3.2 The Poincaré – Reichenbach Strategy

Reichenbach discussed the (in)dispensability of PI in §§ 68–71 of his classic *Theory of Probability* (1971/[1949]). The subject of Reichenbach's analysis is the familiar PI-type inference of equal probabilities of 'red' and 'black' outcomes in the game of roulette. This inference has a familiar ring: apparently, we know of no reason why the rotating needle should stop on red more often than on black, so PI says they should occur equally often. They *do* in fact, and the eliminativist has to 'explain away' the success of this inference, i.e., to show that the same prediction can be made without using PI. The explanation is as follows (see Reichenbach, 1971, 356–358; Poincaré, 1912, 149–5, 1952a, 201–202).

Let us first assume that before each spinning the rotating needle is brought back to the same initial position. The needle is spun and its rest final position is described by an angle θ, which is counted in multiples of 2π. Further, let's assume that the probability the rotating indicator needle stops at a particular position θ is given by a probability function $\varphi(\theta)$. One might ask what is the explicit form of φ. As Reichenbach points out, the function $\varphi(\theta)$ is *not* known, and not necessarily to be known.[8] Only the existence of such a function is assumed. One possible graph of φ is depicted below in Fig. 3.1

As usual, the roulette is divided in red and black sectors, and this corresponds to intervals of equal width $\Delta\theta$ on the abscissa axis. (For simplicity, we assume that unlike a real roulette this one is divided only in red and black sectors, so no sector is designated as profit for the bank.) Figure 3.1 shows a succession of 'red' and 'black' intervals. The corresponding probability of red or black is given by the area of the corresponding stripe, shaded for black, and un-shaded for red.

The probability of the rotating needle stopping at one of these two colours is equal to the total area covered by the shaded and un-shaded stripes, respectively. The result to be proved is that the total area covered by the shaded stripes is equal to the total area covered by the un-shaded stripes.

[8] Accordingly, this method of reasoning was called, after Poincaré (1912/[1896]), 'the method of arbitrary functions'. For a review of its philosophical relevance, see von Plato (1983).

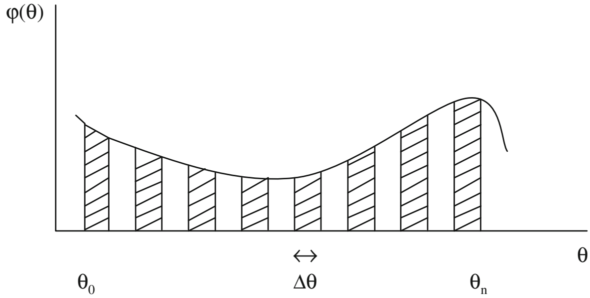


Fig. 3.1 After Reichenbach (1971, 356)

Reichenbach notes that three assumptions are needed in order to complete the proof (1971, 357). First, we have to assume that the function φ is continuous. Second, we assume that the $\Delta\theta$ intervals are equal in length (which is the case). And third, we assume that 'the size of the intervals $\Delta\theta$ is small with respect to the oscillations of the function $\varphi(\theta)$' (357) – or, as Poincaré put it (1952b, 84), we assume 'that probability [can] be regarded as constant in a small interval', which is that '$\varphi(\theta)$ does not oscillate too much' (Reichenbach, 1971, 356). Following Strevens (1998), call such a function 'smooth'. In this context, then, a real function φ is said to be 'smooth' if it is not *too* steep[9] (that is, infinitely steep); more precisely, if there is a finite positive C such that for any x_i, x_j in the function's domain, we have that $\left|\varphi(x_i) - \varphi(x_j)\right| \leq C\left|x_i - x_j\right|$. Obviously, the two sums of areas are *exactly* equal in the (unrealistic) situation when $\Delta\theta = 0$, regardless of the properties of φ. In this case, which corresponds to the limit case when the roulette consists of an infinite number of painted sectors, the only assumption needed to show that $P(\text{red}) = P(\text{black})$ will be that $\varphi(\theta)$ is continuous. But this case is unrealistic, so the smoothness assumption is still needed to complete the proof.

As noted, in order to show that the two probabilities are equal in more realistic cases, we have to show that the sum of the black areas is approximately equal to the sum of red areas. The idea is to estimate the upper bound of the difference of these sums, and then to show that this upper bound is (goes to) 0, as n increases. The estimation proceeds as follows. Consider the finite section between θ_0 and θ_n, and consider all pairs of two consecutive stripes. For each of these pairs, consider the difference between the greatest value of the ordinate in that pair and the smallest value. One of these pairs will feature the maximum value for this difference. If we take M to be the greatest value of the ordinate in that pair and m the smallest value, call their difference $(M-m)_{\max}$; then, regardless of the areas of those two stripes, the difference of areas can't be larger than $(M-m)_{\max}\Delta\theta$. Since $[\theta_n - \theta_0]$ is constant, we have $\Delta\theta = [\theta_n - \theta_0]/n = c/n$, where c is a finite positive constant.

[9] Poincaré (1912, 148–150) doesn't use the word 'smooth'. He describes such a function as having 'une derivée limitée'; more precisely, its derivative is bounded, i.e., there is a positive C such that $\left|\varphi'\right| \leq C$.

The next step is to find an upper bound for $(M - m)_{\max}$. Suppose that α and β are those angle values on the abscissa such that, for a given n, $(M - m)_{\max} = |\varphi(\alpha) - \varphi(\beta)|$. Assuming that φ is smooth, there is a constant $C \geq 0$ such that $|\varphi(\alpha) - \varphi(\beta)| \leq 2C\Delta\theta$. Proof: the smoothness of φ gives us a positive C such that $|\varphi(\alpha) - \varphi(\beta)| \leq C|\alpha - \beta|$. Given how α and β were chosen, we also have $|\alpha - \beta| \leq 2\Delta\theta$, hence $|\varphi(\alpha) - \varphi(\beta)| \leq 2C\Delta\theta$. QED. Thus we have found an upper bound: $(M - m)_{\max} \leq 2C\Delta\theta$. Now, since there are $n/2$ pairs of stripes, the maximum value of the difference between the sum of black areas and red areas is smaller than $(n/2)(M - m)_{\max}\Delta\theta$. Given that $(M - m)_{\max} \leq 2C\Delta\theta$, we note that

$$0 \leq (n/2)(M - m)_{\max}\Delta\theta \leq (n/2)(2C\Delta\theta)\Delta\theta = nC(\Delta\theta)^2 = nC(c^2/n^2) = Cc^2/n$$

Therefore,

$$0 \leq (n/2)(M - m)_{\max}\Delta\theta \leq Cc^2/n$$

Obviously, as n increases, Cc^2/n decreases, hence $(n/2)(M - m)_{\max}\Delta\theta$ approaches 0. But this product upper-bounds the difference of sums of black and red stripes respectively, so this difference must go down to 0 too.

This completes the explanation of the equiprobable distribution.[10] As Reichenbach emphasizes:

> In this sense the theory supplies an *explanation* of the equiprobability. (...) This way of handling the problem carries the advantage that the principle of "no reason to the contrary" [that is, PI, cf. p. 353] is completely eliminated. The equiprobability does not appear as following from the *absence* of reasons, but as a result of the *existence* of definite reasons (...) (1971, p. 358, emphasis in original)

The relevant question is, of course, whether the use of PI is 'completely eliminated'. The crucial aspect to focus on below is the justification of the assumption regarding the smoothness of $\varphi(\theta)$.[11] My claim is that once we unpack what is actually assumed here, we come to realize that this assumption relies in fact on PI – hence Reichenbach's explanation fails to eliminate PI.

---

[10] Many authors have analyzed the roulette set up in a manner similar to Poincaré; see Borel (1909, 117–21), Hopf (1934), Fréchet (1952, 3–8), Savage (1973), Diaconis and Engel (1986), Kechen (1990) and Engel (1992).

[11] Strevens (1998, 238) insightfully remarks that such a justification might be missing. Here is what Reichenbach actually says about this assumption, on pages 357–358 of his (1971): 'The assumption 3 represents a certain rough appraisal of the metrical properties of the probability function that may, however, remain undetermined within wide limits.' Moreover: 'Even the rough appraisal of degrees of probability, employed in assumption 3, is used in many cases that apparently have nothing to do with probability. We always make use of such appraisals in daily life when we regard statements about future events as "practically certain".' But smoothness is not as unproblematic as Reichenbach might want us to believe. In fact, it ain't hard being non-smooth: consider a trivial example, f: $[0, \infty) \to [0, \infty)$, where f is defined as follows: $f(x) = x$, if $0 \leq x < 1$, $f(x) = (x - 1)^{1/2} + 1$, if $x \geq 1$. While f is continuous everywhere (including $x = 1$), it becomes infinitely steep as it descends toward $x = 1$ (as the ratio $[f(x) - f(1)]/[x - 1]$ goes off to infinity).

Strevens' (1998) analysis of this reasoning attempts to elucidate, on Reichenbach's behalf, where the smoothness assumption could come from:

> The difficult question for Poincaré and Reichenbach concerns our knowledge of φ(θ). How do we reach the conclusion that φ(θ) is relatively smooth? The intuitive answer is that we expect φ(θ) to be smooth because the roulette wheel has perfect circular symmetry. In fact we do far better than this: from the symmetry of the wheel, we infer that the probability distribution over θ is *uniform*, that is, that the probability is the same that any θ will be the wheel's final resting place. (Strevens, 1998, 237–238; emphasis in original)

On this account, Reichenbach's smoothness assumption is justified by deriving it from the assumption of another property of φ, namely its uniformity. On one hand, this is not surprising, since uniformity is a more intuitive and more powerful property than smoothness. If we assume that $\varphi(\theta)$ is uniform, this amounts to assuming that *any* angle θ on the wheel's circumference is an equally probable rest point for the rotating needle. The next step is to remark that, trivially, the uniformity of φ entails its smoothness. Yet, obviously, the worrying consequence of assuming uniformity is that, in fact, it amounts to assuming PI. So, if Strevens' suggestion (that Reichenbach's justification of the smoothness assumption is to be reconstructed, or understood, in this way) is correct, it follows that Reichenbach's assumption subtly relies – by assuming the uniformity of φ – on PI; hence, PI is not eliminated from the picture.[12]

The key point (as well as the weakest point) of the eliminativist argument is clearly the solution to the smoothness problem. While I think that Strevens is correct to signal that 'Reichenbach does not attempt to solve the problem' (1998, 238), he is mistaken to say the same thing about Poincaré.[13] It is true that Poincaré thought that the smoothness assumption would become irrelevant when the number of painted sectors was very large, perhaps infinite – as noted above, if $\Delta\theta \to 0$, all we need to run the argument is a continuous φ. (In fact, a further complication, which I ignore here, arises: since a realistic roulette spin corresponds to a relatively small n, it is no longer clear whether the calculations work even in a realistic case.) Yet, as we'll see, Poincaré did attempt to advance a solution to the smoothness problem.[14] In what follows, I first present this solution and then I signal some of its shortcomings.

---

[12] Strevens observes that this justification is not satisfactory, so he notes that 'to make progress, Poincaré and Reichenbach must explain the basis of the inference that $\varphi(\theta)$ is smooth.' (1998, 238).

[13] See Strevens (1998, 238).

[14] Poincaré presents the arguments I'll discuss here in his *Science and Method* (1952b/[1908]), Chapter I.IV ('Chance'). Interestingly, Reichenbach seems unaware of what Poincaré said about the roulette in *Science and Method*, as he cites only Poincaré's *Calcul de Probabilités* (1912/[1896]) – see Reichenbach (1971, 355, fn 1). Strevens (1998), too, only mentions what Poincaré said in his 1902 *Science and Hypothesis* (which is essentially the argument rehearsed by Reichenbach and taken from Poincaré, 1912), but not what Poincaré argued in his 1908 *Science and Method*, Chapter I.IV. My exegetical hypothesis is thus that both Reichenbach and Strevens just didn't know about Poincaré's further elaborations on the smoothness issue in *Science and Method*.

Not surprisingly, Poincaré realized that the solution to the roulette equiprobability problem can't be complete without addressing the smoothness issue, i.e., the constancy of probability in a small interval. He first formulates the problem,

> Why can that probability be regarded as constant in a small interval? It is because we admit that the law of probability is represented by a continuous curve, not only continuous in the analytical sense of the word, but *practically* continuous, as I explained above [when he offered the solution discussed above]. This means not only that it will present no absolute hiatus, but also that it will have no projections or depressions too acute or too much accentuated. (*Science and Method*, 1952b, p. 84; emphasis in original)

and then he asks the crucial question:

> What gives us the right to make this hypothesis? (*Science and Method*, 1952b, 84)

I reproduce his answer bellow and then I'll make a couple of remarks on (what I take to be) his central point. So, immediately after asking this question Poincaré writes:

> As I said above, it is because, from the beginning of the ages, there are complex causes that never cease to operate in the same direction, which cause the world to tend constantly toward uniformity without the possibility of ever going back. It is these causes which, little by little, have levelled the projections and filled up the depressions, and it is for this reason that our curves of probability present none but gentle undulations. In millions and millions of centuries we shall have progressed another step towards uniformity, and these undulations will be ten times more gentle still. The radius of mean curvature of our curve will have become ten times longer. And then a length that today does not seem to us very small, because an arc of such a length cannot be regarded as rectilineal, will at that period be properly qualified as very small, since the curvature will have become ten times less, and an arc of such a length will not differ appreciably from a straight line. (*Science and Method,* 1952b, p. 85)

Therefore, the explanation of the smoothness of the probability function appears to proceed as follows. First of all, the smoothness of φ is not derived directly from its uniformity. This time φ's smooth behaviour is justified by understanding the behaviour of the roulette as a physical phenomenon following the general tendency of the entire universe 'toward uniformity'. The occurrence of red and black outcomes must follow this universal tendency, towards equalizing frequencies. This tendency toward equilibrium applies to *all* physical processes in nature, hence it applies to the spins of the needle, and it is further reflected in levelling the peaks and valleys of the probability function.

Poincaré's fundamental insight here seems to be that the ultimate explanation of the smoothness phenomenon has to proceed by assuming the hypothesis of an increasing *entropy* universe. This is of course hard to claim with certainty, not only because the term 'entropy' does not actually occur in those passages, but also because Poincaré is not as forthcoming as one might expect in developing this subtle thought. Yet we know that he was involved in the development of Statistical Mechanics and, when invoking the 'causes that never cease to operate in the same direction, which cause the world to tend constantly toward uniformity without the possibility of ever going back', he seems to have in mind (at least an analogy with) the irreversible, thermodynamic-like effects of smoothing-out various physical differences in a system (the roulette system in this case). The value of entropy would thus be an indication of how far this 'ironing-out' process has progressed, hence his references to the position of the temporal moment when the roulette mechanism is analyzed relative to the cosmological temporal axis ('In millions and millions of centuries we shall have progressed another step towards uniformity, and these undulations will be ten times more gentle still.')

Now let me stress that it is not my intention to belittle this attempt at a physical justification of the smoothness assumption. I shall confine myself to noting what seems to be an obvious point: this explanation manages to trade one mystery for another. In other words, I doubt that framing the explanation of smoothness in no less than cosmological terms advances the eliminativism cause significantly. It seems that in order to solve one kind of local mystery (why is φ's first derivative upper-bounded?), this explanation must assume facts about the entropy of the whole Universe, which, it seems, are even more pressingly in need of an explanation themselves. As even a cursory glance at some recent discussions in the philosophy of physics reveals, entropy and its explanatory import are among the most problematic issues debated these days. The entropy concept fuels unsettled controversies, the motivation of some of them being the concerns that triggered the suspicion over PI in the first place – namely, general worries that we live in a human-friendly universe. More concretely, one example of the debate generated by some of these issues is the ongoing controversy between Huw Price and Craig Callender over whether or not the 'Past Hypothesis' is in need of explanation. Roughly, this is the hypothesis that entropy was low in the past in our region of the universe, in order to be possible for human species to inhabit a still increasing entropy cosmological epoch. (For more details, see Price (2004), Callender (2004) and Sklar's (1993) classic.)

Given this uncertain state of (conceptual) affairs, I submit that one is entitled to doubt that the appeal to an explanatory framework in which the entropy of the whole universe plays a central role would lead to a substantial clarification of the smoothness issue. This seems to be a typical case in which the putative *explanans* seem harder to fathom than the *explananadum* itself. I now turn to the second eliminativist strategy.

3.3 The Gillies Strategy

As noted in the introduction, a different kind of eliminativist strategy can be discerned in the recent literature, and one of the clearest expositions of it can be found in Donald Gillies (2000, 48–49). This strategy comprises two aspects. First, PI is regarded as an important heuristic device, useful in *discovering*, or suggesting probabilistic hypotheses. Yet, second, PI is denied any role in *justifying* probability hypotheses. Using the terminology I introduced here, I shall call this strategy 'eliminativist' as well. Although PI is not eliminated directly or completely, its epistemic role in the justification of our hypotheses is seriously downplayed. The principle is granted a rather insubstantial epistemic role, by being rendered epistemically inert, so to speak; it is in fact eliminated. I now examine the strengths and the weaknesses of this strategy.

Gillies underscores the principle's incapacity to show hypotheses to be correct 'independently of experience' (2000, 48). After mentioning the usefulness of PI in various physics contexts (e.g., in calculating the viscosity of a gas in statistical mechanics, its employment in connection to the Bose-Einstein statistics), Gillies notes:

> However, this seems to me to show the fruitfulness of the principle as a *heuristic* principle *not* its validity as a *logical* principle. The Principle of Indifference, together with additional considerations such as invariance requirements arguments about the indistinguishability or otherwise of particles, etc. has been, and perhaps will be in the future, very useful for suggesting hypotheses in physics but the principle does not establish the truth of these hypothesis. (2000, p. 48)

Gillies' point assumes the familiar distinction between simply conjuring a hypothesis, and confirming (or falsifying) it by running experiments.[15] It follows that the assignment of the probability values via PI should be regarded as any scientific hypothesis, which has to be 'tested empirically like any other hypotheses in physics' (Gillies, 2000, 48).

This Popperian (falsificationist) picture would clearly offer a simple and elegant solution to the epistemic problems motivating the empiricists' dislike for PI. Yet this strategy is not completely satisfactory unless it deals with the following problem. As Popper pointed out a long time ago, probabilistic hypotheses have a special status among scientific hypotheses: they can't be strictly speaking falsified or confirmed. In Section 66 of his *Logic of Scientific Discovery*, he notes:

> Probability estimates are *not* falsifiable. Neither, of course, are they verifiable, and this for the same reasons as hold for other hypotheses, seeing that no experimental results, however numerous and favourable, can ever finally establish that the relative frequency of 'heads' is 1/2, and will *always* be 1/2. (Popper, 1959, 191, emphasis in original.)

[15] Two clarifications are in order. First, I'm not claiming that Gillies is a frequentist (of any kind). He labels himself as a 'long run propensity' theorist later on in his 2000 book (Chapter 7). Second, Reichenbach himself, a paradigmatic frequentist, is rather cautious when discussing the refutation/confirmation of PI by appeal to relative frequencies. He says: 'The only possible defence of the *a priori* determination [via PI – my note], therefore, consists in the attempt to restrict it to a meaning of probability that is not expressible in terms of frequencies [. . .].' (1971, 354) It seems then that he leaves open the issue whether other interpretations of probability are more permissive to this defence, and stresses that frequentism is the main enemy of PI: one who accepts the frequency interpretation of probability can't defend PI as a 'logical' principle, acting beyond the so-called 'context of discovery'. (However, as known, he'll argue that frequentism is actually the only admissible interpretation of probability, so his hesitation is only temporarily). So, a defender of PI should now inquire how convincing is frequentism as an interpretation of probability; if frequentism is so robust that we *must* embrace it, then we'd lose any hope to defend PI, by Reichenbach's own lenses at least. Yet the recent philosophy of probability literature seems to converge toward the point that frequentism is *not* very convincing as an interpretation, or definition of probability. See Hajek (1997) for a collection of fifteen arguments against frequentism.

Popper's argument is laid out in the previous Section 65 and uses the example of coin-tossing.[16] Suppose we advance hypothesis H that the probability to obtain a 'head' is Pr (head) $= p = 1/2$. The question is how should we proceed to corroborate or reject this hypothesis. Of course *in practice* the problem is taken much lightly: we think we corroborate or falsify H by tossing the coin several times, recording the outcomes and then analyzing them. Yet, as Popper points out, even so 'there can be no question of falsification in logical sense.' (Popper, 1959, 190). Furthermore, he goes on and points out that 'Only an infinite sequence of events – defined intensionally by a rule – could contradict a probability estimate.' (Popper, 1959, 190)

At this point, a serious epistemic problem crops up. Obviously, the production and inspection of such a sequence is highly problematic for an empiricist, hence it should be regarded as epistemologically suspect – one would add, as suspect as accepting PI in the first place as a justificatory device. So, at this point one ceases to see the epistemological advantages of justifying probability assignments by the empirical (frequentist) procedure advocated by Gillies, rather than by the a priori procedure involving PI.

To this it might be replied that the advantages of the empirical approach to justification can still be retained by appealing to a more sophisticated version of the argument from practice mentioned above. Indeed, Popper does offer a second solution along these lines, which Gillies endorses and develops (Gillies, 2000, 145–150). As is well known, Popper did not give up his falsificationism so easily, and distinguished between 'logical falsifiability' and 'methodological falsifiabiality' (See Gillies's, 2000, 146). While logical falsifiability is problematic, methodological falsifiability is apparently not. Gillies takes up this point and notes that, in practice, a natural thing we do when we test a probabilistic hypothesis is (a) take the random variable $X =$ 'number of heads out of the number of tosses' as a test-statistic, (b) toss the coin a number of times and (c) analyze the data by subjecting it to usual statistical significance tests.[17]

However, the appeal to statistical significance tests is highly questionable when presented in this context. More precisely, it is very problematic when presented as motivating the elimination of PI, given that one of the reasons motivating the elimination was that PI leads to the Bertrand-type paradox(es). From now on then, I shall argue that the problem faced by the Gillies approach is this: *the objections*

---

[16] This argument, however, is not the one advanced by Kuhn and Quine (and perhaps earlier on by Duhem) and based, roughly speaking, on considerations about underdetermination and addition of auxiliary assumption. A sketch of Popper's argument is as follows (see his 1959, §65). Suppose we toss a coin, and suspect Pr (heads) $= p$. We know that in general the probability to get m 'heads' in n successive tosses of a coin is Pr $(m : n) = {}^{n}C_{m}p^{m}(1-p)^{n-m}$. Since this value is never zero (regardless of the value of p, and of how big n and m are), our hypothesis can't be refuted. For a more detailed presentation of this argument, followed by a discussion, see Gillies (2000, 146).

[17] Gillies interprets Popper as making this point in the following remark: "(…) a physicist is usually quite able to decide whether he may, for the time being accept some particular hypothesis as "empirically confirmed", or whether he ought to reject it as "practically falsified"." (Popper 1959, p. 191, quoted in Gillies (2000, p. 146).

against the use of PI (stressing its connection to the paradox) are identical to the objections raised against the use of statistical tests. Hence it is no longer clear why appealing to statistical tests to justify/confirm a probabilistic hypothesis represents any improvement over the use of PI.

If this is so, it follows that it is just inconsistent to adopt the statistical tests solution to the problem of hypothesis confirmation/rejection, in so far as this solution is subject to the same objections raised against the solution based on PI. I shall deal with this point in detail below, splitting the explanation in two parts. First, I'll remind the reader the role of PI in deriving the paradox and second, I call attention to the overlooked similarity between the objections to PI and those to the underlying logic of statistical tests.

As is well known, the paradox occurs because PI seems to allow the derivation of different probabilities for the same event. More precisely, it turns out that if PI is applied to different (partitions of the) outcome spaces, it yields different results. In so far as none of these partitions stands out as the 'correct' one, we have an inconsistency. More concretely, a version of the Bertrand paradox is as follows (van Fraassen, 1989, Chapter 12).

Suppose a factory makes cubes of side-length L, and all we know is that $1 \leq L \leq 3$. What is the probability that the factory will make a cube such that $1 \leq L \leq 2$? Given our ignorance of the possible values of L, the application of PI yields the answer

$P(1 \leq L \leq 2) = (2 - 1)/(3 - 1) = 1/2$. Yet the question can be given an equivalent formulation, since $1 \leq L \leq 2$ is equivalent to $1 \leq S \leq 4$, where S is the area of a face of a cube of side-length L (hence $1 \leq S \leq 9$). So, this time we are interested in the probability that the factory makes a cube whose face has area S, such that $1 \leq S \leq 4$. The application of PI gives $P(1 \leq S \leq 4) = (4 - 1)/(9 - 1) = 3/8$. Since it is uncontroversial that $1 \leq L \leq 2$ is the same event as $1 \leq S \leq 4$, we have an inconsistency.[18]

As is easy to see, the origin of this difficulty can be traced back to the initial decision involved in choosing a parameter (L or S), which is further reflected in the way the outcome space is constituted. More precisely, the difficulty is that there is no rule that could somehow objectively indicate which parameter should be chosen (side-length L or face area S) to formulate the problem,[19] and thus no rule indicating how the possible outcomes space should look like.

How is all this relevant for the Gillies strategy? In advancing the alternative method of justification/acceptance of probabilistic hypotheses (essentially, the method of statistical significance tests), Gillies overlooks the fact that the central problem faced by this method has the same conceptual origin as the problem identified above in connection to PI. The problem for the statistical tests method is the

[18] As is perhaps clear by now, I'm sympathetic to PI; see Bangu (2010) for an attempt to deal with the paradox.

[19] Jaynes (1973) proposes a clever but controversial solution to this type of problem, based on symmetry considerations; yet Jaynes doesn't deal with this particular cube factory scenario, but with Bertrand's original chord problem.

absence of a rule determining the constitution of the outcome space. Now, as has long been discussed in the statistical literature, the problem for statistical tests is this: different initial configurations of the outcome space determine two inconsistent final decisions (acceptance/rejection of H at the same significance level, say 5%), *and* there is no epistemically motivated rule for deciding which of those spaces should be chosen.[20]

I'll explain this in more detail below; however, my exposition will assume some knowledge of statistical tests, as here I'll omit all technical aspects. (A clear and complete presentation of these calculations can be found in Howson and Urbach (2006, Chapter 5), and I reproduce them partially, for readers' convenience, in the Appendixes.) My main intention is to stress the philosophical point that the attempt at a justification by appealing to statistical tests faces the same difficulties as those encountered when using PI.

Let us begin by rehearsing how the method of significance tests proceeds. In order to subject a hypothesis H to a statistical significance test, one first needs to specify, in addition to the test-statistic employed and the significance level, the space of possible outcomes. Conceptual complications occur immediately; as Howson and Urbach point out, 'the space of possible outcomes is created, in part, by what is called the *stopping rule*; this is the rule that fixes in advance the circumstances under which the experiment should stop.' (2006, 156)

Now the problem is that different stopping rules yield different outcome spaces and thus different decisions. Here is a very simple example.[21] Suppose we test our $H : p = 1/2$. (This is the hypothesis that the coin is fair). One stopping rule we might choose is

SR_1: stop the trial after 20 tosses[22]

Simple calculations show that the result (6 heads, 14 tails) is not significant at the 5% level. Hence *H*, treated as the null hypothesis, should *not* be rejected at this level.

Yet one might choose a different stopping rule:

SR_2: stop the trial as soon as 6 heads occur

The consequence of SR_2 is the creation of a different outcome space[23] and, again, calculations show that the same result (6 heads, 14 tails) comes out as significant at the 5% level. Hence H should be rejected at this level, contrary to what we decided in the first case!

---

[20] This holds for all statistical tests, including the heavily used chi-square test, which is based on arranging population into cells – see the remark by Hays and Winkler who acknowledge that *the arrangement into population class intervals is arbitrary* (1971, 791, emphasis in original). Taking note of this, Howson and Urbach (2006, Chapter 5) signal the amazing "complacency" of Fisherian statisticians when confronted with this problem.

[21] From Howson and Urbach (2006, 157–158), who in turn adapt an example by Lindley and Philips (1976).

[22] The outcome space associated with SR1 can be described as pairs (heads, tails): (20, 0), (19,1), (18, 2), . . . , (0, 20). See Appendix 1.

[23] The new outcome space is thus (6, 0), (6,1), (6, 2), . . .etc. See Appendix 2.

The decision over *H* is thus essentially influenced by the way the space of possible outcomes is determined, by choosing of a stopping rule. But, as it should be clear by now, this is the same kind of objection raised to the use of PI: namely the existence of an initial dilemma over the constitution of the outcome space, which immediately generated the Bertrand paradox. Since both approaches lack an answer to the question ‘What’s the rule determining the outcome space?’, one can no longer see the advantages of the statistical tests alternative method.

The logical impossibility of refuting/confirming probabilistic hypotheses should be alarming for a supporter of this kind of heuristic eliminativism. (Of course, this issue has more general implications about testing probabilistic hypotheses, but I’ll not touch upon them here.) The heuristic eliminativist strategy was based on a distinction (between what it takes to invent, or to discover a hypothesis, and what it takes to confirm/accept or reject it) that turns out to involve more problematic aspects than were taken into account when it was proposed. One can of course argue that PI is a mere heuristic tool and then deny its role in justifying probabilistic hypotheses. Yet, when pressed about justification (or acceptance), one *can’t* go further an also argue that PI’s role in justification can be replaced by the method of statistical significance tests – this is so since this method is vulnerable to the same kind of objections motivating the elimination of PI in the first place. Thus I claim that Gillies’ heuristic eliminativism is incomplete in a fundamental sense. It can become complete only if it comes equipped with a procedure of justification, or acceptance/rejection of hypotheses which is not only logically independent of PI, but also immune to the type of objections raised against the principle.

3.4 Conclusion

My aim in this paper was entirely critical: to show that the attempts to dispense with PI, either via the Poincaré – Reichenbach strategy or via the Gillies strategy, are not compelling. Both implementations of the first eliminative strategy turned out to be flawed: Reichenbach’s eliminativist argument was simply unable to offer an account of the smoothness assumption – or, if Strevens’ suggestion is to be followed, it needed to assume the uniformity of the distribution, which was to assume PI. Poincaré’s attempt to justify the smoothness was unsatisfactory too, since it involved an assumption (the increasing entropy of the universe) no less in need of explanation than the workings of PI.

The second kind of eliminativism (the heuristic approach) fell prey to a somewhat similar difficulty. Gillies’ alternative solution to the confirmation/acceptance problem (the appeal to statistical tests) was found to be vulnerable to the same kinds of objections as those that motivated the elimination of PI in the first place. While the a priori flavour of the principle can’t be denied (hence the empiricists’ dislike of it), PI seems to be indispensable to probabilistic reasoning, in one form or another.[24]

[24] Strevens (1998) ends up by accepting a (more empirically-flavored) form of PI as well.

Finally, note that the arguments here don't amount to any substantial direct support for PI, as I didn't propose any further justification or explanation of this apparent indispensability. The present arguments are meant to challenge the supporters of the empiricist eliminativism to devise more sophisticated versions of this view, able to avoid the pitfalls of the strategies examined here.

Acknowledgments I thank Margaret Morrison, Peter Urbach, Colin Howson, Bob Batterman, Anjan Chakravartty, Ranpal Dosanjh and Kaave Lajevardi for discussions and critical comments on earlier versions of this paper. I also thank Mauricio Suarez for encouraging feedback and for stirring up my interest in these issues. Any mistakes that remain are mine. Financial support for working on this paper was kindly provided by the Rotman Postdoctoral Fellowship in Philosophy of Science at the University of Western Ontario.

Appendix 1

According to Stopping Rule 1, we toss the coin and stop after 20 tosses. Let Pr (head in a coin-tossing experiment) $=$ p, Pr (tail) $=$ q. We have Pr (r heads in n tosses) $= {}^nC_r p^r q^{n-r}$ (since this is a Bernoulli process), where ${}^nC_r = (n!)/[(n-r)!\ r!]$. The following table describes the possible outcomes.

X (# of heads /20 tosses)	Pr (X)	X (# of heads/20 tosses)	Pr (X)
0	9×10^{-7}	11	0.1602
1	1.9×10^{-5}	12	0.1201
2	2×10^{-4}	13	0.0739
3	0.0011	14	0.0370
4	0.0046	15	0.0148
5	0.0148	16	0.0046
6	0.0370	17	0.0011
7	0.0739	18	2×10^{-4}
8	0.1201	19	1.9×10^{-5}
9	0.1602	20	9×10^{-7}
10	0.1762		

Let's assess (the null) hypothesis

$$H\colon \text{Pr (head)} = 1/2$$

Suppose that the actual result is (6h, 14t). If H is true, Pr (6h, 14t) $= 0.0370$. This is so since $p = q = 1/2, n = 20$. The method (invented by Fisher) of assessing this result proceeds in the familiar way. First, we look at those results with *less or equal* probability to 0.0370. They are obtained for $X = 6, 5, 4, 3, 2, 1, 0$ and $x = 14, 15, 16, 17, 18, 19, 20$. Then we find the probability that any of these results will occur. This is the *p-value* for the result (6h, 14t), or p^*:

$$p^* = 2 \times (0.0370 + 0.0148 + 0.0046 + 0.0011 + 2 \times 10^{-4} + 1.9 \times 10^{-5} + 9 \times 10^{-7})$$

So, $p^* = 0.115$. Now, there is the convention to reject H (null hypothesis) only if $p^* < \alpha$, where, by another convention, $\alpha = 0.05 = 5\%$. This is called *the significance level* of the test. Since $p^* = 0.115 > 0.050$, we decide that H should *not* be rejected at 5% level.

Appendix 2

According to Stopping Rule 2, we toss the coin and stop after the first 6 heads occur. SR_2 produces (6, k) whenever (5, k), appearing in any order, is succeeded by a head.

If H: Pr (h) $= 1/2$ is true, then Pr $(6, \mathrm{k}) = {}^{k+5}C_5(1/2)^5(1/2)^k \times 1/2$. The following table describes the possible outcomes.

Outcome(h, t)	Pr(h, t)	Outcome(h, t)	Pr(h, t)
6, 0	0.0156	6, 11	0.0333
6, 1	0.0469	6, 12	0.0236
6, 2	0.0820	6, 13	0.0163
6, 3	0.1094	6, 14	0.0111
6, 4	0.1230	6, 15	0.0074
6, 5	0.1230	6, 16	0.0048
6, 6	0.1128	6, 17	0.0031
6, 7	0.0967	6, 18	0.0020
6, 8	0.0786	6, 19	0.0013
6, 9	0.0611	6, 20	0.0008
6, 10	0.0458	6, 21	0.0005
		Etc.	Etc.

Let's assess H. Like above, suppose we get the result (6 h, 14t). Hence, for $\mathrm{k} = 14$, Pr $(6, 14) = 0.0111$. Applying the same method as above, results with less or equal probability to 0.0111 are (6, 14), (6, 15), (6, 16), ... So, the probability of either of any of these occurring is their sum, the p-value or $p^* = 0.0111 + 0.0074 + 0.0048 + 0.0031 + 0.0020 + \cdots = 0.0319 = (1 - [0.0156 + 0.0469 + 0.0820 + \cdots + 0.0163])$

Since $p^* = 0.0319 < 0.0500$, H *should be rejected* at 5%.

References

Bangu, S. (2010), On Bertrand's paradox, *Analysis* 70, 30–35.

Bartha, P. and Johns, R. (2001), Probability and symmetry, *Philosophy of Science* 68 (Proceedings), S109–S122.

Bertrand, J. (1889), *Calcul des probabilitiés*. Paris: Gauthier-Villars.

Borel, E. (1909), *Elements de la Theorie des Probabilites*. Paris: Librairie Scientifique.

Castell, P. (1998), A consistent restriction of the principle of indifference, *British Journal for the Philosophy of Science* 49(3), 387–395.

Callender, C. (2004), Is there is a puzzle about the low entropy past? In Hitchcock, C. (ed.), *Contemporary Debates in the Philosophy of Science,* Oxford: Blackwell.

Diaconis, P. and Engel, E. (1986), ‘Comment,’ *Statistical Science*, 1(2), 171–174.

Engel, E. (1992), *A Road to Randomness in Physical Systems*. Springer Lecture Notes in Statistics No. 71, New York, NY: Springer.

Fréchet, M. (1952), Methode des functions arbitraries. Theorie des evenements en chaine dans le cas d’un nombre fini d’etats possibles In E. Borel (ed.), *Traite du Calcul des Probabilites et de ses Applications* (Tome I, Fascicule III, Second livre). Paris: Gauthier-Villars.

Gillies, D. (2000), *Philosophical Theories of Probability*. London: Routledge.

Hacking, I. (1975), *The Emergence of Probability*. Cambridge: Cambridge University Press.

Hajek, A. (1997), ‘Misses Redux’ *Redux*. Fifteen Arguments Against Finite Frequentism. *Erkenntnis* 45, 209–227.

Hays, W. and Winkler, R. L. (1971), *Statistics: Probability, Inference and Decision.* New York, NY: Holt, Rinehart and Winston.

Hopf, E. (1934) On causality, statistics and probability, *Journal of Mathematics and Physics* 17, 51–102.

Howson, C. and Urbach, P. (2006), *Scientific Reasoning. The Bayesian Approach.* Chicago, IL: Open Court (3rd edition).

Jaynes, E. T. (1973), The Well-Posed Problem, *Foundations of Physics* 3, 477–493.

Kechen, Z. (1990), Uniform distribution of initial states: The physical basis of probability, *Physical Review* A 41, 1893–1900.

Kittel, Ch. and Kroemer, H. (1980), *Thermal Physics*. New York, NY: W. H. Freeman and Company.

Lindley, D. V. and Philips, L. D. (1976), Inference for a Bernoulli process (a Bayesian view) *American Statistician*, 30, 112–119.

Marinoff, L. (1994), A Resolution of Bertrand’s Paradox, *Philosophy of Science* 61, 1–24.

Mikkelson, J. M. (2004), Dissolving the wine/water Paradox, *British Journal for the Philosophy of Science* 55, 137–145.

Norton, J. (2008), Ignorance and indifference, *Philosophy of Science* 75, 45–68.

Poincaré, H. (1912), *Calcul de probabilités*. Paris: Gauthier-Villars (1st edition 1896).

Poincaré, H. (1952a), *Science and Hypothesis*. Reprint, Dover, New York, NY. Translation of *La science et l’hypothèse* (1st edition 1902).

Poincaré, H (1952b), *Science and Method*. Reprint, Dover, New York, NY. Translation of *Science et méthode*. (1st edition 1908).

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

Price, H. (2004), On the origins of the arrow of time: why there is still a puzzle about the low entropy past, In C. Hitchcock (ed.), *Contemporary Debates in the Philosophy of Science*. Oxford: Blackwell.

Reichenbach, H. (1971/[1949]), *Theory of Probability.* Berkeley, CA: University of California Press.

Reif, F. (1965), *Fundamentals of Statistical and Thermal Physics.* New York, NY: McGraw-Hill.

Savage, L. J. (1973) Probability in science; a personalistic account In P. Suppes et al. (eds.), *Logic, Methodology and Philosophy of Science IV*. North Holland: Amsterdam. pp. 467–483.

Shackel, N. (2007), Bertrand Paradox and the principle of indifference *Philosophy of Science* 74, 150–175.

Sklar, L. (1993) *Physics and Chance*. Cambridge: Cambridge University Press.

Strevens, M. (1998), Inferring probabilities from symmetries, *Noûs* 32, 231–246.

Van Fraassen, B. (1989), *Laws and Symmetry*. Oxford: Clarendon Press.

Von Plato, J. (1983), The method of arbitrary functions, *British Journal for the Philosophy of Science* 34, 37–47.

Chapter 4
Why Typicality Does Not Explain the Approach to Equilibrium

Roman Frigg

4.1 Introduction

A gas that is confined to the left half of a container uniformly spreads over the entire available space as soon as the confining wall is removed. Yet we never observe the reverse process of a uniformly distributed gas suddenly concentrating in the left half of the container. Such irreversible behaviour is characteristic of many processes and is enshrined in the so-called Second Law of thermodynamics, which, roughly, states that entropy cannot decrease in isolated systems. Statistical mechanics (SM) aims to explain irreversible behaviour in terms of the dynamical laws governing the individual molecules of which the gas is made up. What is it about molecules and their motions that leads them to spread out when the wall is removed? And crucially, what accounts for the fact that the reverse process never happens?

An important answer to these questions was suggested by Boltzmann (1877), and variants of it are currently regarded by many as the most promising option among the innumerable of approaches to statistical mechanics. An important contemporary version of the Boltzmannian approach, originating in the work of Joel Lebowitz (1993a, b), differs from traditional approaches in that it explains irreversibility in terms of the notion of 'typicality'. Intuitively, something is typical if it happens in the 'vast majority' of cases: typical lottery tickets are blanks, typical olympic athletes are well trained, and in a typical series of a thousand coin tosses the ratio of the number of heads and the number of tails is approximately one. The leading idea of a typicality-based approach to SM is to show that thermodynamic behaviour is typical; that is, that the entropy in a system typically increases.

This approach has grown increasingly popular in recent years and has been advocated by a number of authors (references will be given below). The problem with understanding this approach is that it comes in different versions, which are, however, not recognised as such, much less clearly distinguished. We often find different

R. Frigg (✉)
Department of Philosophy, Logic and Scientific Method, London School of Economics, London WC2A 2AE, UK
e-mail: r.p.frigg@lse.ac.uk

M. Suárez (ed.), *Probabilities, Causes and Propensities in Physics*, Synthese Library 347, DOI 10.1007/978-1-4020-9904-5_4,

arguments pursued side by side and eventually we end up not having a clear picture of the claims being made. The aim of this paper is to disentangle different versions of typicality-based explanations of thermodynamic behaviour and evaluate their respective success. My somewhat sober conclusion will be that the boldest version fails for technical reasons (having to do with the mathematical structure of the theory), while more prudent versions leave unanswered essential questions.

Before delving into the discussion, two disclaimers are in order. First, this paper only deals with the role typicality plays in explaining the approach to equilibrium; what typicality has to offer in response to other problems in SM, in particular to the question of how to reconcile the Gibbsian with the Boltzmannian approach, needs to be discussed elsewhere. Second, typicality has also been invoked in other contexts, for instance in Bohmian mechanics (Dürr et al., 1992; Dürr, 2001; Galvan, 2006) and in quantum SM (Goldstein et al., 2006). The use of typicality in these theories is beyond the scope of this paper, which is concerned only with classical Boltzmannian SM.

4.2 Classical Boltzmannian SM

Consider a system consisting of n classical particles with three degrees of freedom each. The state of this system is specified by a point x, also referred to as the system's microstate, in its $6n$-dimensional phase space Γ, which is endowed with the 'standard' Lebesgue measure μ_{L}.[1] The dynamics of the system is governed by Hamilton's equations, which define a measure preserving flow ϕ_t on Γ, meaning that for all times t, $\phi_t : \Gamma \rightarrow \Gamma$ is a one-to-one mapping such that $\mu(R) = \mu(\phi_t(R))$ for all regions $R \subseteq \Gamma$. The system's microstate at time t_0 (its 'initial condition'), $x(t_0)$, evolves into $x(t) = \phi_t(x(t_0))$ at time t. In a Hamiltonian system energy is conserved and hence the motion of the system is confined to the $6n - 1$ dimensional energy hypersurface Γ_{E}. The measure μ_{L} can be restricted to Γ_{E}, which induces a natural invariant measure μ on Γ_{E}.

To each macrostate M_i, $i = 1, ..., m$ (where m is finite), of the system, which is characterised by the values of macroscopic parameters such as volume, local pressure and local temperature, there corresponds a set of so-called micro-regions Γ_{M_i} consisting of all $x \in \Gamma$ for which the macroscopic variables assume the values characteristic for M_i. The Γ_{M_i} together form a partition of Γ_{E}, meaning that they do not overlap and jointly cover Γ: $\Gamma_{M_i} \cap \Gamma_{M_j} = \oslash$ for all $i \neq j$ and $i, j = 1, ..., m$, and $\Gamma_{M_1} \cup ... \cup \Gamma_{M_m} = \Gamma_E$, where '$\cup$', '$\cap$' and '$\oslash$' denote set theoretic union, intersection and the empty set respectively.

The Boltzmann entropy of a macrostate M_i is defined as $S_{\mathrm{B}}(M_i) := k_{\mathrm{B}} \log[\mu(\Gamma_{M_i})]$, where k_{B} is the so-called Boltzmann constant. Given this, we define the Boltzmann entropy *of a system* at time t, $S_{\mathrm{B}}(t)$, as the entropy of the system's macrostate at t: $S_{\mathrm{B}}(t) := S_{\mathrm{B}}(M_{x(t)})$, where $x(t)$ is the system's microstate at t and

[1] For compact presentations of Boltzmann's account see Goldstein (2001), Goldstein and Lebowitz (2004), Lebowitz (1993a, b, 1999).

$M_{x(t)}$ is the macrostate corresponding to $x(t)$ (i.e. $M_{x(t)}$ is that M_i for which it is the case that $x(t) \in \Gamma_{M_i}$ at t).

The idea now is that the behaviour of $S_B(t)$ mirror the behaviour of the thermodynamic entropy S_{TD}; that is, it should increase with time t and reach its maximum at equilibrium. Explaining why and how this happens is the central question the Boltzmann approach needs to answer.[2]

Explaining why entropy increases makes sense only if it is far below its equilibrium value to begin with. That this is the case is the subject matter of the so-called past hypothesis, the postulate that the system starts off in a low entropy macrocondition, the 'past state'. Depending on one's stance on reductionism one either takes, with the grand majority of Boltzmannians, the past state to be the Big Bang and the system under investigation to be the entire universe, or, in keeping with the spirit of laboratory physics, one regards states brought about in experimental setups (such as the gas being confined to the left half of the container) as the past state and takes the relevant system to be the gas in the box. How this issue is resolved is an important question in its own right, but it is inconsequential for my discussion of typicality.[3] All that is assumed in what follows is that the system under investigation (whatever it is) be governed by classical Hamiltonian mechanics, isolated from its environment and come into being in a low entropy state. For this reason I adopt a neutral language and from now talk about 'systems', rather than 'the universe', and the 'past state', rather than the 'Big Bang'.

Let M_p and M_{eq} be the past and the equilibrium macrostate, and Γ_{M_p} and $\Gamma_{M_{eq}}$ the respective micro regions (for ease of notation later on I assume, without loss of generality, that macrostates are labelled such that $M_p = M_1$ and $M_{eq} = M_m$). The explanandum then is this: given that the system's macrostate at t_0 is M_p (i.e. given that the system's microstate $x(t_0)$ lies within Γ_{M_p} at t_0), why does the Boltzmann entropy increase as time unfolds and why does the system eventually reach equilibrium (i.e. why does the system's microstate $x(t)$ eventually wind up in $\Gamma_{M_{eq}}$)?

The standard Boltzmannian response is to introduce a probability measure over the M_i and to argue that these probabilities come out such that the system is, in one way or another, overwhelmingly likely to evolve in such a way that entropy increases and it eventually reaches M_{eq} (see Frigg (2008, Section 2) for a discussion of this approach). The main problem with this response is that at some point it inevitably has to invoke ergodicity, a notion which is notoriously beset with problems (Earman and Rédei, 1996). Typicality approaches promise to eschew such commitments and provide an explanation of the approach to equilibrium free of unmanageable notions like ergodicity.

[2] This 'mirroring' need not be perfect and occasional deviations of the Boltzmann entropy from its thermodynamic counterpart are no cause for concern (Callender, 1999, 2001).

[3] If one takes the past state to be the state at the beginning of the universe, there is the further question of whether or not one needs to explain why the world came into being in such a special state. For opposite views on that matter see the contributions of Callender and Price to Hitchcock (2004).

4.3 Typicality and the Approach to Equilibrium

Consider an element e of a set Σ. Typicality is a relational property of e, which e posses with respect to Σ, a property P and a measure ν, often referred to as 'tyicality measure'.[4] Roughly speaking, e is typical if most members of Σ have property P and e is one of them. More precisely, let Π be the subset of Σ consisting of all elements that have property P. Then the element e is typical iff $e \in \Pi$ and $\nu_{\Sigma}(\Pi) := \nu(\Pi)/\nu(\Sigma) \geq 1 - \varepsilon$, where ε is a finite but small positive real number; $\nu_{\Sigma}(\,\cdot\,)$ is referred to as the 'measure conditional on Σ', or simply 'conditional measure'.[5] Derivatively, one can then refer to Π as the 'typical set (with respect to Σ and ν)' and to those elements that possess property P (i.e. the members of Π) as 'typical elements (with respect to Σ, P, and ν)'. Conversely, an element e is atypical iff it belongs to the complement of Π, $\Omega := \Sigma \setminus \Pi$, in which case we refer to Ω as the 'atypical set' and to its members as 'atypical elements'. For instance the number $\pi/4$ is typical with respect to the interval [0, 1], the property 'not being specifiable by a finite number of digits' and the usual Lebesgue measure on the real numbers because it is a theorem of number theory that the set of all numbers that have this property has measure one. Similarly, numbers in the interval $[1/2, 1/2 + \varepsilon/2]$ are atypical in [0, 1] because $[0, 1] \setminus [1/2, 1/2 + \varepsilon/2]$ has Lebesgue measure greater than $1 - \varepsilon$.

The element of interest in SM is a microstate x. Different approaches to SM disagree about the choice of the set Σ and about the selection of a relevant property P; they all agree that the typicality measure is the Lebesgue measure μ (I discuss this assumption in the next Section). In this Section I show that typicality is used in (at least) three different ways to explain why a system like a gas approaches equilibrium and argue that none of them is successful.

Before discussing these approaches an important technical result needs to be stated. Under certain circumstances (I come back to these in Section 4.4) it is the case that $\Gamma_{M_{eq}}$ is the largest of all Γ_{M_i} (relative to the Lebesgue measure μ); in fact, for large n it is vastly larger than the area of all other regions (Ehrenfest and Ehrenfest, 1912, 30). Numerical considerations show that the ratio $\mu(\Gamma_{M_{eq}})/\mu(\Gamma_{M_i})$, where M_i is a 'standard' non-equilibrium macrostate (e.g. one of the kind in which the gas is confined to the left half of the container), is of the magnitude of 10^n (Goldstein, 2001, 43; Penrose, 1989, 403). For want of a better term I refer to this matter of fact as the 'dominance of the equilibrium macrostate'.

---

[4] Tyicality measures often are, but need not be, probability measures (Zanghì, 2005, 188).

[5] This definition of typicality is adapted from Dürr (1998, Section 2), Lavis (2005, 258), Zanghì (2005, 185), and Volchan (2007, 805). Strictly speaking one should refer to this notion as 'ε-typicality' because the definition depends on the choice of ε and elements that are typical with respect to one choice of ε need not be typical with respect to another. However, nothing in what follows depends on a particular choice of ε and so there is no need to make this dependence explicit. Furthermore, there is an alternative definition of typicality which is stricter than the one adopted here in that it requires $\nu(\Pi)/\nu(\Sigma) = 1$. This definition is unsuitable in the present context because it classifies as atypical certain elements that, from a physics point of view, clearly are typical.

This dominance is then often glossed as implying (or being equivalent to the fact) that for large n, Γ_{E} is almost entirely taken up by equilibrium microstates; in other words, it is glossed as the fact that equilibrium microstates are typical with respect to Γ_{E} and the Lebesgue measure μ (Bricmont, 1996, 146; Goldstein, 2001, 43; Zanghì, 2005, 191, 196). As we shall see in Section 4.4, this gloss is not generally true. However, for the sake of argument I assume throughout this Section that we are dealing only with systems for which this gloss is correct.

Account 1. A first account of why systems behave thermodynamically is suggested by Goldstein (2001) and explains this fact in terms of the dominance of the equilibrium macrostate:

> [Γ_E] consists almost entirely of phase points in the equilibrium macrostate [$\Gamma_{M_{eq}}$], with ridiculously few exceptions whose totality has volume of order $10^{-10^{20}}$ relative to that of [Γ_E]. For a non-equilibrium phase point [x] of energy E, the Hamiltonian dynamics governing the motion [$x(t)$] would have to be ridiculously special to avoid reasonably quickly carrying [$x(t)$] into [$\Gamma_{M_{eq}}$] and keeping it there for an extremely long time – unless, of course, [x] itself were ridiculously special. (Goldstein, 2001, 43–44)[6]

Some pages further down he summarises his view as follows:

> Suppose a system, e.g. a gas in a box, is in a state of low entropy at some time. Why should its entropy tend to be larger at a later time? The reason is basically that states of large entropy correspond to regions in phase space of enormously greater volume than those of lower entropy. (Goldstein, 2001, 49).

These passages allow for two readings. On the first – and more obvious – reading, Goldstein suggests that a system approaches equilibrium *simply because* the overwhelming majority of states in Γ_{E} are equilibrium microstates; in other words, it approaches equilibrium *simply because* equilibrium microstates are typical and non-equilibrium microstates are atypical (with respect to Γ_{E} and μ). This also seems to be Zanghì's view when he writes that

> reaching the equilibrium distribution in the course of the temporal evolution of a system is inevitable due to the fact that the overwhelming majority of microstates in the phase space have this distribution; a fact often not understood by the critics of Boltzmann [...] (Zanghì, 2005, 196; my translation)

This point of view contrasts with one that explains the approach to equilibrium by appeal to specific dynamical properties such as ergodicity or mixing. Goldstein dismisses the view that either of these properties could play any role in the foundation of SM as 'thoroughly misguided' (2001, 45)[7]:

[6] Square brackets indicate that Goldstein's notation has been replaced by the notion used in this paper. I will use this convention throughout.

[7] Albert takes a similar stance and dismisses approaches to the foundations of SM that appeal to ergodicity as 'sheer madness' (2000, 70) and ergodic theory as an enterprise that has 'produced beautiful mathematics' but is ultimately, if we are interested in the foundation of SM, 'nothing more nor less [...] than a waste of time' (*ibid.*).

> Boltzmann's key insight was that, given the energy of a system, the overwhelming majority of its phase points on the corresponding energy surface are equilibrium points, all of which look macroscopically more or less the same. This means that the value of any thermodynamic quantity is, to all intents and purposes, constant on the energy surface, and averaging over the energy surface will thus reproduce that constant value, regardless of whether or not the system is ergodic. (Goldstein, 2001, 45)

This criticism is not specific to ergodicity and could just as well be levelled against any other dynamical property that a system could posses. This suggests that dynamical considerations are regarded as irrelevant for an explanation of the approach to equilibrium and a system eventually reaches equilibrium just because equilibrium conditions are typical.

This is not so. In general there is no reason to assume that points in an atypical set have to evolve into a typical set; typical states do not per se 'attract' atypical states. Uffink (2007, 979–980) provides the following example. Consider a trajectory $x(t)$, i.e. the set $\{x(t) = \phi_t(x(t_0)) \mid t \in [t_0, \infty)\}$, a set of measure zero in Γ_{E}. Its complement, the set $\Gamma_{\mathrm{E}} \setminus x(t)$ of points not laying on $x(t)$, has measure one. Hence the points on $x(t)$ are atypical while the ones not on $x(t)$ are typical (with respect to Γ_{E}, μ, and the property 'being on $x(t)$'). But from this we cannot conclude that a point on $x(t)$ eventually has to move away from $x(t)$ and end up in $\Gamma \setminus x(t)$; in fact the uniqueness theorem for solutions tells us that it does not (for a discussion of uniqueness theorems see Arnold (2006)). The moral is that non-equilibrium states do not evolve into equilibrium states simply because there are overwhelmingly more of the latter than of the former, i.e. because the former are atypical and the latter are typical. It does not somehow lie in the 'nature' of atypical states to evolve into typical ones.

One might reply that this example does not fit the mould because the claim is not that *any* typical set is such that trajectories having atypical initial conditions eventually wind up in the typical set; the claim rather is that this is a special feature of the set that is typical with respect to the property of being an equilibrium state.

But why should this be so? Equilibrium is defined solely in terms of macroscopic quantities and without any reference to the system's dynamics. Why, then, should it be the case that the micro-dynamics is such that it carries atypical points into the typical set? The fact that the there are many more typical than atypical points does not in any way imply that the latter have to evolve towards the former. In other words, if a system is in an atypical microstate (which it is by the Past Hypothesis), it does not evolve into an equilibrium microstate *just because* the latter are typical. Whether or not this happens depends on the dynamics of the system, and whether the dynamics is of the right kind is a question that cannot be answered by appeal to measure-theoretic arguments about the system's macrostate structure.

Account 2. If a given non-equilibrium microstate eventually evolves into an equilibrium microstate this happens due to the dynamics of the system, which is determined by equations of motion and the system's Hamiltonian. Hence an account that disregards dynamical consideration and tries to explain the approach to equilibrium *solely* by appeal to considerations having to do with the measures of macrostates is doomed to failure. So the question remains: what dynamical conditions does the system have to satisfy for it to approach equilibrium? On the second reading of the first of the above quotations, Goldstein offers at least the beginning of an answer to exactly this question when he restricts his claim that systems reach equilibrium quickly to a dynamics that is not 'ridiculously special' and to initial conditions that are not 'ridiculously special' either. This clearly is a condition on the dynamics of the system, albeit not a very informative one because Goldstein does not tell us what he means by 'ridiculously special'. The only indication of what non-ridiculously-specialness could consist in is contained in the following remark:

> The dynamics of the system prefers a given equilibrium point neither more nor less than it prefers any other given phase point, even a specific far-from equilibrium phase point, corresponding say to the leftmost snapshot. (Goldstein, 2001, 42)

Stripped of its anthropomorphisms, this passage might be read as saying that sooner or later $x(t)$ visits every point in Γ_{E}, which is just Boltzmann's original definition of ergodicity (see Sklar, 1993, 160). However, as is well known, there are no trajectories that satisfy this condition (in phase spaces of more than one dimension). An obvious way to fix the problem would be to substitute the modern definition of ergodicity (roughly that the system's state visits every subset of finite measure at some point and spends an amount of time in it that is proportional to the subset's volume) for Boltzmann's. However, given Goldstein's polemic against ergodicity this can hardly be the dynamical condition that he envisages.

So the crucial question is still unanswered: what are the properties of the dynamics of a system that exhibits the right kind of entropy increasing behaviour? Surprisingly, this question has hardly attracted any attention so far; in fact, I am aware of only two proposed answers. The first is due to Bricmont, who tentatively puts forward the suggestion

> that some form of mixing is important for the approach to equilibrium to take place (after all, for the harmonic oscillator we have neither approach to equilibrium nor any form of mixing), but only in some kind of reduced phase space ($\mathbf{R}^2$ here [i.e. in the example of a system of N uncoupled anharmonic oscillators of identical mass]), determined by the macroscopic variables. (Bricmont, 2001, 16)

Bricmont himself is clear that this is only a 'suggestion' that he does 'not know how to formulate precisely' (ibid.), and that it is still an open question whether, and if so how, this suggestion can be generalised to yield a general condition that would do the work that ergodicity (with respect to the entire phase space) was supposed to in the orthodox approach to SM.

The second suggestion departs from Lavis' (2008, Section 2) observation that the Kac ring model, which, as is well known, behaves theormodynamically while failing to be ergodic (see also Bricmont 2001, 10–14), in fact has an ergodic decomposition. This suggests that having such a decomposition plays a part in explaining the approach to equilibrium. Again, the difficulty is that this observation is made in the context of a particular example and it is not at present clear whether, and if so how, it could be generalised to yield a general necessary condition for the approach to equilibrium to take place.

These two suggestions point in the right direction. The question is whether they can be given a precise and general formulation, and whether it is possible to show that realistic systems actually obey one of them. A further question concerns the relation between these (and potential other) conditions. Is one a special case of the other? If not, do they belong to a family of conditions that have certain important features in common? These are important questions that should be addressed in the future.

Account 3. An altogether different line of argument can be found in Lebowitz (1993a, b, 1999) and Goldstein and Lebowitz (2004) and (possibly) Zanghì (2005, Section 2.4.4). The difference lies in the fact that what I refer to as Account 3 focusses on the internal structure of the micro-regions Γ_{M_i} rather than the entire phase space. The core of this view is captured in the following quotation:

> By "typicality" we mean that for any $[\Gamma_{M_i}]$ [...] the relative volume of the set of microstates $[x]$ in $[\Gamma_{M_i}]$ for which the second law is violated [...] goes to zero rapidly (exponentially) in the number of atoms and molecules in the system. (Goldstein and Lebowitz, 2004, 57)[8]

This definition contains different elements that need to be distinguished for the discussion to follow. Let us begin by introducing some notation. $\Gamma_{M_i}^{(++)}$ is the subset of Γ_{M_i} containing all those x that lie on trajectories that come into Γ_{M_i} from a macrostate of higher entropy and that leave Γ_{M_i} entering into a macrostate of higher entropy; $\Gamma_{M_i}^{(+-)}$, $\Gamma_{M_i}^{(-+)}$ and $\Gamma_{M_i}^{(--)}$ are defined accordingly. These four subsets form a partition of Γ_{M_i}.[9] Furthermore, $\Gamma_{M_i}^{(+)} := \Gamma_{M_i}^{(++)} \cup \Gamma_{M_i}^{(-+)}$ and $\Gamma_{M_i}^{(-)} := \Gamma_{M_i}^{(+-)} \cup \Gamma_{M_i}^{(--)}$ are the subsets of Γ_{M_i} that have a higher and lower future entropy respectively.

The microstate $x \in \Gamma_{M_i}$ has the property 'being entropy increasing' ('*I*' for short) iff it lies on a trajectory that moves into a microstate of higher entropy when leaving Γ_{M_i}. Hence, x has property I iff $x \in \Gamma_{M_i}^{(+)}$. Entropy increasing states are typical in Γ_{M_i} iff $\mu_i(\Gamma_{M_i}^{(+)}) \geq 1 - \varepsilon$, where $\mu_i(\cdot) := \mu(\cdot)/\mu(\Gamma_{M_i})$ is the Lebesgue measure relative to Γ_{M_i}.

A *system* possesses the property of being 'globally entropy increasing' ('*GI*' for short) iff entropy increasing states are typical in every Γ_{M_i} except the equilibrium macrostate itself (because, trivially, once the system has reached equilibrium entropy cannot further increase). Goldstein and Lebowitz's explication of typicality (quoted above) amounts to saying that the system is *GI*. This can be seen as follows. In technical terms, Goldstein and Lebowitz's condition is $\lim_{n\to\infty} \mu_i(\Gamma_{M_i}^{(-)}) = 0$ for all i. Since the $\Gamma_{M_i}^{(++)}$, etc., form a partition of Γ_{M_i}, this is equivalent to $\lim_{n\to\infty} \mu_i(\Gamma_{M_i}^{(+)}) = 1$ for all macrostates M_i except the equilibrium macrostate. If we now assume (reasonably) that for $n \simeq 10^{23}$ we are already 'close' to the limit it follows that $\mu_i(\Gamma_{M_i}^{(+)}) \geq 1 - \varepsilon$ for some small but finite ε.

---

[8] Explications of typicality very similar to this one can be found in Lebowitz (1993b, 7–8; 1999, 348).

[9] I neglect the possibility that there maybe x that come from or move into microstates of the *same* entropy. These cases could be accounted for by introducing the subsets $\Gamma_{M_i}^{(0+)}$, etc., and rephrasing the argument accordingly. One can easily see that this would not alter the conclusions that I reach and I therefore neglect them in the interest of ease of discussion and notion.

We now face two questions. First, under what circumstances is it the case that a system is *GI*? Second, assuming we have a satisfactory answer to the first question, do we then have a good explanation for why the system approaches equilibrium? I discuss these questions in turn.

Goldstein and Lebowitz offer the following answer to the question of when a system is *GI*:

> Boltzmann then argued that given this disparity in sizes of different *M*'s [i.e. the above-mentioned dominance of the equilibrium macrostate], the time evolved [$M_{x(t)}$] will be such that [$\mu(M_{x(t)})$] and thus [$S_B(t)$] will *typically* increase in accord with the law. (2004, 57)

They do not reference the work of Boltzmann they have in mind and so we have to work with their paraphrase of what they take to be Boltzmann's view. The argument seems to be that if it is the case that the ratio $\mu(\Gamma_{M_{\mathrm{eq}}})/\mu(\Gamma_{M_i})$, where M_i is a 'standard' non-equilibrium macrostate, is large (i.e. is of the magnitude of 10^n), then the system is *GI*.

This is incorrect. Dominance of the equilibrium macrostate and being *GI* are compatible with each other, but the latter does not follow from the former. From the fact that Γ_{E} *as a whole* is almost entirely filled with equilibrium microstates and that therefore the measure of $\Gamma_{M_{\mathrm{eq}}}$ is 10^n times the one of other macro-regions, it just does not follow that *within* every macro region $\Gamma_{M_i}^{(+)}$ is typical. In fact, the dominance of the equilibrium macrostate is compatible, in principle, with it being the case that $\mu_i(\Gamma_{M_i}^{(+)}) \ll \mu_i(\Gamma_{M_i}^{(-)})$ for many low entropy macrostates M_i, in which case the system would fail to be *GI*. And the point is not one about there being the possibility of one or two macrostates behaving strangely and the system being 'a little bit non-*GI*'; it could be the case equilibrium microstates are typical with respect to Γ_E *as a whole*, while entropy increasing behaviour is *atypical* in *all* low entropy macrostates.

That Account 3 fails is no surprise; whether or not a system is *GI* depends both on its dynamics and the construction of the macrostates and so it would be something of a miracle if one could prove systems to be *GI* without even mentioning either of the two.

Given that we do not have a general argument for the claim that relevant systems are *GI*, the best we can do is look at examples. And here the evidence is mixed. One can show that the Kac ring model is *GI* (Lavis, 2005, 259). However, *GI* seems to fail in other examples. Numerical considerations show that entropy increasing microstates are not typical within the low entropy macrostates of the baker's gas (as David Lavis pointed out to me in personal communication). So *GI* is not a trivial condition and there is a substantial question under which circumstances it holds.

There are also problems as regards the second question. To begin with, even if a system were *GI* it could still be the case that an approach to equilibrium would not take place. The problem is the following. Assume that the system is in macrostate M_i at time t_1 and evolves into a macrostate M_j of higher entropy at time t_2 (without passing through any other macrostates in-between). Furthermore assume that in both M_i and M_j entropy increasing microstates are typical. By construction, all states that evolve into M_j from M_i have to be either in $\Gamma_{M_j}^{(-+)}$ or in $\Gamma_{M_j}^{(--)}$. In which one of these a particular $x \in M_i$ ends up is determined by the dynamics of the system, and it is possible that under certain dynamical laws most $x \in M_i$ end up moving into $\Gamma_{M_j}^{(--)}$. In this case most trajectories that are compatible with the system's actual past history move towards macrostate of *lower* entropy after t_2, despite the fact that $\Gamma_{M_j}^{(+)}$ is typical in Γ_{M_j}.

So we need to add the further constraint that the dynamics of the system is such that for all (or at least most) contiguous macrostates M_i and M_j, where M_i has lower entropy than M_j, it be the case that the overwhelming majority of microstates in $\Gamma_{M_i}^{(+)}$ move into $\Gamma_{M_j}^{(-+)}$. What condition could assure that this is the case? A possible answer to this question (or, rather, part of an answer) might be that the system has to show Goldilocks mixing (Earman, 2006, 406). Although Earman discusses Goldilocks mixing in a different context and does not suggest that it is a solution to the current problem, it might at least be worth considering whether Goldilocks mixing, probably in conjunction with other conditions, proves useful in solving the problem at hand.

Furthermore there is the problem that most of the states that lie on trajectories that move towards higher entropy macrostates also have a *high* entropy *past*, i.e. behave un-thermodynamically.[10] This can be seen as follows. By assumption $\Gamma_{M_i}^{(+)}$ is typical, i.e. $\mu(\Gamma_{M_i}^{(+)}) \geq 1 - \varepsilon$, and hence $\mu(\Gamma_{M_i}^{(-)}) < \varepsilon$. Since $\Gamma_{M_i}^{(-)} = \Gamma_{M_i}^{(--)} \cup \Gamma_{M_i}^{(+-)}$, we also have $\mu(\Gamma_{M_i}^{(+-)}) < \varepsilon$. The time reversal invariance of the Hamiltonian dynamics implies $\mu(\Gamma_{M_i}^{(-+)}) = \mu(\Gamma_{M_i}^{(+-)})$ and therefore $\mu(\Gamma_{M_i}^{(-+)}) < \varepsilon$. With $\Gamma_{M_i}^{(+)}) = \Gamma_{M_i}^{(-+)} \cup \Gamma_{M_i}^{(++)}$ we obtain $\mu(\Gamma_{M_i}^{(++)}) \geq 1 - 2\varepsilon$. Hence the typicality of $\Gamma_{M_i}^{(+)}$ is hardly relevant to thermodynamic behaviour because the overwhelming majority of states in $\Gamma_{M_i}^{(+)}$ do not exhibit the desired behaviour (i.e. they belong to $\Gamma_{M_i}^{(++)}$ and hence have a high entropy past).

Remedy can be found in Albert (2000, Chapter 4), who suggests solving the problem by conditionalising on the past hypothesis (Albert does not put his argument in terms of typicality and uses probability language instead; what I am presenting here is an adaptation of his point to the present context). In technical terms that means that rather than pondering the question of whether microstates with high entropy future are typical with respect to the entire set Γ_{M_i} we should require that this be the case with respect to $\Gamma_{M_i} \cap \phi_t(\Gamma_{M_\mathrm{p}})$. The question now is whether states which evolve into macrostates of higher entropy are typical within *that* set.

And now we are back to the above problem, namely that this question cannot be answered without taking the dynamics of the system into account. There is nothing, in principle, to rule out that all states that satisfy this condition evolve into $\Gamma_{M_j}^{(--)}$ once they leave Γ_{M_i}, in which case the system's entropy decreases once the

[10] A point to this effect was first made by Ehrenfest and Ehrenfest-Afanassjewa (1912, 32–34). However, their argument is based on an explicitly probabilistic model and so its relevance to deterministic dynamical system is tenuous.

states move from Γ_{M_j} into the next macrostate. Albert (2000, 67, 81–85, 94–96) suggest ruling out that this happens by requiring that microstates that lead to unthermodynamic behaviour are scattered in tiny clusters all over Γ_{M_i}. This is an interesting suggestion, but, again, there are neither a priori reasons nor plausibility arguments to suggest that this generally is the case in relevant systems. Whether or not this 'scattering condition' holds depends on the details of the dynamics and the construction of the macrostates, and merely asserting that the condition does hold is simply begging the question.

4.4 Further Qualms

There are five further problems for an approach to SM based on the notion of typicality: the justification of the Lebesgue measure as the relevant typicality measure, that the equilibrium macrostate may not be typical, that in interacting systems the largest macrostate may not be the equilibrium macrostate, the reliance on measures in general, and objections to the explanatory power of typicality even where it can be had. I will discuss each of these in turn.

First. Typicality judgements in all three accounts I have distinguished are made relative to the Lebesgue measure μ. How can this be justified? Dürr (1998, Section 3) emphasises that the crucial criterion for the choice of a typicality measure is invariance over time. What is typical at some time t also has to be typical at some earlier or later time t'. In the context of SM this means that the typicality measure has to be invariant under the dynamics of the system (given by the flow ϕ_t). As we have seen in the Section 4.2, the Lebesgue measure satisfies this criterion and therefore seems to be a natural choice.

Things are more involved, however. As Zanghì (2005, 189) points out, the Lebesgue measure μ may not be the only invariant measure in a *particular* system. For any specific Hamiltonian (equivalently for any specific ϕ_t) there could also be invariant measures other than the Lebesgue measure whose explicit form depends on the details of the dynamics. Zanghì then points out that what makes the Lebesgue measure special is the fact that it is the only generic invariant measure, meaning that it is the only measure that is invariant under *all* Hamiltonian flows.

It is not clear, however, that this fact is relevant for the problem at hand. Each system is governed by one, and only one, Hamiltonian and it is therefore not clear why the fact that the Lebesgue measure is the only measure that is invariant under *all* Hamiltonians is relevant for typicality judgements in *this* system. If it happens that there is a measure μ' which is invariant under the dynamics of the system under investigation and which is non-equivalent to the Lebesgue measure, why should we not make typicality judgments about this system with respect to μ'? This question is particularly pressing for those – like most Boltzmannians – who take the relevant system to be the universe as a whole and the past state the Big Bang. There is only one universe and there is only one Hamiltonian flow in this universe. What reason could there be to prefer μ to μ' to make typicality judgements in *this* universe?

There is no obvious answer to this question. But maybe none is needed. A similar issue arises in the case of the Galton board. Maudlin (2007) points out that atypical initial conditions have measure zero and hence typicality judgments remain unaltered under a change of measures as long as the alternative measure μ' is absolutely continuous with the Lebesgue measure μ.[11] So there is actually no need to worry about the question of picking the 'right' measure because under all choices the same sets come out as typical, which is all we need.

It is not clear whether this strategy is available in SM. First, Maudlin's argument only applies to measures that are absolutely continuous with the Lebesgue measure. So we would need an argument for the conclusion that all invariant measures have this property. This may or may not be the case; at any rate it is not *a priori* clear that this is so.[12] Second, one would have to show that it is indeed the case that all atypical sets have Lebesgue measure zero. Again, this is not evidently so. Even in a simple system like the Galton Board a host of drastic idealisations are needed to reach this conclusion (for instance, one has to assume that the board is infinitely long and that all the nails are perfectly symmetrical), which then still is only supported by a plausibility argument and not a rigorous proof. It is not clear that idealisations of this sort can be made of our universe, and even if they can this may not yield the desired result because the dynamics of our universe is much more complex than the one of the Galton Board. Hence, it is at least a possibility that some sets of finite measure are atypical. If this is the case and if there is an invariant measure μ' (which could even be absolutely continuous with μ), it might be the case that μ' assigns high weights to sets that come out small under μ, which would reverse typicality judgements. Hence what is typical with respect to μ would come out to be atypical with respect to μ' and vice versa. There is no a priori reason to rule out this possibility.

Second. A further difficulty concerns the dominance of the equilibrium macrostate. As I have briefly mentioned above, from the fact that the equilibrium macrostate is larger than any other macrostate one cannot infer that it is typical. Lavis (2005, 255–258) points out that entropy levels can be degenerate, meaning that there may be more than one macrostate for which the Boltzmann entropy assumes a particular value. More precisely, consider a particular macrostate M_j, construct the set $\{M_i | S_B(M_i) = S_B(M_j), i = 1, ..., m\}$ of all macrostates that have the same entropy as M_j, and let ω_j be the number of macrostates in this set; ω_j is the degeneracy of the entropy value $S_B(M_j)$. The important point is that these degeneracies may be large enough for it to be the case that the non-equilibrium macrostates associated with a particular entropy value *together* take up a larger chunk of the phase space than the equilibrium macrostate; that is, it may be the case

[11] A measure μ' is absolutely continuous with μ iff for any measurable region $A \subseteq \Gamma_E$: if $\mu(A) = 0$ then $\mu'(A) = 0$. More colloquially, a measure μ' is absolutely continuous with another measure μ if it assigns measure zero to all sets that are assigned measure zero by μ, while, possibly, assigning different values to the sets to which μ assigns non-zero measure.

[12] Maybe an defence along the lines of Malement and Zabell (1980) would fit the bill, but this would need to be argued in detail.

that $\omega_j \, \mu(\Gamma_{M_j}) > \mu(\Gamma_{M_{\mathrm{eq}}})$, for some non-equilibrium macrostate M_j. Lavis shows that this is not only a theoretical possibility. He points out that it is exactly what happens in the case of the baker's gas (*ibid.*) and in the Kac ring model (Lavis, 2008, Section 2), in which the proportion of the phase space occupied by the maximum entropy state even decreases as n becomes large. Of course, real systems are neither baker's gases nor Kac rings and so this problem with degeneracies may not surface in more 'realistic' systems. However, whether or not this is the case depends on the details of the system and one would have to show that in the systems of interest no such degeneracies crop up.

Third. So far we have assumed that the equilibrium macrostate is the largest of all macrostates (and the second problem concerns the question of whether this state is typical in Γ_{E}). Although this is usually stated as if it were a general truism, it is proven only for an ideal gas, i.e. a system of non-interacting particles. In broad outline, the reasoning, invented by Boltzmann in 1877 and now usually referred to as the 'combinatorial argument', is as follows (for an in-depth discussion see Uffink (2007, 974–983). Consider the phase-space of one gas molecule; the state of the entire gas (consisting of n molecules) is specified by n labeled points in this space. Now put a grid-like partition on it with the border of the cells running in the directions of the momentum and position axes. Every one of the n points comes to lie within a particular cell of the partition. A specification of which point lies in which cell is called an 'arrangement'; a specification of how many points (no matter which ones) are in each cell is a 'distribution'. Boltzmann then considered how many arrangements are compatible with each distribution and associated the logarithm of this number, W, with the entropy of the system (this can be shown to be equivalent to the definition of the Boltzmann entropy given in Section 4.2). One can then prove that W is proportional to the Lebesgue measure of the region of the n-particle phase space, corresponding to the distribution. By construction it follows that largest macrostate is associated with the largest Boltzmann entropy, and this macrostate is then considered to be the equilibrium macrostate.

However, we should not be mislead by the suggestive use of the word 'entropy'; the argument so far is just a combinatorial exercise and its physical relevance yet needs to be shown. And this is where the crucial assumptions enter. Suppose that the energy of a molecule only depends on the cell in which *it* is (but not on where all the other molecules are) and that the total energy of the system is the sum of these 'individual' energies. Under this assumption (and the further assumption that the number of molecules in each cell is far greater than one) one can prove that the velocity distribution of those phase points that are in the maximum Boltzmann entropy region is the Maxwell-Boltzmann equilibrium distribution. For this reason it is indeed legitimate to associate equilibrium with maximum Boltzmann entropy.

The crucial assumption in this proof is that the entropy of a molecule only depends on the cell in which it is, as this amounts to nothing less than the assumption that there is no interaction between the molecules; in other words, it amounts to assuming that the system is an ideal gas (Uffink, 2007, 976). Hence, for systems that are not ideal gases there is at least a question of whether their equilibrium macrostate can be associated with the largest macrostate. And this is more than an academic point. Most systems, not least the universe as a whole, are not ideal gases, not even approximately, and it is not clear whether in such systems the equilibrium macrostate can legitimately be associated with the largest macrostate (i.e. the one for which the Boltzmann entropy is maximal).

In fact, it is a real option that this is not the case. Consider a system of gravitating particles. These particles attract each other and hence have the tendency to clump together. So if it happens that a large amount of these are distributed evenly over a bounded space, then they will move together and eventually form a lump. However, the volume corresponding to a lump is much smaller than the one corresponding to the original spread out state, and hence it seems that the system evolves from a high to a low entropy state. This conclusion is usually blocked by pointing out that the loss in volume in configuration space is compensated by a corresponding increase in volume in momentum space, and as a result entropy does not decrease after all. But whether this is true depends on the details of the system at hand. There are situations in which this is not the case, for instance one in which all particles end up moving around with almost the same velocity and hence occupy only a small volume of momentum space. So one would need to argue that the systems of interest are not of this kind.[13]

Fourth. One of the main objections against approaches to SM that invoke ergodicity is the so-called 'measure zero problem' (see van Lith (2001) for a discussion). The results of ergodic theory come with the qualification 'almost everywhere' – i.e. everywhere except, perhaps, for a set of measure zero – which is commonly understood as suggesting that sets of measure zero can be ignored because they are somehow 'sparse'. This piece of common wisdom has been criticised as untenable. Sets of measure zero need not be 'small' at all (e.g. the rational numbers have measure zero within the real numbers and yet there are 'many' of them) and, as Sklar (1993, 182–188) points out, a set of measure zero need not be (or even appear to be) negligible if sets are compared with respect to properties other than their measures. For instance, we can judge the 'size' of a set by its cardinality or Bair category rather than by its measure which may lead to different conclusions about a set's 'size'.

This point has to do with the use of measures in general and is not specific to ergodic theory. In fact, because typicality is determined with respect to a measure, approaches to SM appealing to typical behaviour face a very similar problem: sets of measure zero (like the rational numbers) are classified as atypical and it is suggested that these can therefore be neglected. However, echoing Sklar's point, sets that come out as atypical when compared to other sets with respect to their measures may not come out as atypical when compared with respect to some other property (such as their Bair category). So we face the question of what conveys upon measures a privileged status when it comes to judging typicality.

Fifth. The basic strategy of typicality-based approaches is to explain X by pointing out that X is typical. For instance, when asked why a system approaches

[13] See Callender (2010) for a further discussion of the problems that arise in connection with gravity.

equilibrium the proponent of Approach 2 answers that this is because initial conditions that lie on trajectories that approach equilibrium are typical in the set of all initial conditions. It is questionable whether this answer is satisfactory, even if the desired behaviour in fact turns out to be typical. The problem is, again, parallel to one that threatens the ergodic approach. As Sklar (1973, 210–211) points out in his critique of this approach, from the fact that an initial condition lies within a set of measure zero we cannot infer that this initial condition does not occur. Whether the system has a particular initial condition is a factual question, and as such it has to be settled by an appeal to matters of fact and not measures of sets; to explain why the system exhibits entropy increasing behaviour we need an argument for the conclusion that the system indeed started out in a typical initial condition, but that these are of measure (close to) one does not give us such an argument.

But now the significance of typicality seems to have evaporated entirely. All we need to explain a system's *actual* behaviour is that its *actual* initial condition is one which, under the dynamical law governing the system's evolution, evolves in a thermodynamic way. Whether or not this initial condition is also typical is simply irrelevant. So typicality does not play a role in explaining the behaviour of a particular system (like, for instance, our universe).

One could reply that the notion of explanation that underlies this criticism is too metaphysical (in that it implicitly assumes that an explanation of *X* has to show that *X must* happen under the given circumstances) and that a different, less assuming, notion should be applied. An obvious candidate is rational expectability. On this conception of explanation we explain *X* by showing that it is rationally expectable that *X* occurs. This seems to square well with typicality, because if a behaviour is typical we are surely rationally justified in believing that it occurs most of the time. This also squares well with the intuition driving the (probabilistic version of) the covering law account of scientific explanation, according to which we explain *X* if we can show that *X* is very likely to occur.

But even if we are willing to set all the well-known problems of accounts of this sort aside (see Salmon (1992) for survey), such an account would not sit well with the general hostility towards epistemic approaches that permeates this literature, in particular the flamboyant rejection of an epistemic interpretation of probability (see for instance Albert (2000, 64), Loewer (2001, 611), and Goldstein (2001, 48)). But if we reject an epistemic notion of explanation, it remains unclear how we can explain the behaviour of a particular system (*this* universe) by appeal to typicality.

4.5 Conclusion

I have distinguished three different ways in which typicality is used to explain why systems approach equilibrium and argued that none of them is successful. The first is false for mathematical reasons, while the latter two prima facie provide a restatements of the problem rather than a solution because they do not provide dynamical conditions. But even if these difficulties can be solved, there are further conceptual problems. First, all accounts attribute a special status to the Lebesgue measure, but the justifications of this choice do not seem to be conclusive. Second, it is not clear whether the equilibrium macrostate is typical in Γ_E. Third, typicality arguments are usually put forward in the context of ideal gases, and there are serious questions about whether they can be carried over to gravitating systems. Fourth, like approaches based on ergodicity, typicality arguments dismiss sets of measure zero as 'negligible'. It is not clear, however, how this can be justified. Finally, it is questionable whether an appeal to what typically happens has any explanatory force at all when it comes to explaining what happens in a particular system.

Acknowledgements Special thanks goes to David Lavis for many illuminating discussions on SM in general, and the Boltzmannian approach in particular. I also would like to thank Craig Callender, Stephan Hartmann, Carl Hoefer, Wolfgang Pietsch, Charlotte Werndl, and two anonymous referees for valuable comments on earlier drafts. Thanks to Jean Bricmont for a helpful email conversation on his mixing condition discussed in Section 4.3, and to Detlef Dürr for drawing my attention to omissions in my first bibliography. Many thanks to Flavia Padovani for helping me with those passages in Zanghì's chapter that were beyond the reach of my 'FAPP Italian'. Thanks to Mauricio Suárez for organising the workshop at which this paper has first been presented, and thanks to the audiences in Madrid and Oxford for stimulating discussions. Finally, I would like to acknowledge financial support from two project grants of the Spanish Ministry of Science and Education (SB2005-0167 and HUM2005-04369).

References

Albert, D. (2000), *Time and Chance*, Cambridge, MA: Harvard University Press.

Arnold, V. I. (2006), *Ordinary Differential Equations*. Heidelberg: Springer.

Boltzmann, L. (1877), 'Über die Beziehung zwischen dem zweiten Hauptsatze der mechanischen Wärmetheorie und der Wahrscheinlichkeitsrechnung resp. den Sätzen über das Wärmegleichgewicht. *Wiener Berichte* 76, 373-435. Reprinted in F. Hasenöhrl (ed.): *Wissenschaftliche Abhandlungen*. Leipzig: J. A. Barth 1909, Vol. 2, 164-223.

Bricmont, J. (1996), Science of Chaos or Chaos in science?, In P. R. Gross, N. Levitt, and M. W. Lewis (eds.), *The Flight from Science and Reason*, Annals of the New York Academy of Sciences, New York, NY Vol. 775, 131-175.

Bricmont, J. (2001), Bayes, Boltzmann and Bohm: Probabilities in physics, in Bricmont et al., 3–21.

Callender, C. (1999), Reducing thermodynamics to statistical mechanics: The case of entropy, *Journal of Philosophy* 96, 348–373.

Callender, C. (2001), Taking thermodynamics too seriously, *Studies in the History and Philosophy of Modern Physics* 32, 539–53.

Callender, C. (2010), The past hypothesis meets gravity, forthcoming In G. Ernst and A. Hüttemann (eds.), *Time, Chance and Reduction. Philosophical Aspects of Statistical Mechanics*. Cambridge: Cambridge University Press 2010, pp. 34–58.

Dürr, D. (1998), Über den Zufall in der Physik, manuscript presented at the 1998 Leopoldina Meeting in Halle. Available at http://www.mathematik.uni-muenchen.de/%7Eduerr/Zufall/zufall.html

Dürr, D. (2001), *Bohmsche Mechanik als Grundlage der Quantenmechanik*. Berlin: Springer.

Dürr, D., S. Goldsein and N. Zanghì (1992), Quantum Equilibrium and the Origin of Absolute Uncertainty, *Journal of Statistical Physics* 67, 843–907.

Earman, J. (2006), The “past hypothesis”: Not even False, *Studies in History and Philosophy of Modern Physics* 37, 399–430.
Earman, J. and Rédei, M. (1996), Why Ergodic Theory Does Not Explain the Success of Equilibrium Statistical Mechanics, *British Journal for the Philosophy of Science* 47, 63–78.
Ehrenfest, P. and Ehrenfest-Afanassjewa, T. (1912/1959), *The Conceptual Foundations of the Statistical Approach in Mechanics*. New York NY: Dover 2002. (First published in German in 1912; first English Translation 1959.)
Frigg, R. (2008), A Field Guide to Recent Work on the Foundations of Statistical Mechanics, In Rickles, D. (ed.), *The Ashgate Companion to Contemporary Philosophy of Physics*. London: Ashgate, 99–196.
Galvan, B. (2006), Typicality vs. probability in trajectory-based formulations of quantum mechanics, *Foundations of Physics* 37, 1540–1562.
Goldstein, S. (2001), Boltzmann’s Approach to Statistical Mechanics, in: Bricmont et al. 2001, 39–54.
Goldstein, S. and Lebowitz, J. L. (2004), On the (Boltzmann) entropy of non-equilibrium systems, *Physica* D 193, 53–66.
Goldstein, S., Lebowitz, J. L. Tomulka, R. and Zanghì, N. (2006), Canonical Typicality, *Physical Review Letters* 96, Issue 5.
Hitchcock, C. (ed.) (2004) *Contemporary Debates in Philosophy of Science*. Malden, MA: Blackwell.
Lavis, D. (2005), Boltzmann and Gibbs: An attempted reconciliation, *Studies in History and Philosophy of Modern Physics* 36, 245–273.
Lavis, D. (2008), Boltzmann, Gibbs and the concept of equilibrium, forthcoming in *Philosophy of Science*, 75 (December 2008), pp. 682–696.
Lebowitz, J. L. (1993a), Boltzmann’s entropy and time’s arrow, *Physics Today*, September Issue, 32–38.
Lebowitz, J. L. (1993b), Macroscopic laws, microscopic dynamics, time’s arrow and Boltzmann’s entropy, *Physica* A 194, 1–27.
Lebowitz, J. L. (1999), Statistical mechanics: A selective review of two central issues, *Reviews of Modern Physics* 71, 346–357.
Loewer, B. (2001), Determinism and chance, *Studies in History and Philosophy of Modern Physics* 32, 609–629.
Malament, D. B. and Zabell, S. L. (1980), Why Gibbs phase averages work, *Philosophy of Science* 47, 339–349.
Maudlin, T. (2007), What could be objective about probabilities?, *Studies in History and Philosophy of Modern Physics* 38, 275–291.
Penrose, R. (1989), The *Emperor’s New Mind*. Oxford: Oxford University Press.
Salmon, W. (1992), Scientific Explanation, In M. Salmon, et al. (eds.), *Introduction to the Philosophy of Science*. Cambridge: Hackett, 7–23.
Sklar, L. (1973), Statistical explanation and ergodic theory, *Philosophy of Science* 40, 194–212.
Sklar, L. (1993), *Physics and Chance. Philosophical Issues in the Foundations of Statistical Mechanics*. Cambridge: Cambridge University Press.
Uffink, J. (2007), Compendium of the foundations of classical statistical physics, In J. Butterfield and J. Earman (eds.), *Philosophy of Physics.* Amsterdam: North Holland, 923–1047.
van Lith, J. (2001), Ergodic theory, interpretations of probability and the foundations of statistical mechanics, *Studies in History and Philosophy of Modern Physics* 32, 581–594.
Volchan, S. B. (2007), Probability as Typicality, *Studies in History and Philosophy of Modern Physics* 38, 801–814.
Zanghì, N. (2005), I Fondamenti concettuali dell’approccio statistico in Fisica, In V. Allori, M. Dorato, F. Laudisa and N. Zanghì (eds.), *La Natura Delle Cose. Introduzione ai Fundamenti e alla Filosofia della Fisica.* Roma: Carocci.

Part II
Causes

Chapter 5
From Metaphysics to Physics and Back: the Example of Causation

Federico Laudisa

5.1 Introduction

It was around a century ago when a ban was heard to come from several different philosophical voices. The target was the very notion of causation. It seemed as though philosophy, held to be born in the ancient Greece exactly as that sort of investigation which finally emerges out of myth and religion and becomes *cognitio per causas*, reached its full maturity only to realize that our understanding of science can and must do without causes, effects and all that. We can hardly underestimate the extent to which the development of the mathematical and physical sciences between the end of the nineteenth century and the beginning of the twentieth century contributed to such circumstance. Quantum mechanics, in particular, or better the ideology that underpinned the development of its standard theoretical formulation strongly enhanced an acausal (or even anticausal) attitude toward the foundations of science in general, and of physics in particular.

Fortunately, we are at present far from that old ban. The causality Renaissance we have been experiencing since the early eighties of the last century has had a twofold consequence. On the one hand we feel justified in attempting to assess still the fruitfulness of causal concepts and frameworks, on the other hand we are aware that in doing this we are not chasing after *the* Holy Grail of Causation but rather we are developing a vast number of causal models, with varying degrees of applicability and formalization. Namely, there seems to be a growing consensus on the need to entertain a modest, *pluralist* and open attitude toward the role of causality in science, a need on the basis of which causality itself acquires more and more the character of a cluster and context-dependent concept[1]. This consensus itself leans toward – no matter what kind of view of causation one might adopt – an epistemic attitude rather than a metaphysical one If in other words we agree to be sensitive about the

F. Laudisa (✉)
Department of Human Sciences, University of Milan-Bicocca, Piazza dell'Ateneo Nuovo 1, 20126 Milan, Italy
e-mail: federico.laudisa@unimib.it

[1] See for instance Hitchcock (2007).

M. Suárez (ed.), *Probabilities, Causes and Propensities in Physics*, Synthese Library 347, DOI 10.1007/978-1-4020-9904-5_5,

context-dependence of causal notions, in order for these notions to be meaningful in the philosophical investigations in the foundations of special sciences, there appears to be less and less ground for defending strongly metaphysical theories of causation, although it must be stressed that the epistemic bias is by no means a necessity.

In the spirit of what Hitchcock defines as *methodological pluralism* in its scientific version, a pluralist attitude should also prevent from conceiving the relevance or irrelevance of causal modelling in a uniform way irrespective of the different scientific domains in which causal reasonings are employed. It appears much more reasonable to think that every scientific theory – or every class of interrelated theories in a given broad scientific domain – sets somehow its own standards of explanatory value whenever some sort of causal talk is put to work in order to account for phenomena (be they natural or social). Thus it might be the case that the search for causal models or explanations turns out to be not only useful but essential in a certain scientific domain, and irrelevant or indeed deeply misplaced in others.

Of course the question is particularly pressing for the foundations of physics and of quantum mechanics in particular. As far as the latter is concerned the situation is especially controversial. For on the one hand the formal development of quantum mechanics in the first half of the last century gave rise – for motivations that are highly controversial themselves – to a dismissal of any investigation on the ontological status of the entities that the theory was supposed to be about, and hence to a dismissal of any interest for causality in quantum mechanics, both in a general and in a more specific sense (namely, in terms of specific theories of causation). This dismissal was grounded to a large extent on the ill-based argument according to which it was Nature itself that proved that we cannot describe a 'quantum reality' – whatever it might turn out to be – so that any causal attitude toward what quantum mechanics was meant to lead to a dead end. In our times it has become much more evident that such arguments were not forced by Nature but were rather the outcome of a 'brainwashing' (as Gell-Mann used to call it) with no firm theoretical grounds.

If causal talk in quantum mechanics is no more anathema, however, this need not mean that it crystal-clear what it takes exactly to approach quantum phenomena in a causal vein. The main obstacle is that quantum mechanics, unlike relativistic theories, has no basic formulation that is completely free from problems and ambiguities: as is well known, if we assume the theory be a complete description of the states and properties of micro-objects – let me be vague on what is to be a *micro*-object – then we have to tell some convincing stories about several hot points, the first of which are of course the measurement problem and the role of entanglement and nonlocality.

Also on the basis of recent classifications of theories of causation, in Section 5.2 I will draw some general remarks mainly of a methodological character, concerning the sort of questions that seem to naturally arise when the relation between nonlocality and causation is taken into account, whereas in the Section 5.3 I will review the conditions under which nonlocality can be shown to seriously challenge the no-action-at-a-distance requirement that special-relativistic theories are usually thought to embody. In this connection I will turn then to recent work on causal models of EPR. Over and above the specific merits of these models – mainly concerning the refutation of ‘impossibility claims’ about causal models of quantum correlations – a question arises: what sort of conceptual advantage do we obtain in producing causal models for such correlations *in absence of a deeper understanding of the overall structure of the theory*? I will argue that the only way toward such an understanding may be to cast in advance the problems in a clear and well-defined interpretational framework – which means primarily to specify the ontology that quantum theory is supposed to be about – and *after* to wonder whether problems that seemed worth pursuing still are so in the framework.

As a consequence, in the last two sections I will refer to GRW and Bohmian formulations and quantum mechanics, in order to emphasize essentially two points:

(i) the discussion on causality in quantum mechanics should be cast by using the conceptual resources allowed by ontologically unambiguous interpretations of quantum mechanics and *not* on the background of its ‘orthodox’ – hence vague – formulation;
(ii) the interpretation-dependence of causal reasoning in quantum mechanics implies different approaches to causality in (the different versions of) GRW and Bohmian formulations.

5.2 Prolegomena on Causation and Quantum Nonlocality

In the natural domain (as well as in the social or psychological one), puzzling phenomena call for an explanation, and there is little doubt that the connection among quantum events across spacetime – known as *non-locality* – is indeed puzzling. Events that we might reasonably consider mutually independent, according to our best theory of space and time, turn out to influence each other. But as soon as we try to understand what this ‘influence’ could amount to, we find ourselves in deep physical and philosophical troubles, and if we attempt to investigate the connection between non-locality and causation, the situation may become even more complicated. For if for the sake of the argument we assume we have a vague intuition of what non-locality might be, several are the questions worth asking. Is a *causal* view of non-locality itself possible and useful? In particular, can the nature of quantum non-locality be somehow clarified by viewing it as grounded in some (perhaps unfamiliar) sort of causation? Which properties should this sort of causation satisfy?

There are two preliminary and general circumstances that need to be taken into account but that, at the same time, contribute to make the picture unclear. First, there seem to be different ways in which non-locality is manifested in quantum mechanics. Second, the notion of causation itself is far from being understood in an univocal and uncontroversial sense. The intuition according to which the occurrence of a physical event A determines (produces, brings about, raises the probability of, ...) the occurrence of a distinct physical event B – in which case A is said to be the ‘cause’ of B – can be represented differently in different causal theories.[2] Within the

[2] See for instance Williamson (2007).

physicists' community, for instance, it is assumed – tacitly or not – that events recognized to be causes must be temporally prior to their alleged effects, and the causal doctrine based on this assumption is sometimes referred to as 'relativistic causality'. This terminology is itself biased, however, since it takes for granted that special relativity provides the strongest possible support for this assumption. In fact, a rich philosophical debate has shown that if, more generally, the only requirement to be satisfied is the impossibility of generating causal paradoxes, several causal theories may be developed without assuming any temporal priority of causes. Moreover, different causal theories may have a differing degree of adequacy when applied to the domain of microphysics. The evaluations that may be made of their basic causal principles according to different formulations and applications of the principles themselves may widely differ, so that when one claims to defend or counteract a causal view of non-locality, he should specify in advance what is the causal theory in terms of which that view is supposed to be 'causal'. A clear demonstration of the interpretation-dependent character of causal notions is the debate on Reichenbach's common cause principle, according to which when two events *A* and *B* are correlated, either there is a direct connection between *A* and *B* producing the correlation or there is a different event *C* which causes the correlation. On the basis of different intuitions and formal definitions, opposite conclusions have been drawn on whether explanations of nonlocal quantum correlations in terms of probabilistic common causes are an option or not (more on this point below[3]). This circumstance strongly supports in my opinion the view according to which what is usually called the common cause principle 'is not really a principle but a schema of principles that calls for interpretation' (Berkovitz, 2000b, 53), a circumstance that again turns out to be compatible with the methodological pluralism about causation that we mentioned above.

The variety of formulations that both the notion of (non-)locality and the notion of causation may assume in different theoretical frameworks can be considered primarily as a *logical* problem. In the assessment of the status and significance of a causal view of non-locality, however, we have first to take into account its *physical* background, namely we have to take into account the investigations on the physical meaning of non-locality in quantum mechanics. The standard framework is that of EPR-Bell correlation experiments, involving a two spin-1/2 particles' system S_1+S_2 prepared in the singlet state, and such that the spin measurements are performed when the two subsystems S_1 and S_2 occupy two space-like separated spacetime regions R_1 and R_2, respectively, after leaving the source. The common feature of these investigations is basically an assumption of incompleteness for the purely quantum description of physical states; on the basis of such assumption a 'finer' state description is postulated via the introduction of extra ('hidden') variables that 'add up' to the quantum state. In this vein the first step was to introduce *deterministic* hidden variable models, in which the source state λ is postulated to be complete and assumed to determine with certainty the outcome of any measurement that can

[3]For a recent survey on this issue see Suarez (2007).

be performed on the two distant subsystems. Later the condition of determinism for hidden variables has been relaxed. Stochastic hidden variable models were then introduced, in which the state description λ allowed by the model enables one to determine not the measurement outcome but only its probability of occurrence.

Both in the deterministic and stochastic frameworks, a locality condition is usually motivated by a prescription of 'lack of influence' between the spacetime regions in which the measurement events are localized, although the specific condition of locality that was assumed in deterministic hidden variables models had to be reformulated in order to comply with the stochastic character of the more general model. The locality condition was then formulated as an independence constraint on the statistical predictions generated by the complete descriptions of the single particles' states (when the particles themselves are spatially well separated). Namely, the assumption of the mutual independence between the relevant spin measurement events was formulated as the invariance of the probabilities prescribed by λ for any outcome in one wing of the experiment under the change of some relevant parameter in the distant wing. Consequently, several discussions focused on what different locality conditions obtained when such parameter was taken to represent different things, typically parameters pertaining either to apparatus settings or to outcomes of the measurements.[4] The greater generality of these stochastic hidden variables models should make the conclusions drawn from them stronger. If locality is violated in these models, the existence of non-local influences is strongly supported, and thus their significance for the notion of causation can be investigated. However, even this more general framework provides no clear answer to the following central questions:

(a) How should the causal meaning of non-locality be assessed by the point of view of the *spacetime structure* in which non-local correlations display themselves?
(b) Provided we adopt the most natural interpretation of probability in physics, namely the relative frequency interpretation, and we do not turn to more controversial notions such as chances, propensities or dispositions, what might non-local correlations tell us about *single* events confined in bounded spacetime regions? [5]

---

[4]It is worth emphasizing that I refer here to hidden variables *models*, and not to hidden variables *theories*, for a simple reason. In the history of the hidden variables' issue, the 'theories' in which more and more general locality conditions were assumed – and whose predictions have been shown to be inconsistent with those of quantum mechanics – were in fact theories only as a *façon de parler*; whereas the only full-fledged formal construction deserving the title of theory, namely Bohmian mechanics, is explicitly nonlocal.

[5]In his 1996 paper, Dickson has questioned the adequacy of locality conditions based on probabilistic independence when Bohmian mechanics is taken into account, and he argued that Bohmian mechanics may be shown to satisfy or violate that kind of locality depending on how a specific model of the theory is constructed (Dickson, 1996). This indicates, according to Dickson, that probabilistic independence is not adequate to capture the meaning of locality. It is worth recalling that the Dickson argument concerning the status of locality as probabilistic independence in Bohmian mechanics has been challenged in Maudlin (2000).

This is why in the sequel, when I will discuss the status and significance of causal relations within the issue of non-locality in quantum mechanics, I will assume simply as a working hypothesis that causal relations may be analyzed as holding among single events in spacetime, on the basis of processes that need not refer to any recurrence in order to be considered 'causal'. As every philosopher of causation will immediately acknowledge, this assumption is somehow reminiscent of a *singularist* approach to causation, endorsed among others by such eminent philosophers as C.J. Ducasse and G.E.M. Anscombe. In the singularist view of causation the cause of a particular event [is defined] in terms of but a single occurrence of it, and thus in no way involves the supposition that it, or one like it, ever has occurred before or ever will again. The supposition of recurrence is thus wholly irrelevant to the meaning of cause; that supposition is relevant only to the meaning of law. And recurrence becomes related at all to causation only when a law is considered which happens to be a generalization of facts themselves individually causal to begin with. [...] The causal relation is essentially a relation between concrete individual events; and it is only so far as these events exhibit likeness to others, and can therefore be grouped with them into kinds, that it is possible to pass from individual causal facts to causal laws. (Ducasse, 1926, 129–130).

I wish to stress, however, that I am not embracing a preliminary philosophical position on causation, namely singularism, and then turning to argue that causation in quantum mechanics can only make sense if interpreted in singularist terms. This attitude would point exactly in what I take to be a wrong direction by the methodological viewpoint, namely the attempt to select a priori a philosophical doctrine in advance and then to try to accommodate the physics accordingly. Moreover, as is well known, non-locality in quantum mechanics involves a fundamental reference to counterfactual situations, and since non-trivial counterfactuals are usually supposed to be grounded in laws supporting them, an orthodox singularist might be already suspicious. The meaning I attach to singularism is rather general and so is the motivation for adopting such a viewpoint. Even if one allows the *a priori* plausibility of investigating new forms of causation, that might explain the 'action at-a-distance' allegedly entailed by non-locality (I briefly review the modalities of such 'action' in Section 5.3), I still conceive it to involve physical processes connecting single events. That is, I incline to interpret this hypothetical causation as a sort of singular phenomenon, that is enhanced by the actualization of a property instantiated by a physical event and that affects the actualization of different properties pertaining distant events. The causal action displayed by this phenomenon should thus be understood as *taking place* in spacetime in some well-specified sense, although clearly not as a process propagating *continuously* in spacetime (Berkovitz, 2000a). So the question is: how and to what extent can this unfamiliar causation be interpreted consistently with the more familiar spacetime structure in which – according to our well-established physical theories – single physical events live?

Within a formulation of quantum mechanics with state reduction, a reasonable starting point for addressing the problem is in my opinion to consider the implications of this singularist-like view on non-locality and causation when the state reduction is taken into due account. In the usual interpretation of quantum mechanics, state reduction is not only included among the basic postulates of the theory but is also assumed to be a real physical process. In this interpretation, it is state reduction that is supposed to actualize most properties of quantum systems, and this is a very general motivation for pursuing an analysis of the conceptual link between causation and state reduction. But there is also a more specific motivation for the study of such link. The events that might be causally connected are assumed to be located at *space-like* separated regions: thus if we take seriously – as we should – the spacetime geometry that underlies this assumption, then we also have to take into account at least some ways out of the problem of the non-covariance of the state reduction process in *relativistic* quantum mechanics. In particular, in view of this problem, the Section 5.4 is devoted to the exploration of some of the implications that different assumptions on *where* the state reduction occurs may have on the link causation-reduction.[6]

In following this line of analysis I do not assume, however, that a causal view of non-locality cannot be evaluated in a quantum theory *without* state reduction. Although for obvious reasons I will not take into account all no-collapse interpretations of quantum mechanics, in the last section I will consider how causation (understood in very general terms) might fare in Bohmian mechanics.

5.3 Nonlocality, Supeluminal Influence and Causation

Having reasons to believe that, given two events A and B, their occurrences depend on (or influence or affect) one another, is not sufficient in general to claim that A and B are causally connected. On the other hand, a mutual dependence between A and B is a good reason for us to search whether such dependence is grounded in some underlying causal mechanism, so far unknown to us. In the context of the EPR-Bell correlations in quantum mechanics, the events under consideration are assumed to be space-like separated, so that the search for causation in this context is a search for a *superluminal* causation, pursued under the assumption that our quantum-mechanical events display at least a superluminal dependence.

In order then to investigate whether long distance correlations in EPR-Bell experiments deserve to be called causal, it is convenient to briefly review the reason why in ordinary quantum mechanics such correlations can be in fact regarded as an instance of superluminal dependence between events that in a purely relativistic perspective should be taken to be mutually independent. For the sake of simplicity, I will assume here that performing a *measurement* and detecting an *outcome* are not distinct events: the terms of the hypothetical causal connection that I wish to investigate are then to be meant as *measurement-and-outcome* events.

[6]For the sake of the present discussion, I assume such notions as property or emergence as uncontroversial. Of course they are not, but in my opinion it is anyway doubtful that a purely philosophical analysis of such notions could substantially contribute to a better understanding of the main issues in the foundations of quantum mechanics.

In a standard EPR-Bell correlation experiment involving a two spin-1/2 particles' system S_1+S_2 prepared in the singlet state, we know that the spin measurements are supposed to be performed when S_1 and S_2 occupy two space-like separated space-time regions R_1 and R_2, respectively. Under the hypothesis that quantum predictions are correct, S_1 and S_2 exhibit a perfect spin correlation, namely if the outcome of an actual measurement of the spin up along any direction x for the particle S_1 is +1, the probability of obtaining -1 as outcome of the measurement of the spin up along the direction x for the particle S_2 equals 1. Hence, we may say that had the measurement of the spin up along any direction x for the particle S_1 come out -1, we would have obtained with certainty +1 for S_2. However, in ordinary quantum mechanics the measurement process is stochastic, namely from identical preparations we may obtain different outcomes: the spin of S_1 can be either +1 or -1 in different runs also when the whole set of events causally relevant to obtaining +1 or -1, localized in the backward light cone of the that event, is exactly the same. But if S_1 and S_2 are shown to be perfectly correlated in their outcomes, either there is a direct dependence between the two measurements, performed in the space-like separated regions R_1 and R_2, or there is an dependence between the measurement of $S_1[S_2]$ and some event in the backward light cone of $S_2[S_1]$, and in both cases the dependence holds between space-like separated events, namely it is superluminal (cfr. Maudlin, 1996, 285–289, 2002, 87 ff,). Moreover, due to the Bell theorem, any theory assuming the existence of events or factors that (i) are causally relevant to obtaining +1 or -1 for $S_1[S_2]$, (ii) are located in the backward light cone of $S_1[S_2]$, and (iii) screen off the causal relevance which is in the backward light cone of $S_2[S_1]$ (but not in the overlap of the backward light cones of S_1 and S_2), is bound to give predictions that disagree with those of quantum mechanics.[7]

Before going on, a pair of remarks concerning possible objections to the above argument in favor of superluminal dependence. First, it is worth stressing that in the above argument counterfactuals are involved just to express the content of the spin strict correlation property, whereas the locality condition that is presupposed is expressed in terms of the invariance – across possible different runs of the experiment – of the light cone structure of the events that are causally relevant to obtaining a given outcome. The latter condition is independent in principle from any sort of counterfactual locality condition, such as 'the outcome of a measurement on S_2 of the spin x-component would have been still +1, had the spin component been measured on S_1 in the z-direction instead of the x-direction', a formulation which is exposed to the objection of having non-contextuality tacitly built in: in fact there is no reason why the measurement on S_2 of the spin component in a given direction should have the same outcome when in different runs of the experiment it is measured with observables of S_1 that are mutually incompatible. Second, the above

[7] As should be clear from the above account, the stochastic nature of the measurement process makes the instance of superluminal dependence even more perspicuous. On the difficulties of making sense of locality – and of the superluminal dependence that its violation would imply – in a strictly deterministic theory, see Dickson (1996) and Maudlin (2000).

argument does not rely on the assumption that, after a measurement has been performed and an outcome obtained, there is necessarily a value of the measured system that corresponds to the outcome (and hence that, after the completion of the measurement, the measured system satisfies the definite property of having that value). Namely, the argument holds also if we just assume that, after the completion of the measurement, the outcomes +1 and −1 are definite properties *of the measuring apparatuses*.

The relation between outcomes of spin measurements in EPR-Bell correlation experiments is then an instance of superluminal dependence. Such terms as ‘dependence’ or ‘influence’ are admittedly vague, however, so that the attempt to elaborate arguments by which we could legitimately interpret superluminal dependence as a form of causation appear at first completely reasonable. In addition, there are already well-developed theories of causation at our disposal, and in principle we are able to analyze the viability of their main assumptions and conditions by the particular viewpoint of the nature of the dependence between distant quantum events.

According to Maudlin’s terminology, for instance, correlated events like the outcomes of EPR-Bell correlation experiments are *causally implicated with each other*, a formulation that is supposed to suggest that the causal implication need not distinguish causes from effects, and it may hold between events neither of which is a direct cause of the other (Maudlin, 2002, 128). The generality of the definition has a non trivial justification. If we decide to adopt or develop a more sophisticated theory of causation, in which more stringent conditions on the identification of causes and effects are required, we immediately run into difficulties: due to the space-like separation between the dependent events, the time ordering between them is non-invariant across different Lorentz frames.

A first option is trying to dissolve the problem, rather than solving it, by arguing that the very distinction between cause and effect is hardly applicable to EPR-Bell frameworks. This position, albeit logically consistent, seems to imply that we do not need even to stipulate what are the terms of the allegedly causal relation that we are investigating. I will not discuss this option further since I doubt that anything relevant to a decent notion of causation is left in it. A second option is to retain the distinction between cause and effect, but to argue that it is the very time ordering associated to *any* Lorentz frame that *defines* which is the cause and which the effect. In this option the cause-effect distinction is thus not rejected but is remarkably weakened, since it acquires itself the status of a frame dependent distinction.

No matter which of the first two preceding options is adopted, however, the superluminal dependence between EPR-Bell outcomes appears to be ‘causal’ in such a weak sense as to prompt the question: having acknowledged that EPR-Bell outcomes are somehow connected across spacetime, do we really obtain any deep insight by calling ‘causation’ that connection? Or rather what we are doing when we say that the EPR-Bell outcomes are ‘causally implicated with each other’ is nothing but saying that ‘connected events are connected’? If the non-invariance of the time ordering between the connected events forces us to abandon such typical conditions on causal relations as the temporal priority of the cause, or less typical but still reasonable conditions such as simultaneity between cause and effect, the features of this link are themselves so vague that we should not be worried by the vagueness of the non-causal terms – namely 'dependence', 'influence' and the like – that we might use to denote it: using causal concepts in this case appears then to be a mere labeling devoid of any real physical and philosophical significance.

In recent times the discussion developed again on causation in EPR-Bell frameworks through the investigation on the consistency of specific *causal models* of the statistical correlations involved. I refer in particular to Suarez (2007), in which the author challenges several conclusions drawn from different approaches but that seem to converge on the claim that it seems in principle impossible to causally model EPR-Bell correlations. As a matter of logic, however, Suarez shows convincingly that the possibility of giving a causal account of the EPR-Bell correlations is perfectly consistent, since it is possible to construct several models whose causal structure is not in principle ruled out by relativistic considerations. Therefore, in terms of strict consistency the line of research of causally modelling the EPR-Bell correlations is still open and connected with the idea of a causal structure of the physical world, that is to be investigated in probabilistic terms. By this point of view, the highly sensible conclusion of Suarez is that 'the question "are the EPR correlation causal?" in general has no informative answer. To answer this question we have to engage with the details both of the different theories of causation and the different possible models for the EPR correlations. Different combinations of causal theories and empirical models will yield different answers to this question.' (Suarez, 2007, 104).

A general question, however, arises. Any specific model chosen out of the above mentioned plurality of causal models can be associated to the EPR-Bell framework and its correlations, but the latter are formulated within the formalism of standard quantum mechanics, which is plagued by very deep problems. In itself, this formalism does not enable us to specify what is the basic minimal ontology the theory is supposed to be about, so that it is not clear how the theoretical framework describing an alleged causal structure of the world should match with standard quantum formalism. The upshot is that, before diving into the details of an alleged causal structure of the world, we would better define *first* an ontologically unambiguous interpretational framework for quantum mechanics, something that can be pursued for instance in the Bell spirit: 'It would be foolish to expect that the next basic development in theoretical physics will yield an accurate and final theory. But it is interesting to speculate on the possibility that a future theory will not be *intrinsically* ambiguous and approximate. Such a theory could not be fundamentally about "measurements", for that would again imply incompleteness of the system and unanalyzed interventions from outside. Rather it should again become possible to say of a system not that such and such may be *observed* to be so but that such and such *be* so. The theory would not be about "*observ*ables" but about "*be*ables". [...] The idea that quantum mechanics is primarily about "observables" is only tenable when such beables are taken for granted. Observables are *made* out of beables.' (Bell, 2004, 41)

5.4 Causation in Quantum Mechanics with State Reduction and in Its GRW Formulation

In the discussion developed so far, the role of the state reduction process has not been taken into due account. For in EPR-Bell frameworks, the mark of a relation that we might consider causal between the two outcomes is the emergence of actual properties of the system on one wing, as a consequence of obtaining a certain outcome after measuring an observable of the system on the other wing. But in ordinary quantum mechanics the process itself through which such properties emerge is exactly the state reduction, so that it is sensible to investigate this notion of causation at-a-distance yet to be characterized on the background of the reduction process. Under the assumption that the state reduction is a real physical process (that, as it stands, lacks Lorentz covariance), there are different options on where the state reduction might take place and, in view of the above mentioned causation-reduction link, we should take into account how a notion of causation – even very general – fares with respect with the different accounts on where the state reduction occurs.

In an early investigation on the non-covariance of the state reduction process, Bloch argued that the hypersurface on which the state reduction may be taken to occur can be chosen arbitrarily, since that choice will not affect the probability distribution of all (local) observables. This prescription is clearly non-covariant, but in a relativistic quantum theory of measurement 'it appears that either causality or Lorentz covariance of wave functions must be sacrificed [...] Covariance seems the smaller sacrifice, since it is apparently not required for the calculation of invariant probabilities.' (Bloch, 1967, 1384). This argument might provide a motivation for one of the above mentioned options, according to which it is the very time ordering associated to any Lorentz frame that defines which is the cause and which the effect. If for instance one performs a measurement in an EPR-Bell correlation experiment, it can be assumed in the Bloch spirit that the state reduction occurs along a space-like hyperplane containing the measurement event in the frame of the observer who performed the measurement. In a later paper Hellwig and Kraus, although still emphasizing that what matters are just probability distributions since these are insensitive to the Lorentz frame adopted to order the events, have proposed a prescription according to which the reduction occurs along the backward light cone of the measurement event (Hellwig and Kraus, 1970).

In a series of papers Aharonov and Albert have shown that, although Lorentz-covariant, the Hellwig-Kraus prescription turns out to be inadequate when *non-local* observables are taken into account, namely observables of just such composite systems as those considered in EPR-Bell correlation experiments (Aharonov and Albert, 1980, 1981, 1984). But also without addressing the Aharonov and Albert criticisms (for more recent debates cfr for instance Ghirardi, 2000), the very fact which the Bloch non-covariant prescription and the Hellwig-Kraus covariant one rely on is unsatisfactory by my specific point of view. Namely, the fact that the expectation values of the considered observables – be they local or non-local – are invariant across different Lorentz frames tells us nothing that might be relevant to explaining the superluminal dependence between *single* events and perhaps to interpreting it in causal terms. Moreover, as far as just expectations values are taken into account, quantum mechanics *does* satisfy *statistical* locality in the sense that in a typical EPR-Bell correlation experiment, for instance, the expectation value of a spin observable pertaining one subsystem is completely unaffected by any kind of operation performed on the distant subsystem (Eberhard, 1978; Ghirardi, et al., 1980). Therefore, should we confine our attention to the level of expectation values, the very non-locality problem (and the correlated one of attempting a causal interpretation of it) would not even arise.

But there is a further consequence of the Aharonov-Albert analysis that turns out to be relevant by our viewpoint, namely the revision of the usual meaning ascribed to the wave function in a relativistic context. According to their proposal, when a local measurement is performed at a spacetime point S, the state reduction should be taken to occur along *every* space-like hyperplane intersecting S. In addition to the Lorentz-covariance that this proposal allows one to achieve, it implies that the state of the system in a relativistic quantum-mechanical context must be represented as a functional defined on the set of space-like hyperplanes, so that in turn the ordinary wave function takes on different values at a given spacetime point according to which space-like hyperplane is considered (Aharonov and Albert, 1984, 231–232). By the point of view of causal relations between events, however, this implies that certain events – that might play the role of 'causes' and that are given by wave functions taking on definite values at spacetime points – are actual in certain hyperplanes and not in others. This leads us back to the starting point: also in a relativistic account of the state reduction process such as the Aharonov-Albert one, there seem to be no room for a characterization of causation that goes beyond a merely verbal elaboration of the circumstance that certain events manifest a mutual connection different in important respects from all other physical forces known in nature.[8]

More generally, in the shift to relativistic quantum mechanics, there is a circumstance that Aharonov and Albert emphasize and that seems to be forced upon us by the attempt of finding a Lorentz-covariant formulation of the measurement process: the theory preserves the capacity of prescribing the correct probabilities for measurement outcomes, but not the capacity of attributing definite states to the physical systems whose outcome probabilities are evaluated. If this is the case, the prospects of a causal view of the relation holding between the correlated outcome events in EPR-Bell correlation experiments appear rather dim also in a relativistic quantum-mechanical context: it is problematic to think of EPR-Bell events as causally connected when these events should be represented as instances of

[8]The fact that the ordinary wave function takes on different values at a given spacetime point according to which space-like hyperplane is considered, following from the generalization of the state as represented by a functional on the set of space-like hyperplanes, has analogies with the Fleming hyperplane dependence approach to quantum states (Fleming, 1989, 1996). It seems to me that the status of causation in the Fleming approach would be similar to that in the Aharonov-Albert approach, but this point deserves further investigations.

properties satisfied by the suitable physical systems, but in fact no definite ordinary state can be attributed to the latter.

The considerations developed so far refer to a very general quantum theoretical framework that encompasses state reduction. The theory proposed in 1986 by Ghirardi, Rimini and Weber (GRW henceforth) introduced a quantitatively detailed model of how a state reduction process can be incorporated into quantum mechanics such that typically quantum phenomena on the microscale coexist with what GRW used to call the *macro-objectification* of physical properties pertaining to apparatuses in measurement interactions. As is well known, the heart of the formulation lies in a nonlinear stochastic modification of the evolution law for wavefunctions, a modification that is supposed to induce spontaneous collapse processes for the wavefunctions themselves.[9] In more recent times, the GRW research program evolved into a class of theories attempting to formulate a clear ontology into which the state reduction is to be framed. This is of primary relevance to the issue of causal talk in quantum mechanics, since there seems to be no hope for a fruitful discussion of that issue unless it becomes reasonably clear which is the basic furniture of the physical world quantum mechanics is supposed to be about. Two different proposals have been introduced as to the sort of ontology GRW models are held to deal with, the *matter density ontology* and the *flash ontology*.[10] In the former the theory assumes a continuous ontology, consisting essentially in a field on three-dimensional space that, for a given t, is to be understood as the density of matter in space at time t (Benatti et al., 1995, Bassi and Ghirardi, 2003); in the latter the theory assumes a *discrete* ontology, in which matter is made up by discrete points ('flashes') in spacetime such that to each of these flashes there correspond one of the spontaneous collapses of the wavefunction, and the spacetime location of the flashes is the spacetime location of the collapses. In the flash ontology – whose original proposal is due to J.S. Bell – flashes are the *local beables* of the theory (Bell, 1987).[11]

According to the view discussed in the preceding sections, the connection between causality and the spacetime structure is crucial. Therefore, any possible investigation on the role of a certain notion of causality in GRW models is to confront with the problem of finding a relativistically satisfactory formulation of those models.[12] So far, only a GRW model endowed with a flash ontology has been shown to be Lorentz-invariant in a well specified way, namely by a coordinate-free law that is able to prescribe probability distribution of all flashes (which we recall are the

[9]For a wide survey of the dynamical reduction model see Bassi and Ghirardi (2003), whereas for a recent and extremely lively recollection of the intellectual path leading to the development of that model, see Ghirardi (2007).

[10]For a very sharp discussion of these ontologies and their implications, see Tumulka (2007).

[11]For a thorough discussion of the implications of the two GRW ontologies and their relation to Bohmian mechanics see Allori et al. (2006).

[12]As Tumulka aptly stresses, the problem itself of a possible relativistic extension of dynamical reduction models crucially depends on the clarification of the primitive ontology underlying the models (Tumulka, 2007, 3260).

basic elements of the ontology) for a space-like surface previously fixed as the initial conditions (Tumulka, 2006). Since the model is stochastic, there seems to be room for investigating probabilistic theories of causality that might shed light on the causal structure of a flash world whose probabilistic relations are determined by a Lorentz-invariant law.

5.5 Causation and Spacetime Foliation in Bohmian Mechanics

In ordinary quantum mechanics (i.e. quantum mechanics with state reduction), there would be in principle a further option that we did not consider so far: a preferred foliation of spacetime might be explicitly assumed, with respect to which it would be perfectly determinate which events are causes and which effects. This would amount, however, to a violation *de facto* of the above mentioned relativistic constraint, since in this case the space-like separation between the causally connected events would be only a phenomenological relation. Moreover, the only reason for such a strong assumption would be just to make room for space-like causation. This move would also have the somewhat ironic consequence that, in order to explain a deeply non-classical feature like a fundamental physical relation between space-like separated events, deeply pre-relativistic features like the absoluteness of the time ordering between the events themselves are reintroduced.

The situation is different for Bohmian mechanics, whose overall structure may provide independent (and deeper) reasons for justifying the assumption of a preferred foliation of spacetime. In its basic formulation, Bohmian mechanics is not a Lorentz-invariant theory. In the general case of a N-particles system, the guidance equation – the only dynamical law added by the theory to the Schrödinger equation – concerns the positions of the N particles at a common and absolute time: this presupposes the assumption of a foliation of spacetime into space-like hyperplanes that, however, turns out to be impossible to determine. In this context, a causal interpretation of non-locality appears rather natural, since a causal relation between the EPR-Bell events might be then assumed to hold just with respect to the foliation. If one is willing to accept that the democracy reigning among Lorentz frames, prescribed by special relativity, is to be meant just as a phenomenological circumstance – as David Albert put it, 'taking Bohm's theory *seriously* will entail being *instrumentalist* about special relativity' (Albert, 1992, 161, emphasis in the original) – then a notion of space-like causation linking EPR-Bell events may very well be accommodated into Bohmian mechanics. Admittedly, it would be quite an unconventional sort of causation, since it would share with the foliation (and with all quantities defined with respect to it) an *epistemic* inaccessibility: in this picture, we know that causation is there, although we are bound to remain ignorant about its mechanisms and about which events are 'causes' and which 'effects'. It is also true, however, that a supporter of Bohmian mechanics need not being particularly worried by this circumstance, rather disturbing for others. In fact he should find it relatively easy to accommodate it within the framework of Bohmian mechanics, since fundamental beables of the theory – in Bell's terminology – like the particles' positions and trajectories are themselves out of reach. As aptly pointed out by Maudlin, 'if the existence of empirically inaccessible physical facts is fatal, then Bohmian mechanics is a non-starter even before Relativity comes into play' (Maudlin, 1996, 296).

It has been argued that in Bohmian mechanics the essential symmetry of the possibly causal relations between two typical EPR-Bell subsystems – when for instance the position of one particle causes the velocity of the other, which is space-like separated from the first – makes it difficult to speak of a serious 'causal' influence between the systems (Dickson, 1996, 325). This argument presupposes that in order for a notion of causal influence to be meaningful, a direction is to be selected along which the influence is supposed to act. Once we assume a preferred foliation of spacetime in the spirit of Bohmian mechanics, however, we need not hold on to a notion of causation in which an event A, in order to be causally related to a different event B, must temporally precede B with respect to the preferred time ordering. We can envisage a sort of causal implication between events similar to that discussed by Maudlin for ordinary quantum mechanics, the only relevant difference being that in Bohmian mechanics we can correctly interpret it as a *simultaneous* and *mutual* causal influence, since we have assumed a privileged time ordering.

A final remark is in order about the role Lorentz invariance should play in Bohmian mechanics (touching upon, indirectly, any investigation on causality in Bohmian mechanics). Although the basic formulation of Bohmian mechanics is explicitly non-Lorentz-invariant, there is no compelling argument (even less a 'no-go' theorem) as to the impossibility in principle of a Lorentz-invariant extension of the theory. An interesting suggestion in this direction was made by Dürr, Goldstein, Münch-Berndl and Zanghì, according to whom Lorentz invariance can be an ingredient of the theory provided one turns the preferred foliation of spacetime into a *dynamical* object, governed by an evolution law that can be formulated as a Lorentz-invariant law (Dürr et al., 1999).[13]

5.6 Conclusions

Is causal talk meaningful in the foundations of quantum mechanics? The question does not admit a straightforward and model-independent answer. On one hand, I think that we should welcome the acknowledgement – nowadays largely agreed upon – that quantum mechanics does not per se prevent any causal approach: quantum mechanics is not intrinsically 'less causal' than any other possible well-defined physical theory and any conclusion concerning causality in quantum mechanics must be evaluated within a well-specified causal framework and with respect a well-defined domain of phenomena. On the other hand, most investigations on causality in the foundations and philosophy of quantum mechanics fail to refer to an

[13]The problems with other proposals of relativistic extensions of Bohmian mechanics are briefly discussed in Tumulka (2007, 3257–3259).

unambiguous interpretation of quantum mechanics itself, namely an interpretation that– unlike the standard formulation – clearly specifies the intended *ontological* content of the theory. Shifting to ontologically unambiguous interpretations of quantum mechanics – such as Bohmian mechanics or GRW theory in one of its variants – might contribute in this respect to a deeper understanding of an hypothetical causal structure of the quantum world, by paying for instance a special attention to the complex ways in which causality is connected to the issue of Lorentz invariance in these interpretations.

References

Aharonov, Y. and Albert, D. (1980), States and observables in relativistic quantum field theories. *Physical Review D*, 3316–3324.

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

Aharonov, Y. and Albert, D. (1984), Is the usual notion of time evolution adequate for quantum-mechanical systems? II. Relativistic considerations, *Physical Review D* 228–234.

Albert, D. (1992), *Quantum Mechanics and Experience.* Cambridge, MA: Harvard University Press.

Allori, V., Goldstein, S., Tumulka, R., and Zanghì, N. (2006), On the common structure of Bohmian mechanics and the Ghirardi-Rimini-Weber theory, http://arxiv.org/quant-ph/0603027. Accessed Apr 2008.

Bassi, A and Ghirardi, G. (2003), Dynamical reduction models, *Physics Reports* 379, 257–427.

Bell, J. S. (1987), Are there quantum jumps? In C. W. Kilmister (ed.), *Schrödinger. Centenary of a Polymath*, Cambridge: Cambridge University Press, pp. 41–52 (reprinted as chapter 22 of Bell 2004).

Bell, J. S. (2004), *Speakable and Unspeakable in Quantum Mechanics*, Cambridge: Cambridge University Press, 2nd edition.

Benatti, F., Ghirardi, G. and Grassi, R. (1995), Describing the macroscopic world: closing the circle within the dynamical reduction program, *Foundations of Physics* 25, 5–38.

Berkovitz, J. (2000a), The nature of causality in quantum mechanics, *Theoria* 15, 87–122.

Berkovitz, J. (2000b), The many principles of the common cause, *Reports on Philosophy* 20, 51–83.

Bloch, I. (1967), Some relativistic oddities in the quantum theory of observation, *Physical Review* 156, 1377–1384.

Dickson, M. (1996), Is the Bohm theory local? In J. Cushing, A. Fine and S. Goldstein (eds.), *Bohmian Mechanics and Quantum Theory: An Appraisal* Dordrecht: Kluwer, pp. 321–330.

Dickson, M. (1998), *Quantum Chance and Non-Locality*, Cambridge: Cambridge University Press,.

Ducasse, C. J. (1926), On the nature and observability of the causal relation, *Journal of Philosophy* 23, 57–68 (reprinted in E. Sosa and M. Tooley (eds.), *Causation*, Oxford, Oxford University Press, 1993, pp. 125–136, page references to the reprinted edition).

Dürr, D., Goldstein, S., Münch-Berndl, K. and Zanghì, N. (1999), Hypersurface Bohm-Dirac models, *Physical Review A* 60, 2729–2736.

Eberhard, P. (1978), Bell's theorem and the different concepts of locality, *Nuovo Cimento* 46B, 392–419.

Fleming, G. N. (1989), Lorentz invariant state reduction and localization, In A. Fine and J. Leplin (eds.), *PSA1988* Vol. 2, E. Lansing, MI: Philosophy of Science Association, pp. 112–126.

Fleming, G. N. (1996), Just how radical is hyperplane dependence? In R.K. Clifton (ed.), *Perspectives on Quantum Reality*, Dordrecht: Kluwer, pp. 11–28.

Ghirardi, G. C. (2000), Local measurements of nonlocal observables and the relativistic reduction process, *Foundations of Physics* 30, 1337–1385.

Ghirardi, G. C. (2007), Some reflections inspired by my research activity in quantum mechanics, *Journal of Physics A* 40, 2891–2917.

Ghirardi, G. C., Rimini, A. and Weber, T. (1980), A general argument against superluminal transmission through the quantum mechanical measurement process', *Lettere al Nuovo Cimento* 27, 293–298.

Ghirardi, G. C., Rimini, A. and Weber, T. (1986), Unified dynamics for microscopic and macroscopic systems, *Physical Review* D 34, 470–496.

Hellwig, K. E. and Kraus, K. (1970), Formal description of measurements in local quantum field theory, *Physical Review* D1, 566–571.

Hitchcock, C. (2007), How to be a causal pluralist, In P. Machamer and G. Wolters (eds.), *Thinking About Causes: From Greek Philosophy to Modern Physics*, Pittsburgh: University of Pittsburgh Press, pp. 200–221.

Maudlin, T. (1996), Space-time and the quantum world, In J. Cushing, A. Fine and S. Goldstein (eds.), *Bohmian Mechanics and Quantum Theory: An Appraisal*, Dordrecht: Kluwer, pp. 285–307.

Maudlin, T. (2000), Review of Dickson 1998, *British Journal for the Philosophy of Science* 51, 875–882.

Maudlin, T. (2002), *Quantum Non-Locality and Relativity.* Oxford:Blackwell (2nd edition).

Suarez, M. (2007), Causal inference in quantum mechanics: a reassessment, In F. Russo, and J. Williamson (eds.), *Causality and Probability in the Sciences*, London: College, pp. 65–106.

Tumulka, R. (2006), A relativistic version of the Ghirardi-Rimini-Weber model, *Journal of Statistical Physics* 125, 821–840.

Tumulka, R. (2007), Two unromantic pictures of quantum theory, *Journal of Physics A* 40, 3245–3273.

Williamson, J. (2007), Causality, In D. Gabbay, and F. Guenthner (eds.), *Handbook of Philosophical Logic*, *vol. 14*, Dordrecht: Springer, pp. 89–120.

Chapter 6
On Explanation in Retro-causal Interpretations of Quantum Mechanics

Joseph Berkovitz

6.1 Retro-Causal Interpretations of Quantum Mechanics: Background and Motivations

In quantum phenomena, there are curious correlations between distant events. A famous example is the Einstein-Podolsky-Rosen experiment. In Bohm's (1951) version of this experiment (henceforth, the EPR/B experiment), particle pairs are emitted from a source in opposite directions in the singlet state for spin:

$$\text{(Singlet state)}\, |\psi\rangle = \frac{1}{\sqrt{2}} \left(|n+\rangle_1 \, |n-\rangle_2 - |n-\rangle_1 \, |n+\rangle_2\right);$$

where '1' and '2' are the indexes of the particles, and $|n-\rangle_1$ ($|n-\rangle_2$) and $|n+\rangle_1$ ($|n+\rangle_2$) are the states of the first (second) particle having respectively spin 'up' and spin 'down' along the direction n. When the particles are far away from each other, they encounter apparatuses that can be set to measure spin properties along various directions. Each of the measurements occurs outside the backward light cone of the other measurement, so that there could not be any subluminal or luminal influences between them. According to orthodox quantum mechanics, the outcome of each of the distant measurements is a matter of sheer chance. Yet, the measurement outcomes are curiously correlated: the probability of the outcome spin 'up' along the direction n in one wing of the experiment will depend on whether the outcome in the other wing is spin 'up' or spin 'down' along the direction m. These correlations suggest the existence of superluminal influences between the measurement outcomes, and indeed the orthodox interpretation of standard quantum mechanics seems to predict the existence of such non-local influences.

Famously, Einstein et al. (1935) argued that the correlations between the distant outcomes are due to the incompleteness of the orthodox interpretation, and they

J. Berkovitz (✉)
IHPST, Victoria College, University of Toronto, 91 Charles St. West, Toronto, ON, Canada M5S 1K7
e-mail: joseph.berkovitz@utoronto.ca

M. Suárez (ed.), *Probabilities, Causes and Propensities in Physics*, Synthese Library 347, DOI 10.1007/978-1-4020-9904-5_6,

thought that a 'complete' quantum theory would explain these correlations without postulating any non-local influences. Their idea may be explained by an appeal to common presuppositions about the relations between causation and correlation, which are embodied in Reichenbach's (1956) principle of the common cause. It is commonly assumed that a systematic correlation between distinct events/states has a causal explanation. Such correlation may be due to a 'direct' causal connection between them, or a common cause. According to Reichenbach's principle, if the correlation is due to a common cause, it is explained away by this cause. That is, conditioning on the common cause, the correlated events become probabilistically independent: their joint probability factorizes into their single probabilities.[1] In the EPR/B experiment, the emission of the particle pair occurs at the intersection of the backward light cones of the distant measurement outcomes, and accordingly the pair-state at the emission is a natural candidate for a local common cause of the outcomes (i.e. a common cause that exerts only subluminal or luminal influences on them). But, the singlet state does not render the measurement outcomes probabilistically independent. Thus, following Reichenbach's principle, if this state constituted the relevant, complete common-causal past of the outcomes (i.e. the common-causal past that screens the outcomes off from any other part of their common-causal past), the correlations between them would imply the existence of superluminal causation.[2] EPR thought that the quantum-mechanical pair-state is incomplete, and that a complete pair-state would render the distant outcomes probabilistically independent, and accordingly account for the correlations between them without postulating any superluminal influences. The idea is that the complete pair-state and the setting of the local apparatus to measure a certain quantity would determine the local measurement-outcome and/or its probability; and the joint probability of the distant outcomes would factorise into their single probabilities, thus rendering the outcomes probabilistically independent.

As John Bell (1964) demonstrated in his celebrated theorem, granted very natural assumptions about causation and the physical realm, such a local common-cause explanation of the correlations in the EPR/B experiment is impossible. Bell's theorem focuses on the local models EPR had in mind. In these models, it is supposed that in different runs of the EPR/B experiment with the (incomplete) quantum-mechanical pair-state, the particle pair may have different complete pair-states λ. The model specifies the distribution of these complete states, ρ, in any quantum-mechanical pair-state and settings of the measurement apparatuses.

It is assumed, though, that the probability distribution of the complete pair-state is independent of the setting of the measurement apparatuses: for any quantum-mechanical state ψ and apparatus settings l and r,

---

[1] There are some dissenting views. In particular, Fine (1981, 1986, 59–60, 1989) denies that non-accidental correlations must have causal explanation, and Cartwright (1989, Chapters 3 and 6) and Chang and Cartwright (1993) challenge the assumption that common causes render their joint effects probabilistically independent.

[2] Recall that an event C screens off event E from event F if given C the probability of E is independent of F: $P(E/C \,\&\, F) = P(E/C)$.

$$(\lambda\text{-independence}) \quad \rho(\lambda/\psi \;\&\; l \;\&\; r) = \rho(\lambda/\psi).$$

The complete pair-state and the setting of the local measurement apparatus jointly determine the local measurement outcome or its probability, and the joint probability of the distant outcomes given the complete pair-state and the apparatus settings factorizes into the single probabilities of the outcomes given the complete pair-state and the local apparatus setting. More formally, let l (r) be the setting of the L- (R-) apparatus to measure spin along the direction l (r), X_l (X_r) denote the outcome of such a measurement, and 'i' ('j') be a variable that denotes either the outcome 'up' or the outcome 'down'. Then, for any complete pair-state λ, apparatus settings l and r, and values of i and j:

$$(\text{Factorizability}) \quad P(X_l = i \;\&\; X_r = j/\lambda \;\&\; l \;\&\; r) = P(X_l = i/\lambda \;\&\; l) \cdot P(X_r = j/\lambda \;\&\; r).$$[3]

Although the model probabilities are different from the quantum-mechanical probabilities, the model is supposed to reproduce the probabilities of standard quantum mechanics, and accordingly the correlations between the measurement outcomes, as statistical averages over the model probabilities of outcomes in various complete pair-states. That is, the quantum-mechanical probabilities of measurement outcomes are obtained as sum-averages of the probabilities of outcomes in all the potential complete pair-states, according to the distribution of these states in the given quantum-mechanical pair-state. More formally, for any quantum-mechanical pair-state ψ, apparatus settings l and r, and values of i and j:

$$(\text{Statistics}) \quad P(X_l = i/\psi \;\&\; l) = \int_\lambda P(X_l = i/\lambda \;\&\; l)\, d\rho(\lambda)$$

$$P(X_r = j/\psi \;\&\; r) = \int_\lambda P(X_r = j/\lambda \;\&\; r)\, d\rho(\lambda)$$

$$P(X_l = i \;\&\; X_r = j/\psi \;\& l \;\&\; r) = \int_\lambda P(X_l = i \;\&\; X_r = j/\lambda \;\&\; l \;\&\; r)\, d\rho(\lambda);$$

[3] There are different ways to interpret the meaning of conditional probability. The common way is along Kolmogorov's axiomatization, where the probability of B given A is defined as the ratio of unconditional probabilities: $P(B/A) \equiv P(B \;\&\; A)\big/P(A)$. A different approach is to interpret conditional probability as a 'primitive', $P_A(B)$, so that the conditioning events A are outside the probability space. On this alternative approach, conditional probability could be understood as a conditional with a probabilistic consequent, so that $P_A(B) = p$ denotes the conditional "if A, then the probability of B is p" or "if A had been the case, then the probability of B would have been p". For ease of presentation, we shall follow the first approach. For a general discussion of the second approach and its relation to the first approach, see Hajek (2003) and Berkovitz (2009a,b), and for a discussion of the advantages of the second approach in the context of Bell's theorem, see Butterfield (1992) and Berkovitz (2002).

where $\rho(\lambda)$ is the distribution of λ in ψ, and the values of '*i*' and '*j*' may be either 'up' or 'down'.

Bell (1964, 1966, 1971, 1975a,b) demonstrated that models that satisfy *Factorizability*, *λ-independence* and *Statistics* predict probabilities of outcomes that are constrained by the so-called 'Bell inequalities' – inequalities that impose constraints on the values that joint and single probabilities of outcomes may have. But these inequalities contradict the predictions of standard quantum mechanics. Granted the empirical adequacy of this theory, Bell assumes *Statistics*, and he held that *λ-independence* is a very plausible assumption. Thus, he concluded that *Factorizability* fails, and he interpreted this failure as indicating non-local influences between the distant (space-like separated) measurement events.[4] The nature of these influences varies according to the interpretation of quantum mechanics or the alternative quantum theory under consideration. And while there is no proof that all these different kinds of influences are incompatible with relativity theory, it has been very difficult to reconcile quantum mechanics with relativity.[5]

A way around Bell's conclusion is to reject λ-independence. While this premise seems very plausible, some have considered interpretations of quantum mechanics that circumvent non-locality by violating it. These interpretations postulate local influences from the measurement events backward to the source at the emission – influences that cause the distribution of complete pair-state at the emission to be dependent on the measured quantities. For advocates of such interpretations, see Costa de Beauregard (1953, 1977, 1979, 1985), Davidon (1976), Cramer (1980, 1983, 1986, 1988), Sutherland (1983, 1998, 2008), Price (1984, 1994, 1996, Chapters 3, 8 and 9, 2008), Reznik and Aharonov (1995), Miller (1996, 2008), Gruss (2000), Aharonov and Gruss (2005), Aharonov and Tollaksen (2007), Aharonov and Vaidman (2007) and Argaman (2007).

The main motivation for local, retro-causal interpretations of quantum mechanics is that they provide good prospects for reconciling quantum mechanics with relativity theory. Other motivations include the new features of quantum mechanics that these interpretations reveal, and the theoretical developments that they may bring about (for some examples, see Aharonov and Tollaksen, 2007). Although these are important objectives of current physics, a number of objections may be raised against such interpretations. First, many find the postulation of backward influences unappealing. Second, backward causation may complicate scientific methodology. For example, the control over various relevant factors is an important part of experimental science, and it may be argued that backward causation may

---

[4] Although Bell's conclusion is widely accepted, there are some dissenting views. In particular, Fine (1981, 1986, pp. 59–60, 1989), Cartwright (1989, Chapters 3 and 6) and Chang & Cartwright (1993) deny that the failure of *Factorizability* entails non-locality, and Fine (1982a, p. 294) argues that what the Bell inequalities are all about is the dubious requirement of making "well defined precisely those probability distributions for non-commuting observables whose rejection is the very essence of quantum mechanics" (cf. Berkovitz 1995, 2009a, Sections 1–2).

[5] For recent review essays on Bell's theorem and its implications for non-locality, see Shimony (2006) and Berkovitz (2007) and references therein.

limit the possibility of such control. Third, in retro-causal interpretations of quantum mechanics the measurement events are causes of the states of the measured systems before the measurements and these states are causes of the measurement events. In particular, in some retro-causal models of EPR/B experiments the measurement outcomes are causes of the complete state of the particles before the measurements and this state is a cause of the measurement outcomes, and the worry is that the causal explanation of the outcomes in these models may be vacuous. Fourth, many believe that retro-causality may give rise to causal paradoxes, i.e. inconsistent closed causal loops in which an event causes other events that in turn render its occurrence impossible. Finally, there is the worry that although the influences postulated by retro-causal models are local, such models may allow for superluminal signalling of information, i.e. a signalling of information between space-like separated locations. If the setting of the R-apparatus during the R-measurement could influence the complete pair-state at the source and in turn influence the L-measurement outcome, a change in the setting of the R-apparatus may cause a change in the statistics of the distant (space-like separated) L-measurement outcome. Such a signalling would be incompatible with the predictions of standard quantum mechanics, which is considered the benchmark of empirical adequacy in its domain of application, and many believe that it would also run against the teachings of relativity.

Theories that predict causal paradoxes and superluminal signalling would be untenable. However, retro-causal interpretations are designed to reproduce the statistical predictions of standard quantum mechanics; and in any retro-causal interpretation that reproduces these predictions, backward causation should be neither observable nor controllable in a way which could give rise to superluminal signalling, causal paradoxes, or any other observable effect that is not predicted by standard quantum mechanics. Furthermore, while such interpretations predict the occurrence of causal loops, all the existing arguments for the inconsistency or impossibility of such loops are based on disputable premises. Indeed, various retro-causal models do not predict inconsistent causal loops and accordingly could not be excluded on the grounds of causal paradoxes.

Concerns about backward causation and causal loops and their implications for scientific methodology are a different matter, as they do not render the theories that predict them untenable. The question is whether the challenges that such curious patterns of causal connections pose justify the exclusion of the retro-causal interpretations of quantum mechanics as serious candidates. In what follows, we shall consider the challenges that causal loops pose for the predictive and explanatory power of retro-causal interpretations of quantum mechanics. In our analysis, we shall focus on interpretations that assign Bell-like retro-causal models for EPR/B experiments. These models are similar to the models that Bell considers in his theorem, but in addition postulate backward-causal influences from the apparatus settings or the measurement outcomes to the complete pair-state at the emission.

Section 6.2 briefly explains why retro-causal interpretations of quantum mechanics predict the existence of closed causal loops, and Section 6.3 spells out briefly the relevant concepts involved in the analysis of such loops. Section 6.4 reviews the main arguments for the impossibility of backward causation and causal loops and explains why they fail to exclude retro-causal interpretations of quantum mechanics. Sections 6.5, 6.6, 6.7, 6.8, and 6.9 discuss the challenges that causal loops pose for the predictive and explanatory power of local retro-causal models of EPR/B-like experiments.

6.2 Causal Loops in Retro-Causal Interpretations of Quantum Mechanics

Consider the following EPR/B-like experiment (henceforth, Experiment X). The R-measurement occurs before the L-measurement. The R-apparatus is set to measure spin along the direction *r*. The outcome of the R-measurement determines by a subluminal signal the setting of the L-apparatus (see Figs. 6.1 and 6.2): if the R-outcome is spin 'down' along the direction *r*, the L-apparatus is set to measure spin along the direction *l*, which is the same as the direction *r*; and if the R-outcome is spin 'up' along the direction *r*, the L-apparatus is set to measure spin along a different direction, *l**.

Retro-causal interpretations of quantum mechanics predict the occurrence of closed causal loops in this experiment (see Figs. 6.2 and 6.3). Interpretations that postulate backward-causal influence from the settings of the apparatuses during the measurements to the complete pair-state at the emission predict Loop I: the complete pair-state at the emission (jointly with the fixed setting of the R-apparatus) causes the R-outcome, the R-outcome determines the setting of the L-apparatus, and this setting (jointly with the setting of the R-apparatus and the initial conditions of the

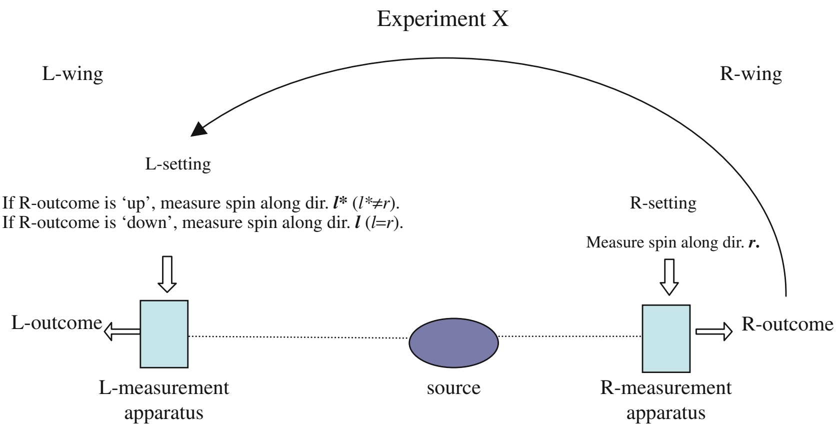


Fig. 6.1 An EPR/B-like experiment. The R-measurement occurs before the L-measurement, the R-apparatus is set fixed to measure spin along the direction *r*, and the R-measurement outcome determines in a perfectly local manner the setting of the L-apparatus: if the R-outcome is 'down' along the direction *r*, the L-apparatus is set to measure spin along the direction $l (l = r)$; and if the R-outcome is 'up' along the direction *r*, the L-apparatus is set to measure spin along a different direction, l^*

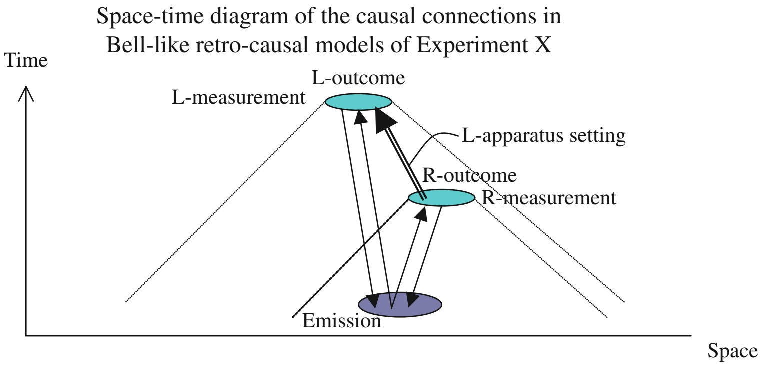


Fig. 6.2 Ovals denote events. Dotted lines denote the boundaries of the backward light cones of the measurement outcomes. Arrows denote causal connections. Retro-causation from the R- and the L-measurement events convey information about the L- and the R-apparatus settings or the L- and R-measurement outcomes, according to the model (see Fig. 6.3)

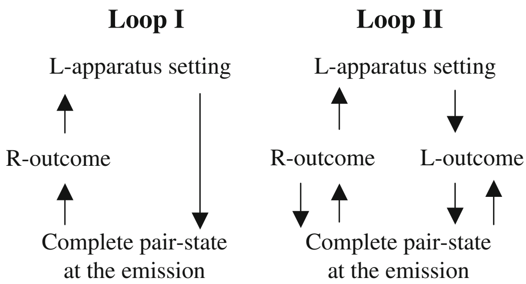


Fig. 6.3 Loop I (Loop II) occurs in retro-causal models of Experiment X that postulate influences from apparatus settings (measurement outcomes) to the complete pair-state at the emission. Here and henceforth, for simplicity's sake the causal circumstances of these loops, e.g. the fixed R-setting, are suppressed

experiment) causes the complete pair-state at the emission. Interpretations that postulate backward-causal influence from the measurement outcomes to the complete pair-state at the emission predict Loop II: the complete pair-state at the emission (jointly with the setting of the R-apparatus) causes the R-outcome, the R-outcome determines the setting of the L-apparatus, this setting and the pair-state jointly cause the L-outcome, and the L-outcome and the R-outcome (and the initial conditions of the experiment) jointly cause the complete pair-state at the emission.[6]

Two remarks: (i) Here and henceforth, in describing causal loops, the factors in brackets denote the relevant causal circumstances of the causal connections in the

[6] In more general models, the complete pair-state at the emission may also be influenced by other factors, such as other final boundary conditions of the experiment. But while these models will substantially complicate our analysis, they will not alter significantly its main conclusions.

loop. These circumstances are supposed to influence, but not to be influenced by, the events in the loop. (ii) Unless said otherwise, by 'spin measurement outcomes', we shall mean 'specific' outcomes, e.g. spin 'up' in the direction l (rather than just the 'non-specific' outcomes spin 'up' and 'down'), and by retro-causal models of EPR/B experiments in which the measurement outcomes influence the complete pair-state, we shall mean models in which such specific outcomes influence the complete pair-state at the emission. In such models (whether they are models of spin or other measurements), the information about the measurement outcomes embodies information about the apparatus settings.

The exact nature of Loop I and Loop II, and accordingly the characteristics of the loops depend on whether the model is deterministic or indeterministic. Deterministic retro-causal models predict loops with deterministic causal connections, whereas indeterministic retro-causal models predict causal loops with some indeterministic causal connections. As we shall see in Sections 6.5, 6.6, and 6.9, the challenges that indeterministic causal loops pose are somewhat different from those of deterministic loops.

6.3 Causal Loops: The Basic Concepts

Before turning to consider the challenges that causal loops pose for retro-causal models, we first need to introduce the notion of causal loop in some more detail. A causal loop is a closed chain of causal connections. The causes in the loop are partial causes: they cause their effects in the actual circumstances and may fail to cause them in other circumstances. In deterministic causal connections, the cause and the relevant circumstances determine the effect. In indeterministic causal connections, the cause and the relevant circumstances determine the probability of the effect, and the effect occurs as a matter of sheer chance.

The probability in indeterministic causal connections may have different interpretations in different accounts of causation. We shall focus on single-case objective interpretations, in particular on single-case propensity. According to this interpretation, probability is thought of as a propensity, or disposition, or tendency of a given type of circumstances to yield an outcome of a certain kind. The exact nature of these propensities and their source varies from one version of this interpretation to another, but common to all versions is the assumption that propensities depend on certain conditions. We shall suppose that the cause and the causal circumstances of an effect determine its single-case propensity. As we shall see in Section 6.6, our analysis of the challenges that causal loops pose for retro-causal interpretations of quantum mechanics will not depend on the single-case propensity interpretation. The main conclusions of this analysis will also apply to other probabilistic accounts of causation, e.g. accounts in which the probabilities of effects are interpreted along the frequency or the epistemic interpretations.

While intuitively the idea that the cause and the causal circumstances jointly determine the effect and/or its probability may seem sufficiently clear, it is difficult to give a general, uncontroversial analysis of this idea. Luckily, our consideration will not require such an analysis. We shall suppose that any retro-causal interpretation of quantum mechanics, which is sufficiently complete, specifies the relevant ontology. In particular, we shall suppose that any such interpretation will specify the interpretation of the probabilities assigned, the dynamical laws of states and properties, and the causal relations between states/events/systems.

We shall assume the ontological framework of the 'block universe', arguably the dominant framework of thinking in modern physics. In this framework, the universe is represented as a four-dimensional block, where three dimensions represent space and the fourth dimension represents time. Events may be thought of as the properties or states of space-time regions in the block, or properties or states of things in such regions. We shall henceforth call these events 'space-time' events. In the ontological framework of the block universe and the context of local retro-causal interpretations of quantum mechanics, it is natural to think of events as space-time events.

All past, present and future events exist, and the division between past, present and future is relational to a standpoint. There is no ontological difference between the past, present and the future. Time is represented like space. Just as New York, London and Moscow all exist but not at the same place, the past, present and future all exist but not at the same time. While the four-dimensional space-time events in the block never change, changes are accounted for in terms of the patterns of, and the relations between such events. For example, the motion of an object is characterized in terms of the events of the object being at different locations at different times.

Causes, effects and causal circumstances may be thought of as space-time events. Causation between events is explicated in terms of possibilities – namely, the way things could or would have been in similar block universes in which the cause does not exist. The exact account is a matter of controversy, but Lewis's (1986, Chapter 21) counterfactual account of causation may provide an example of the general idea. In Lewis's theory of deterministic causation, an event E in the block universe w is said to be causally dependent on a distinct event C in w just in case in the block universes that are the most similar to w and in which C does not occur, E does not occur. In Lewis's theory of indeterministic causation, E in the block universe w is said to be causally dependent on C in w just in case the single-case objective probability of E in w is higher than the single-case objective probability that E would have had in the most similar universes in which C does not occur.

Similarly, single-case objective probabilities of events in the block universe are (partially) explicated in terms of possibilities. Suppose, for example, that the single-case propensity of 'heads' in a genuinely indeterministic coin-toss in the block universe w is $\frac{1}{2}$. In any toss, the coin turns either 'heads' or 'tails', i.e. one of these events inhabits w. Yet, the propensity $\frac{1}{2}$ of this occurrence is partially explicated in terms of long-run frequencies of 'heads' in the same type of tosses with the same type of coin in block universes that are the most similar (in the relevant respects) to w. In particular, according to the law of the large numbers, in a long series of independent tosses of the same type in the most similar universes, the frequency of 'heads' will almost certainly be $\frac{1}{2}$.

In the ontological framework of the block universe and the context of the local retro-causal interpretations of quantum mechanics, we may think of causal loops as sets of space-like events that are related to each other in a cyclical way. In a consistent loop, each event in the loop's set (together with the causal circumstances) determines the next event in the set and/or its probability according to the postulates of the retro-causal interpretation under consideration. Unlike inconsistent loops, in such loops the application of these postulates to the events in the loop's set never leads to inconsistency, i.e. to events that are incompatible with the set. Some may hesitate to call these loops 'causal'. Never mind terminology. However you call them, they pose a challenge for retro-causal interpretations.

The block-universe ontological framework will be helpful in evading some misconceptions about the nature of causal loops and common misguided arguments for their impossibility. But we need first to clear a common misconception about the block-universe. It is frequently claimed that the block-universe dictates determinism and excludes the possibility of chancy events. The idea is that if all future events exist, the future is predetermined. If future events exist and it is a fact about the future that event E will occur at some future time t, then E is bound to occur at t, and no past, present or future event could cause E not to occur at t. Similarly, if future events exist and it is a fact about the future that event E will not occur at t, then E is bound not to occur at t, and no past, present or future event could cause E to occur at t. Generalizing this idea, the block universe framework seems to imply that the future is pre-determined and is not genuinely open to different possibilities, as indeterminism requires. On the basis of similar reasoning, it is also argued that the block universe entails fatalism. If the future exists, it is pre-determined, and nothing that we do now could change the future from what it is going to be. So why bother?

The reasoning above is based on a failure to distinguish between 'the impossibility to *change* the future' and the 'impossibility to *influence* the future'. It is impossible to change the future from what it is going to be, but it is not impossible to influence it. If it is true now that I will sleep tomorrow until noon, i.e. if it is a fact about the actual future, then I will sleep tomorrow until noon, and it is impossible to cause this future event not to happen. But that is not say that the fact that I will sleep tomorrow until noon is not influenced by present, past and perhaps also future events. For instance, my decision now to wake up tomorrow late is going to cause me to wake up tomorrow noon. Had I decided to wake up earlier, I could have done so. I could have gone to sleep earlier, or could have set my alarm for earlier time, etc. And each of these potential actions could have caused a different future in which I wake up before noon.

Put it another way, the four-dimensional block that represents our universe reflects the actual past, present and future states of affairs in that universe, and does not reflect all the possibilities that were not realized. That this block is fixed does not entail that things in our universe could not have been different, or that they were bound to happen. Had things been different, a different fixed block would have represented our universe. If the causal relations in our universe are indeterministic, the actual past is compatible with different possible futures, which are represented by different four-dimensional blocks. In such indeterministic universe, present (and perhaps past and future) events determine the single-case propensities that future actual events have now, and then as a matter of chance these events occur. The block that represents such universe only represents the events that occurred as a matter of chance. It does not represent all the possible events that could have occurred as a matter of chance.

Obviously, everything that we said about future also applies to the past and the present. The fact that the block that represents our universe is fixed entails that the actual past cannot be changed. Thus, the present and the future could not change the past from what it was. But that is not to say that present and future events could not possibly be deterministic or indeterministic causes of past events – causes that influence the past to be what it was.

Finally, it is noteworthy that the argument that the block universe entails fatalism is also based on a failure to distinguish between the 'impossibility to change the future' and the 'impossibility to influence the future'. The fact that the block that represents our universe is fixed entails that nothing that I decide or do now could change the actual future from what it is going to be. But that is not say that my actions or decisions now could not influence the future. I am going to wake up tomorrow at noon because I decided to set up my alarm clock for 11:55 am. Had I decided to wake up earlier, I would have set the alarm clock for an earlier time – an event that would have caused me to wake up earlier. Had I set up the alarm clock for an earlier time, our universe would have been slightly different and a slightly different four-dimensional block universe would have represented it.

6.4 Arguments for the Impossibility of Backward Causation and Causal Loops

There are various arguments for the impossibility of backward causation and causal loops. A common view has it that in our universe causes precede their effects. Moreover, in the literature there are various arguments for the impossibility of backward causation. Some arguments exclude backward causation by virtue of definition, the most obvious example being Hume's famous account of causation, where causes precede their effects as a matter of definition. Other arguments exclude backward causation by an appeal to certain theories of time or causation. (Mellor, 1981, 123; Tooley, 1997, 49, 118–119) All these kinds of arguments are based on disputable conceptions of causation or time. More important to our consideration, these conceptions are different from the ones we shall focus on.

Yet, other arguments for the impossibility of backwards causation rely on more neutral premises. In particular, the popular 'bilking arguments' attempt to demonstrate that backward causation is impossible, or at least that the belief in its existence could never be justified. (Flew, 1954; Dummett, 1964) The basic idea is that if it is possible to prevent events from occurring, and an effect *E* occurred before its cause *C*, one could observe *E* and then try to prevent *C* and thus 'bilk' the alleged backward causation from *C* to *E*. If *C* can be prevented, then it cannot be a cause of *E*;

and if it is never possible to prevent C when E occurs before C, then there are no grounds to claim that there is backward causal influence from C to E.

In consistent retro-causal models bilking is impossible: Further, the presupposition of the bilking arguments that the only grounds to believe in the existence of backward causation are concerned with the capability to manipulate it, is disputable. Indeed, recall that the main grounds for considering retro-causal interpretations of quantum mechanics are theoretical, especially the search for interpretations of this theory which do not postulate non-local influences and accordingly improve the prospects of reconciling it with relativity theory. If one believes that quantum mechanics and relativity are approximately true and that a promising way to reconcile between them is to adopt a retro-causal interpretation of quantum mechanics, then one has grounds to consider seriously the existence of backward causation in our universe. Obviously, the strength of such grounds will depend on how promising this way of reconciling quantum mechanics and relativity is in comparison to other attempts.

Most contemporary philosophers agree that the idea of backwards causation is coherent. In particular, in the context of the ontological framework of the block universe, there are no compelling reasons to exclude the possibility of backward causation on the grounds of conceptions of causation or time. Yet, there are arguments that attempt to exclude causal loops on other grounds. One line of argument is to try to exclude causal loops on the basis of specific theories of causation or time. For example, it is sometimes argued that causal relations are asymmetric and transitive, but in causal loops transitivity entails symmetry. For another example, in some causal theories of time (where the direction of time is determined by the direction of causation) causal loops are excluded since they would give rise to opposite time orders between events: causes would occur both before and after their effects. (Mellor, 1981, 175–177, 1998, 125–126) The theories employed in this kind of arguments are controversial, as their advocates seem to recognise. In any case, these arguments do not show that causal loops are impossible in the ontological framework we outlined above.

Another line of argument for the impossibility of causal loops is that such loops could create causal paradoxes. The main idea is that if loops were possible, effects could undercut the existence of their causes: an event C will cause an event E, and E will render C impossible. If C is a cause of E in our universe, then both of these events actually occur, and no causal influence from E to C could cause C not to occur. In the block-universe framework, this idea is represented by the fact that the fixed block that represents our universe includes both C and E. That is not to deny, of course, that some theories may predict inconsistent causal loops, which no consistent block could represent. Obviously, any retro-causal interpretation that predicts the existence of such loops is untenable.

Other more subtle lines of arguments do not intend to demonstrate the impossibility of causal loops, but rather to show that such loops would involve anomalies, like improbable or causally inexplicable coincidences between events. (Horwich, 1987; Smith, 1997; Dowe, 2003) For example, Sarah hates her grandmother, who died before she was born. Sarah travels to the past with the intent to kill grandma and waits outside her house with the gun ready to shoot. When the opportunity comes, Sarah pulls up the trigger but she slips on a banana skin and thus misses grandma, who a few years later gives birth to Sarah's mother, an event that leads through a long chain of causal connections to Sarah's time travel. Furthermore, no matter how many times Sarah attempts to kill her grandmother, she is destined to fail for some commonplace reasons. Thus, if causal loops were possible, there would be such improbable and/or causally inexplicable coincidences – coincidences that would be explained by neither direct causal connections nor common causes.

All these improbability/inexplicability arguments, as well as some other impossibility arguments, are based on presuppositions that are common in linear causation but unwarranted in causal loops. (Lewis, 1986, Chapter 18; Horwich, 1987, Chapter 7; Smith, 1997; Berkovitz, 2001, 2002; Dowe, 2003) Indeed, as we shall see below, overlooking the differences between linear causation and causal loops may lead to the wrong conclusions about the challenges that causal loops pose for retro-causal models.

6.5 On Causal Loops in Bell-like Retro-Causal Models

Deterministic and indeterministic Bell-like retro-causal models of EPR/B experiments may be divided into two types, according to whether the settings of the measurement apparatuses or the measurement outcomes influence the complete pair-state at the source. Both types of models face challenges in Experiment X, though the challenges are somewhat different. For the sake of simplicity, in what follows we shall mainly focus on models of the first type, and we shall only discuss briefly models of the second type. For a more detailed discussion of models of the second type, see Section 6.7.2 and Berkovitz (2008, Sections 5.1.2, 5.2.2, and 9).

6.5.1 Deterministic Models

Similarly to conventional deterministic Bell models, we may suppose that in deterministic retro-causal models of the EPR/B experiment complete pair-states prescribe outcomes for all possible spin measurements, and in any such state the outcomes are independent of each other.[7] When the particle pair is in the singlet state and the measurements in both wings are of the same spin quantities, the outcomes are anti-correlated, and accordingly there are two possible joint outcomes; and when the measurements are of different spin quantities, all the four combinations of joint outcomes are possible. Thus, letting l, l^* and r,$l \neq l^*$ and $l = r$, be some

[7] The main arguments to follow will hold even if we assume that complete pair-states are only compatible with certain apparatus settings, and accordingly prescribe outcomes just for the corresponding measurements.

specific directions, and X_l, X_{l^*} and X_r denote respectively the outcomes of spin measurements along these directions, complete pair-states that prescribe the following dispositions will be characteristic:

λ_{D1}: If the L- and the R-apparatus are set to measure spins along the directions l^* and r respectively, the L-outcome will be $X_{l^*} =$ 'up' and the R-outcome will be $X_r =$ 'down'.

If the L- and the R-apparatus are set to measure spins along the directions l and r respectively, the L-outcome will be $X_l =$ 'up' and the R-outcome will be $X_r =$ 'down'.

. . ..

λ_{D2}: If the L- and the R-apparatus are set to measure spins along the directions l^* and r respectively, the L-outcome will be $X_{l^*} =$ 'down' and the R-outcome will be $X_r =$ 'up'.

If the L- and the R-apparatus are set to measure spins along the directions l and r respectively, the L-outcome will be $X_l =$ 'down' and the R-outcome will be $X_r =$ 'up'.

. . ..

where '. . . . . .' refers to the dispositions in other spin measurements.

We shall argue below that in deterministic models, λ_{D1} and λ_{D2} give rise to causal loops. But first we need to look at how complete pair-states are selected, focusing on deterministic models in which the apparatus settings influence the complete pair-state at the source (henceforth, Model DS). In general, in 'hidden-variables' models of EPR/B experiments there are both controllable and uncontrollable factors that determine complete pair-states. In conventional deterministic Bell models, the controllable factors are the controllable initial conditions that bring about the quantum-mechanical pair-state at the source; and the uncontrollable factors are the relevant uncontrollable initial conditions that exist at the preparation of the source and which jointly with the controllable factors determine the complete pair-state at the emission. The uncontrollable initial conditions, and accordingly the complete pair-state, may vary from one run of the experiment to another, and the distribution of the complete pair-state reflects ignorance over these initial conditions. It is supposed that this distribution depends only on the quantum-mechanical pair-state. In Model DS, the distribution of the complete pair-state is also dependent on the settings of the measurement apparatuses. The apparatus settings during the measurements influence the complete pair-state at the emission, and accordingly λ-independence fails.

Model DS and the corresponding indeterministic model, Model IS (see Section 5.2), do not predict the existence of causal loops in conventional EPR/B experiments; for in these models of conventional EPR/B experiments, the apparatus settings influence complete pair-states at the emission but these states do not influence the settings. Things are different in models in which the measurement outcomes influence complete pair-states. In such models, complete pair-states at the emission influence the measurement outcomes and the outcomes influence these states in all EPR/B experiments. We shall discuss these models below and in Sections 6.7.2 and 6.8.

But in Experiment X, Model DS and Model IS predict causal loops, and the loops that Model DS predict may be inconsistent. Suppose, for simplicity's sake, that the relevant uncontrollable initial conditions of Experiment X are independent of the settings and the complete pair-state. Then, for λ-independence to be violated and for this violation to be the result of causal influence, there would have to be some uncontrollable initial conditions and some settings l, l^* and r, such that the settings l^* and r cause the state λ_{D1} and the settings l and r cause the state λ_{D2}. But in such circumstances Model DS predicts the occurrence of the following inconsistent causal loops (see Fig. 6.4). Loop III: if λ_{D1} obtained, the R-measurement outcome would be spin 'down' along the direction r, this outcome would set the L-apparatus to measure spin along the direction l, and thus (given the setting of the R-apparatus and the controllable and uncontrollable initial conditions) λ_{D2} should obtain. Loop IV: if λ_{D2} obtained, the R-outcome would be spin 'up' along the direction r, this outcome would cause the L-apparatus to measure spin along the direction l^*, and thus (given the setting of the R-apparatus and the controllable and uncontrollable initial conditions) λ_{D1} should obtain.

For simplicity's sake, we assumed above that the relevant uncontrollable initial conditions of Experiment X influence the complete pair-state, but are not influenced by it or by the settings of the measurement apparatuses. If this assumption is relaxed, Loop III and Loop IV may not occur; for the uncontrollable initial conditions may change according to the apparatus settings or the complete pair-state,

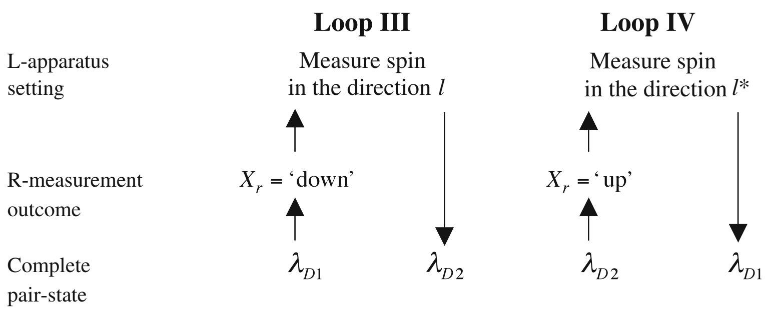


Fig. 6.4 The inconsistent causal loops that some versions of Model DS – a deterministic, Bell-like retro-causal model in which the apparatus settings influence the complete pair-state at the emission – predict in Experiment X. λ_{D1} and λ_{D2} are two characteristic complete pair-states in Model DS. These states may occur in certain apparatus settings (and initial conditions). In particular, λ_{D1} (λ_{D2}) occurs when the L- and R-apparatus are set to measure spins along the directions l^* and r (l and r), respectively; where $l \neq l^*$ and $l = r$. $X_r =$ 'down' ($X_r =$ 'up') is the R-measurement outcome being spin 'down' ('up') along the direction r

so as to exclude the circumstances that give rise to these loops. But relaxing this assumption may not be sufficient. The problem with Model DS is that while in Experiment X it predicts the existence of both forward and backward causation between the complete pair-state at the emission and the R-measurement outcome, it has no mechanism for securing the compatibility of these causal connections. It is difficult to see how the postulation of causal connection from apparatus settings or complete pair-states to uncontrollable initial conditions per se would solve this problem.

Wheeler and Feynman (1945) claim that the fact that nature is continuous could be used to argue that causal paradoxes may be avoided (for a discussion of this claim, see Maudlin, 1994 and Arntzenius and Maudlin, 2009, Section 3). Applying this idea to Model DS, Loop III and Loop IV could be avoided if the range of the L-apparatus settings and the distribution of complete pair-state are continuous. Yet, the resulting models will predict the existence of Loop I (see Fig. 6.3).

The challenge from inconsistent loops does not arise in deterministic models in which complete pair-states at the emission depend on the measurement outcomes, henceforth Model DO (for an example of such model, see Section 6.7.2). In Model DS, complete pair-states at the emission influence the measurement outcomes and the measurement outcomes may indirectly influence these states by influencing the apparatus settings. That is the case in Experiment X. The complete pair-state at the emission (jointly with the setting of the R-apparatus) causes the R-measurement outcome, and the R-outcome influences the setting of the L-apparatus, which in turn influences the complete pair-state at the emission. As we have seen, such forward and backward causal influences may be incompatible with each other and give rise to inconsistent loops. This cannot happen in Model DO, where the measurement outcomes impose further constraints on the selection of complete pair-states. Complete pair-states at the emission have to be consistent not only with the apparatus settings but also with the measurement outcomes. That is, complete pair-states have to prescribe measurement outcomes that are the same as the actual ones. Accordingly, inconsistent loops, like Loop III and Loop IV, could not occur.

However, Model DS and Model DO face another challenge in Experiment X. Unlike in conventional Bell models of the EPR/B experiment, in these models the quantum-mechanical pair-state does not determine the distribution of the complete pair-state at the emission. In Model DS of Experiment X, the distribution of the complete pair-state depends on the apparatus settings, and the distribution of the L-apparatus setting depends on the distribution of the R-outcome, which in turn depends on the distribution of the complete pair-state. The problem is that due to these dependencies, it is impossible to predict the distribution of the L-apparatus setting, and accordingly the distribution of the complete pair-state, from the quantum-mechanical pair-state. Similarly, in Model DO the distribution of the complete pair-state depends on the measurement outcomes, and the distribution of the measurement outcomes depends on the distribution of the complete pair-state, and the problem is that due to these dependencies it is impossible to determine the distribution of complete pair-state from the quantum-mechanical pair-state. In order to reproduce the predictions of standard quantum mechanics, these models have to be supplemented with postulates that determine the distribution of complete pair-states or measurement outcomes in any given quantum-mechanical pair-state and experimental set-up. The question is whether such postulates could be motivated on non-ad hoc grounds. A failure to motivate them on non-ad hoc grounds will have implications for the explanatory power of the deterministic retro-causal models.

It may be argued that (more) conventional interpretations of quantum mechanics also include analogous probabilistic postulates, and it is similarly difficult to motivate them on non-ad hoc grounds. For example, the orthodox interpretation assumes the Born rule and Bohmian mechanics presupposes that the distribution of the possible position configurations of the particles is determined by the quantum-mechanical wavefunction, and the challenge is to provide a non-ad hoc justification for these postulates. So the problem of justifying the above probabilistic postulate in retro-causal models may seem on a par with the problem of justifying the probabilistic postulates in more conventional interpretations of quantum mechanics. Yet, as we shall suggest in Sections 6.6, 6.7.2, and 6.9, there are reasons to think that retro-causal interpretations of quantum mechanics may require two independent probabilistic postulates, and accordingly render the problem of justification more challenging.

6.5.2 Indeterministic Models

Like conventional indeterministic Bell models, indeterministic Bell-like retro-causal models of EPR/B experiments satisfy *Factorizability*: the probability of joint measurement outcomes factorizes into the product of their single probabilities. Since in the singlet state outcomes of measurements of the same spin quantities in both wings have to be anti-correlated, the probabilities of such outcomes have to be either one or zero. On the other hand, when the measurements are of different spin quantities, the outcomes need not be anti-correlated, and accordingly their probabilities may be strictly between zero and one. Thus, complete pair-states that prescribe the following probabilities will be characteristic in indeterministic Bell-like retro-causal models of the EPR/B experiment with the quantum-mechanical pair-state being the singlet:

λ_{I1}: If the L- and the R-apparatus are set to measure spin along the directions l^* and r respectively, the probability of $X_{l^*} =$ 'up' is p_{l^*} and the probability of $X_r =$ 'up' is p_r; where p_{l^*} and p_r are strictly between 0 and 1.

λ_{I2}: If the L- and the R-apparatus are set to measure spin along the directions l and r respectively, the probabilities of $X_l =$ 'up' and of $X_r =$ 'down' are both 1.

where l, l^* and r, $l \neq l^*$ and $l = r$, are some particular directions, and X_l, X_{l^*} and X_r denote the outcomes of spin measurements along these directions. Note that on pain of trivializing the probabilities of outcomes (so that they will always be either zero or one), in indeterministic Bell-like retro-causal models complete pair-states cannot assign probabilities for all possible measurements.

We shall argue below that the states λ_{I1} and λ_{I2} generate causal loops in Experiment X. To prepare the ground, we need first to consider the nature of the probability distribution of the complete pair-state and the causal mechanism that gives rise to it. Recall that in deterministic Bell-like models of EPR/B experiments, the probability distribution of the complete pair-state is epistemic, reflecting ignorance about the actual complete pair-state. In indeterministic models, this distribution may also have a non-epistemic interpretation, reflecting the idea that the actual complete pair-state is the outcome of indeterministic causal process. The causes of this state depend (among other things) on the ontological status of quantum-mechanical states. Quantum-mechanical states may be interpreted as either representing physical states of systems, or states of information, knowledge or ignorance about systems. To simplify things, we shall focus on the non-epistemic interpretation, and assume that the quantum-mechanical pair-state is an incomplete pair-state, which is a partial cause of the complete pair-state at the emission. We shall focus on indeterministic, Bell-like retro-causal models in which the quantum-mechanical pair-state and the apparatus settings jointly constitute an indeterministic cause of the complete pair-state, and the distribution of the complete pair-state reflects this causal connection (henceforth, Model IS). We shall assume that this distribution denotes the single-case propensities that the particles have to be in various complete pair-states in a single run of the EPR/B experiment. This interpretation is plausible in indeterministic models. Moreover, as we shall see in Section 6.6, our analysis of the challenges that causal loops pose for retro-causal interpretations will also be applicable to the frequency and epistemic interpretations of probability. In general, the distribution of the complete pair-state need not be discrete, and accordingly the probability of these states may have zero measure. Yet, to simplify things, we shall suppose that the distribution is discrete and moreover that the propensities of the complete pair-states λ_{I1} and λ_{I2} are non-zero in certain settings. More specifically, we shall suppose that in the singlet state and some settings l, l^* and r, $l \neq l^*$ and $l = r$, λ_{I1} has non-zero propensity of occurring in the settings l^* and r, and zero propensity of occurring in the settings l and r; and λ_{I2} has non-zero propensity of occurring in the settings l and r. Nothing essential in our analysis will hinge on the above simplifications.

Granted these assumptions, Model IS predicts the occurrence of Loop V and Loop VI in Experiment X (see Fig. 6.5). In Loop V, the complete pair-state λ_{I1} (jointly with the setting of the R-apparatus to measure spin along the direction r) is an indeterministic cause of the R-outcome spin 'up' along the direction r, this outcome is a deterministic cause of the setting of the L-apparatus to measure spin along the direction l^*, and this setting (jointly with the setting of the R-apparatus and the singlet state) is an indeterministic cause of λ_{I1}. In Loop VI, the complete pair-state λ_{I2} (jointly with the setting of the R-apparatus to measure spin along the direction r) is an indeterministic cause of the R-outcome spin 'down' along the direction r, this outcome is a deterministic cause of the setting of the L-apparatus

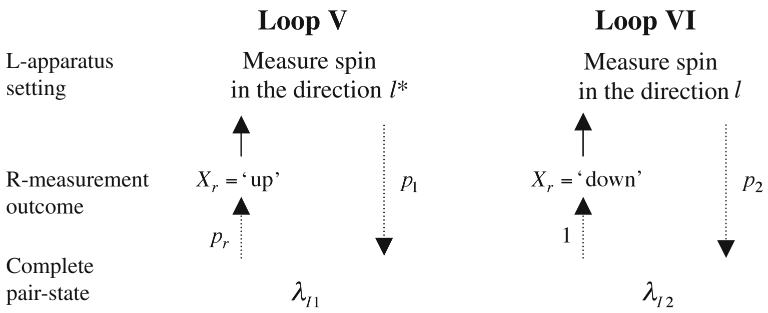


Fig. 6.5. Loop V and Loop VI are among the causal loops that Model IS – an indeterministic, Bell-like retro-causal model in which the apparatus settings influence the complete pair-state at the emission – predicts in Experiment X. Arrows denote deterministic causal connections. Dashed arrows denote indeterministic causal connections, and the associated probabilities denote the probabilities that causes give to their effects in the loop's circumstances. The probabilities p_r, p_1 and p_2 are strictly between zero and one. λ_{I1} and λ_{I2} are characteristic states in Model IS, and X_r = 'down' (X_r = 'up') is the R-measurement outcome being spin 'down' ('up') in the direction r

to measure spin along the direction l, and this setting (jointly with the setting of the R-apparatus and the singlet state) is an indeterministic cause of λ_{I2}.

Note that in Loop VI the causal connection between the state λ_{I2} and the outcome X_r = 'down' is depicted as indeterministic. The idea is that the propensity one of X_r = 'down' in λ_{I2} does not entail that X_r = 'up' is impossible in this state, and that there may be a zero measure of such outcomes.[8] If the causal connection between λ_{I2} and the outcome X_r = 'down' is deterministic, Loop VI will have only one indeterministic connection; whereas if this causal connection is indeterministic, the loop will have two indeterministic causal connections. As we shall see in Sections 6.6 and 6.9, loops with more than one indeterministic causal connection pose a more acute challenge for the explanatory power of indeterministic retro-causal interpretations. In any case, it is also noteworthy that in indeterministic universe the causal connection between the R-measurement outcome and the L-apparatus could be indeterministic, thus giving rise to a causal loop with two indeterministic causal connections even if the causal connection between λ_{I2} and the outcome X_r = 'down' is considered deterministic.

Although Loop V and Loop VI are consistent, they pose a challenge for Model IS. In each of these loops, the model assigns probabilities that the loop's causes give to their effects in the circumstances (where causes, effects and causal circumstances are the states of the relevant systems). In particular, in Loop V the state λ_{I1} (jointly with the setting of the R-apparatus to measure spin along the direction r) determines the propensity of the R-measurement outcome X_r = 'up' to be p_r, and

[8] I am grateful to Mauricio Suarez for pointing out to me the importance of emphasizing here the distinction between deterministic causal connections and indeterministic causal connections with propensity one.

the setting of the L-apparatus to measure spin along the direction l^* (jointly with the setting of the R-apparatus and the singlet state) determines the propensity of the state λ_{I1} to be p_1. Yet, as we shall see in Section 6.6, in Model IS there is no known way to compute from these (and the other model) probabilities the long-run frequency of λ_{I1}, or the long-run frequency that the outcome $X_r =$ 'up' has in this state; and similarly for other complete pair-states. The same is true for indeterministic, Bell-like retro-causal models in which the measurement outcomes influence the complete pair-states (henceforth, Model IO). Thus, indeterministic Bell-like retro-causal models fail to provide definite statistical predictions of measurement outcomes in Experiment X.

For simplicity's sake we supposed that characteristic complete pair-states have non-zero probability. In general, the distribution of complete pair-states in indeterministic retro-causal models may not be discrete, and accordingly states like λ_{I1} and λ_{I2} may have a measure zero probability of occurring. Yet, the reasoning above could be slightly modified so as to apply to models with non-discrete distribution of complete pair-states. The problem with Model IS and Model IO is that in Experiment X, the model's probability distribution of complete pair-states does not determine their long-run frequencies, and the model's probabilities of outcomes in complete pair-states do not determine their long-run frequencies. It is difficult to see how the postulation of a continuum of complete pair-states per se could solve this problem.

Like in the deterministic retro-causal models, in order to reproduce the predictions of standard quantum mechanics Model IS and Model IO have to be supplemented with a Born-like postulate. Unlike in the deterministic models, a postulate that determines the distribution of complete pair-states in any quantum-mechanical pair-state and experimental set-up would not do; for in the indeterministic retro-causal models, there is no known way to compute the long-run frequencies of outcomes in complete pair-states from the probabilities that these states assign to the outcomes. In the indeterministic models, the probabilistic postulate has to specify the distribution of measurement outcomes in any given quantum-mechanical pair-state and experimental set-up. As in the deterministic retro-causal models, the question is whether such a postulate could be motivated on non-ad hoc grounds. And, again, a failure to motivate such postulate on non-ad hoc grounds would have implications for the explanatory power of the indeterministic retro-causal models.

Moreover, since there is no known way to compute the long-run frequencies of outcomes in complete pair-states from the probabilities that these states assign to the outcomes, the challenge for the predictive power of these models has further consequences for their explanatory power. Unlike in conventional Bell models and deterministic retro-causal Bell-like models, the probability of outcomes in any given quantum-mechanical pair-state could not be obtained as an average-sum over the probabilities of outcomes in the complete pair-states. So although Model IS and Model IO do not postulate any non-local influences, they fail to fulfil one of the main objectives of retro-causal interpretations of quantum mechanics – namely, to explain how the quantum-mechanical correlations between the measurement outcomes could be explained by local common-causes. Thus, curiously, although the probabilities of outcomes in these models are factorizable, they still fail to live up to the standards of explanation dictated by Reichenbach's principle of the common cause.

6.6 On Probabilities and Predictions in Indeterministic Causal Loops

In order to see why indeterministic Bell-like retro-causal models of Experiment X fail to provide a common-cause explanation of the quantum-mechanical correlations between measurement outcomes, we need to reflect for a moment on the nature of probabilities in indeterministic causal loops. Consider the following curious coin toss. The tossing momentum exerted by my finger is an indeterministic cause of the coin landing on either 'heads' or 'tails'. That is, the tossing momentum and the relevant circumstances (the air friction, the coin structure, the landing surface, etc.) jointly determine the single-case propensities of 'heads' and 'tails', and as a matter of sheer chance the coin lands on either 'heads' or 'tails'. The propensities of 'heads' ('tails') in tosses with momentum m and with momentum m^* are $\frac{1}{2}$ and $\frac{1}{4}$ ($\frac{1}{2}$ and $\frac{3}{4}$), respectively. The coin landing on 'heads' ('tails') is a deterministic cause of my perception of 'heads' ('tails'), and my perception of 'heads' ('tails') is a deterministic cause of my finger exerting momentum m (m^*) at the earlier tossing time.

Each coin toss generates either Loop VII or Loop VIII (see Fig. 6.6). In Loop VII, the propensity of 'heads' in tosses with a momentum m is $\frac{1}{2}$, but the frequency of 'heads' in the reference class of such tosses is 1; for the fact that a toss with momentum m is a deterministic effect of 'heads' dictates that 'heads' occurs whenever a toss with momentum m occurs. In Loop VIII, the propensity of 'tails' in tosses with a momentum m^* is $\frac{3}{4}$, but the frequency of 'tails' in the reference class of such tosses is 1; and, again, this conditional frequency is a consequence of the loop's consistency conditions, which dictate that 'tails' occurs whenever a toss with momentum m^* occurs.

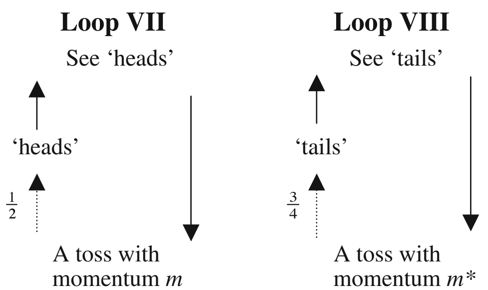


Fig. 6.6 The indeterministic causal loops that occur in the curious coin-toss example. Arrows denote deterministic causal connections, and dashed arrows denote indeterministic causal connections with the associated probabilities being the probabilities that causes give to their effects in the circumstances

These inequalities between propensities and the corresponding frequencies are curious, but do not entail any inconsistency. The law of the large numbers relates single-case propensities to long-run frequencies in the following way: the long-run frequency of ‘heads’ in a *non-biased* reference class of independent tosses with the same momentum (and the same relevant circumstances) will almost certainly be equal to the propensity of ‘heads’ in such tosses. But in Loop VII and Loop VIII, the reference classes of such tosses are *biased*. In Loop VII (Loop VIII), it is a class of tosses that occur only when the coin lands on ‘heads’ (‘tails’). Thus, the assumption that the long-run frequency of ‘heads’ in Loop VII (‘tails’ in Loop VIII) will almost certainly display the corresponding propensity is unwarranted.

In Loop VII and Loop VIII, all the causal connections but one are deterministic. Accordingly, it is possible to compute the *conditional* frequency of ‘heads’ in the reference classes of tosses with momentum m (m^*) from the loop’s constraints – namely, from the fact that a toss with momentum m (m^*) occurs just in case the coin lands on ‘heads’ (‘tails’). But, there is no known way to compute the *unconditional* frequencies of the occurrence of these loops, and accordingly the *unconditional* long-run frequency of ‘heads’ in a series of independent tosses.

Things are even more complicated in loops that have more than one indeterministic causal connection. Consider a slightly different version of our curious coin-toss in which the causal connection between ‘heads’ (‘tails’) and perceiving ‘heads’ (‘tails’) is indeterministic. In this case, coin tosses will give rise to Loop IX – Loop XII (see Fig. 6.7), which are similar to Loop VII and Loop VIII but have an additional indeterministic causal connection. The constraints of these loops do not

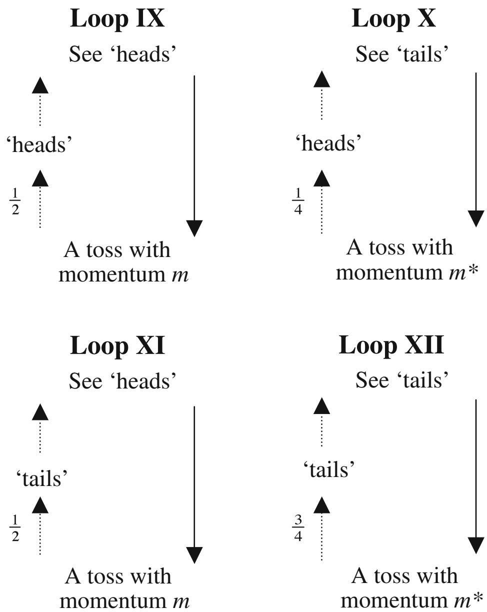


Fig. 6.7 The causal loops that occur in the revised version of the curious coin-toss example, where ‘heads’ (‘tails’) is an indeterministic cause of seeing ‘heads’ (‘tails’)

determine the conditional frequency of 'heads' in tosses with momentum m (m^*); and there is no known way to compute this conditional frequency from the propensities that causes give to their effects in these loops. More generally, in loops with more than one indeterministic causal connection, the loop's constraints do not determine the frequency of effects in the reference class of their indeterministic causes, and there is no other known way to compute these frequencies.[9] The indeterministic Bell-like retro-causal models we considered in Section 6.5.2 predict the occurrence of causal loops with two indeterministic causal connections. Thus, in such models there is no known way to compute the (distribution of the) long-run frequencies of complete pair-states, or the long-run frequencies of outcomes in a given complete pair-state.

It may be tempting to attribute the problems above to the interpretation of the model probabilities as propensities, and to suggest that these problems would not arise under other common interpretations of probabilities, like the frequency and epistemic interpretations. But such reinterpretations of the model probabilities would not do. To see why, let us reconsider our coin-toss examples, and this time interpret the model probabilities as prescribing long-run frequencies of 'heads' and 'tails' in the reference classes of tosses with momentum m and with momentum m^*. That is, suppose that our model of the coin postulates that the long-run frequency of 'heads' in the reference class of tosses with momentum m (m^*) is $\frac{1}{2}$ ($\frac{1}{4}$), and the long-run frequency of 'tails' in the reference class of tosses with momentum m (m^*) is $\frac{1}{2}$ ($\frac{3}{4}$). While these frequencies characterize the nature of the coin and the tossing set-up, they fail to determine the long-run frequencies of 'heads' and 'tails' in our coin-toss examples. In our first coin-toss example, where Loop VII and Loop VIII occur, the long-run frequency of 'heads' ('tails') in the reference class of tosses with momentum m (m^*) will still be one, and this frequency is determined by the loop's constraints rather than the model probabilities for the coin. Further, the model probabilities and the constraints of Loop VII and Loop VIII do not determine the unconditional long-run frequencies of these loops and, accordingly, the unconditional long-run frequencies of 'heads' and 'tails'. Turning to the second coin-toss example, where Loop IX – Loop XII occur, suppose (for simplicity's sake) that the imperfect perception of the outcomes of tosses is also characterized by long-run frequencies, so that the long-run frequency of 'heads' ('tails') in the reference class of the coin landing on 'heads' ('tails') is less than 1. Like in the propensity interpretation of the model probabilities, the long-run frequencies that characterize the coin and the imperfect perception fail to determine the long-run frequencies of 'heads' and 'tails' in tosses with momentum m (m^*), and there is no known way to compute these frequencies from the constraints of these loops.

Similarly, the model probabilities fail to determine the long-run frequencies of 'heads' and 'tails' when they are interpreted as epistemic. The inability to predict the

---

[9] The constraints of causal loops with two indeterministic causes may, however, limit the range of the possible long-run frequencies of effects in the reference class of their indeterministic causes. For an example, see Berkovitz (2002, Section 4).

long-run frequencies of ‘heads’ and ‘tails’ in the coin-toss examples is not due to the interpretation of the model probabilities per se. These probabilities are probabilities that causes give to their effects in the local circumstances – namely, probabilities that are determined by the states of the systems that are involved in these causal connections and which are independent of the other causal connections in the loop. For example, the probability $\frac{1}{2}$ of the coin landing on ‘heads’ in a toss with momentum m is assumed to reflect the coin’s characteristics and the local tossing circumstances rather than the characteristics of Loop VII. This probability is supposed to be the same, independently of whether the toss occurs in causal loops or chains of linear causal connections. In indeterministic causal loops, such probabilities do not determine the long-run frequencies that effects have in the reference class of their causes, independently of whether these probabilities are interpreted as propensities, long-run frequencies or epistemic probabilities. In indeterministic causal loops with one indeterministic causal connection, like Loop VII and Loop VIII, the loop’s constraints determine the long-run frequencies of effects in the reference class of their causes. But in causal loops with more than one indeterministic connection, like Loop V, Loop VI and Loop IX – Loop XII, the loop’s constraints are not sufficiently strong to determine these frequencies, and there is no other known way to compute them on any of the above interpretations of probability. Further, in all the indeterministic causal loops above, the probabilities that causes give to their effects in the local circumstances and the loop’s set-up do not determine its long-run frequency, and accordingly the long-run frequencies of events in it; and, again, this failure is independent of whether these probabilities are interpreted as single-case propensities, long-run frequencies or epistemic probabilities.

6.7 Retro-Causal Theories and the Measurement Problem

One of the main objectives of current interpretations of quantum mechanic is to solve the measurement problem that mars the orthodox interpretation. Lewis (2006, 375) proposes that retro-causal interpretations of quantum mechanics may suffer from a ‘measurement problem’. Their dynamical laws may depend on whether or not a measurement occurs. In Bell-like retro-causal models of EPR/B experiments, the complete pair-state at the emission depends on the measurement events. If measurements as *such* were the trigger of the backward-causal mechanism, “then whether a measurement occurs would have a dynamical effect on the behavior of the system. But, as stressed before, since there is no fundamental distinction between measurement processes and non-measurement processes, any theory that gives a dynamical role to measurement as such is ill-founded.”

Whether or not a retro-causal interpretation is subjected to the measurement problem will depend on its ontology. In what follows, we shall consider the ontology of two retro-causal theories: Cramer’s (1980, 1986, 1988) ‘transactional’ interpretation of quantum mechanics, and Sutherland’s (1998, 2008) causally symmetric Bohmian model. Our main focus will be on the measurement problem. (For a detailed discussion of the challenges that causal loops raise for predictions in these theories, see Berkovitz 2008, Sections 5, 8 and 9.) We shall argue that whether a retro-causal interpretation of quantum mechanics is subjected to the measurement problem does not depend on the postulation of backward-causal mechanism per se.

6.7.1 The Transactional Interpretation of Quantum Mechanics

Cramer's theory is a retro-causal interpretation of orthodox quantum mechanics, which was inspired by Wheeler and Feynman's (1945, 1949) Absorber theory. The Wheeler-Feynman theory was originally conceived as a time-symmetric alternative to conventional electromagnetism. The basic idea of this theory is that electromagnetic interactions involve time-reversed 'advanced-wave' solutions as well as the usual 'retarded-wave' solutions to the electromagnetic wave equation. Cramer extends this idea to the quantum domain. In his theory, the basic causal mechanism, which constitutes the fundamental quantum-mechanical interactions, is a 'transaction' between 'emitters' and 'absorbers.' (Cramer, 1986, 665) In this transaction, there are 'retarded' waves that propagate forward in time, and the corresponding 'advanced' waves that propagate backward in time. The usual solution of the Schrödinger equation, ψ, represents a retarded wave, and the solution of the complex conjugate of the Schrödinger equation, ψ^*, represents an advanced wave. Emitters emit both ψ and the time-reversed counterpart wave ψ^*, which is exactly out of phase with ψ. Cramer calls both of these waves 'offer waves'. Figure 6.8a provides a simple, two-dimensional example of offer waves. The bold post-emission sinusoidal wave is the retarded offer wave, and the dashed pre-emission sinusoidal

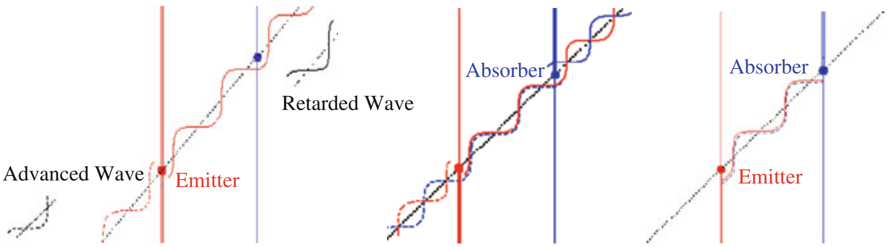


Fig. 6.8 A simple, two-dimensional illustration of the transaction between 'emitters' and 'absorbers' in Cramer's theory. Figure 6.8a shows the 'offer' waves that the emitter sends. The bold post-emission sinusoidal wave is the retarded offer wave and the dashed pre-emission sinusoidal wave is the advance offer wave, which is exactly out of phase with the retarded wave. Figure 6.8b shows both of these offer waves and the 'confirmation' waves that the absorber produces in response to the retarded offer from the emitter. The confirmation waves are of the same amplitude as the offer waves. The post-absorption bold sinusoidal wave is the retarded confirmation wave, and the pre-absorption dashed sinusoidal wave is the advance confirmation wave, which is exactly out of phase with the retarded confirmation wave. Figure 6.8c shows the net effect of the offer and confirmation waves. The pre-emission and the post-absorption waves are cancelled and the only non-zero wave is spanned in the space-time region between the emission and the absorption, where the dashed advanced confirmation wave reinforces the bold retarded offer wave

wave is the advance offer wave. The retarded offer wave interacts with the absorber, which absorbs it and in response emits advanced and retarded ‘confirmation waves’, and again these waves are exactly out of phase. Figure 6.8b provides an example of such waves, where the post-absorption bold sinusoidal wave is the retarded confirmation wave, and the pre-absorption dashed sinusoidal wave is the advance confirmation wave. The confirmation and offer waves extending forward in time beyond the absorption and backward in time beyond the emission are exactly out of phase. The amplitudes of the confirmation waves are the same as those of offer waves. Accordingly, the ‘pre-emission’ and the ‘post-absorption’ waves are cancelled, and the only non-zero wave is in the space-time region between the emission and the absorption, where the advanced confirmation wave reinforces the retarded offer wave (see Fig. 6.8). The final amplitude of the standing wave between the emitter and the absorber is $\psi^*\psi$, and the probabilities of outcomes are determined by this amplitude according to the Born rule.

Cramer describes the transaction between the emitter and the absorbers as occurring in cycles that repeat until the transaction is completed. The transaction are constituted by cycles of a four-step sequence, where (1) the emitter sends offer waves, (2) the absorbers absorb the retarded offer waves and produces retarded and advanced confirmation waves, (3) the advanced confirmation waves are sent back to the emitter, and (4) the emitter responds to these confirmation waves. This cycle repeats itself until all of the quantum boundary conditions are satisfied, at which point the transaction is completed, a wavefunction collapse occurs, and the only traces of the transaction are the resulting standing wave, $\psi^*\psi$. (Cramer 1986, 661–663)

Cramer (1986, 661, footnote 14) says that the cycles of transactions occur in pseudo-time. He considers this account only as a heuristic device, and emphasizes that the process is atemporal. This raises questions as to the explanatory value of the transactional interpretation. If the description in pseudo-time were the whole story, the transactional interpretation would be like a standard collapse interpretation with an associated tale about how collapses come about.

Cramer (1986, 663) also suggests an alternative account of the transaction, where the four-dimensional vector (or, in short, the four-vector) ‘standing wave’ spanned between the emitter and the absorber is supposed to embody the transaction between them. “As a three-space standing wave is a superposition of waves travelling to the right and left, this four-vector standing wave is a superposition of advanced and retarded components. It has been established between the terminating boundaries of the emitter, which blocks passage of the advanced wave further down the time stream, and the absorber, which blocks passage of the retarded wave further up the time stream. This space-time standing wave is the transaction we shall use as a basis for the discussion that follows.” Figure 6.8 provides a simple example of the general idea in two-dimensional space-time: the standing wave between the emitter and the absorber in Fig. 6.8c is a superposition of the offer and confirmation waves in Fig. 6.8b.

In order to determine whether Cramer’s theory is subjected to the measurement problem, we first need to clarify the ontology of the transaction. Cramer’s theory is intended to be a retro-causal interpretation of standard quantum mechanics. As such, it is a collapse theory, and the transaction between emitters and absorbers is supposed to provide an explanation, or account of how the collapse comes about. In the general case, the "emitter is presented with echoes [i.e. confirmation waves] from potential absorbers which form a weighted list of transactions, from which only one may be chosen. The future absorbers can influence the past emission event only through the strength of their echo entry on the list, but cannot influence which entry is chosen for the transaction." (Cramer 1986, 668) Now, "[b]ecause of the quantum-mechanical boundary conditions, the transaction is only completed between a single emitter and absorber in one quantum effect." (Cramer 1986, 667) The transaction is supposed to be completed with the collapse of the wavefunction (Cramer 1986, 665), but Cramer's theory does not provide any account of the nature of the collapse and how it comes about. Like the pseudo-time account, the standing-wave account of the transaction fails to depict the physical process that leads to the collapse and to the measurement outcomes. The theory only postulates that the quantum-mechanical wavefunction ψ is the initial 'offer wave' of the transaction; the superposition of the offer and confirmation waves $\psi + \psi^*$, which is represented by the standing wave between the emitter and absorber, embodies the transaction; and the collapse wavefunction is identical to the completion of the transaction. Yet, it is not clear how the standing wave between the emitter and absorber is related to the collapsed quantum-mechanical wavefunction. Cramer's (1986, 670–1) discussion of Renninger's (1953) 'negative-result' thought experiment seems to suggest that standing waves exist only between emitters and absorbers between which energy is transferred. If so, these waves represent the outcome of the transaction, i.e. the collapsed wavefunction, rather than the communication between the emitter and the various potential absorbers. On the other hand, if standing waves exist between the emitter and all its potential absorbers, then they represent the part of communication that precedes the collapse. Either way, the relation between the standing waves and the collapsed wavefunction remains a mystery. The transactional interpretation postulates the existence of wavefunction collapse but fails to account for it as a real physical process. So similarly to the orthodox interpretation, it is subjected to a measurement problem.

In a later review of his theory, Cramer (1988, Section 6) seems to recognize this difficulty when he remarks that the problem of "accommodating collapse for a single quantum event, is one that must be addressed by the formalism" and that the transactional interpretation "cannot supply mechanisms missing from the formalism." He holds that the "nonlocal collapse mechanism is strictly at the interpretational level." And while he rejects the orthodox interpretation and its suggestion that wavefunction collapses are related essentially to measurements as such (Cramer, 1986, Section I), his transactional interpretation is subjected to the same measurement problem (cf. Marchildon, 2006, 13–14). Indeed, the measurement problem in the transactional interpretation is related to the fact that it is formulated within the framework of the orthodox interpretation, rather than to the postulation of retro-causal mechanism per se.

Finally, it is noteworthy that while the transactional interpretation is a retro-causal interpretation of quantum mechanics, its model for EPR/B-like experiments is different from the Bell-like retro-causal models we discussed in Section 6.5. In particular, unlike the Bell-like retro-causal models, the transactional interpretation violates *Factorizability*.

6.7.2 Causally Symmetric Bohmian Model

Lewis is well aware of the fact that a retro-causal mechanism per se does not require an appeal to measurements *as such*. Indeed, in his discussion of Price's (1996) retro-causal model for the EPR/B experiment, he suggests that the model could avail itself to Bohmian-like mechanism where the cause of the correlation between the complete pair-state and apparatus settings "is not the act of the measurement as such, but the motion of the particles during the measurements." (Lewis, 2006, 375) In fact, the idea that the ontological framework of Bohmian mechanics could serve as a basis for a retro-causal interpretation of quantum mechanics has been pursued by Rod Sutherland. Sutherland (1998, 2008) proposes that the main postulates of Bohmian mechanics can be revised so as to give rise to a causally symmetric Bohmian model, where all influences are local: the causes of any event are confined to its backward and forward light cones.

Recall that unlike orthodox quantum mechanics, in conventional Bohmian mechanics wave functions always evolve according to the Schrödinger equation, and thus never collapse. Wave functions do not represent the states of systems. Rather, they are states of a 'quantum field (on configurations space)' that influence the states of systems. The theory is deterministic. Systems always have definite positions (the so-called 'hidden variables'), and their positions and the wavefunction at a given time jointly determine their trajectories at all future times. Thus, the positions of systems and their wavefunction determine the outcomes of any measurements so long as these outcomes are recorded in the positions of some physical systems, as in any practical measurement. In particular, in EPR/B-like experiments the quantum-mechanical wavefunction of the particles and their position configuration constitute their complete state, and this state and the apparatus settings jointly determine the measurement outcomes.

Wave functions govern the trajectories of systems according to the so-called 'guidance equation', which expresses the velocities of systems at any time t, $\mathbf{v}(\mathbf{x}, t)$, in terms of their wavefunction at that time:

$$\text{(Guidance)} \quad \mathbf{v}(\mathbf{x}, t) \equiv \frac{d\mathbf{x}}{dt} = \frac{\hbar}{2im} \frac{\psi^* \overset{\leftrightarrow}{\nabla} \psi}{\psi^* \psi};$$

where m is the system's mass, $\mathbf{x}$ is the system's position configuration, $\hbar$ is Planck's constant, $\overset{\leftrightarrow}{\nabla}$ stands for $\overset{\rightarrow}{\nabla} - \overset{\leftarrow}{\nabla}$ and the grad operators $\overset{\rightarrow}{\nabla}$ and $\overset{\leftarrow}{\nabla}$ act to the right and to the left, respectively.

Since Bohmian mechanics is a deterministic theory, its predictions in individual measurements are different from those of standard quantum mechanics. Yet, the theory reproduces the observable predictions of standard quantum mechanics by postulating that the distribution $\rho(\mathbf{x}, t)$ of the possible position configurations, $\mathbf{x}$, at any time t is determined by the wavefunction ψ at that time:

$$\text{(Distribution)} \quad \rho(\mathbf{x}, t) = \psi^* \psi.$$

Given *Distribution*, the predictions of standard quantum mechanics are obtained as statistical averages over the outcomes that Bohmian mechanics predicts in each of the position configurations that the particles may be, according to the distribution of these configurations in ψ.

In the causally symmetric Bohmian model, the velocity of a system depends on two wave functions, which are supposed to reflect its initial and final boundaries conditions. The 'initial' wavefunction, ψ_i, which is supposed to record the initial boundary conditions, evolves forward in time, and the 'final' wavefunction, ψ_f, which is supposed to record the final boundary conditions, evolves backward in time. The nature of these boundary conditions depends on the exact ontology of the model. Sutherland does not specify this ontology, but we may think about the boundary conditions as determining the local fields that guide the particles' trajectories. Both wave functions are solutions of the time-dependent Schrödinger equation, and the final wavefunction is not to be confused with the time-evolved initial wavefunction, or the time-evolved quantum-mechanical wavefunction to a later time. Sutherland proposes that *Guidance* and *Distribution* could be reformulated along the following lines, so as to yield a causally symmetric Bohmian model:

$$\text{(Guidance}^{\text{S}}\text{)} \quad \mathbf{v}(\mathbf{x}, t) = \text{Re}\left(\frac{\hbar}{2ima} \frac{\psi_f^* \overleftrightarrow{\nabla} \psi_i}{\psi_f^* \ \psi_i}\right)$$

$$\text{(Distribution}^{\text{S}}\text{)} \quad \rho(\mathbf{x}, t) = \text{Re}\left(\frac{1}{a} \psi_f^* \ \psi_i\right);$$

where a is a normalization factor:

$$\text{(Normalization)} \quad a \equiv \int_{-\infty}^{\infty} \psi_f^* (\mathbf{x}, t) \, \psi_i (\mathbf{x}, t) \, d^3\mathbf{x}.$$

While conventional Bohmian mechanics reproduces the correlations between the distant measurement outcomes in EPR/B experiments by postulating non-local influences between them, the causally symmetric Bohmian model accounts for these correlations by a local, common-cause: the complete pair-state at the emission. The complete pair-state is constituted by the positions of the particles and their initial and final wave functions. But since the measurement outcomes are determined by the initial and final wave functions of the particles, the relevant complete pair-state is constituted by these wave functions. The final wave functions of the particles between the emission and spin measurements are eigenstates of spin quantities that correspond to the measurement outcomes. The initial wavefunction of the particle pair – the non-separable quantum-mechanical wavefunction – and the final wavefunction of each of the particles between the emission and its measurement jointly determine the initial wavefunction of the other particle at the emission (see Fig. 6.9). Formally, the initial wave functions of the particles between the emission and the measurements are obtained as follows:

$$\text{(Initial)} \quad \psi_i\,(x_1) = \frac{1}{N_1}\int_{-\infty}^{\infty} \psi_f^*\,(x_2)\;\psi_i\,(x_1,x_2)\,d^3x_2$$
$$\psi_i\,(x_2) = \frac{1}{N_2}\int_{-\infty}^{\infty} \psi_f^*\,(x_1)\;\psi_i\,(x_1,x_2)\,d^3x_1;$$

where $\psi_i\,(x_1,x_2)$ is the (non-separable) wavefunction of the particle pair, $\psi_i(x_1)$ and $\psi_i(x_2)$ are the initial wave functions of the particles, and $\psi_f(x_1)$ and $\psi_f(x_2)$ are their final wave functions.

So the initial and the final wave functions of both particles before the measurements are separable, and the measurement outcomes are determined in a perfectly local way. Accordingly, *Factorizability* obtains: the probability of joint measurement outcomes factorizes into the single probability of outcomes. That is, the joint probability of the outcomes given the (relevant) complete pair-state (i.e. the initial and the final wave functions of the particles), and the apparatus settings is equal to the product of the single probabilities of outcomes given the complete pair-state and

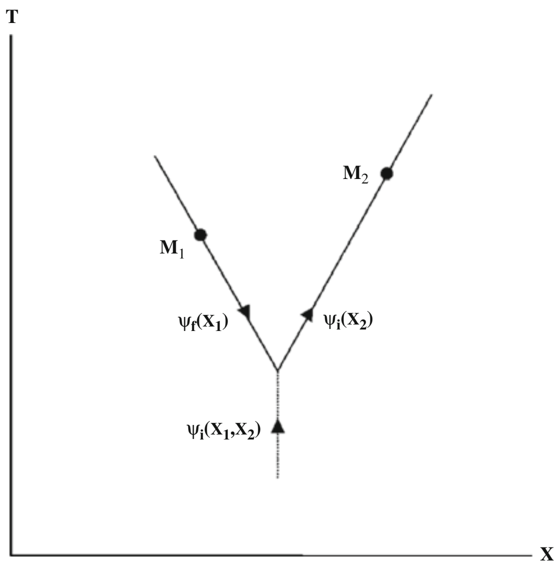


Fig. 6.9 A schematic diagram of how the initial wavefunction of the R-particle is formed in the EPR/B experiment. The formation of the initial wavefunction of the L-particle is similar

the local apparatus setting. In fact, the causally-symmetric Bohmian model satisfies a stronger factorizability condition: the joint probability of measurement outcomes given the complete pair-state and the apparatus settings factorizes into the single probabilities of each of the outcomes given the complete state of the local particle, which is constituted by its initial and final wave functions, and the local apparatus setting.[10]

It is noteworthy, though, that for the causally symmetric Bohmian model to be genuinely local, the non-separable quantum-mechanical wavefunction has to have an epistemic interpretation. As we shall see below, this wavefunction could be interpreted as a state that provides information about the distributions of the separable initial and final wave functions of the particles.

As is not difficult to see, measurements *as such* play no role in the dynamical laws of the causally symmetric Bohmian model. In particular, the backward-causal mechanism that accounts for the correlations between distant measurement outcomes in EPR/B experiments is not triggered by measurements as such. It is an integral part of the ontology of this theory. Yet, measurements as such do play a role in reproducing the observable predictions of standard quantum mechanics. Recall (Section 5) that unlike conventional Bell models of EPR/B experiments, in Bell-like retro-causal models the quantum-mechanical (incomplete) pair-state does not determine the distribution of the complete pair-state. In the causally symmetric Bohmian model, and more generally in Model DO, the distribution of complete pair-states at the emission depends on the distribution of the measurement outcomes. Recall also that in order to reproduce the statistical predictions of standard quantum mechanics, Model DO has to be supplemented with postulates that determine the distribution of complete pair-states or measurement outcomes in any given quantum-mechanical pair-state and experimental set-up. In the causally symmetric Bohmian model, the quantum-mechanical wavefunction of the particle pair fails to determine the distribution of the final wave functions, and accordingly the distribution of the complete pair-state. The model addresses this complication by postulating the following probabilistic relations between initial and final wave functions.

Predictions. Let ψ_i and ψ_f be respectively the initial and final wave functions of a system at some time t. If ψ_f corresponds to one of the *possible outcomes of a subsequent measurement*, the conditional probability distribution of ψ_f given ψ_i, $\rho\left(\psi_f/\psi_i\right)$, is:

$$\text{(Predictions)} \quad \rho\left(\psi_f/\psi_i\right) = |a|^2\,;$$

where a is as defined in *Normalization*.

Granted *Predictions*, quantum-mechanical wave functions determine the distribution of initial and final wave functions and thus the distribution of the

[10] The same is true for the complete pair-state, i.e. the state that is constituted by the initial and final wavefunctions of the particles and their position configuration; for the initial and final wavefunctions of the particles determine the measurement outcomes.

(relevant) complete pair-state. Accordingly, the causally symmetric Bohmian model reproduces the statistical predictions of standard quantum mechanics. And although this postulate appeals to measurements *as such*, it does not entail any dependence of the dynamical laws on measurements. The dynamical laws are always the same, and the behavior of systems is always governed by *Guidance*^{S}. *Predictions* could be thought of as an epistemic postulate which provides information about the distribution of final wave functions of systems in any quantum-mechanical wave function.

Finally, it is noteworthy that while in standard Bohmian mechanics there is only one probabilistic postulate, *Distribution*, in the causally symmetric Bohmian mechanics there are two independent probabilistic postulates, *Distribution*^{S} and *Predictions*. It is also noteworthy that *Distribution*^{S} does not play any role in reproducing the statistical predictions of standard quantum mechanics. This suggests that, in theory, deterministic retro-causal interpretations of quantum mechanics in which the measurement outcomes influence the complete pair-state at the emission could reproduce these predictions by a single probabilistic postulate. Yet, *Distribution*^{S} is important for the ontology of the causally symmetric Bohmian model, and in particular for its solution to the measurement problem. More generally, while logically such deterministic retro-causal interpretations could reproduce the statistical predictions of standard quantum mechanics by a single probabilistic postulate, ontological and theoretical considerations may require some additional, independent probabilistic postulates.

6.8 Are Retro-Causal Interpretations of Quantum Mechanics Explanatory Vacuous?

In Section 6.6, we argued that the indeterministic Bell-like retro-causal models fail to explain the correlation between the distant measurement outcomes in EPR/B experiments in terms of local, common causes; and in Sections 6.5, 6.6 and 6.7.2, we argued that, in contrast to standard quantum mechanics and conventional Bohmian mechanics, indeterministic retro-causal models face the explanatory challenge of motivating multiple, independent probabilistic postulates, and deterministic retro-causal models may face the same challenge. Lewis (2006, 376) poses another challenge for the explanatory power of retro-causal models. He poses the challenge in the context of retro-causal models of the EPR/B experiment that assume the ontological framework of Bohmian mechanics (and accordingly are not subjected to the measurement problem). In such models, "the device settings explain the motions of the particles, which in turn explain the hidden variables of the particles [i.e. their 'complete' state in our terminology]. But the hidden variables, presumably, themselves explain the motions of the particles on measurement; a particle moves upward under the influence of a magnetic field precisely *because* its spin-value is 'up'. The worry here is that the backwards-causal mechanism makes the causal explanation viciously circular; the particle moves up because it moves up."

Lewis suggests that the key point is to notice the distinction between "the axis along which the particle moves, and whether they move up or down along this axis." And he argues that in explanatory retro-causal models, the "backwards-causal mechanism determines the axis alone; the settings of the measuring devices explain why each particle moves along a particular axis, but not why the particle moves up rather than down along its axis. The two particles carry the axis information backwards to the their joint source, and this information enable some mechanism at the particle source to arrange the hidden variables for these two axes so as to satisfy the Bell correlations. But it is the mechanism at the source, whatever it may be, that provides the causal explanation for the actual value of the hidden variables, and hence for the actual motion of the particles, up or down, on measurement. This causal story avoids the circularity of the causal story above. The fact that a given particle moves up rather than down on a measurement is explained, as it should be, by the hidden variables of the particle, which are in turn explained by the process by which the particle is produced in the source. The fact that a given particle moves along this axis rather than some other, though, is explained, as it should be, by the setting of the measuring device, which is in turn explained by the procedure in which the device is set." (Lewis, 2006, 376)

Lewis seems to appeal implicitly to the distinction between (1) retro-causal models in which the apparatus settings influence the complete pair-state at the source, e.g. Model DS and Model IS, and (2) retro-causal models in which the measurement outcomes influence this complete state, e.g. Model DO and Model IO (for a discussion of these models, see Sections 6.2, 6.5, 6.6, and 6.7.2). In the former models, the backward-causal mechanism sends to the source information about the apparatus settings, whereas in the latter models the backward-causal mechanism sends information about the measurement outcomes, which (at least in ideal measurements) embodies information about both the actual apparatus settings and the actual measurement outcomes. And Lewis's argument seems to be that a causal explanation of the measurement outcomes and the correlations between them will be vacuous in the latter models, but not in the former. The question is what licenses this argument.

One notable distinction between the two kinds of models is that in conventional EPR/B experiments, which are the focus of Lewis's discussion, models of the second kind predict the existence of causal loops (see Loop XIII in Fig. 6.10), whereas models of the first kind do not (see Fig. 6.11). In Model DO and Model IO of

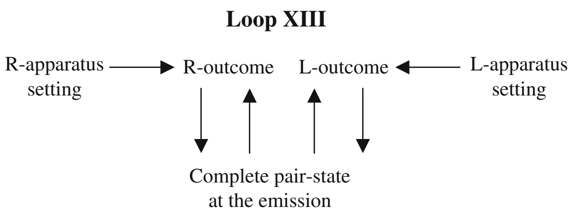


Fig. 6.10 The causal loops that Model DO and Model IO – retro-causal models in which the measurement outcomes influence the complete pair-state – predict in 'conventional' EPR/B experiments. Unlike in previous figures, the influence of the fixed apparatus setting is made explicit

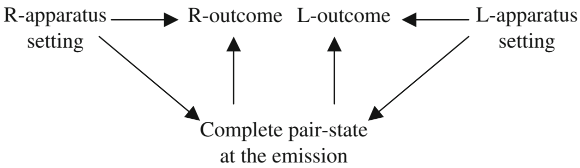


Fig. 6.11 The causal connections that Model DS and Model IS – retro-causal models in which the apparatus settings influence the complete pair-state – predict in conventional EPR/B experiments

conventional EPR/B experiments, the measurement outcomes are causes of the complete pair-state and this state is a cause of the outcomes; whereas in Model DS and Model IS of such experiments, the apparatus settings are causes of the complete pair-state and this is a cause of the outcomes, but the measurement outcomes do not cause the apparatus settings. It may be suggested then that retro-causal models that predict causal loops in which the measurement outcomes are causally explained by their very existence are explanatory vacuous.

It is noteworthy, however, that while Model DS and Model IS do not predict causal loops in conventional EPR/B experiments, they do predict the existence of Loop I in Experiment X (see Fig. 6.3 in Section 6.2).[11] In this loop, the R-measurement is a cause of the complete pair-state at the emission, and this state is a cause of that R-outcome. Thus, if Model DO and Model IO are explanatory vacuous, so will be all other retro-causal models of Experiment X.

Although retro-causal interpretations predict causal loops in which the measurement outcomes are causes of their own existence, it does not follow that they are explanatory vacuous. First, the measurement outcomes are only indirect causes of their own existence: the outcomes influence the complete pair-state at the emission, which in turn influences them. In fact, in Model DS and Model IS, the outcomes are even more indirect causes of their own existence: the R-outcome causes the L-apparatus setting, this setting causes the complete pair-state at the emission, which in turn causes the R-outcome. So the retro-causal interpretations do not involve self-causation, which are open to the charges of explanatory vacuity and in fact disallowed by the main current accounts of causation. Second, the measurement outcomes are only partial causes of their own existence (a feature that is disguised by the fact that in the causal loops in Figs. 6.3–6.7 the loop's causal circumstances are suppressed). That is, the measurement outcomes or apparatus settings are only partial causes of the complete pair-state at the emission: the complete pair-state is also influenced by other factors, such as the initial boundary conditions of the source; and the complete pair-state is only a partial cause of the measurement outcomes: the outcomes are also influenced by the settings of the measurement apparatuses. For

[11] In Loop I, the differences between Model DS and Model IS are suppressed. Model DS predict Loop I with all the causal connections being deterministic, whereas Model IS predict Loop I with the causal connections between the complete pair-state and the R-outcome and between the L-setting and the complete pair-state being indeterministic.

example, in the causally symmetric Bohmian model, the measurement outcomes determine the final wave functions of the particles, which are only partial causes of the complete pair-state: the complete pair-state also depends on the particles' initial wave functions. And the complete pair-state is only a partial cause of the measurement outcomes. This state only determines the particles' dispositions to spin in various directions, and the measurement outcomes also depend on the settings of the measurement apparatuses: different measurements will realize different spin dispositions.

That the apparatus settings are partial causes of the measurement outcomes in all 'conventional' and retro-causal models of EPR/B experiments provides another reason to doubt the idea that the distinction between the two kinds of retro-causal models has any relevance for the question of their explanatory power. In standard quantum mechanics the apparatus settings in ideal EPR/B experiments determine the axes along which the particles spin. Thus, in any model of these experiments that reproduces the empirical predictions of this theory, the apparatus settings will determine the axes along which the particles spin in such measurements. That is, in any such a model, conventional or retro-causal, the apparatus settings are partial causes of the measurement outcomes; and in ideal measurements, they cause the axes along which the particles spin. So the claim that in explanatory retro-causal models the settings of the measuring devices should explain why during measurements each particle moves along a particular axis, could not have any bearing on the distinction between retro-causal models that are explanatory vacuous and those that are not.

It is also noteworthy that in both Model DS and Model DO, the backward-causal mechanism explains why a particle moves 'up' rather than 'down' in a certain direction, and in both Model IS and Model IO the backward-causal mechanism explains why the probability of such motions are of specific values rather than others. In particular, for some directions l, l^* and r in Experiment X, in Model DS (see Section 6.5.1) the settings of the L- and the R-apparatus to measure spins in the directions l and r (l^* and r) determine the complete pair-state to be λ_{D1} (λ_{D2}); in Model IS (see Section 6.5.2) the settings of the L- and the R-apparatus to measure l^* and r determine the probability of the complete pair-state λ_{I1} to be p_1 (see Loop V in Fig. 6.5); and in Model DO that the causally symmetric Bohmian mechanics prescribes, the measurement outcomes (jointly with the quantum-mechanical pair-state) determine both the initial and final wave functions of the particles, and accordingly their complete pair-state.

6.9 Conclusions

We have argued that retro-causal interpretations of quantum mechanics predict the existence of closed causal loops, and considered the challenges that these loops pose for the predictive and explanatory power of these interpretations. Our main focus was on retro-causal interpretations that prescribe Bell-like retro-causal models of EPR/B experiments (Sections 6.1, 6.2 and 6.5). Like in the conventional Bell models, it is supposed that the quantum-mechanical pair-state is incomplete, and that the probabilistic dependence between the distant measurement outcomes, and accordingly the apparent non-local influences between them, are due to this incompleteness. That is, it is postulated that for any quantum-mechanical pair-state, there may be various different complete pair-states. Each of these states assigns probabilities of outcomes for various possible measurements, and the probabilities of the outcomes are independent of each other, and accordingly the joint probability of outcomes factorizes into their single probabilities. The probabilities of standard quantum mechanics are interpreted as statistical averages over the model's probabilities according to the distribution of the complete pair-state. In conventional Bell models, this distribution depends only on the quantum-mechanical pair-state. In the retro-causal models it also depends on the settings of the measurement apparatuses or the measurement outcomes. This modification of the conventional Bell models is quite radical. But, as we have argued, it may not be sufficient.

Some deterministic Bell-like retro-causal models in which the apparatus settings influence the distribution of complete pair-states predict inconsistent causal loops (Section 6.5.1). Other deterministic and indeterministic Bell-like retro-causal models predict the existence of consistent causal loops, and due to these loops the models fail to provide definite statistical predictions of measurement outcomes (Sections 6.5 and 6.6). In particular, this problem arises in Experiment X – an EPR/B-like experiment in which the measurement outcome in the R-wing determines in a perfectly local way the setting of the L-measurement apparatus (Section 6.2). In this experiment, deterministic retro-causal Bell-like models predict the existence of causal loops, each with a different complete pair-state (Sections 6.2 and 6.5.1). The distribution of these states depends on the measurement outcomes or the apparatus settings, and the outcomes and the settings depend on the complete pair-state. The problem is that due to these dependencies, quantum-mechanical pair-states fail to determine the distribution of the various loops, and accordingly the distribution of the complete pair-states. Thus, the conventional way of reproducing the statistics of measurement outcomes in Bell models – namely, as an average-sum over the probabilities of outcomes in various complete pair-states, according to the distribution of these states in the given quantum-mechanical pair-state – is inapplicable. As a remedy, deterministic retro-causal models need to be equipped with a probabilistic postulate that assigns a distribution of complete pair-state or measurement outcomes in any given quantum-mechanical pair-state and experimental set-up. For example, Sutherland's (2008) causally symmetric Bohmian model for EPR/B-like experiments is a deterministic retro-causal model in which the measurement outcomes influence the complete pair-state at the emission (see Section 6.7.2). In this model, the time-evolution of a system is governed by two wave functions – the system's initial wavefunction which evolves forward in time, and its final wavefunction which evolves backward in time. The separable initial and the final wave functions of the particles constitute their relevant complete pair-state. The problem is that the quantum-mechanical wavefunction of the particle pair does not determine the distribution of the initial and final wave functions of the particles.

Yet, the model postulates that in measurements the probabilistic relation between initial and final wave functions is determined according to a Born-like rule (see *Predictions* in Section 6.7.2). Granted this probabilistic postulate, the distribution of the complete pair-state and measurement outcomes can be computed in any given quantum-mechanical wavefunction and experimental set up.

Indeterministic Bell-like retro-causal models of Experiment X predict the existence of indeterministic causal loops, and due to these loops it is possible to predict neither the distribution of complete pair-states nor the distribution of measurement outcomes in complete pair-states (Sections 6.5.2 and 6.6). Thus, these models fail to provide definite statistical predictions for measurement outcomes in this experiment. A possible remedy is to reinforce the indeterministic retro-causal models by a probabilistic postulate that assign a distribution of measurement outcomes in any given quantum-mechanical state and experimental set-up. Unlike in the deterministic retro-causal models, in the indeterministic retro-causal models postulates that only assign a distribution of complete pair-state in any quantum-mechanical pair-state and experimental set-up would not do; for in these models, there is no known way to compute from the probabilities of outcomes that complete pair-states assign the long-run frequencies of these outcomes.

In both the deterministic and indeterministic retro-causal Bell-like models, the additional probabilistic postulate is designed to salvage their predictive power. The question is whether it could be motivated on non-ad hoc grounds. It may be argued that conventional interpretations of quantum mechanics include some analogous postulates, such as the Born rule, which are also not easy to motivate on non-ad hoc grounds. Yet, there are important differences between conventional and retro-causal Bell-like models. In the indeterministic retro-causal Bell-like models, there are two independent probabilistic postulates – the postulate that assigns a distribution of complete pair-states in a given quantum-mechanical pair-state and measurement outcomes or apparatus settings (depending on the model), and a postulate that assigns a distribution of measurement outcomes in a given quantum-mechanical pair-state and experimental set up. So the problem of justification is multiplied.

Furthermore, unlike conventional Bell models and deterministic retro-causal models, in the indeterministic retro-causal models the statistical predictions of measurement outcomes could not be obtained as an average-sum over the probabilities of outcomes in complete pair-states; for the model's probabilities of outcomes in any given complete pair-state fails to determine the long-run frequencies of the outcomes in that state. Thus, although these models do not postulate any non-local influences, they fail to fulfil one of the main objectives of retro-causal interpretations of quantum mechanics – namely, to explain how the quantum-mechanical correlations between distant measurement outcomes could be explained by local common-causes. So, curiously, while the probabilities of outcomes in these models are factorizable, they still fail to live up to the standards of explanation dictated by Reichenbach's principle of the common cause.

In deterministic retro-causal Bell-like models, things are less clear-cut. In Sutherland's (2008) causally symmetric Bohmian model there are two independent probabilistic postulates: the postulate that determines the distribution of the position configuration of the particles according to their initial and final wave functions, and the Born-like postulate that relates the distribution of final wave functions to initial wave functions. But only the latter postulate plays a role in reproducing the statistical predictions of standard quantum mechanics. This seems to suggest that there is no a priori reason to assume that deterministic retro-causal models could not reproduce the predictions of this theory by one probabilistic postulate. Yet, the other probabilistic postulate in Sutherland's model is important for the ontology of the model and the way it resolves the measurement problem. More generally, while logically deterministic retro-causal interpretations may reproduce the statistical predictions of standard quantum mechanics by a single probabilistic postulate, theoretical and ontological considerations may well require some additional, independent probabilistic postulates. So the challenge of motivating multiple, independent probabilistic postulates on non-ad hoc grounds also exist for the deterministic Bell-like retro-causal interpretations of quantum mechanics, though it seems less acute than in the indeterministic retro-causal models.

We also considered two other potential challenges for the explanatory power of the retro-causal interpretations. The first is that retro-causal may suffer from a measurement problem (Section 6.7). In the Bell-like retro-causal models, the measurement events influence the complete pair-state at the source, and the question is whether this backward-causal influence depends on measurements as *such*. If measurements as such were the trigger of the backward-causal mechanism, then whether a measurement occurs would have a dynamical effect on the behaviour of the system. But, there seems to be no fundamental distinction between measurement processes and non-measurement processes. The second challenge is that in retro-causal models of Experiment X, the measurement outcomes are causes of their own existence: the measurement outcomes are causes of the complete pair-state at the emission and this state is the cause of the measurement outcomes. The worry is then that the causal explanation that these models provide is vacuous, e.g. that this explanation is of the following kind: the L-particle moves 'up' in the direction z because it moves 'up' in this direction (Section 6.8).

Whether a retro-causal interpretation is subjected to the measurement problem does not depend on the postulation of backward-causal mechanism per se, but rather on whether its ontology provides a solution to the 'conventional' measurement problem. We considered two retro-causal interpretations: the transactional interpretation of standard quantum mechanics (Section 6.7.1), and the causally symmetric Bohmian model (Section 6.7.2). The transactional interpretation is a retro-causal interpretation of quantum mechanics. As such, it is a collapse interpretation. It postulates that in measurements, states of systems seize to follow the dynamics dictated by the Schrödinger equation: measurements cause wavefunction collapse. Yet, similarly to the orthodox interpretation, it does not account for this collapse as a real physical process. Accordingly, it has a measurement problem. The causally symmetric Bohmian model is a no-collapse, retro-causal interpretation of quantum mechanics that embodies much of the ontology of conventional Bohmian mechanics. Like conventional Bohmian mechanics, it postulates that the time-evolution of a state of a system always follows the Schrödinger equation, and the dynamical laws that govern this evolution in measurements are not different from those in non-measurement contexts. Also, like in conventional Bohmian mechanics, particles always have definite positions and accordingly the outcomes of measurements are always definite. Thus, the causally symmetry Bohmian model is not subjected to the measurement problem.

Finally, we argued that the fact that retro-causal interpretations of quantum mechanics predict causal loops in which the measurement outcomes are causes of their own existence does not render them explanatory vacuous. The measurement outcomes are only partial, 'indirect' causes of their own existence. Thus, although the causal explanations that these interpretations propose are curious, they are not cases of self-causation, which are open to the charge of being explanatory vacuous and are in fact excluded by all the main current accounts of causation.

Acknowledgements The work on this paper was prompted by an invitation to contribute to this volume. I am very grateful to the editor, Mauricio Suarez. Parts of this paper were presented at the Time-Symmetric Interpretations of Quantum Mechanics Workshop and the Summer Foundations Conference 2006, Centre for Time, Department of Philosophy, University of Sydney; CREA, Polytechnique, Paris; the Department of Philosophy, Universidad de Barcelona; the Department of Philosophy, Universidad Complutense de Madrid; and the Sigma Club, Centre for the Philosophy of Natural and Social Sciences, London School of Economics. These conferences and colloquia were instrumental in the development of the paper, and I thank the organizers, audiences, and in particular Rod Sutherland, Mauricio Suarez, Huw Price, David Miller, Carl Hoefer, Roman Frigg and Guido Bacciagaluppi. For support, I am very grateful to the Department of Philosophy, University of Sydney, and the Institute for History and Philosophy of Science and Technology, University of Toronto.

References

Aharonov, Y. and Gruss, E. (2005), Two-time interpretation of quantum mechanics. http://arxiv.org/pdf/quant-ph/0507269

Aharonov, Y. and Tollaksen, J. (2007), New insights on time-symmetry in quantum mechanics. arXiv:0706.1232VI

Aharonov, Y. and Vaidman, L. (2007), The two-state vector formalism: An updated review. *Lectures Notes in Physics* 734, 399–447. arXiv:quant-ph/0105101v2

Argaman, N. (2007), On Bell's theorem and causality. Unpublished manuscript.

Arntzenius, F. and Maudlin, T. (2009). Time travel and modern physics. In E. N. Zalta (ed.), *The Stanford Encyclopaedia of Philosophy* (*Summer 2005 Edition*), URL = <http://plato.stanford.edu/archives/spr2010/entries/time-travel-phys/>

Bell, JS. (1964), On the Einstein-Podolsky-Rosen paradox. *Physics*, *1*, 195–200. Reprinted in Bell (1987), pp. 14–21.

Bell, JS. (1966), On the problem of hidden variables in quantum mechanics. *Reviews of Modern Physics* 38, 447–452. Reprinted in Bell (1987), pp. 1–13.

Bell, JS. (1971), Introduction to the hidden-variable question. In Espagnat B. (ed.), *Foundations of Quantum Mechanics*. Proceedings of the International School of Physics 'Enrico Fermi', Academic Press, New York-London, pp. 171–181. Reprinted in Bell (1987), pp. 29–39.

Bell, JS. (1975a). The theory of local beables. TH-2053-Cern. Reprinted in *Epistemological Letters* March 1976 and Bell (1987), pp. 52–62.

Bell, JS. (1975b), Locality in quantum mechanics: reply to critics. *Epistemological Letters*, Nov. 1975, 2–6. Reprinted in Bell (1987), pp. 63–66.

Bell, JS. (1987). *Speakable and Unspeakable in Quantum Mechanics*. Cambridge: Cambridge University Press.

Berkovitz, J. (1995), What econometrics cannot teach quantum mechanics. *Studies in History and Philosophy of Modern Physics* 26, 163–200.

Berkovitz, J. (2001), On chance in causal loops. *Mind*, *110*, 1–23.

Berkovitz, J. (2002), On causal loops in the quantum realm. In T. Placek, and Butterfield, J. (eds.), *Non-locality and Modality*. Proceedings of the NATO Advanced Research Workshop on Modality*,* Probability and Bell's Theorems, Kluwer, pp. 235–257.

Berkovitz, J. (2007). Action at a distance in quantum mechanics. In Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (*Spring 2007 Edition)*, URL = <http://plato.stanford.edu/archives/spr2007/entries/qm-action-distance/>.

Berkovitz, J. (2008), On predictions in retro-causal interpretations of quantum mechanics. *Studies in History and Philosophy of Modern Physics* 39, 709–735.

Berkovitz, J. (2009a), The world according to de Finetti. Unpublished manuscript.

Berkovitz (2009b), The propensity interpretation: Reply to critics. Unpublished manuscript.

Bohm, D. (1951), *Quantum theory*. Englewood Cliffs, NJ: Prentice-Hall.

Butterfield, J. (1992), Bell's Theorem: What It Takes. *British Journal for the Philosophy of Science* 48, 41–83.

Cartwright, N. (1989), *Nature's Capacities and their Measurements*. Oxford: Clarendon Press.

Chang, H. and Cartwright, N. (1993), Causality and realism in the EPR experiment. *Erkenntnis* 38, 169–190.

Costa de Beauregard, O. (1953), Une réponse à l'argument dirigé par Einstein, Podolsky et Rosen contre l'interprétation Bohrienne des phénomènes quantiques, *Comptes Rendus de l'Académie des Sciences* 236 (1953), pp. 1632–1634.

Costa de Beauregard, O. (1977), Time symmetry and the Einstein paradox. *Il Nuovo Cimento* 42B, 41–64.

Costa de Beauregard, O. (1979), Time symmetry and the Einstein paradox – II. *Il Nuovo Cimento* 51B, 267–279.

Costa de Beauregard, O. (1985), On some frequent but controversial statements concerning the Einstein-Podolsky-Rosen correlations. *Foundations of Physics* 15, 871–887.

Cramer, J. (1980), Generalised absorber theory and the Einstein-Podolsky-Rosen paradox. *Physical Review D* 22, 362–376.

Cramer, J. (1983), The Arrow of Electromagnetic Time and Generalized Absorber Theory. *Foundations of Physics* 13, 887–902.

Cramer, J. (1986), The transactional interpretation of quantum mechanics. *Reviews of Modern Physics* 58, 647–687.

Cramer, J. (1988), An overview of the transactional interpretation of quantum mechanics. *International Journal of Theoretical Physics* 27, 227–236.

Davidon, W. C. (1976), Quantum physics of single systems. Il Nuovo Cimento, 36B, 34–39.

Dowe, P. (2003), The coincidences of time travel. *Philosophy of Science* 70, 574–589.

Dummett, M. (1964), Bringing about the past. *Philosophical Review* 73, 338–359.

Einstein, A., Podolsky, R., and Rosen, N. (1935), Can quantum-mechanical description of physical reality be considered complete? *Physics Review* 47, 777–780.

Fine, A. (1981), Correlations and physical locality. In P. Asquith and Giere, R. (eds.), *PSA 1980*, vol. 2. East Lansing, Michigan: Philosophy of Science Association, pp. 535–562.

Fine, A. (1982a), Hidden variables, joint probability, and the Bell inequalities. *Physics Review Letters* 48 (1982), pp. 291–295.

Fine, A. (1986), *The Shaky Game*. Chicago: The University of Chicago Press.

Fine, A. (1989), Do correlations need to be explained? In J. Cushing and E. McMullin (eds.), *Philosophical Consequences of Quantum Theory: Reflections on Bell's Theorem*, South Bend: The University of Notre Dame Press, pp. 175–194.

Flew, A. (1954), Can an effect precede its cause? *Proceedings of the Aristotelian Society Supplement* 28, 45–52.

Gruss, E. (2000), A Suggestion for a teleological interpretation of quantum mechanics. http://arxiv.org/abs/quant-ph/0006070

Hajek, A. (2003), What conditional probability could not be, *Synthese* 137, 273–323.

Horwich, P. (1987), *Asymmetries in Time*. Cambridge MA: MIT Press.

Lewis, D. (1986), *Philosophical papers Vol. 2*. Oxford: Oxford University Press.

Lewis, P. J. (2006), Conspiracy theories of quantum mechanics. *British Journal for the Philosophy of Science* 57/2, 359–381.

Marchildon, L. (2006), Causal loops and collapse in the transactional Interpretation of quantum mechanics. http://arxiv.org/pdf/quant-ph/0603018v2

Maudlin, T. (1994), *Quantum Non-locality and Relativity*. Oxford: Blackwell.

Miller, D. (1996), Realism and time symmetry in quantum mechanics. *Physics Letters A* 222, 31–36.

Miller, D. (2008), Quantum mechanics as a consistency condition on initial and final boundary conditions. *Studies in History and Philosophy of Modern Physics* 39, 767–781.

Mellor, D. H. (1981), *Real Time*. Cambridge: Cambridge University Press.

Mellor, D. H. (1998), *Real Time II*. London: Routledge.

Price, H. (1984), The philosophy and physics of affecting the past. *Synthese* 61/3, 299–323.

Price, H. (1994), A neglected route to realism about quantum mechanics. *Mind* 103, 303–336.

Price, H. (1996), *Time's Arrow and Archimedes' Point*. Oxford: Oxford University Press.

Price, H (2008), Toy models for retrocausality. *Studies in History and Philosophy of Modern Physics* 39, 752–761.

Reichenbach, H. (1956), *The Direction of Time*. University of California Press.

Renninger, M. (1953), Zum wellen-korpuskel-dualismns. *Zeit Schrift Für Physik*, 251–261.

Reznik, B. and Aharonov, Y. (1995), On a time symmetric formulation of quantum mechanics. http://arxiv.org/pdf/quant-ph/9501011

Shimony, A. (2006), Bell's theorem. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (*Fall 2006 Edition*), URL = <http://plato.stanford.edu/archives/fall2006/entries/bell-theorem/>.

Smith, N. (1997), Bananas enough for time travel? *British Journal for the Philosophy of Science* 48, 363–389.

Sutherland, R. I. (1983), Bell's theorem and backwards-in-time causality. *International Journal of Theoretical Physics* 22, 377–384.

Sutherland, R. I. (1998), Density formalism for quantum theory. *Foundations of Physics* 28/7, 1157–1190.

Sutherland, R. I. (2008), Causally symmetric Bohm model. *Studies in the History and Philosophy of Modern Physics* 39, 782–805.

Tooley, M. (1997), *Time, tense, and causation*. Oxford: Clarendon Press.

Wheeler, J. A. and Feynman, R. P. (1945), Interaction with the absorber as the mechanism of radiation. *Reviews of Modern Physics* 17, 157–181.

Wheeler, J. A. and Feynman, R. P. (1949), Classical electromagnetics in terms of direct infer particle action. *Reviews of Modern Physics* 21, 425–434.

Chapter 7
Causal Completeness in General Probability Theories

Balazs Gyenis and Miklós Rédei

7.1 Informal Formulation of the Problem of Causal Completeness

The aims of this paper are (i) to define causal completeness of general probability theories, (ii) to raise the problem of when a probability theory is causally complete and (iii) to recall the known results together with the open problems about causal completeness.

Causal completeness will be defined in terms of a general probability space $(\mathcal{L},\phi)$, where $\mathcal{L}$ is an orthocomplemented (not necessarily distributive) σ-lattice and where ϕ is a σ-additive probability measure on $\mathcal{L}$. Roughly, causal completeness of $(\mathcal{L},\phi)$ means that for every correlation between (compatible) variables $\underline{A},\underline{B}$ in $\mathcal{L}$ that stand in a causal *independence* relation ($\mathcal{R}(\underline{A},\underline{B})$, in notation) the lattice $\mathcal{L}$ contains elements that can be regarded as (Reichenbachian) common causes of the correlation. The problem is then under what condition on $(\mathcal{L},\phi)$ and $\mathcal{R}$ is the probability space $(\mathcal{L},\phi)$ causally complete.

Little is known about this problem in general. Most of the known results concern the case when $(\mathcal{L},\phi)$ is a classical (Kolmogorovian) probability space; i.e. when $\mathcal{L}$ is an orthocomplemented distributive lattice (Boolean algebra). In this Kolmogorovian case, even probability theories with an $\mathcal{L}$ of finite cardinality can be causally complete under weak assumptions on $\mathcal{R}$ (see the Propositions 3, 5 and 6); however, causal completeness is not typical. Causal completeness of quantum probability spaces, i.e. where $\mathcal{L}$ is a von Neumann lattice, and ϕ is a (normal) state, is an almost completely open problem, the only result known spells out causal completeness of $(\mathcal{L},\phi)$ that are atomless (see Proposition 8), a special case of which is the lattice $\mathcal{L}$ being the projection lattice of a type III von Neumann algebra.

The structure of the paper is the following. In Section 7.2 we recall some notions of general probability theory together with some related definitions that will be used in the paper. In Section 7.3 the notion of a generalized Reichenbachian

B. Gyenis (✉)
Department of History and Philosophy of Science, University of Pittsburgh, Pittsburgh, PA, USA
e-mail: gyepi@pitt.edu

M. Suárez (ed.), *Probabilities, Causes and Propensities in Physics*, Synthese Library 347, DOI 10.1007/978-1-4020-9904-5_7, © Springer Science+Business Media B.V. 2011

common cause is given (Definition 3.2); this is followed by the definition of causal completeness in Section 7.4. The known results on causal completeness are summarized in Section 7.5. The paper is closed with Section 7.6 in which we comment on the philosophical interpretation of the results and indicate some further open problems.

7.2 General Probability Spaces – Definitions and Notations

Let $(\mathcal{L}, \leq, \vee, \wedge, \perp)$ be an orthocomplemented, bounded lattice (with zero 0 and unit 1 elements) with respect to the lattice operations $\vee$, $\wedge$ related to the partial ordering $\leq$ in the standard way ($A \vee B$ being the least upper bound of A and B and $A \wedge B$ being the greatest lower bound of A and B with respect to $\leq$). Recall that $A, B \in \mathcal{L}$ are called *orthogonal* if $A \leq B^{\perp}$. The lattice $\mathcal{L}$ is called *distributive* if

$$A \vee (B \wedge C) = (A \vee B) \wedge (A \vee C) \ \textit{for any} \ A, B, C \in \mathcal{L} \tag{7.1}$$

modular if

$$\textit{if} \ A \leq B \ \textit{then} \ A \vee (B \wedge C) = (A \vee B) \wedge (A \vee C) \tag{7.2}$$

orthomodular if

$$\textit{if} \ A \leq B \ \textit{and} \ A^{\perp} \leq C \ \textit{then} \ A \vee (B \wedge C) = (A \vee B) \wedge (A \vee C) \tag{7.3}$$

Elements $C_k \in \mathcal{L} (k \in K)$ are called *compatible* if they are from a distributive sublattice of $\mathcal{L}$. They are said to form a partition ($\{C_k\}_K$, in notation) in $\mathcal{L}$ if C_k and $C_{k'}$ are orthogonal whenever $k \neq k'$ and $\vee_k C_k = 1$. In what follows I, J and K will always denote finite index sets of cardinality I, J and K respectively, i.e.

$$\begin{aligned} I &= \{1, 2, \ldots I\} \\ J &= \{1, 2, \ldots J\} \\ K &= \{1, 2, \ldots K\} \end{aligned}$$

Index variables i, j and k will consistently be used to run through I, J and K, respectively. $\underline{A}_I$, $\underline{B}_J$ and $\underline{C}_K$ will denote *ordered* partitions of cardinality I, J and K respectively, i.e.

$$\begin{aligned} \underline{A}_I &= (A_1, A_k \ldots A_i \ldots A_I) \\ \underline{B}_J &= (B_1, B_k \ldots B_j \ldots A_J) \\ \underline{C}_K &= (C_1, C_k \ldots C_j \ldots C_K) \end{aligned}$$

An ordered partition is called a *variable*. Sometimes we are going to further simplify the notation by dropping I, J and K from $\underline{A}_I$, $\underline{B}_J$ and $\underline{C}_K$. Note that we can think of the elements $A, A^{\perp} \in \mathcal{L}$ as members of the two-element partition $\underline{A}_2$ with $A = A_1$

and $A^{\perp} = A_2$. In what follows, variables $\underline{A}_I$, $\underline{B}_J$ and $\underline{C}_K$ will always be assumed to be compatible.

Let $\mathcal{L}$ be an orthocomplemented lattice. The map

$$\phi\colon \mathcal{L} \to [0, 1]$$

is called *generalized probability measure* or *state* if $\phi(0) = 0$ and $\phi(1) = 1$ and if it is σ-additive on orthogonal elements, i.e. if

$$\phi(\vee_i A_i) = \sum_i \phi(A_i) \qquad A_i \leq A_{i'}^{\perp} \qquad i \neq i'$$

$(\mathcal{L}, \phi)$ is called a (general) probability measure space. If $\mathcal{L}$ is distributive, then it is a Boolean algebra and $(\mathcal{L}, \phi)$ is a classical (Kolmogorovian) probability measure space. In this case we write $(X, \mathcal{S}, p)$ where X is a set, $\mathcal{S}$ is a set of subsets of X and p a classical probability measure. If $\mathcal{S}$ is finite and generated by n atoms, and p assigns non-zero probability to these atoms, then we write $(\mathcal{S}_n, p)$. If $\mathcal{N}$ is a von Neumann algebra, then its projection lattice $\mathcal{P}(\mathcal{N})$ is an orthomodular lattice and $\big(\mathcal{P}(\mathcal{N}), \phi\big)$ with a (normal) state ϕ on $\mathcal{N}$ is a non-commutative (quantum) probability space. If $\mathcal{N}$ is the set of all bounded operators $\mathcal{B}(\mathcal{H})$ on a Hilbert space $\mathcal{H}$, then $\Big(\mathcal{P}\big(\mathcal{B}(\mathcal{H})\big), \phi\Big)$ is the quantum probability space of the standard Hilbert space quantum mechanics. If $\mathcal{H}$ is finite dimensional then $\mathcal{P}\big(\mathcal{B}(\mathcal{H})\big)$ is a modular lattice (Rédei, 1998). In what follows, $(\mathcal{L}, \phi)$ is assumed to be a general probability space, $\mathcal{L}$ is not assumed to be either distributive, modular, or orthomodular.

We say that $\underline{A}_I$ and $\underline{A}'_I$ (in notation: $\underline{A}_I \cong \underline{A}'_I$) are the *same up to measure zero* when $\phi(A_i \wedge {A'}_I^{\perp}) = \phi(A_i^{\perp} \wedge A'_I) = 0$ for all $i = 1, .., I$. When $\underline{A}_I$ and $\underline{A}'_I$ are not the same up to measure zero, we write $\underline{A}_I \ncong \underline{A}'_I$.

If $A, C \in \mathcal{L}$ are compatible then we use the notation

$$\phi_C(A) = \phi(A|C) = \frac{\phi(A \wedge C)}{\phi(C)} \tag{7.4}$$

For later purposes we need the following notion of logical independence of subsets $\mathcal{L}_1, \mathcal{L}_2$ of $\mathcal{L}$:

Definition 2.1 Two subsets $\mathcal{L}_1, \mathcal{L}_2$ *of* $\mathcal{L}$ are called *logically independent if*

$$A \wedge B \neq 0 \qquad \textit{for} \quad 0 \neq A \in \mathcal{L}_1 \qquad 0 \neq B \in \mathcal{L}_2 \tag{7.5}$$

Logical independence is hereditary: if $\mathcal{L}'_1$ and $\mathcal{L}'_2$ are subsets of $\mathcal{L}_1$ and $\mathcal{L}_2$, respectively, then logical independence of $\mathcal{L}_1, \mathcal{L}_2$ entails logical independence of $\mathcal{L}'_1, \mathcal{L}'_2$. A particularly important case is when $\mathcal{L}_1, \mathcal{L}_2$ are sublattices of $\mathcal{L}$. The pair $(\mathcal{L}'_1, \mathcal{L}'_2)$ of sublattices is called a *maximal* logically independent pair, if logical independence of sublattices $\mathcal{L}_1$ and $\mathcal{L}_2$ containing respectively $\mathcal{L}'_1$ and $\mathcal{L}'_2$ implies $\mathcal{L}'_1 = \mathcal{L}_1$ and $\mathcal{L}'_2 = \mathcal{L}_2$. (For a detailed analysis of the notion of logical independence see (Rédei

1995a, b) and Chapter 7 in (Rédei, 1998).) For later purposes we also need the following notions:

Definition 2.2 $\{A_i\}_I$ and $\{B_j\}_J$ are *probabilistically independent* if for all $i \in I, j \in J$:

$$\phi(A_i \wedge B_j) = \phi(A_i)\phi(B_j). \tag{7.6}$$

$\{A_i\}_I$ and $\{B_j\}_J$ of are *related in a genuinely probabilistic way if for all* $i \in I, j \in J$:

$$\phi(A_i \wedge B_j) > 0. \tag{7.7}$$

A map h from $\mathcal{L}$ into $\mathcal{L}'$ is called a *lattice homomorphism* if it preserves all lattice operations, including orthocomplementation. A homomorphism is called *embedding* if $X \neq Y$ implies $h(X) \neq h(Y)$. The probability space $(\mathcal{L}', \phi')$ is called an *extension* of $(\mathcal{L}, \phi)$ if there exists an embedding h of $\mathcal{L}$ into $\mathcal{L}'$ such that $\phi'(h(X)) = \phi(X)$ for every X in $\mathcal{L}$.

Definition 2.3 $(\mathcal{L}, \phi)$ is called *atomless* if for any $A \in \mathcal{L}$ with $\phi(A) > 0$ and for any $0 < r < \phi(A)$ there is $0 \neq B \in \mathcal{L}$, $B \leq A$ such that $\phi(B) = r$.

7.3 A General Notion of Reichenbachian Common Cause

Before turning to a discussion of causal completeness, we need to clarify the notion of a 'common cause of a correlation between two variables.' To do this, we need first a notion of correlation between variables. There are several ways of characterizing the correlation between two variables $\underline{A}$ and $\underline{B}$ in state ϕ. Most of the literature on the Reichenbachian common cause focuses on definitions of correlations that depend only on the prior probabilities $\phi(A_i), \phi(B_j)$ and $\phi(A_i \wedge B_j)$. While this can be physically well motivated, and the notion of correlation with respect to which causal completeness results will be presented in Section 7.5 are of this sort, we wish to leave the notion of correlation unspecified in the Definition 3.1 of common cause and treat it as a variable of the concept of common cause.

Let $(\mathcal{L}, \phi)$ be a general probability space and let $\mathrm{Corr}_\phi(\underline{A}, \underline{B})$ be some measure of correlation between compatible variables $\underline{A}$ and $\underline{B}$ in state ϕ. If $C \in \mathcal{L}$ is compatible with $\underline{A}$ and $\underline{B}$ and $\phi(C) \neq 0$, then the correlation $\mathrm{Corr}_{\phi_C}(\underline{A}, \underline{B})$ in the state ϕ_C obtained from ϕ by conditioning ϕ with respect to C (see (7.4)) will be denoted by $\mathrm{Corr}_\phi(\underline{A}, \underline{B}|C)$. The definition below is a generalized version of Reichenbach's original definition of common cause of a correlation:

Definition 3.1 $\underline{C}_K$ is called a *generalized Reichenbachian common cause* of the correlation $\mathrm{Corr}_\phi(\underline{A}_I, \underline{B}_J) > 0$ between $\underline{A}_I$ and $\underline{B}_J$ if $\phi(C_k)\ /0$ for $k \in K$ and the following conditions hold:

$$\mathrm{Corr}_\phi(\underline{A}_I, \underline{B}_J|C_k) = 0 \text{ for all } k \in K \tag{7.8}$$

$$\mathrm{Corr}_\phi(\underline{A}_I, \underline{C}_K) > 0 \tag{7.9}$$

$$\mathrm{Corr}_\phi(\underline{B}_J, \underline{C}_K) > 0 \tag{7.10}$$

The cardinality of the index set K is called the *size* of the GRCC.

Definition 3.1 is more general than Reichenbach's in the sense that (i) it is formulated in terms of multi-valued variables and multi-valued 'common causes' rather than two-valued ones; (ii) it leaves the specific measure $\mathrm{Corr}_\phi(\underline{A}, \underline{B})$ of correlation between (multi-valued) variables $\underline{A}, \underline{B}$ undetermined. To see how Reichenbach's original definition of common cause can be obtained as special case, consider the following definition of correlation:

$$\mathrm{Corr}_\phi^{\tau_a} \doteq p_c - p_d, \tag{7.11}$$

where

$$p_c = \sum_{i,j} \phi(A_i \wedge B_j) \sum_{i'>i, j'>j} \phi(A_{i'} \wedge B_{j'}) \tag{7.12}$$

$$p_d = \sum_{i,j} \phi(A_i \wedge B_j) \sum_{i'>i, j'<j} \phi(A_{i'} \wedge B_{j'})\,. \tag{7.13}$$

$\mathrm{Corr}_\phi^{\tau_a}$ is motivated by the so-called τ_a association measure, frequently used in the statistical literature. In case of two-valued variables

$$\underline{A}_2 = (A_1, A_2) = (A, A^\perp)$$
$$\underline{B}_2 = (B_1, B_2) = (B, B^\perp)$$

the correlation $\mathrm{Corr}_\phi^{\tau_a}(\underline{A}_2, \underline{B}_2)$ reduces to the (standard) notion of correlation $\mathrm{Corr}_\phi^R(A, B)$ of events used by Reichenbach:

$$\begin{array}{rcl} Corr_\phi^{\tau_a} & = & \phi(A_1 \wedge B_1)\phi(A_2 \wedge B_2) - \phi(A_1 \wedge B_2)\phi(A_2 \wedge B_1) = \\ & = & \phi(A \wedge B)\phi(A^\perp \wedge B^\perp) - \phi(A \wedge B^\perp)\phi(A^\perp \wedge B) = \\ & = & \phi(A \wedge B) - (\phi(A \wedge B) + \phi(A \wedge B^\perp))(\phi(A \wedge B) + \phi(A^\perp \wedge B)) = \\ & = & \phi(A \wedge B) - \phi(A)\phi(B) = \\ & = & \mathrm{Corr}_\phi^R(A, B) \end{array}$$

using

$$\phi(A^\perp \wedge B^\perp) = 1 - \phi(A \wedge B) - \phi(A^\perp \wedge B) - \phi(A \wedge B^\perp).$$

Consequently, taking a (generalized) common cause of size 2:

$$\underline{C}_2 = (C_1, C_2) = (C, C^\perp)$$

(7.8), (7.9), (7.10) reduce to

$$\phi(A \wedge B|C) = \phi(A|C)\phi(B|C) \tag{7.14}$$

$$\phi(A \wedge B|C^{\perp}) = \phi(A|C^{\perp})\phi(B|C^{\perp}) \tag{7.15}$$

$$\phi(A|C) > \phi(A|C^{\perp}) \tag{7.16}$$

$$\phi(B|C) > \phi(B|C^{\perp}) \tag{7.17}$$

This is Reichenbach's original notion of common cause: when (7.14), (7.15), (7.16), (7.17) holds, Reichenbach calls C the *common cause* of the correlation $\mathrm{Corr}^R_\phi(A, B) > 0$. Since it is an important special case, we are going to refer to generalized Reichenbachian common causes of size 2 as *event-type common causes*.

Reichenbach's original definition and its generalization in form of Definition 3.1 specifies common causes of *positive* correlations – Reichenbach's equations (7.14), (7.15), (7.16), (7.17) already entail $\mathrm{Corr}^R_\phi(A, B) > 0$. Common causes of negative correlations can also be straightforwardly characterized and results pertaining to common causes of positive correlations can be translated to the case of negative correlations (see (Gyenis and Rédei, 2004) for details). In some cases one might also be interested in defining common causes of zero correlations. As an example, one might be tempted to say that A and B are 'perfectly correlated' if $A \cong B$; however, a quick calculation shows that in this case $\mathrm{Corr}^R_\phi(A, B) = 0$ and thus when we seek a common cause of such a 'perfect correlation' Reichenbach's original definition of common cause is of no help. In this case it seems to be natural to identify an event $C \cong A$ or $C \cong B$ as a common cause. This strategy works for positive (and negative) correlations as well: according to Reichenbach's conditions $C \cong A$ or $C \cong B$ always qualifies as a common cause of the correlation $\mathrm{Corr}^R_\phi(A, B) > 0$. However, we are interested to find common causes of correlations where $A \ncong B$, and thus to avoid trivialization, we require C to be a *proper* common cause, that is, $C \ncong A$ and $C \ncong B$. In general, $\underline{C}$ is called a *proper* common cause of the correlation $\mathrm{Corr}(\underline{A}, \underline{B}) > 0$ when $\underline{C} \ncong \underline{A}$ and $\underline{C} \ncong \underline{B}$. In what follows we restrict our attention to proper common causes.

Definition 3.1 also covers the recent generalization of Reichenbach's common cause to Reichenbachian common cause *systems* defined and investigated in (Hofer-Szabó and Rédei 2004, 2006):

Definition 3.2 Let *A, B* and members of the (non-ordered) partition $\{C_k\}_K$ be compatible elements of $\mathcal{L}$. $\{C_k\}_K$ is said to be a *Reichenbachian common cause system* for the correlation $Corr^R_\phi(A, B) > 0$ if $\phi(C_k) \neq 0$ for all $k \in K$ and the following conditions hold:

$$\phi(A \wedge B|C_k) = \phi(A|C_k)\phi(B|C_k) \quad \textit{for all } k \in K \tag{7.18}$$

$$\left(\phi(A|C_k) - \phi(A|C_{k'})\right)\left(\phi(B|C_k) - \phi(B|C_{k'})\right) > 0 \quad \textit{whenever } k \neq k'. \tag{7.19}$$

The cardinality of the index set K is called the *size* of the RCCS.

The notion of a Reichenbachian common cause system (Definition 3.2) can be obtained as a special case of Definition 3.1 by taking the following Corr^S_ϕ as the measure of correlation:

$$\mathrm{Corr}^S_\phi(\underline{A},\underline{C}) \doteq \min_{k\in K} \min_{k'\in K, k'>k} \{\phi(A_1 \wedge C_k)\phi(C_{k'}) - \phi(A_1 \wedge C_{k'})\phi(C_k)\}, \tag{7.20}$$

In the case of two valued variables, Corr^S_ϕ also reduces to Corr^R_ϕ. It is easy to see that if $\underline{C}_K$ is a GRCC for the correlation between variables $\underline{A}_2$ and $\underline{B}_2$, then $\{C_k\}_K$ is a RCCS for the correlation $\mathrm{Corr}^R_\phi(A,B) > 0$. Conversely, if $\{C_k\}_K$ is a RCCS of finite size for the correlation $\mathrm{Corr}^R_\phi(A,B) > 0$, and we order this unordered partition $\{C_k\}_K$ in such a way that the series $\phi(A|C_1)$, $\phi(A|C_2)$, ... monotonously decreases, then the resulting $\underline{C}_K$ is a GRCC for the correlation between variables $\underline{A}_2$ and $\underline{B}_2$.

Note that it is not obvious whether a generalized Reichenbachian common cause or a Reichenbachian common cause system exist. We mention two results that spell out the existence of Reichenbachian common causes that are more general than Reichenbach's original notion; both results concern classical probability spaces:

Proposition 1 *Let $(X,\mathcal{S},p)$ be a classical probability space and $\underline{A}_2,\underline{B}_2$ two two-valued variables in $\mathcal{S}$ that are correlated in a genuinely probabilistic way according to $Corr^R_\phi$. Then for any $\infty > K \geq 2$ there exist an extension $(X',\mathcal{S}',p')$ of $(X,\mathcal{S},p)$ such that there exist in $(X',\mathcal{S}',p')$ a Reichenbachian common cause system of size* K *of the correlation $Corr^R_p(A,B) > 0$. If $\mathcal{S}$ is finite, then $\mathcal{S}'$ is also finite.*

(See (Hofer-Szabó and Rédei, 2004) and (Hofer-Szabó and Rédei, 2006) for more details.) It is not known whether Proposition 1 remains true if K is replaced by 'countably infinite' and it is not known if it is valid for probability spaces with a non-distributive $\mathcal{L}$. We conjecture a positive answer to both of these problems.

To formulate the second result, we need this feature of a correlation function: Corr_ϕ is called *independence signaler* if $\mathrm{Corr}_\phi(\underline{A},\underline{B}) = 0$ whenever $\underline{A}$ and $\underline{B}$ are probabilistically independent. Practically all correlation functions used in the statistical literature – $\mathrm{Corr}^{\tau_a}_\phi$ is an example – satisfy this criterion. We have:

Proposition 2 (Gyenis, 2005). *Let $(X,\mathcal{S},p)$ be a classical probability space, $Corr_\phi$ an independence signaler correlation function, and $\underline{A},\underline{B}$ two correlated variables of finite size, related in a genuinely probabilistic way. Then for any $\underline{K}$ satisfying* $\min\{I,J\} \leq K < \infty$ *there exists an extension $(X',\mathcal{S}',p')$ of $(X,\mathcal{S},p)$ such that there exist in $(X',\mathcal{S}',p')$ an (unordered) partition $\{C_k\}_K$ of size $\underline{K}$ satisfying the screening off conditions (7.8).*

For the partition $\{C_k\}_K$ mentioned in Proposition 2 to become a generalized Reichenbachian common cause for the correlation one has to order it in such a way that conditions (7.9), (7.10) are satisfied. There are many ($K!$) ways of ordering this partition. This suggests that the desired ordering can be achieved for many types of correlation functions; however, concise necessary or sufficient conditions on the correlation function entailing such a result are not known. To show case by case that certain correlation functions are having this property would be very tedious and is omitted here. The same remarks apply to Proposition 7 in Section 7.5.

7.4 Notions of Causal Completeness of General Probability Theories

Reichenbach's Common Cause Principle (RCCP) is the claim that if two variables $\underline{A}$ and $\underline{B}$ are correlated, then the correlation is either due to some causal connection between the variables, or, if they are *causally independent* (denoted by $\mathcal{R}(\underline{A},\underline{B})$), then there exists a common cause of the correlation. The origins of this Principle go back to Reichenbach's 1957 book (Reichenbach, 1956) and its status has been extensively discussed in the literature, especially by Arntzenius (Arntzenius, 1993), Butterfield (Butterfield, 1989); Cartwright (Cartwright, 1987); Placek (Placek, 2000a, b); Salmon (Salmon, 1978, 1980, 1984); Sober (Sober, 1984, 1988, 2001); Spohn (Spohn, 1991); Suppes (Suppes, 1970); Uffink (Uffink, 1999) and Van Fraassen (Fraassen, 1977, 1982, 1989), see also (Rédei, 1997) (Hofer-Szabó et al., 1999, 2000a, b) and (Redei-Summers, 2002).

There is no consensus in the literature as to whether RCCP is a true statement characterizing the causal structure of the world. But if one assumes that the Common Cause Principle is valid, it is natural to ask whether our probabilistic theories can be compatible with RCCP in the sense of being 'causally rich' enough to contain the causes of the correlations they predict. The notion of causal completeness intends to express causal richness of a probabilistic theory:

Definition 4.1 The probability space $(\mathcal{L},\phi)$ is called *causally complete* with respect to a causal independence relation $\mathcal{R}$ and correlation function $Corr_\phi$, if for any two compatible variables $\underline{A}_I,\underline{B}_J$ in $\mathcal{L}$ such that $Corr_\phi(\underline{A}_I,\underline{B}_J) > 0$ and $\mathcal{R}(\underline{A}_I,\underline{B}_J)$ holds, there exists a generalized Reichenbachian common cause $\underline{C}_K$ of some size $K \geq 2$ in $\mathcal{L}$ of the correlation. If, moreover, there is a fixed number $N \geq 2$ such that the correlation between any two compatible, correlated, causally independent variables $\underline{A}_I,\underline{B}_J$ in $\mathcal{L}$ has a generalized Reichenbachian common cause $\underline{C}_K$ of size $K = N$ in $\mathcal{L}$, then we call the probability space *causally N-complete.*

Note that causal N-completeness entails causal completeness, but the converse is not true in general. When a probability space is causally 2-complete with respect to an $\mathcal{R}$ and Corr_ϕ we also say that the probability space is *causally event-complete*.

The general problem of causal completeness is then under what conditions on $(\mathcal{L},\phi)$, $\mathcal{R}$ and Corr_ϕ is $(\mathcal{L},\phi)$ causally complete and N-complete with respect to $\mathcal{R}$ and Corr_ϕ. Before formulating open problems and some results on causal completeness with respect to specific choices of $\mathcal{R}$ and Corr_ϕ, we need to make some further preparations to avoid trivialization of the problem.

Note that there exist *trivially* causally complete probability spaces: ones that do not predict any correlation between any variables $\underline{A}$ and $\underline{B}$ in $\mathcal{L}$. This is the case for instance if $(X,\mathcal{S},p)$ is a classical probability space and p is a Dirac measure. In what follows, when we talk about causal completeness, we always mean *non-trivial* causal completeness. Also note that one can easily make an arbitrary probability space $(\mathcal{L},\phi)$ even causally N-complete for arbitrary N by declaring $\underline{A}_I,\underline{B}_J$ causally *dependent* whenever they are correlated but there exists no generalized Reichenbachian common cause of size N in $\mathcal{L}$. This is however an uninteresting way of making probability spaces causally complete – clearly, in order to get an interesting notion of causal completeness, we need a disciplined definition of R.

In general, the causal independence relation $\mathcal{R}$ will depend on the characteristics of the probabilistic theory predicting the correlations. However, on the basis of general considerations, some necessary conditions can be imposed on $\mathcal{R}$. Consider the case when $A \leq B$: then $A \wedge B = A$ and so $\phi(A \wedge B) = \phi(A)$, hence $\phi(A \wedge B)\ /\phi(A)\phi(B)$ if $\phi(B) \neq 1$ and $\phi(A) \neq 0$. That is to say, elements $A \leq B$ are typically correlated in the sense of (7.14); however, this correlation arises *a priori*, purely from the algebraic relation between A and B, and for such a correlation one does not expect to have a common cause explanation. In other words, $\mathcal{R}(A,B)$ should be strong enough to exclude $A \leq B$ (and by symmetry also $B \leq A$); also, by a similar argument, $\mathcal{R}(A,B)$ should exclude $A \leq B^{\perp}$, and by symmetry, also $B^{\perp} \leq A$. These requirements can be expressed compactly by saying that $\mathcal{R}(A,B)$ implies that A and B are logically independent; equivalently, that

$$\{\emptyset, A, A^{\perp}, I\} \text{ and } \{\emptyset, B, B^{\perp}, I\}$$

are logically independent sublattices of $\mathcal{L}$ in the sense of Definition 2.1. More generally, causal independence, $\mathcal{R}(\underline{A}_I, \underline{B}_J)$, of variables $\underline{A}_I$ and $\underline{B}_J$ should entail that the sets $\{A_i\}_{i\in I}$ and $\{B_j\}_{j\in J}$ are logically independent in the sense of Definition 2.1.

These considerations motivate to focus on a particular type of causal independence relation. Let $(\mathcal{L}, \phi)$ be a probability space. $\mathcal{R}_{\mathcal{L}}$ denotes the following causal independence relation on $\mathcal{L}$: $\mathcal{R}_{\mathcal{L}}(\underline{A}, \underline{B})$ if and only if there exists a pair $(\mathcal{L}_1, \mathcal{L}_2)$ of logically independent sublattices of $\mathcal{L}$ such that $\underline{A}$ is in $\mathcal{L}_1$ and $\underline{B}$ is in $\mathcal{L}_2$. Note that causal completeness with respect to $\mathcal{R}_{\mathcal{L}}$ is equivalent to the following: however one gives two logically independent sublattices in $\mathcal{L}$, every correlation between variables belonging to these sublattices has a generalized Reichenbachian common cause in $\mathcal{L}$.

$R(\underline{A}, \underline{B})$ may hold for compatible $\underline{A}, \underline{B}$ variables of $\mathcal{L}$ irrespective of their size. However, in some cases we are only interested in explaining correlations between compatible variables of a given size; for instance, one might be interested in answering the question: can correlations between *events* be explained by a common cause? In order to be able to ask such questions concisely we introduce the following notation: for given natural numbers M_1, M_2 we denote by $\mathcal{R}^{M_1 M_2}$ the causal independence relation restricted to variables of size M_1 and M_2 respectively; i.e.: $\mathcal{R}^{M_1 M_2}(\underline{A}_I, \underline{B}_J)$ if and only if $\mathcal{R}(\underline{A}_I, \underline{B}_J)$ and $I = M_1$, $J = M_2$. The causal independence relation $\mathcal{R}^{22}_{\mathcal{L}}$ (more precisely $\mathcal{R}^{22}_{\mathcal{L}}(\underline{A}, \underline{B})$) holds then between variables $\underline{A}$ and $\underline{B}$ if and only if $\underline{A} = \underline{A}_2 = (A, A^{\perp})$, $\underline{B} = \underline{B}_2 = (B, B^{\perp})$ and $A, A^{\perp} \in \mathcal{L}_1$ and $B, B^{\perp} \in \mathcal{L}_2$ with $\mathcal{L}_1$ and $\mathcal{L}_2$ being some logically independent sublattices of $\mathcal{L}$.

In the next section we recall results that spell out causal (in)completeness, for certain specific classical probability spaces and mainly for the particular correlation given by (7.14) and causal independence relation $\mathcal{R}^{22}_{\mathcal{L}}$.

7.5 Some Results on Causal Completeness

Most of the propositions in this section involve causal event-completeness, and in all propositions in this section, unless explicitly said otherwise, causal completeness will be understood with respect to the causal independence relation $\mathcal{R}^{22}_{\mathcal{L}}$ and correlation $\mathrm{Corr}^R_\phi(A,B)$ between two-valued, compatible variables $\underline{A}_2 = (A, A^\perp)$ and $\underline{B}_2 = (B, B^\perp)$:

$$\mathrm{Corr}^R_\phi(A,B) = \phi(A \wedge B) - \phi(A)\phi(B) \tag{7.21}$$

Note that $\mathrm{Corr}^R_\phi(A,B) > 0$ if and only if $\mathrm{Corr}^R_\phi(A^\perp, B^\perp) > 0$.

Proposition 3 (Gyenis and Rédei 2004) *Let* $(\mathcal{S}_5, p_u)$ *be the probability space with the Boolean algebra* $\mathcal{S}_5$ *generated by 5 atoms and with* $\mathrm{p_u}$ *being the probability measure defined by the uniform distribution on atoms of* $\mathcal{S}_5$. *Then* $(\mathcal{S}_5, p_u)$ *is causally event-complete.*

Proposition 3 tells us that, however one gives two logically independent sub-Boolean algebras in $\mathcal{S}_5$, every correlation (given by p_u) between elements belonging to these sub-Boolean algebras has a Reichenbachian common cause in $\mathcal{S}_5$. Our next proposition shows that this property of the probability space $(\mathcal{S}_5, p_u)$ described in Proposition 3 is exceptional.

Proposition 4 (Gyenis and Rédei, 2004) *If the probability space* $(\mathcal{S}_n, p)$ *is not* $(\mathcal{S}_5, p_u)$*, then it is* not *(non-trivially) causally event-complete.*

Proposition 4 says that a probability space with event algebra generated by $n \neq 5$ atoms or with a non-uniform probability measure is such that it *does* contain a pair of logically independent Boolean subalgebras such that there exist correlations between *elements* in the respective subalgebras that *cannot* have an event-type (i.e. of size 2) common cause explanation.

One can ask how serious is this non-completeness for probability spaces with finite event structures. One known result is:

Proposition 5 (Gyenis Rédei, 2004) *For any* $n \geq 5$*, there exists a probability measure* p *on* $\mathcal{S}_n$ *and there exist two logically independent Boolean subalgebras* $\mathcal{L}_1, \mathcal{L}_2$ *of* $\mathcal{S}_n$ *such that there exists an event-type common cause for every correlation between elements of* $\mathcal{L}_1$ *and* $\mathcal{L}_2$.

In view of the propositions above, the best one can generally hope in the case of finite event structures is that *some* correlations between events belonging to logically independent Boolean subalgebras have common causes explaining them. Causally incomplete probability theories might however be made 'locally causally complete' in the sense of the next definition. To state that definition we need the notion of extension of the causal independence relation: Let $(\mathcal{L}', \phi')$ be an extension of $(\mathcal{L}, \phi)$ via the homomorphism h and $\mathcal{R}$ be causal independence relation on $\mathcal{L}$. We call the causal independence relation $\mathcal{R}'$ on $\mathcal{L}'$ an *extension* of the causal independence relation $\mathcal{R}$ if $\mathcal{R}'(h(\underline{A}), h(\underline{B}))$ whenever $\mathcal{R}(\underline{A}, \underline{B})$, and not $\mathcal{R}'(h(\underline{A}), h(\underline{B}))$ whenever not $\mathcal{R}(\underline{A}, \underline{B})$.

Definition 5.1 $(\mathcal{L}, \phi)$ is *causally (N-)completable* with respect to a correlation function $Corr_\phi$ and a causal independence relation $\mathcal{R}$ if there exists an extension $(\mathcal{L}', \phi')$ of $(\mathcal{L}, \phi)$ and an extension $\mathcal{R}'$ of $\mathcal{R}$ such that $(\mathcal{L}', \phi')$ is causally *(N-)*complete with respect to $R\prime$ and $Corr_{\phi'}$. If $\mathcal{L}'$ can be chosen to be finite, then we say that $(\mathcal{L}, \phi)$ is *finitely causally (N-)completable.*

Note that the notion of causal N-completability is stronger than causal completability and finite causal (N-)completability is stronger than causal (N-)completability.

Let's consider again the specific case of $\mathcal{R}^{22}_{\mathcal{L}}$ and Corr^R_ϕ. We have:

Proposition 6 *Every* $(\mathcal{S}_n, p)$ *is finitely causally event-completable.*

Proposition 6 is a direct consequence of Proposition 1: to see this, one just has to take $\mathcal{R}'$ to be the extension of $\mathcal{R} = \mathcal{R}^{22}_{\mathcal{S}_n}$ for which $\mathcal{R}'(h(\underline{A}_2), h(\underline{B}_2))$ holds only when $\mathcal{R}^{22}_{\mathcal{S}_n}(\underline{A}_2, \underline{B}_2)$. However, Proposition 4 entails that no finite classical probability space is finitely causally event-completable if we allow the extension $\mathcal{R}'$ of $\mathcal{R}^{22}_{\mathcal{S}_n}$ to be $\mathcal{R}^{22}_{\mathcal{S}'}$. Can we say something informative about the intermediate cases; i.e. when $\mathcal{R}'$ is stronger than $\mathcal{R}^{22}_{\mathcal{S}'}$ but weaker than the extension of $\mathcal{R}^{22}_{\mathcal{S}_n}$? The problem is open.

Causal completeness is not impossible for probability spaces with infinite $\mathcal{S}$, as the next proposition indicates:

Proposition 7 (Gyenis, 2005). *Let* $(X, \mathcal{S}, p)$ *be a classical, atomless probability space,* $Corr_\phi$ *an independence signaler correlation function, and* $\underline{A}, \underline{B}$ *two correlated variables of finite size, related in a genuinely probabilistic way. Then for any* K *satisfying* $\min\{I, J\} \leq K < \infty$ *there exists an (unordered) partition* $\{C_k\}_K$ *of size* K *in* $\mathcal{S}$ *satisfying the screening off conditions (8).*

For the case $\mathcal{R}^{22}_{\mathcal{L}}$ and Corr^R_ϕ we can show the completeness of atomless general probability spaces:

Proposition 8 *Every atomless general probability space is causally event-complete.*

Specifically, in von Neumann algebra setting we have:

Proposition 9 *If* $\mathcal{N}$ *is a type III von Neumann algebra and* ϕ *a faithful normal state on* $\mathcal{N}$ *then* $\big(\mathcal{P}(\mathcal{N}), \phi\big)$ *is causally event-complete (with respect to* $\mathcal{R}^{22}_{\mathcal{L}}$ *and* $Corr^R_\phi$*).*

Proposition 9 is a direct corollary of Proposition 8 because $(\mathcal{P}(\mathcal{N}), \phi)$ is an atomless probability space (Lemma 4. in (Rédei and Summers, 2002)). Probability spaces of this sort occur in models of relativistic quantum fields and in quantum statistical mechanics (see (Rédei and Summers, 2007) for a review of quantum probability spaces).

Note that atomless classical probability spaces are not rare: if $\mathcal{S}$ is the Borel σ-algebra of real numbers, then p given by a density function (with respect to the Lebesgue measure) yields an atomless probability space. Probability spaces of this sort occur frequently in applications.

7.6 Closing Comments

The notions of causal completeness specified by Definition 4.1 have a number of variables: the correlation function Corr_ϕ, the causal independence relation R, and the sizes I, J, K of variables $\underline{A}_I, \underline{B}_J, \underline{C}_K$ for which the the common cause relation is required. Thus the concept of causal completeness as specified in this paper is both very general and flexible; as a result, the notion can be adjusted to different particular applications.

The results on causal completeness in Section 7.5 show that the causal behavior of probability spaces (especially probability spaces with a finite $\mathcal{L}$) that emerge in particular situations might differ from case to case. Since the notion of a Reichenbachian Common Cause is rather subtle, there is no straightforward test with the help of which one could tell if a probability theory is causally complete.

The status of causal completeness in probabilistic theories is relevant for the philosophical problem of falsification of Reichenbach's Common Cause Principle: One may, in principle, try to falsify the Common Cause Principle by claiming that if it were true then it must be possible for a probabilistic theory to be causally complete; hence, by proving that it is mathematically impossible to have causally complete probabilistic theories one would show that the Principle is not tenable.

The defence of Reichenbach's Common Cause Principle against such a falsifying attack requires to prove that causally complete theories are *not* impossible. As the results in Section 7.5 show, it *is* possible for a probabilistic theory to be causally complete in *some* sense of causal completeness in the hierarchy of notions of causal completeness; hence the Principle has to be interpreted very strongly for such an attempt at falsification to have a chance of succeeding.

Given a Corr_ϕ and a causal independence relation $\mathcal{R}$, a very strong interpretation of causal completeness (hence of the Common Cause Principle) could be to require causal N-completeness (with respect to Corr_ϕ and $\mathcal{R}$) for *all* natural numbers N for which such a requirement is meaningful in a given probability theory. It is not known if probability theories with $\mathcal{L}$ having *finite* cardinality can be causally complete in this strong sense. While such a strong causal completeness might be possible, there does not seem to be any reason to expect it to be typical in case of probability spaces with finite $\mathcal{L}$, nor can one expect to see a discernible, regular pattern of the causal behavior of $(\mathcal{L}, \phi)$ as a function of (finite) cardinality n of $\mathcal{L}$: As n grows, the possible sizes of the common cause variable grows as well, which makes finding in $\mathcal{L}$ a common cause for a correlation more likely; however, both the number and the sizes of possible correlated pairs $(\underline{A}_I, \underline{B}_J)$ grows with n as well, which requires even more common causes – unless the causal independence $\mathcal{R}$ controls the growth of the number of correlated pairs for which a common cause has to exist in $\mathcal{L}$. Thus there is a complicated interplay between the variables of the problem of common cause completeness.

We have seen however that even the (quite strong) causal 2-completeness, i.e. event-completeness is *not* impossible: if $(\mathcal{L}, \phi)$ is non-atomic (and hence $\mathcal{L}$ has an *uncountably* infinite number of elements) then $(\mathcal{L}, \phi)$ is causally event-complete. We conjecture that atomless probability spaces, by virtue of their being extremely rich in events, are not only causally event-complete but causally complete in a much stronger sense: we expect atomless probability spaces to be causally N-complete for every N (with respect to $\mathcal{R}_{\mathcal{L}}$ and with respect to correlation functions satisfying some minimal requirements such as independence signalling). This expectation motivates focusing on variables of finite size in finite spaces since the casual behavior of such spaces is more varied than those of an atomless space, and the probability spaces needed to represent variables with infinitely many values are frequently atomless.

Atomless probability spaces are however highly non-constructive and hence quite far from what is strictly empirically accessible. It would therefore be interesting to know whether causal event-completeness (or causal N-completeness for *all* natural numbers N) is possible in classical probability spaces $(X, \mathcal{S}, p)$ with a *countably* infinite set X of elementary random events (with respect to some Corr_ϕ and some causal independence relation $\mathcal{R}$). More generally, it would be desirable to know something about the status of causal completeness in this case.

Setting up a probabilistic model $(\mathcal{L}, \phi)$ of certain physical situations requires expressing particular features of the physical system in terms of $(\mathcal{L}, \phi)$, features that might entail conditions that a common cause has to satisfy in addition to the ones in Definition 3.1. This happens in the case of correlations predicted by local relativistic quantum field theory (QFT (Haag, 1992)) between spacelike separated quantities (see (Summers, 1990) for a review of the relevant results on spacelike correlations in QFT). Since by construction QFT is both local and compatible with the basic causality principle of special relativity, QFT both prescribes the location of the common cause and determines a causal independence relation automatically: $\mathcal{R}(\underline{A}, \underline{B})$ holds if $\underline{A}$ and $\underline{B}$ are located in spacelike separated spacetime regions V_1, V_2 and the common cause of correlations between such causally independent variables must lie in the common causal past of V_1 and V_2. These additional requirements result in a particular definition of causal completeness of the probability theory describing QFT, and make the problem of status of causal completeness of QFT mathematically well-defined. The problem is open, only partial results are known that locate the common cause of correlations between spacelike separated variables in the *union* of the causal pasts of V_1 and V_2 (rather then in the *intersection*) (see (Rédei, 1997, 2002; Rédei and Summers, 2002, 2005) for details).

We close with a remark regarding the sufficiency of Definition 3.1 to capture our intuitive notion of common cause. For simplicity, let's consider first Reichenbach's original definition. As we mentioned earlier, Reichenbach's equations (7.14), (7.15), (7.16), (7.17) entail that events A and B are positively correlated: $\mathrm{Corr}^R_\phi(A, B) > 0$. The first two conditions (7.14), (7.15) express that this positive correlation is screened off by event C. The last two conditions (7.16), (7.17) were meant to express that event C is causally related to both A and B, hence the wording common *cause*. However, the mere presence of a correlation between A and C (B and C) can not establish that there is a causal relationship between A and C (B and C) – this is the main moral from recognizing that correlation does not entail causation. Thus generalizing Reichenbach's original equations in the form of Definition 3.1, in which all conditions are expressed in terms of correlations, also makes it clear that (7.8), (7.9), (7.10) (and hence (7.14), (7.15), (7.16), (7.17), in themselves, can not provide a sufficient condition for a variable (or for an event) to be a common *cause* of a correlation. We also need to insure that the correlation between $\underline{A}$ and $\underline{C}$ ($\underline{B}$ and $\underline{C}$) itself doesn't have a common cause.

There are two possible ways to achieve this goal: we either impose additional conditions in terms of correlations, or we impose an additional mathematical structure on our space of events. The former approach leads either to a circularity in the definition of common cause or to an infinite regress of conditions of correlations of variables. The latter approach can also be pursued in different ways. Since in our discussion of causal completeness we introduce an additional structure in the form of a causal independence relation $\mathcal{R}$, it seems natural to require $\underline{A}$ and $\underline{C}$ ($\underline{B}$ and $\underline{C}$) not to be causally independent according to $\mathcal{R}$, and hence their correlation not to be in need of an explanation in terms of common causes. Considerations about direction of causal influence might suggest to subscribe to a more detailed additional structure, such as a directed graph, featured in many approaches to Bayesian causal nets. In general, it is an open question whether such refinements of the notion of the Reichenbachian common cause could be informatively carried out, and it is also an open question how much these refinements would change the results known in the literature.

References

Arntzenius, F. (1993), The common cause principle, *PSA 1992* 2, 227–237.

Butterfield, J. (1989), A space-time approach to the Bell inequality, In J. Cushing and E. McMullin (eds.), *Philosophical Consequences of Quantum Theory*. Notre Dame: University of Notre Dame Press, pp. 114–144.

Cartwright, N. (1987), How to tell a common cause: Generalization of the conjunctive fork criterion. In J. H. Fetzer (ed.), *Probability and Causality*, Dordrecht: Reidel, pp. 181–188.

Fraassen, B. V. (1977), The pragmatics of explanation. *American Philosophical Quarterly* 14, 143–150.

Fraassen, B. V. (1982), Rational belief and the common cause principle, In R. McLaughlin (ed.), *What? Where? When? Why?* Dordrecht: Reidel, pp. 193–209.

Fraassen, B. V. (1989), 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, pp. 97–113.

Gyenis, B. (2005), *On formal, purely probabilistic theories of evidence*, Masters Research Paper, Department of History and Philosophy of Science, University of Pittsburgh.

Gyenis, B. and Rédei, M. (2004), When can statistical theories be causally closed?, *Foundations of Physics* 34, 1285–1303.

Haag, R. (1992), *Local Quantum Physics*. Berlin: Springer.

Hofer-Szabó, G. and Rédei, M. (2004), Reichenbachian Common Cause Systems, *International Journal of Theoretical Physics* 43, 1819–1826.

Hofer-Szabó, G. and Rédei, M. (2006), Reichenbachian common cause systems of arbitrary finite size exist, *Foundations of Physics Letters* 35, 745–746.

Hofer-Szabó, G., Rédei, M. and Szabó, L. (1999), On Reichenbach's common cause principle and Reichenbach's notion of common cause, *The British Journal for the Philosophy of Science* 50, 377–398.

Hofer-Szabó, G., Rédei, M. and Szabó, M. (2000a), Common cause completability of classical and quantum probability spaces, *International Journal of Theoretical Physics* 39, 913–919.

Hofer-Szabó, G., Rédei, M. and Szabó, L. (2000b), Reichenbach's common cause principle: Recent results and open questions, *Reports on Philosophy* 20, 85–107.

Placek, T. (2000a), *Is Nature Deterministic?* Cracow: Jagellonian University Press.

Placek, T. (2000b), Stochastic outcomes in branching space-time. An analysis of Bell theorems, *The British Journal for the Philosophy of Science* 51, 445–475.

Rédei, M. (1995a), Logical independence in quantum logic, *Foundations of Physics* 25, 411–422.

Rédei, M. (1995b), Logically independent von Neumann lattices, *International Journal of Theoretical Physics* 34, 1711–1718.

Rédei, M. (1997), Reichenbach's common cause principle and quantum field theory, *Foundations of Physics* 27, 1309–1321.

Rédei, M. (1998), *Quantum Logic in Algebraic Approach*, Vol. 91 of *Fundamental Theories of Physics*. Dordrecht: Kluwer.

Rédei, M. (2002), Reichenbach's common cause principle and quantum correlations, In T. Placek and J. Butterfield (eds.), *Modality, Probability and Bell's Theorems*, Vol. 64 of *NATO Science Series, II.* Dordrecht: Kluwer pp. 259–270.

Rédei, M. and Summers, S. (2002), Local primitive causality and the common cause principle in quantum field theory, *Foundations of Physics* 32, 335–355.

Rédei, M. and Summers, S. (2005), Remarks on causality in relativistic quantum field theory, *International Journal of Theoretical Physics* 44, 1029–1039.

Rédei, M. and Summers, S. (2007), Quantum probability theory, *Studies in the History and Philosophy of Modern Physics* 38, 390–417.

Reichenbach, H. (1956), *The Direction of Time*. Los Angeles MA: University of California Press.

Salmon, W. (1978), Why ask "Why?"? *Proceedings and Addresses of the American Philosophical Association*, 51, 683–705.

Salmon, W. (1980), Probabilistic causality, *Pacific Philosophical Quarterly* 61, 50–74.

Salmon, W. (1984), *Scientific Explanation and the Causal Structure of the World*, Princeton NJ: Princeton University Press.

Sober, E. (1984), Common cause explanation, *Philosophy of Science* 51, 212–241.

Sober, E. (1988), The principle of the common cause, In J. Fetzer (ed.), *Probability and Causality*. Boston MA: Reidel, pp. 211–228.

Sober, E. (2001), Venetian sea levels, British bread prices, and the principle of common cause, *The British Journal for the Philosophy of Science* 52, 331–346.

Spohn, W. (1991), On Reichenbach's principle of the common cause, In W. Salmon and G. Wolters (eds.), *Logic, Language and the Structure of Scientific Theories*, Pittsburgh PA: University of Pittsburgh Press.

Summers, S. (1990), Bell's inequalities and quantum field theory, In *Quantum Probability and Applications V.*, Vol. 1441 of *Lecture Notes in Mathematics*. Berlin, Heidelberg, New York: Springer, pp. 393–413.

Suppes, P. (1970), *A Probabilistic Theory of Causality*. Amsterdam: North-Holland.

Uffink, J. (1999), The principle of the common cause faces the Bernstein paradox, *Philosophy of Science, Supplement* 66, 512–525.

Chapter 8
Causal Markov, Robustness and the Quantum Correlations

Mauricio Suárez and Iñaki San Pedro

8.1 Introduction

Questions regarding the status of causation in quantum mechanics are as ancient as the discipline itself. The founding parents of quantum mechanics often identified causation with determinism and consequently understood the emergence of the fundamentally probabilistic quantum mechanics as the demise of a causal picture of the world. As a consequence quantum theory is often presented as non-causal.[1] The identification of causality and determinism was rather universal: even those who regretted the demise of a causal picture attempted to restore a causal understanding of quantum mechanics precisely by restoring determinism. For instance, David Bohm showed von Neumann's theorem against hidden variables to involve essentially questionable premises, thus paving the way for hidden variables. But while Bohm and von Neumann disagreed regarding the status of causation in quantum mechanics, they agreed that the fortunes of causation and determinism were essentially linked. Bohm's theory is in essence a programme to endow quantum mechanics with an underlying deterministic dynamics.

The identification causality = determinism (let us call it the '$c = d$ *identity*') has continued in different, not always explicit, guises. For example Bell's theorem and the work leading up to it during the 1960s presupposes the notorious *factorizability* condition as a criterion of local causality. *Factorizability* is applicable to the correlations between measurement outcomes of spatially separated systems in EPR-like set-ups. Bell's theorem demonstrated that no 'factorizable' theory can reproduce the quantum correlations. It is thus concluded that Bell's work shows that not only quantum mechanics but any other empirically indistinguishable theory would be non-causal in this sense. But philosophers have shown that Bell's theorem does not entail a departure from the $c = d$ *identity*. Some brilliant work by philosophers of

M. Suárez (✉), I. San Pedro (✉)
Department of Logic and Philosophy of Science, Complutense University of Madrid,
28040 Madrid, Spain
e-mail: msuarez@filos.ucm.es
e-mail: inaki.sanpedro@files.ucm.es

[1] See (Heisenberg, 1958) and (von Neumann, 1955).

M. Suárez (ed.), *Probabilities, Causes and Propensities in Physics*, Synthese Library 347, DOI 10.1007/978-1-4020-9904-5_8, © Springer Science+Business Media B.V. 2011

physics in the early 1980s showed that the *factorizability* condition implies determinism when applied to the EPR perfect anti-correlations.[2] So in the end it turns out that the rejection of local causality promoted by Bell's theorem also presupposes a rejection of determinism, and is hence compatible with the $c = d$ *identity*.

Many physicists have continued to presuppose the $c = d$ *identity*, sometimes unquestioningly so. Philosophers of science by contrast long ago started to work out the details of a stochastic view of causality. On this view causation is essentially probabilistic association, and hence supposedly divorced from determinism. One of the earliest and most influential attempts is Hans Reichenbach's *The Direction of Time*, where the Principle of the Common Cause (PCC) is first stated. The programme gains its full and most developed expression in Patrick Suppes' epoch-making 1970 book, *A Probabilistic Theory of Causality*. In spite of the fact that these were both explicit attempts at building a stochastic theory of causality, it remains controversial just how much they depart as a matter of fact from the $c = d$ *identity*. In particular regarding the PCC some philosophers have gone on to argue that the assumption of screening-off is only valid for deterministic, or quasi-deterministic common causes, but does not hold for probabilistic causes. Hence philosophers have for a very long time now considered that the $c = d$ *identity* is controversial, although they have disagreed among themselves as to whether it should be rejected altogether, or weakened in some interesting sense.[3]

The disagreement over rejection *vs.* weakening goes a long way to explaining why the status of causation in quantum mechanics also remains controversial. A weak version of the $c = d$ *identity* is at the heart of a condition that was widely discussed among philosophers of physics in the early 1990s in connection with the Einstein-Podolsky-Rosen correlations, namely Michael Redhead's robustness. The consensus reached then was by and large that robustness is too strong a condition on probabilistic causality. So the failure of robustness in the EPR set-up is uninformative, and a causal account of the EPR correlations remains an open option.

The current debate on causal inference has moved to a discussion of the Causal Markov Condition (CMC), a sophisticated version of the PCC for directed acyclic graphs.[4] The condition employs a similarly weak version of the $c = d$ *identity*, and remains equally controversial. But it has not been systematically applied to the EPR case, nor has the connection been made explicit to the robustness condition discussed in the early 1990s. Our main aim in this paper is to make an explicit link between CMC and robustness in the context of the EPR correlations. Thus we aim to show that the application of CMC to the EPR correlations is exactly as informative (or uninformative, depending on taste) as robustness. Both conditions hold or fail for the same types of systems. So a defender of the weak version of the $c = d$ *identity*

---

[2] See (Fine, 1982a, b) and (van Fraassen, 1982). The original theorems are due to Suppes and Zanotti (1981).

[3] And several philosophers have gone as far as to defend that causality and determinism in fact exclude each other. See (Hoefer, 2004) for a recent example.

[4] Cf. (Hausman and Woodward, 1999), (Cartwright, 2002) and (Steel, 2005).

will find the failure of both CMC and robustness in the EPR correlations revealing of a striking failure of causality in quantum mechanics – and there is a sense in which this result vindicates the founding parents' suspicion that the probabilistic nature of quantum mechanics is what underlies the failure of causality. But those who are inclined to reject the $c = d$ *identity* altogether are likely to draw rather the opposite lesson: the failure of CMC and robustness is precisely what is to be expected given the probabilistic nature of quantum mechanical causation. Although there is thus no essential superiority, in the context of the EPR correlations, to discussing CMC over robustness we aim to show that the application of CMC is sharper and less cumbersome. Thus we shall urge that the debate over the causal status of the EPR correlations is best continued in the new terms laid down by the Causal Markov Condition.

8.2 EPR and Quantum Correlations

Einstein, Podolsky and Rosen first introduced the so-called EPR thought experiment in 1935[5] as an argument to suggest that the then young and emergent quantum theory did not provide a complete description of reality. In a later refined version presented by David Bohm, two *entangled* electrons are emitted from a source in opposite directions. The spin component of each of the electrons can be later detected (measured) when the electrons hit a fluorescent screen after having passed through an inhomogeneous magnetic field (produced by a Stern-Gerlach magnet).

Several features of this experiment are potentially relevant. First, we will denote by a and b the value of the spin variable of each electron which, in the singlet state, can be either 'spin-up' ($\uparrow$) or 'spin-down' ($\downarrow$) with probability 1/2. We can then denote the corresponding measurement outcome events on each particle as $\uparrow a$, $\downarrow a$, $\uparrow b$ and $\downarrow b$. Second, it is assumed that the state of the entangled electron pair is the *singlet* state:

$$\Psi = \frac{1}{\sqrt{2}} \left(|\uparrow_a\rangle |\downarrow_b\rangle - |\downarrow_a\rangle |\uparrow_b\rangle \right).$$

Third, it is assumed that measurement events at each wing of the experiment, such as $\uparrow a$, and $\downarrow b$, are space-like separated events, i.e. lie outside each other's light cone. This is best represented in the diagram of Fig. 8.1 as the statement that no time-like world-line can reach from b to a or vice versa. Under a conventional albeit controversial interpretation of special relativity, such events can not be causally connected.[6]

Quantum mechanics allows us to calculate single and joint probabilities for the different possible outcomes on both wings. When those calculations are performed on the singlet state, correlations between these outcomes are derived. The EPR

5 (Einstein et al., 1935).

6 See (Maudlin, 1994) for a critical discussion.

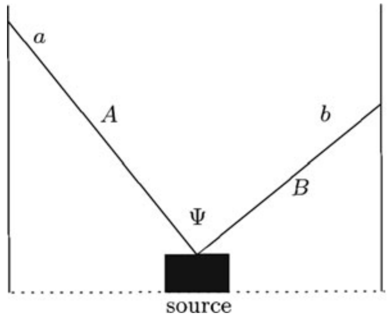


Fig. 8.1 Spacetime representation of a typical EPR experiment

correlations between the different outcome events in both wings of the experiment can be succinctly expressed as:

$$p(a \wedge b) \neq p(a) \cdot p(b).$$

These are the EPR correlations, which have been often positively tested in experiment, and for which we would like to know whether they are the result of underlying causal processes, and which processes. An attempt to determine an answer to these questions was carried out in the late 1980s by the distinguished British philosopher of physics Michael Redhead.

8.3 Redhead's Robustness

Redhead introduced his robustness condition in 1987[7] in order to argue that no *direct* causal relation could be established between the outcome events of an EPR-type experiment. The claim was part of Redhead's attempt at showing that quantum mechanics and relativity can peacefully coexist. Under the presumption that only timelike related events can be causally connected, the measurement outcome events a and b in an EPR experiment can not be causally connected. In particular, Redhead suggested that the EPR correlations were not what he called *robust causal connections*. This in turn entitled him to discard direct causal links between EPR correlated events:[8]

> A stochastic causal connection between two physical magnitudes a and b pertaining to two separated systems A and B is said to be robust if and only if there exist a class of sufficiently small disturbances acting on B (A) such that b (a) screens off a (b) from these disturbances. Denoting the disturbance action on B by d, then the first part of this condition can be rendered formally as

$$\exists D(\forall d \in D[p(a = \varepsilon_a | b = \varepsilon_b \wedge d) = p(a = \varepsilon_a | b = \varepsilon_b)])$$

[7] (Redhead, 1987).

[8] (Redhead, 1987, 102–103).

A similar condition can be written down for disturbances acting on A. The requirement of *robustness* as a necessary condition for a causal relation means that sufficiently small disturbances of either relata do not affect the causal relation.

The intuition that underlies Redhead's robustness is both simple and powerful, and is best brought out by a simplified version of the condition. First, let us simplify Redhead's terminology by identifying physical quantities and the corresponding events. Typically A, B denote a quantity (variable), while a, b denote a value of the corresponding quantity. Hence A, B are capable of entering in causal relations, while a, b are capable of standing in probabilistic relations. However, for convenience, and without loss of generality, we will run them together. Thus a, b will denote indistinctly the quantites and their values; which is which should be clear from the context. We will then say in general that a stochastic causal link between two quantities a and b is robust if and only if the statistical relation $p(a/b)$ is invariant under small disturbances d acting on the putative cause b. In other words b is a robust cause of a if and only if $p(a/b \wedge d) = p(a/b)$. We can see that the intuition behind Redhead's robustness is that it does not matter to the causal link between b and a how the putative cause b comes about, only that it does so (see Fig. 8.2).

It is worth mentioning that initially Redhead apparently took robustness to be both necessary and sufficient for a causal link, but in response to criticism he weakened this to a necessary condition only.[9] In any case robustness is understood to be at least a necessary condition on a causal link. So it becomes superfluous to speak of a robust causal link, since no link that fails to be robust can on this understanding be causal: there is no such a thing as a non-robust causal link. The double terminology points already to what will be the heart of the problem. For Redhead defines robustness as a statistical condition. Hence 'robust causal link' is really a heterogeneous combination of a statistical condition and a causal relation. In stating that robustness is a necessary condition on a causal link Redhead is stipulating that the presence of the causal relation always necessarily implies the statistical condition. So there is a necessary statistical consequence of the existence of a causal relation. As we shall see the critics of robustness quickly pointed out that the statistical condition was not general enough to cover all kinds of probabilistic causes, but rather entailed a

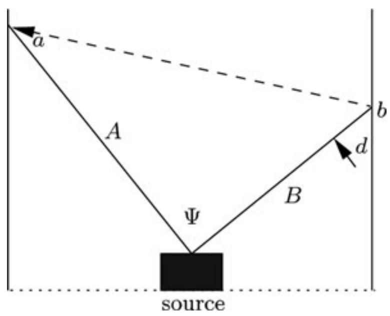


Fig. 8.2 Redhead's Robustness for EPR correlations

[9] For a discussion see (Healey, 1992b).

particular pseudo-deterministic assumption on the working of the cause. The situation is entirely analogous in the recent debate over the Causal Markov Condition.

8.4 Healey on Robustness

The publication of Redhead's work on robustness attracted considerable attention and gave rise to an engaging debate on causation in quantum mechanics in general and in EPR in particular. One of the staunchest critics of robustness is Richard Healey, who discussed and criticised the condition at length in two papers published in the early 1990s.[10] In these papers Healey cast doubts upon the validity of robustness as a criterion for causal inference. His arguments are designed to show that robustness is not a necessary condition in general for a causal link. Our thesis in this paper is that in the context of the EPR correlations the debate over the Causal Markov Condition recapitulates the debate over robustness, so it is worth reviewing the latter in a little detail.

Healey first pointed out that robustness, as defined by Redhead (see Section 8.3), can only be taken to be a necessary condition on *total* causes. In other words, robustness can only be a necessary condition on a causal link between b and a as long as no other causes are operating on a (see Fig. 8.2).[11]

In order to deal with cases in which b is only a *partial* cause of a, Healey introduced a new condition, which he called *internal* robustness:[12]

> A stochastic relation between two events h, k is *internally robust* just in case $p(h/k)$ is invariant under all (sufficiently small) modifications in the causal antecedents of k that leave k fixed and preserve independent causal antecedents of h.

We may rephrase this condition in our terminology as follows. A stochastic causal link between a and b is *internally* robust if and only if the statistical relation $p(a/b)$ is invariant under small disturbances d which leave b and all the independent causal antecedents of a fixed. That is a stochastic causal link between quantities a and b is *internally* robust iff $p(a/b \wedge d \wedge c) = p(a/b \wedge c)$ and d does not causally affect c, where c is the set of all independent causal antecedents of a (see Fig. 8.3).

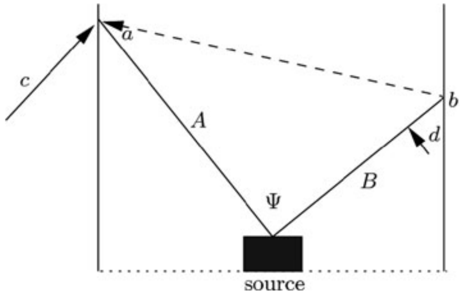
Healey finds both conditions problematic as criteria for causal inference: robustness is problematic because we are very rarely in a position to know that b is the total cause of a, and so a violation of robustness in practice will say nothing informative about whether or not there is a direct causal link between a and b. Robustness may fail because b does not cause a, but it may also fail because we are not accounting for a third partial cause c of a. More specifically in the quantum case it is impossible to know whether the measurement outcome event b in one wing is the only cause of

[10] (Healey, 1992a, b).

[11] We will not here assess this claim, since the aim of the paper is not to evaluate but to compare *robustness* and the Causal Markov Condition, and to show that they face similar difficulties and challenges.

[12] (Healey, 1992b, 183–184).

Fig. 8.3 Healey's Internal Robustness



the measurement outcome event *a* in the other wing, and hence impossible to know in advance whether a failure of robustness implies no causal relation between *b* and *a* or is simply due to the presence of third causes. Similarly for *internal* robustness: we are never in a position to know whether the small disturbances on *b* have in fact no causal effect upon some of the causal antecedents of *a* other than *b*. So as a criterion for causal inference *internal* robustness is just as unhelpful: a failure of *internal* robustness might mean that *b* is not even a partial cause of *a*, but it might also mean that there are other unaccounted partial causes of *a* besides *b* that are in turn effects of causes of *b*.

Redhead's response to these criticisms was to assert that "at some stage in the process of incorporating antecedents in the total cause, robustness must be rescued. Otherwise we would live in a 'marshmallow' world where the notion of cause would not, I believe, be appropriate."[13] In other words, whatever our cognitive and epistemic limitations, a causal relation is properly causal only if robust in actual fact when all other causes have been accounted for. So in other words Redhead's most considered view is that while robustness is not helpful in general as a criterion for positive causal inference, its failure nonetheless allows some minimal negative causal inference. In the EPR case this allows him to say at least that a failure of robustness between the outcome events on both wings *b* and *a* definitely implies that *b* is not the *total* or only cause of *a*. Redhead identified robust causality with action at a distance, and distinguished it from what Abner Shimony[14] called *passion at a distance*, a kind of nomic acausal stochastic link between variables that are 'holistically' implicated – whatever that might mean. The application of the robustness condition to the EPR set-up was designed to show that quantum phenomena exhibits *passion* rather than *action* at a distance. This in turn was argued to be enough to warrant peaceful coexistence with special relativity.

The critics of robustness did not rest content at this point however, but went on to argue against robustness as a necessary condition on causation in general.[15] In other words, they came to dispute the very idea that causal links are Markovian in the way

---

[13] Cf. (Redhead, 1987, vi).

[14] Cf. (Shimony, 1984).

[15] Cf. (Healey, 1992a, b) and (Cartwright and Jones, 1991).

specified by either robustness or *internal* robustness. However, our aim in this paper is not to evaluate robustness and *internal* robustness as necessary conditions on total and partial causes; so we will not review this debate. Our main aim here is to more modestly show that robustness in the EPR case follows logically from the Causal Markov Condition. We consequently argue that discussion regarding causality in EPR is best conducted in terms of the CMC.

8.5 The Causal Markov Condition

The Causal Markov Condition (CMC) is inspired by the Principle of the Common Cause (PCC) and is a keystone and crucial assumption in the most powerful contemporary programs of causal inference. It is intended as a generalised version of the PCC and can be defined, following Hausman and Woodward, as follows:[16]

Causal Markov Condition (CMC): For all distinct variables X and Y in the variable set $\mathbf{V}$, if X does not cause Y then

$$p(X|Y \wedge \mathrm{Par}(X)) = p(X|\mathrm{Par}(X)),$$

where Par(X) (read *parents of* X) is the set of all direct causes of X in $\mathbf{V}$.

The Causal Markov Condition is an extrapolation of the PCC to directed acyclic graphs. The PCC states that a common cause screens-off its effects from each other, as long as there are no direct causal links between these effects. The CMC more generally states that the parents of X (Par(X)) screen-off X from any other variable Y in the variable set $\mathbf{V}$ that is not a direct causal descendent of X. In short: if X does not cause Y in $\mathbf{V}$, then Par(X) will screen them off. The contraposition is rather useful in the EPR set-up: if the putative parent of one of the measurement outcome events, say a, does not screen it off from the other outcome event b then it follows that a does not cause b, or we have not identified the only putative parent. In the EPR scenario it is often assumed that (i) the two measurement outcome events can not be causally connected because of relativistic constraints, and (ii) that the only putative common parent of such measurement outcome events is the singlet state at the source. So the residual correlation between the events a and b which does not disappear when CMC is applied, must be accounted by some rather mysterious nomic and acausal mechanism.[17] This of course is very much in line with Readhead's thought that underlying the EPR correlations are non-robust stochastic links that are unthreatening to special relativity. It will then not come as a surprise that there is a strong formal connection between the CMC and the robustness conditions.

[16] See (Hausman and Woodward, 1999, 523). Note that Hausman and Woodward's definition is distinct in some significant ways from the original in Spirtes, Glymour and Scheines (2000 [1993], 29) – see (Steel, 2006) for a discussion. The distinction makes no difference to our argument, however, so we ignore it here – and instead stick to Hausman and Woodward's definition for consistency.

[17] See (Hausman and Woodward, 1999, 564–567) and (Hausman, 1999).

8.6 Robustness and the Causal Markov Condition

We show in this section that robustness is indeed a consequence of applying the Causal Markov Condition to an EPR setting, given some additional assumptions. In fact we show this for both of Healey's conditions by simply applying the Causal Markov Condition to total and partial causes respectively.

8.6.1 Total Causes and the Causal Markov Condition

Let us first consider robustness. Let us suppose that b is the *total cause* of a. In this case b is the set of *all parents* of a. That is:

(I) If b is the total cause of a, then $\mathrm{Par}(a) = b$.

Let us first assume, following robustness, that there exist a small disturbance d on the putative parents of a, and let us substitute d in for the term Y in the expression of CMC:

(II) $\exists d : d = Y$.

Let us then assume that the measurement outcome event a is not a cause of the small disturbance, i.e. let us assume:

(III) a does not cause d.

And finally let us turn to the definition of the Causal Markov Condition CMC given in the previous section. By substitution it follows from (I), (II), (III) and CMC that:

$$p(a|d \wedge b) = p(a|b),$$

which is an explicit expression of robustness. Thus we have formally shown that under assumptions (II) and (III):

$$(\mathit{TotalCause}) \wedge (\mathit{CausalMarkov}) \Rightarrow (\mathit{Robustness}).$$

In other words under the assumption of the existence of independent distubing causes, robustness is the consequence of applying the Causal Markov Condition to total causes.

8.6.2 Internal *Robustness, Partial Causes and the Causal Markov Condition*

Now let us turn to *internal* robustness and *partial* causes. Let us then suppose that b is a partial cause of a. It follows that there is a non-empty set of additional variables c that represent all other independent causal antecedents of a. Let us simplify by bringing them all under an additional variable c in the causal graph. Then the complete set of parents of a in the graph is the union of c and b:

(I′) If b is *partial cause* of a, then: $\mathrm{Par}(a) = \{b, c\}$.

Let us assume, as before, the existence of a small disturbance d on the putative parents of a in place of Y in the expression of CMC:

(II) $\exists d: d = Y$.

And similarly, that the measurement outcome event a is not a cause of the small disturbance, i.e. that:

(III) a does not cause d.

By substitution, it follows from (I′), (II) and (III) and the CMC that:

$$p(a|d \wedge b \wedge c) = p(a|b \wedge c),$$

which is Healey's *internal* robustness. Thus we have formally shown that under the same assumptions (II) and (III):

$$(\mathit{PartialCause}) \wedge (\mathit{CausalMarkov}) \Rightarrow (\mathit{InternalRobustness}).$$

Under the assumption of independent disturbing causes, *internal* robustness is a consequence of applying Causal Markov to partial causes.[18]

8.6.3 *Robustness Updated*

We have shown that robustness and *internal* robustness are consequences of applying the Causal Markov Condition to the measurement outcome events a and b. If b is taken to be a total cause of a then the CMC together with some special assuptions entails robustness. If on the other hand b is taken to be merely a partial cause of a then the CMC with the same assumptions entails *internal* robustness. So it seems

[18] A referee pointed out that the role of total or partial cause in these proofs is to make sure that d can only cause a via b in the case of total cause, and via $\{c, b\}$ in the case of partial cause. Indeed that would be an alternative definition of Healey's terms.

that the intuition underlining Michael Redhead's conditions is as a matter of fact the Causal Markov Condition. And the contrary intuitions and arguments by their critics are conversely related to doubts regarding the Causal Markov Condition. The CMC backs up Redhead's robustness, so if CMC is false in general as many recent critics believe,[19] then robustness is left without substantial justification. The failure of robustness in EPR established by Redhead would be without any consequences were it not backed up by the CMC.

Moreover we have shown that a failure of Redhead's conditions entails a failure of the CMC regardless of whether the putative link is taken to be a total or a partial cause. So the distinction between total and partial causes that seemed so important in the early 1990s now seems irrelevant. The Causal Markov Condition is what underlies Redhead's intuition regardless. Similarly Healey's subtle distinctions between kinds of robustness are now seen to be irrelevant for a proper assessment of the causal nature of the EPR correlations. The peaceful coexistence between quantum mechanics and relativity so sought after by philosophers in the early 1990s is to be achieved always at the cost of a violation of the CMC, regardless of the underlying causal structure. So philosophers of physics interested in the issue of coexistence would be well adviced to turn to a careful and detailed analysis of the implications of the CMC to the EPR correlations. This is essentially the central claim of our paper, and we find it remarkable that it needs to be made. But indeed it does, for such an analysis has not yet been carried out. We can at best find the outlines in the very brief discussion of EPR in Hausman and Woodward[20], and in a a recent paper by Daniel Steel[21].

8.7 EPR and the Causal Markov Condition

It has been claimed (for example by Salmon[22]) that many genuinely statistical phenomena violate the PCC. Most prominently the EPR correlations are supposed to provide a set of established correlations that can not be explained by either a direct cause or a common cause model under the strictures of PCC.[23]

Yet an important part of Hausman and Woodward's defence of CMC is that EPR is no counterexample.[24] They do not claim that the CMC is satisfied by the EPR correlations, but rather that it is *inapplicable*: it is neither satisfied nor violated, simply inappropriate. The discussion interestingly brings out some crucial differences

[19] (Cartwright, 2002) and (Williamson, 2005).

[20] (Hausman and Woodward, 1999).

[21] (Steel, 2005).

[22] (Salmon, 1984, chapter 7).

[23] One of us has argued against this common lore (Suárez, 2007). However, these arguments do not vindicate the PCC as usually stated but a very different reformulation. We will not review this literature here, but instead refer the reader to that paper.

[24] See (Hausman, 1999) and (Hausman and Woodward, 1999).

between on the one hand the PCC as usually understood and on the other the CMC and the robustness conditions. So we review it briefly here.

8.7.1 Causal Markov, Interventions and Modularity

The key difference between the usual statement of the PCC and the CMC is the assumption of invariance under intervention that, according to Hausman and Woodward underlies and motivates CMC. This is best expressed in the modularity condition:[25]

Modularity (MOD): For all subsets **Z** of the variable set **V**, there is some non-empty range **R** of values of members of **Z** such that if one intervenes and sets the value of the members of **Z** within **R**, then all equations except those with a member of **Z** as dependent variable (if there is one) remain invariant.

Hausman and Woodward take this condition, in conjunction with a few others, to provide the grounds for the CMC. The set **V** is the set of variables in the causal graph, and the equations are the linear regression equations that characterise a causal system. Modularity as a condition on causal systems is then the thought that a relation between two quantities a and b is causal only if (i) it is possible at least in principle to intervene in order to set the values of a and b and their probabilities, and (ii) these interventions – as long as within a permissible range – leave intact the functional connections between the values of a and b, or their probabilities.[26]

The statement of MOD is a conditional with an antecedent that may be false, so a truth-functional interpretation as a material implication would entail that MOD is true by default in all such cases. But the context of the discussion suggests that MOD is meant to be non-applicable in such cases. That is, if interventions are possible in some set **V** and equations do not remain invariant then modularity is false. But if, on the other hand, interventions are not possible for some subset **Z** of **V** then MOD is strictly speaking not false but non-applicable.

Hausman and Woodward's strategy is to attempt to back up CMC by appeal to MOD. But some significant additional assumptions are required to show MOD and CMC equivalent, namely: (i) *causal sufficiency* i.e. that all common causes are included in the set **V**; (ii) the assumption that all correlations have causal explanation; and (iii) the assumption that there exist unrepresented causes which can play the role of interventions. There is no need to get into the details of the equivalence proof, although it is worth mentioning that it has been contested.[27] In this paper we assume for the sake of argument that the proof is valid, and that a failure of CMC entails a failure of either of these conditions.

---

[25] Cf. (Hausman and Woodward, 1999, 545).

[26] The qualification of values *or* probabilities is needed to account for probabilistic causality, which Hausman and Woodward define as deterministic causation of probabilities (Hausman and Woodward, 1999, 570).

[27] See e.g. (Cartwright, 2002).

This has consequences for the discussion of the EPR correlations as we shall see. It also helps to distinguish subtly robustness from the usual statement of the PCC. For the PCC makes no implicit or explicit reference to interventions. By contrast, the notion of 'disturbance' required by robustness is clearly akin to an intervention. Hence a system that allows no interventions at all on any of its variables even in principle (or countenances no small disturbances) might violate the PCC without violating robustness.[28] This is the line defended by Hausman and Woordward with respect to the EPR correlations.[29] Their argument is essentially that there is no possible way to intervene on either of the distant measurement events. Consequently, they argue, it is impossible in this set-up to evaluate the CMC: the EPR correlations can not be shown to be a counterexample. This is precisely the claim we take issue with in this paper.

8.7.2 Interventions in EPR

The main aim of this paper is to urge that the debate over possible causal explanations of the EPR correlations ought to move to a detailed discussion of the CMC and its presuppositions in the context of the EPR experiment. Thus we oppose Hausman and Woodward's thought that CMC is inappropriate for the EPR correlations. On the contrary we believe it is an appropiate kind of condition to apply, but we just do not share the widespread intuition that CMC (or the PCC) necessarily fails for the EPR correlations. We argue instead that whether or not CMC holds depends very much on the details of the precise causal hypothesis under test. The question requires investigation and can not be brushed aside as speedily as Hausman and Woodward would like.

In a sense we believe we have already achieved this aim – it follows from the equivalence proof in Section 8.6. Hence the paper so far may be taken as endorsement of the suitability of CMC for understanding the status of causality in quantum phenomena. However, in this final section we outline our disagreement with the particular conclusions that Hausman and Woodward draw concerning the EPR correlations.

Hausman and Woodward back up the CMC with MOD. So to evaluate their claims we must concern ourselves with whether interventions are possible in the EPR context, and what significance must be attached to this fact. Hausman and Woodward endorse the view that in the EPR set-up there are no distinct mechanisms in the wings of the experiment because in fact there are no different systems to speak of. Both entangled particles are just 'parts' of the same irreducible holistic

[28] The observation is consistent with our results in the previous section, since we showed that CMC entails *robustness* but not that modularity entails *robustness* – the main difference is clear now.

[29] See (Hausman, 1999) and (Hausman and Woodward, 1999).

or non-separable system.[30] Together with the fact that there is no way to control the outcome of the first measurement this indeed seems to entail that interventions to set the values of the outcome events *a* and *b*, separately or jointly, are impossible. They conclude that EPR is no counterexample to MOD or CMC, but rather that these conditions are inappropriate in this context.

However, note that CMC states nothing whatever regarding interventions. It neither requires nor disallows interventions. Hausman and Woodward justify CMC by appeal to MOD, and the latter condition certainly requires interventions. But CMC could in principle be justified by other means that do not require MOD, as long as some of the additional assumptions are forfeited. So, contrary to what Hausman and Woodward seem to claim, the applicability of CMC does not seem to turn on the applicability of MOD and the related availability of interventions.

We have already noted that both MOD and CMC are explicitly stated as conditions on either values *or* probabilities of variables in the variable set **V**. In cases of genuine probabilitistic causation the only relevant factor are the probabilities of the variables, since the causal structure fails to determine the values themselves. And it is of course well known that deterministic local hidden variables are ruled out for quantum mechanics by the Bell inequalities. Hence the EPR correlations are potentially a paradigm but subtle case of probabilistic causality.

Interventions are not impossible to set the *probabilities* of some of these outcomes in the appropriate circumstances. For notice that the experimenter controls the settings of the measurement apparata that determine the direction of spin that gets meausured on each wing. Let us refer to the two wings of the experiment and their corresponding particles as '1' and '2'. It is true that the spherical symmetry of the singlet state entails that the first measurement outcome in the laboratory rest frame always has probability one half, regardless of what direction one measures spin along. Suppose then that in that frame spin is measured on '1' first, and suppose the outcome correspoding to 'spin-up' is found. If this information is provided to the second experimenter on time to set the direction of spin measured on particle '2' she can then easily set her measurement device to definitely get the outcome corresponding to 'spin-down' with probability one (or indeed any other probability but zero). For any value of probability of 'spin-down' on particle '2' she can use quantum mechanics to calculate the appropriate direction of measurement and set her device accordingly.

So it turns out that interventions are possible in particular experimental EPR setups. Notice that the intervention does not just consist in choosing a frame; rather given any frame, an intervention is the setting of a polariser direction. In such setups the question is then whether MOD and CMC hold. We urge in this paper that this is the relevant question to ask for causal modellers of EPR; but we will not

[30] They refer extensively to an old paper by Skyrms that defends this view (Skyrms, 1984); it is worth mentioning that the literature on EPR has moved on a very great deal in the last two decades, particularly on the physics side. Quantum entanglement was not then the area of intense research among physicists that it has become now, and Skyrms' views were much more entrenched twenty five years ago than they are now among both physicists and philosophers.

attempt a comprehensive answer here. The answer is complicated and depends on the details of the causal hypotheses under test.[31] A brief and intuitive argument suggests that CMC may fail here. The EPR correlations are not screened-off by the creation event at the source. Similarly the value of the setting of the measurement device on '2' will not screen-off the outcome event in that wing from the outcome event in the distant wing. But this really says nothing about a direct causal link between the wings. And if CMC failed for indeterministic systems, as some authors argue, then a common cause structure underlying the direct cause link would also be possible, which means that CMC might fail for a and b too. However, this claim requires further investigation in the context of alternative causal hypotheses. For our purposes in this paper this is a side issue, since whatever the correct answer it will already show that CMC *is* applicable to the EPR correlations in spite of Hausman and Woodward's claims to the contrary.

8.7.3 Causal Markov and Other Interpretations

The argument we have just given shows that in any EPR experiment there always exists a subset of the relevant variables that are susceptible to intervention. This leaves open several causal accounts for the EPR correlations. The fate of the CMC very much depends on the details of each account. But we believe that a stronger claim can be made. So far we have been assuming the standard or orthodox interpretation of quantum mechanics. So we have assumed that the violation of the Bell inequalities in the EPR experiments is due to a failure of what is known as *outcome independence*, and correspodingly that the only possible causes of each measurement outcome event are the distant outcome event and the proximate measurement device setting event.

In other words we have assumed that it is meaningless to suppose that the setting events in each wing can be a causal influence upon the *distant* outcome events. But it is well known that on some interpretations of quantum mechanics this is not just allowed but likely. The paradigm case is Bohm's theory. On the account of the EPR correlations provided in Bohmian mechanics,[32] the actual measurement outcome event on one wing has no influence upon the measurement outcome event on the other wing, because in Bohm's theory measurements simply reveal values that are already there, they do not bring these values into being. Yet the setting of the distant device does have a putative causal influence, since it affects the quantum wavefunction of both particles in configuration space, and thus affects the probabilities for outcomes in the distant wing. (The distant setting does not *determine* the proximate outcome of course, which also depends on the initial wavefunction state and the initial complete state of both particles; but it does partly determine the outcomes' probabilities).

[31] For a preliminary account see (Cartwright and Suárez, 2000).

[32] See (Cushing, 1994, pp. 82–95) for a very nice review.

Daniel Steel[33] has claimed that Bohm's theory shows that CMC can fail for deterministic systems. The claim is part of a larger argument in the debate over whether the CMC is satisfied only by deterministic, or more generally 'pseudo-deterministic', systems.[34] Steel argues that the key to the validity of the CMC is not whether the system is deterministic or pseudo-deterministic but rather whether there are exogenous variables that are probabilistically independent from any other variable in the causal strucutre. Bohm's theory is an important part of the argument because it is the only example that Steel provides of a deterministic system that does not satisfy CMC. In other words Bohm's version of the EPR experiment is presented as a plausible counterexample to the claim that determinism grounds the CMC. Presumably, given Steel's argument, this must be the case because there are some probabililistically independent exogenous random variables in Bohm's description of EPR. We shall study this claim closely.

But first let us note some relevant differences between Steel's overall argument for CMC and Hausman and Woodward's defence of CMC by means of the proof of the equivalence of MOD and CMC. Steel does not claim that interventions are required for CMC. (Neither are they required by the letter of CMC, nor are they required to ground CMC).[35] But he does think interventions, by means of controlled experiments, are one way of securing the independence of exogenous variables that does ground CMC. And in his view there is no more reason to expect the method to work in indeterministic contexts. This is a crucial difference between the accounts and it explains Steel's desire to find a counterexample to the $c = d$ *identity* and the related claim that the CMC is linked to determinism. Note in this respect that although we disagree with Steel's claim to have found a counterexample in Bohm's theory, we do not necessarily disagree with the claim that the $c = d$ *identity* is false, nor with the concommittant claim that CMC might be valid for indeterministic and not just deterministic systems. Since it is not the aim of this paper to debate the general validity of CMC we will not assess these general arguments. We are interested though in assessing the chances of CMC for the EPR correlations. And we conjecture that the fate of the CMC in EPR is extremely sensitive to both the details of the causal hypothesis under test *and* the interpretation of quantum mechanics that is adopted.

So does the Bohmian description of the EPR correlations violate CMC? Steel assumes that it does since it predicts the very same EPR correlations. As he writes:

> [...] the EPR example is a problematic basis for the claim that the CMC is a more reliable assumption for deterministic than indeterministic systems for the simple reason that there is a fully deterministic (though heterodox) interpretation of quantum mechanics,

---

[33] Cf. (Steel, 2005).

[34] By pseudo deterministic system we mean a system with causes that do not fix the ocurrence of all their effects, but that can nonetheless be 'embedded in another more complete graph [...] in which the parents of the given effect are sufficient to fix the value of the effect'. (Cartwright, 1999). For a discussion and a reference to the notorious *cheap but dirty factory* example of the presumed failure of CMC in indeterministic systems see (Cartwright, 1993).

[35] (Steel, 2006).

> namely Bohm's. Bohm's quantum theory predicts precisely the same non-local (and hence putatively non-causal) correlations in the EPR example as the standard, indeterministic interpretation. Hence it is far from clear that the blame for the (putative) counter-example can be laid at the door of indeterminism.

In our view this makes the very mistake to suppose that the fate of the CMC is independent of the details of the causal hypothesis under test. There are two versions or interpretations of Bohm's theory: the minimal Bohm theory championed originally by Bell[36], and the causal interpretation defended by Dewdney et al.[37] and Holland[38]. According to the minimal interpretation, particles' only primitive property is position, and there is no such thing as intrinsic 'spin'. Instead the theory manages to produce the same predictions as quantum mechanics for the motion of all particles going through a Stern-Gerlach apparatus simply by means of the influence of the guiding field upon the particle though the so-called 'guidance condition'.[39] The causal interpretation, by contrast, has it that particles are endowed with the intrinsic property of spin, which is understood to be causally reactive to the quantum potential.[40] In both cases the causal structure is rather different and hence there is no real reason to expect CMC's fate to be the same as in orthodox quantum mechanics. On the contrary, we would like to argue that at least in its causal interpretation the Bohmian description of EPR definitely satisfies the CMC.

We have already noted that in the EPR experiment as described by Bohm the measurement outcome events do not cause each other, but the setting events have an influence upon the outcomes. On the minimal interpretation, the settings influence the quantum wave function in configuration space in such a way that the motion of the particles is correspondingly affected after interaction with their respective Stern-Gerlach measuring devices. However, since no intrinsic property of spin is hypothesized, no changes take place ahead of the particle's interaction with their respective measurement devices. So on Bohm's minimal theory, the settings causally influence the outcomes via the measurement process only. It would not be correct to claim on this interpretation that the violation of *parameter independence* entails a causal influence directly from settings to outcomes.

On the causal interpretation by contrast, the settings have a direct and instantaneous causal influence upon both particles' spin values. Indeed the underlying determinism of the theory implies that, on this causal interpretation, the setting events are instantaneous partial causes of the values of spin of the distant particles, which are only later revealed by measurement, if there ever is one, on the

[36] (Bell, 1982).

[37] (Dewdney et al., 1988).

[38] (Holland, 1993).

[39] The full details can be found in (Bohm and Hiley, 1993, chapter 10). See (Berkovitz, 2007, Section 5.3.1) for a brief review.

[40] (Dewdney et al., 1988, pp. 537–539); (Holland, 1993, chapters 10 and 11).

distant wing. So, on this view, my setting the measurement device of particle '1' partially determines not just the probability of an outcome of a measurement on particle '2' – it actually partially determines its value. The reason is that particles on Bohm's theory have well defined values of their dynamical variables at all times – so on the causal interpretation the EPR particles have a value of position and spin from the word go, as they are ejected from the source. This value can change though at any time, and in the case of an entagled particles as in the EPR case, it might do so non-locally as a result of changes in the wavefunction. And the wavefunction is responsive not only to the values of the distant entangled particles but also to the features of the systems those distant particles interact with. Thus although essentially non-local, the causal Bohm theory is indeed also essentially causal, in the strong sense of the $c = d$ *identity* that we mentioned in the introduction. [41]

How do we evaluate CMC then? Since in Bohm's theory measurement outcomes a and b do not cause each other, we can apply CMC fully as follows:

Causal Markov Condition (CMC) for Bohm's theory: For measurement outcome events a and b, since a does not cause b then

$$p(a|b \wedge Par(a)) = p(a|Par(a)),$$

where Par(a) is the set of all direct causes of a in **V**.

There is absolutely no reason to suspect that in Bohm's theory this condition is false, in either the minimal or the causal versions of the theory. On the contrary, since in Bohm's theory the explicit causal antecedents of the measurement outcomes include the quantum wavefunction, the initial complete states of both particles (which includes their spin in the causal interpretation) *and* the distant settings, it follows that Par (a) includes all these. And since these variables jointly determine the value of the outcomes a and b, they jointly determine their probabilities. So the CMC is trivially satisfied in the Bohmian description of the EPR correlations, as long as we include in the set **V** all those variables that according to the theory are effectively causal antecedents of the outcomes a and b.

---

[41] In response to our reasoning at this point Steel has retorted as follows (private correspondence): 'I am not assuming that EPR is a violation of the CMC if Bohm's theory is correct. Rather, I am making the following conditional claim: *if* locality is a necessary condition for causation, *then* EPR is a violation of the CMC according to Bohm's theory'. If this is Steel's more considerate view, it seems to us to worsen his position. For note that the truth of the antecedent of the above conditional claim would make causation impossible by definition on almost any interpretation or version of quantum mechanics – since some form of non-locality is required in any case. But, worse still, the antecedent is false precisely in Bohm's theory, irrespective of interpretation: In both the minimal and the causal interpretations causation is certainly possible, and yet in both cases the theory is explicitly non-local. So the conditional above, if read as a material implication, would turn out to be vacuously true and uninformative about the actual status of CMC in Bohmian mechanics. (If read as an indicative conditional, Steel's statement is just false).

8.8 Conclusions

Our aim in this paper has been to urge more research to be conducted on applying the Causal Markov Condition to the diverse models and interpretations of the EPR correlations. We hope to have shown that questions regarding the causal nature of explanations of the EPR correlations are best explored by means of a detailed and careful analysis of the application of the CMC. This is the right framework to update the debate regarding Michael Redhead's robustness in the early 1990s and to generally conduct the debate. Despite claims to the contrary the answers are not trivial, and the CMC is in principle applicable to the EPR correlation phenomena. But questions remain as to whether CMC is satisfied by these phenomena. We conjecture that the answer to this question is highly sensitive to the details of the causal hypothesis under test. We have also claimed it to be sensitive to the interpretation of quantum mechanics that is adopted, a claim that we have supported by looking at the Bohmian description of the EPR experiment. Contrary to recent claims the Bohmian description of the EPR correlations satisfies CMC.

This suggests that the CMC is a generally valid background or methodological assumption for deterministic or pseudo-deterministic systems.[42] It remains to be seen whether it can be similarly assumed for indeterministic ones such as EPR on the orthodox interpretation of quantum mechanics. Concomittantly it also remains to be seen whether a causal understanding of indeterministic phenomena requires the CMC. Suppose that CMC fails for at least some of the main causal hypotheses for the EPR correlations under the standard or orthodox understanding. If CMC is not required for causation then even the weakest interpretation of the $c = d$ *identity* will have been refuted. If on the other hand CMC is required for causation then quantum mechanical phenomena, on the orthodox interpretation at least, abandons causality as well as determinism, the $c = d$ *identity* is retained, and the intuition of the founding parents is proved correct (for orthodox quantum mechanics at least). The questions are relevant, the stakes are high, and the answers should be informative.

Acknowledgements A preliminary draft of this paper was circulated in discussion paper form at the Centre for the Philosophy of the Natural and Social Sciences, London School of Economics (M. Suárez and I. San Pedro, "EPR, Robustness and the Causal Markov Condition", LSE Philosophy Papers PP/04/07, 19 August 2007). We would like to thank all those who offered comments and suggestions, in particular Daniel Steel and Carl Hoefer. Research towards this paper has been funded throughout by research project HUM2005-01787-C01-03 of the Spanish Ministry of Education and Science. We would like to thank the members of its associated 2005 reading group on causal inference, as well as three anonymous referees for helpful comments and suggestions.

[42] *Modulo* the usually discussed exceptions such as nomic or non-causal inducers of correlations – see e.g. (Hausman, 1999) – , and accidental dependencies such as the one between British bread prices and the water level in Venice in Sober's famous example (Sober, 2001).

References

Bell, J. S. (1982), On the impossible pilot wave, *Foundations of Physics*, 12, 989–999. Reprinted in Bell (1987), *Speakable and Unspeakable in Quantum Mechanics*. Cambridge: Cambridge University Press, pp. 159–168.

Berkovitz, J. (2007), Action at a distance in quantum mechanics, In E.N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2008 Edition), URL = <http://plato.stanford.edu/archives/win2008/entries/qm-action-distance/>.

Bohm, D. (1952) "A suggested interpretation of quantum theory in terms of hidden variables, I and II". *Physical Review*, 85, 166–193 and 369–396.

Bohm, D. and Hiley, B. (1989), Non-locality and locality in the stochastic interpretation of quantum mechanics, *Physics Reports*, 172(3), 93–122.

Bohm, D. and Hiley, B. (1993), *The Undivided Universe*. London: Routledge.

Butterfield, J. (1990), Causal independence in EPR arguments. In *Proceedings of the Bienal Meeting of the Philosophy of Science Association*, vol. I, pp. 213–25.

Butterfield, J. (1992) David Lewis meets John Bell. *Philosophy of Science*, 59, 26–43.

Cartwright, N. (1990) Quantum causes: The lesson of the Bell inequalities. In *Philosophy of the Natural Sciences: Proceedings of the 13th International Wittgenstein Symposium*. Vienna: Hölderlin-Pichler-Tempsky.

Cartwright, N. (1993), Marks and probabilities: Two ways to find causal structure. In F. Stadler, (ed.), *Scientific Philosophy: Origins and Developments,* Yearbook 1/93, Institute Vienna Circle. Dordrecht: Kluwer Academic Publishers.

Cartwright, N. (1999), Causal diversity and the markov condition,. *Synthese*, 121, 3–27.

Cartwright, N. (2002), Against modularity, the causal markov condition, and any link between the two: Comments on Hausman and Woodward. *British Journal for the Philosophy of Science*, 53, 411–453.

Cartwright, N. and Jones, M. (1991), How to hunt quantum causes. *Erkenntnis*, 35, 205–231.

Cartwright, N. and Suárez, M. (2000) A causal model for EPR. Discussion paper 50/00, LSE Centre for the Philosophy of the Natural and Social Sciences.

Chang, H. and Cartwright, N. (1993) Causality and realism in the EPR experiment. *Erkenntnis*, 38, 169–190.

Cushing, J. T. (1994), *Quantum Mechanics*, Chicago: University of Chicago Press.

Dewdney, C., Holland, P., Kyprianidis, A. and Vigier, J. P. (1988) Spin and non-locality in quantum mechanics, *Nature*, 336(8), 536–544.

Einstein, A., Podolsky, B. and Rosen, N. (1935), Can a quantum mechanical description of physical reality be considered complete? *Physical Review*, 47, 777–780.

Elby, A. (1992) Should we explain the EPR correlations causally? *Philosophy of Science*, 56, 16–25.

Elby, A. (1993), Why local realistic theories violate, nontrivially, the quantum mechanical EPR perfect correlations. *The British Journal for the Philosophy of Science*, 44, 213–230.

Fine, A. (1982a), Hidden variables, joint probability, and the Bell inequalities. *Physical Review Letters*, 48, 291–295.

Fine, A. (1982b), Joint distributions, quantum correlations, and commuting observables. *Journal of Mathematical Physics*, 23, 1306–1310.

Glymour, C., Spirtes, P. and Scheines, R. (1991), Causal inference. *Erkenntnis*, 35, 151–189.

Hausman, D. (1999), Lessons from quantum mechanics. *Synthese*, 121, 79–92.

Hausman, D. M. and Woodward, J. (1999), Independence, invariance and the Causal Markov condition. *British Journal for the Philosophy of Science*, 50, 521–583.

Healey, R. (1992a), Causation, robustness, and EPR, *Philosophy of Science*, 59, 282–292.

Healey, R. (1992b), Chasing quantum causes: how wild is the goose? *Philosophical Topics*, 20, 181–204.

Heisenberg, W. (1958), *Physics and Philosophy*. New York, Harper & Row.

Hoefer, C. (2004), Causality and determinism: Tension, or outright conflict? *Revista de Filosofía*, 29, 99–115.

Holland, P. (1993), *The Quantum Theory of Motion*. Cambridge: Cambridge University Press.

Maudlin, T. (1994), *Quantum Non-Locality and Relativity*. Oxford: Blackwell Publishing.

Papineau, D. (1990), Causes and mixed probabilities,. *International Studies in the Philosophy of Science*, 4, 79–88.

Redhead, M. (1987), *Incompleteness, Nonlocality and Realism*. Oxford: Oxford Calrendon Press.

Redhead, M. (1989), The nature of reality, *British Journal for the Philosophy of Science*, 40, 429–441.

Salmon, W. (1984), *Scientific Explanation and the Causal Structure of the World*. Princeton: Princeton University Press.

Shimony, A. (1984), Controllable and uncontrollable non-locality. In Kamefuchi et al. (ed.), *Proceedings of the International Symposium: Foundations of Quantum Mechanics in the Light of New Technology*, pp. 225–30. Tokyo: Physical Society of Japan.

Skyrms, B. (1984), EPR: Lessons for metaphysics. In P. A. French, and Uehling, T. E. Jr (eds.), *Causation and Causal Theories. Midwest Studies in Philosophy,* vol 9, pp. 245–255. Minneapolis: University of Minnesota Press.

Sober, E. (2001), Venetian sea levels, British bread prices and the principle of the common cause, *British Journal for the Philosophy of Science*, 52, 1–16.

Spirtes, P., Glymour, C., and Scheines, R. (eds.), (2000 [1993]), *Causation, Prediction and Search*, 2nd edition, Cambridge, MA: MIT Press.

Steel, D. (2005), Indeterminism and the Causal Markov condition, *British Journal for the Philosophy of Science*, 56, 3–26.

Steel, D. (2006), Comment on Hausman and Woodward on the Causal Markov condition, *British Journal for the Philosophy of Science*, 57, 219–231.

Suárez, M. (2007), Causal inference in quantum mechanics: A reassessment". In Russo, F., and Williamson, J. (eds.), *Causality and Probability in the Sciences*, pp. 65–106. London: College Publications.

Suppes, P. and Zanotti, M. (1981), When are probabilistic explanations possible? *Synthese*, 48, 191–199.

van Fraassen, B. C. (1982), The Charybdis of Realism: Epistemological Implications of Bell's Inequality. *Synthese*, 52, 25–38. Reprinted with corrections in J. Cushing and McMullin, E (eds.), *Philosophical Consequences of Quantum Theory*, Notre Dame: University of Notre Dame Press, 1989.

von Neumann, J. (1955), *Mathematical Foundations of Quantum Mechanics*. Princeton: Princeton University Press. First published in German in 1932 as *Mathematische Grundlagen der Quantenmechanik*, Berlin: Springer.

Williamson, J. (2005) *Bayesian Nets and Causality: Philosophical and Computational Foundations*. Oxford: Oxford University Press.

Part III
Propensities

Chapter 9
Do Dispositions and Propensities Have a Role in the Ontology of Quantum Mechanics? Some Critical Remarks

Mauro Dorato

9.1 Dispositions and the Interpretive Task of Quantum Mechanics

In trying to understand the role of propensities or dispositions, if any, in the interpretations of quantum mechanics (henceforth QM), I think that one can do no better than start from a fundamental question once posed by John S. Bell: 'What are quantum probabilities *probabilities of?*'

As I see it, this question addresses two deeply related issues, both of which are relevant to evaluate the role of dispositions in QM. The first is an *ontological* question, namely an attempt to connect the formal structure of quantum theory with entities in the physical world, in order to try to figure out *what the theory is about*. I take it that, in all generality, interpreting the mathematical formalism featuring in physical theories ought to mean:

1.1 understanding the ontological implications of physical theories ('*the scientific image*');
1.2 connecting the postulated ontology with our pre-theoretical experience of the world ('*the manifest image*').[1]

In the case of QM, however, such an interpretive task is complicated by the fact that there is no agreed-upon 'theory', except operationally of course, or, in Bell's words (Bell, 1990) 'For All Practical Purposes' (FAPP). Therefore, the interpretive task of QM cannot consist, contrary to what it has been often maintained, in figuring out 'what the world must be like if quantum mechanics accurately describes it' (van Fraassen, 1981, 230; Hughes, 1989, 296; Healey, 1989, 7), because we don't know what 'quantum mechanics' *is* without an explicit interpretation in the two senses above. For instance, according to some interpretations, QM should

M. Dorato (✉)
Department of Philosophy, University of Rome 3, Viale Ostiense 234, 00144 Rome, Italy
e-mail: dorato@uniroma3.it

[1] The distinction between scientific and manifest image of the world is Sellars' (1963).

M. Suárez (ed.), *Probabilities, Causes and Propensities in Physics*, Synthese Library 347, DOI 10.1007/978-1-4020-9904-5_9,

be supplemented with a genuine process of collapse of the wave function, while according to others it shouldn't. For my purpose, it follows that the question of discussing the role of propensities or dispositional properties in QM can only have *interpretation-dependent results*.

The second issue raised by Bell's question above has to do with the meaning of the notion of *probability*, a question upon which the philosophical and scientific community so far has reached no agreement (see Hájek, 2010). Are quantum mechanical probabilities – that figure so prominently in the theory – to be regarded as frequencies, propensities, or simply epistemic states of subjects, as in Bayesian accounts?

Given what I said above, it should be clear why having a clear answer to the first issue is essential to figure out a response to the second: as is well known, if one adopts a Bohmian interpretation, probabilities may be regarded as merely epistemic, or due to our ignorance of the positions of the particles, while in collapse theories, or in the Copenhagen interpretation, probabilities are typically regarded as frequencies, chances, or *propensities*, i.e., objective properties or powers of individual states or events of the physical world. Consequently, in what follows, I will dedicate more attention to explore the first issue by defending the following two claims:

(i) In dynamical reduction models *à la* Ghirardi, Rimini and Weber (1986), propensities or dispositions *might* have a role, despite their (temporary?) irreducibility to non-dispositional, categorical properties;
(ii) In no-collapse interpretations, dispositions are *dispensable*: they are either reducible (as in Bohmian mechanics), or their ascription amounts to a mere 're-labelling' of the predictive content of the wave function (Bohr, Heisenberg).

The 'might' in (i) can be interpreted as a concession to prudence, and therefore can be read in a conditional form: *if* there are dispositions in the quantum world, *then* they are at home in collapse theories, and in particular in the dynamical models proposed by GRW. In this paper, I am not trying to argue that GRW type of theories *require* dispositions. For an unconditional defense of this claim, see Dorato, Esfeld (2010). However, I am claiming that a dispositional reading of a particular version of GRW provides an interesting alternative to the so-called 'flash ontology' presented in Tumulka (2006a, 2006b) and in Allori, Goldstein, Tumulka, and Zanghì (2008). Furthermore, the skeptical conclusion in (ii) does not prevent the fact that Bohr's interpretation can be made much clearer by an appeal to dispositions, especially in order to make sense of his somewhat obscure appeal to 'mutually exclusive and jointly exhaustive properties'.

The plan of the paper is as follows. Since a basic question posed by the attempt to introduce dispositions in QM is to clarify the very meaning of the concept of 'dispositional properties', in the *first* section I will briefly review some of the main problems in the metaphysical literature on dispositions, in order to show that the distinction between *occurrent* (i.e., non-dispositional) and *dispositional* properties is not at all clear and sharp. The fact that in ordinary language no clear demarcation criterion is available seems to push the philosopher of QM in two opposite directions: either *all* properties are to be treated as dispositional also in QM – sure-fire dispositions and propensities alike, as maintained by Suárez (2004b) –, or a clear criterion can be found only in the particular context of QM.

In order to justify the legitimacy of linking the metaphysical notion of 'disposition' with the formal structure of QM, in the *second* section I will first discuss Clifton and Pagonis' (1995) proposal to regard the dispositional properties as the contextual properties, and then advance my own view. In the third section, I will review recent *relativistic* extensions of GRW type of dynamic reduction theories, mainly due to Tumulka (2006a,b), and show how the so-called 'flashes ontology' of GRW type of dynamical reduction models could be supplemented by an ontology of irreducibly probabilistic dispositions. In the *fourth*, final section, I will discuss the sense in which non-collapse views might be interpreted in a dispositional fashion, and will conclude by briefly discussing the *selection approach* to QM advocated by Suárez (2004a,b). Suárez's approach deserves in fact a special discussion, as he claims that the passage from the possession of a purely dispositional property (like spin or position) to the manifestation of such a disposition in a measurement setting *is* a real physical process. And yet his theory cannot be classified among the genuine dynamical reduction models, as he does not provide any detailed physical story about how such a process should occur (the *when?, how? and where?* questions that a model like GRW's tries to tackle).

9.2 Is the Distinction between Dispositional and Non-Dispositional Properties *Genuine*?

First of all, and in order to fix terminology, I should state at the outset that in the context of QM 'dispositions' or 'tendencies' are to be interpreted as *qualitative*, *intrinsic* properties of physical systems. *Propensities* are to be regarded as probabilistic, *quantitative measures* of the dispositions that single systems might have, say, to localize in a region of space, as in certain dynamical reduction models.

Both physics and ordinary language are replete with what philosophers call dispositional properties (in short dispositions): think of the paradigmatic cases of 'fragility' 'permeability' or 'irritability'. However, it is much more difficult to characterize the feature of such properties that distinguishes them from non-dispositional, or categorically possessed properties. And yet, a minimal success in this enterprise seems important to all interpretive projects trying to establish some role for dispositions in QM. *If we were not able to distinguish dispositional from non-dispositional properties even in ordinary language, what would we gain by introducing dispositions in the philosophy of QM?* For example, if we had to conclude that, from a general metaphysical viewpoint, all properties, physical and non-physical, turned out to be dispositional,[2] referring to dispositions in QM

[2] As in theories in which properties are regarded as the causal powers of the entities having them (Shoemaker, 1984).

would either be empty (one could simply talk about properties *tout court*) or would deprive the philosophy of quantum mechanics of any vital contact with more general metaphysical issues.[3]

A first attempt at distinguishing the categorical from the dispositional might be suggested by the fact that dispositions typically have a context of manifestation ('glass is fragile because in certain situations it breaks easily'), something pushing us toward the claim that dispositions might be *relational* properties, i.e., properties that are non-intrinsic or extrinsic in Langton and Lewis' sense (1998).[4] However, the attempt of drawing the distinction between the dispositional and the non dispositional in terms of 'intrinsically possessed' vs. relational or extrinsic *fails*: I agree with various scholars that a window pane would count as fragile independently of any breaking context, and even if it will never break (see Mumford, 1998; Suárez, 2004b). So not only is relationality not necessary, but also not sufficient for dispositionality, since 'being a brother' clearly does not count as a *prima facie* dispositional property. The fact that there are some apparently *extrinsic* dispositions, like 'weight', should not prevent us from acknowledging that at least some dispositions look intrinsic and non-relational.

This remark seems to be relevant also for the philosophy of QM, as Popper's relationalism (1982) about dispositions notoriously led him to interpret quantum dispositions as *relational properties of quantum entities*, linking them to the whole experimental setup. According to Popper, who was obviously influenced by Bohr's thesis about the non-separability between quantum entities and classical apparatuses, an isolated particle would have no dispositions whatsoever. It seems plausible to maintain that while the context of manifestation of fragility or permeability is the necessary epistemic ground to believe in the existence of the relevant disposition, a piece of glass would count as fragile even if it never broke, i.e., even if it never manifested its disposition.

All this is well-known and basically agreed upon. However, it could be objected that in a different possible world, made just of glass and liquid stuff that cannot be accelerated beyond the speed that would be sufficient to break a pane of glass,[5] *glass would not count as fragile* because no harder stuff would be present to break it. Wouldn't this show that there is a certain amount of relationality in the property in question, as in all properties? In this case, fragility might seem to depend on the fact that the laws of nature in the 'liquid world' prevent that liquid stuff from being accelerated beyond a certain speed. And if laws are dependent on local facts in this

---

[3] The reason for this second alternative is that it might be possible to define the difference between the dispositional and the categorical just in terms of the formalism of QM, while admitting that such a distinction has no application elsewhere. However, this option would raise some doubts about the faithfulness of the explication of the word 'disposition', since part of its intuitive meaning would be lost.

[4] A property is intrinsic when its attribution to an entity x does not presuppose the existence of any other entity. It is extrinsic or relational when it is not intrinsic.

[5] This is added so as to prevent that the impact between glass and very fast-moving blobs of water could break the glass.

world, as Lewis's Best System Analysis has it, there would still be a dependency of fragility on what else is occurring in the world and on the presence of harder stuff. And fragile objects would seem to be possessing their dispositional property in an extrinsic way. On the other hand, *if* fragility were regarded as a microscopic property of glass – that is, if it were reducible to, or fully explainable in terms of – the microscopic structure and forces of the crystals composing glass, *then* it would seem that the disposition in question could be ascribed to glass as one of its intrinsic properties. But the meaning of the term, in the different possible world made just of glass and liquid stuff in which glass never breaks, would be quite different.

Be that as it may, the possible dependence of (the ascription of the property) 'fragility' on the properties of other materials in the environment simply shows that analyzing the distinction between dispositional and non-dispositional in terms of the distinction between *extrinsic* and *intrinsic* carries the additional risk of attributing any vagueness of the latter distinction to the former. And this is an additional reason against accepting the above analysis.

Another possible attempt at distinguishing dispositional from non-dispositional properties might consist in the fact that the former could be regarded as directly observable only in the context of manifestation, while the latter would be *always* observable. 'Fragility' might be the *intrinsically possessed* property of glass that becomes manifest or directly observable *only* when a piece of glass breaks, while a broken window pane displays the corresponding property at all times, and would therefore be always observable. Analogously, a disposition like *permeability*, unlike the property expressed by 'being wet', is *not* directly observable all the times, but becomes observable only when the entity exemplifying it interacts with water or other fluids.[6]

But also this attempt at distinguishing dispositional from categorical properties fails: the earth's gravitational field manifests the disposition to attract bodies toward the ground at all times, and not just when we observe its manifestation in falling bodies. By exerting an attracting force, the earth infact keeps any object firmly attached to the ground *at all times*.[7] Furthermore, to the extent that fragility and permeability are regarded as being identical with the microscopic, molecular structure of glasses and sponges respectively, one could note that such a structure can be considered to be observable at all times, albeit indirectly with the aid of electronic microscopes. After all, don't we observe through a microscope?

In a word, also this second criterion does not secure any firm ground, and fails.

For a more fruitful attempt at indicating the distinction in question, we could look at the *role* performed in ordinary language by obviously dispositional terms. Consider dispositions like 'irritable' or 'poisonous', which manifest themselves

---

[6] It is possibly for this reason that Carnap argued that dispositional predicates were intermediate between theoretical and observational terms (Carnap, 1936).

[7] After general relativity, we may need to redescribe the situation by saying that the surface of the Earth constantly manifests its disposition to not be penetrated, by pushing objects out of their free fall. Thanks to Carl Hoefer for reminding me the need to take into account the post-Newtonian paradigm of gravitation.

when people get angry and, say, mushrooms poison the blood. From these examples, it would seem that the function of dispositional terms in natural languages is to encode useful information about the way objects around us *would* behave, *were* they subject to specific causal interactions with other entities (often ourselves). This remark shows that *the function of dispositional predicates in ordinary language is essentially predictive.* Consider the evolutionary advantage of classing all animals or people around our ancestors as 'dangerous' or 'innocuous', as 'peaceful' or 'ferocious'. In learning that a particular mushroom is 'poisonous', a child learning the language also learns to stay away from it whenever she recognizes one.

Clearly, this analysis of the distinction between dispositional and categorical properties can be correct only if it can be shown that *prima facie* examples of categorical terms, like 'is broken' (*vs.* 'is fragile') or 'is dissolved in water' (*vs.* 'is soluble') do *not* have a similar predictive function. And it seems to me that a distinction 'of degree' between the categorical and the dispositional can be traced in such cases. I say 'of degree', because any attribution of a property to an entity involves a certain amount of predictability, even if one does not accept Sellar's and Brandom's inferential theory of meaning (Brandom, 1994). If we know that 'salt has dissolved in water', of course we also know a good amount of things about salted water (a categorical state/property of the liquid),[8] and this might be true simply in virtue of the fact that *properties just are the causal powers of entities*. In this case, however, any clear-cut distinction between dispositional properties (powers) and non-dispositional properties is also dissolved.

Despite this remark, I think that it is still fair to say that in dispositional terms the predictive role is much more explicit or evident, a fact which could explain why natural languages and especially folk psychology, are so replete with predicates like jealous, amiable, and peaceful, etc. I say 'more evident' because of the well-known link of dispositions with the modal talk presupposed by causes, counterfactuals and laws. A stone *causes* the manifestation of the disposition 'fragile' (and therefore causes the breaking of the glass) because it causally interacts with its microscopic structure, the categorical basis of fragility itself. Counterfactuality is involved because the attribution of the disposition 'soluble' to salt entails that in the appropriate context, salt melts, while the regularity with which the fragility of glasses is manifested refers to a regularity or a law of nature capturing the behaviour of the micro-constituents of glass.

In a word, the fact that we cannot analyze dispositions by using conditionals does not prevent us from advancing the following claim: *dispositions express and encode, directly or indirectly, those regularities of the world around us that enable us to predict the future.* If this is their main role and function in ordinary language, we understand why their distinction from intuitively non-dispositional properties is just one of degree.

That the distinction between dispositions and categorical properties cannot be so sharp is further confirmed by Mumford's analysis of the problem of the reducibility

[8] Likewise, if the window is broken, we know we shouldn't walk bare foot in that area.

of dispositions to their so-called ‘categorical basis’. According to Mumford (1998), the difference between a dispositional property like fragility and the microscopic property of glass constituting its categorical basis is merely *linguistic*, and not *ontological*. Referring to a property by using a dispositional term, or by choosing its categorical-basis terms, depends on whether we want to focus on, respectively, the functional role of the property (the causal network with which it is connected), or the particular way in which that role is implemented or realized.

But notice that if we agree with Mumford’s analysis, it follows that it makes little sense to introduce irreducible *quantum dispositions* as *ontological* hypotheses. If, by hypothesis, no categorical basis were available, we should admit that we don’t not know what we are talking about when we talk the dispositional language in QM, quite unlike the cases in which we refer to ‘fragility’ or ‘transparency’, in which the categorical bases are available and well-known. Introducing irreducible quantum dispositions would simply be a black-box way of referring to the functional role of the corresponding property, i.e., to its predictive function in the causal network of events.

The upshot of this brief survey on the metaphysics of disposition should be clear. The predictive function of dispositions illustrated above – as well as Mumford’s view about the conceptual, non-ontic distinction between the categorical and the dispositional – should be attentively kept in mind when we will discuss the ‘dispositional nature’ of microsystems before measurement, or the irreducible ‘dispositions to localize’ possessed by microsystems in dynamical reduction models.

In a word, the use of the language of ‘dispositions’ by itself does not point to a clear ontology underlying the observable phenomena. On the contrary, when the dispositions in question are irreducible and their categorical bases are unknown, such a use should be regarded as a shorthand to *refer to the regularity that phenomena manifest and that allow for a probabilistic prediction*. Consequently, attributing physical systems *irreducible* dispositions may just result in a more or less covert *instrumentalism*, unless the process that transforms a dispositional property into a categorically possessed one is explained in sufficient detail.

In a word, friends of dispositions might end be up using an elaborate or fancy metaphysical language to redescribe measurement interactions, especially if they are not ready to provide a precise, exact physical theory about when and how a dispositional property corresponding to a state of a system which is not an eigenstate of the observable turns into a categorically possessed property. As we will see, this is the main difficulty with Bohr’s philosophy interpreted in a dispositional way, or with Suárez’s otherwise brilliant attempt at using dispositions to solve the measurement problem (2004b).

9.3 Dispositions and Categorical Properties in QM

The history of the attempts at introducing dispositions in QM is long and significant (Margenau, 1954; Heisenberg, 1958; Redhead, 1987; Maxwell, 1988), but here I will discuss only the two most recent attempts at linking the language of dispositions to the formalism of QM.[9] The first is due to Clifton and Pagonis (1995) and links dispositionality to *contextuality*. The second is due to Suárez (2004b), and relates dispositions to a *selective interpretation* of QM. If one is careful enough to avoid some misleading associations of the word 'contextual' with 'relational', I think that both are perfectly compatible with each other and with my own view.

As I see the matter, the introduction of a dispositional language in QM is based on the replacement of 'dispositional properties' with 'intrinsically indefinite properties', i.e. properties that before measurement are objectively and actually 'indefinite' (that is, without a precise, possessed value). The following two postulates express what I regard as the essential tenets of a dispositionalist approach to the interpretation of QM, and specify the *meaning* of a dispositional property in the context of the formalism QM:

P1 a property of a system describable by QM is categorically possessed if and only if the state it corresponds to is an eigenstate of the observable. Otherwise it is dispositional. In this sense, mass, charge, spin are to be regarded as intrinsic, categorically possessed properties, since they are always definite.[10]

P2 the passage from dispositional to non-dispositional magnitudes is the passage from the indefiniteness to the definiteness of the relevant properties, due to in-principle *describable*, genuine physical interactions of quantum systems with other systems, that typically possess a much larger number of particles.

P2 in particular refers to the process that transforms an objectively, mind-independently indefinite magnitude into a definite one, and allows me to link the manifestation of dispositions with precisely described measurement interactions between quantum systems and larger physical systems.

This sense of dispositionality will be adopted here in order to interpret the GRW dynamical reduction models from a metaphysical viewpoint, and seems quite close to what Heisenberg had in mind in the following, often quoted passage, which refers to the well-known thesis that QM reintroduces Aristotelian *potentiae* as intermediate 'between full being and nothingness': 'Therefore, the transition from the 'possible' [dispositional] to the 'actual' [categorical] takes place during the act of observation [a correlation] ... we may say that the transition from the 'possible' to the 'actual' takes place as soon as the interaction of the object with the measuring device, and thereby with the rest of the world, has come into play; it is not connected with the act of registration of the result by the mind of the observer.' (Heisenberg, 1958, 54).[11]

The best way to justify these two postulates, and especially the second, is by briefly reviewing the interpretive proposals offered by Clifton and Pagonis and Suárez.

[9] For a review, see Dorato (2007) and Suárez (2007).

[10] Here I respect the standard eigenvalue-eigenvector link.

[11] Words in square brackets have been added by me and reformulate Heisenberg's language by using the key terms adopted here.

9.3.1 Clifton and Pagonis on Dispositionality as Contextualism

P1 is equivalent to Clifton and Pagonis' strong contextuality. The idea of contextuality is simple. Assign a certain value to the square of the operator 'spin in the *z* direction' – call it $S^2{}_z$ – when it is measured together with $S^2{}_x$ and $S^2{}_y$ in the direction *x* and *y*. If $S^2{}_z$ is *not* contextual, we must get the same value if we measure it together with $S^2{}_{x'}$ and $S^2{}_{y'}$, assuming that the direction *x'* and y' are *different* from *x* and *y*. Contextualism is quite widespread a phenomenon in QM, and it seems to entail that some QM 'properties' are not sharply possessed before measurement, since otherwise they could *not* manifest themselves in different ways, according to the type of measurement we perform.

Consequently, in QM we seem to have *two* kinds of intrinsically possessed properties, depending on the way the system has been prepared before measurement. If the system has a definite value also before measurement and the latter just reveals it with probability 1, we have either a non-dispositional, categorically possessed property, (a weakly contextual property), or we have what we could call a deterministic, 'sure-fire' disposition (Suárez, 2007). On the contrary, if the value revealed by measurement causally depends on the interaction, we have a *strong* form of contextualism that, according to Clifton and Pagonis, implies the presence of intrinsic dispositions (Clifton and Pagonis, 1995, 283), or simply probabilistic dispositions, i.e., propensities.

This aspect of Clifton and Pagonis' approach is quite similar to my first postulate above; put it in different words, we could also express the identification of the dispositional with the contextual by noting that in QM we cannot assume that there is a one-to one correspondence between an operator and an observable: contextualism or dispositionalism as expressed in P1 bans a certain form of 'naïve realism about operators' (Daumer et al., 1996).

The proposal to establish a significant link between contextuality *and* dispositionality is open to two objections, which, in my opinion can both be tackled.

The first objection is based on the fact that contextuality itself has recently been the target of various critical remarks, especially by philosophers working in the bohmian group. They claim that, after all, it is quite trivial that if we perform different measurements, we are going to obtain different results (see Goldstein, 2009, Section 12). Why make such a fuss about contextualism? If this objection were correct, however, also dispositionalism as defined by Clifton and Pagonis would be a trifling matter, since it is *defined* in terms of the former notion. The second objection points to the fact that contextualism entails a kind of extrinsic-ness or relationality of dispositions, a position that we have already rejected. Let us analyze these two objections in turns.

I disagree with Goldstein's opinion for two different reasons. First, contextuality has an important role as a premise of fundamental no-go theorems against the possibility of assigning simultaneously definite values to systems whose dimension is greater than 2 (Kochen and Specker, 1967). And these theorems seem certainly important contributions to our understanding not just of the formal structure of QM, but also of the possibility of simultaneously attributing well-definite properties to certain quantum systems, which is part and parcel of the interpretive task spelled out in Section 1. The fact that we can deny the significance of these theorems for QM only by endorsing contextuality/dispositionality (alternatively, by denying non-contextuality) is *not* a trifling matter. In classical physics, measurements typically *do* reveal pre-existing properties, and the fact that in quantum systems before measurement one cannot rely on categorically possessed properties in the sense given above by P1) and P2) cannot be deemed as being without significance.

The second counter-objection is that the superposition principle is *the* distinguishing mark of QM with respect to classical mechanics (Dirac, 1930), and superpositions are not ignorance interpretable. It follows that the passage from 'the dispositional' to 'the actual/categorically possessed' is a very important feature of QM, because it is the passage that takes us from a superposition of states to one particular state with a certain probability. Such a passage is obviously involved in the process of measuring a quantum entity in a superposed state, which is arguably *the* central aspect of the theory that still needs to be explained.

Going now to the second objection, what is instead potentially misleading about identifying dispositionalism with contextualism is the fact that contextual properties seem to have been identified with extrinsic, relational properties, contrary to what was argued in Section 9.2. If the value of the spin measured on S_x^2 depends on whether we observe it with S_y^2 or $S_{y'}^2$, there seems to be a clear sense in which not only is the spin in question not possessed before measurement, but the disposition itself (i.e., the property of manifesting a definite spin in a given direction) depends on other properties of other entities. However, we should notice that it is the *manifestation* of the spin that is relational, not the disposition itself, which is as intrinsic at it may be.

A simple example, made by Albert (1992), will help us to make the point. If we reverse the polarity of a magnet of a Stern-Gerlach apparatus, and measure the spin of a particle that is in a superposition of spin in the *z* direction, we obtain a result that is opposite to what we obtained before the reversal. The very same intrinsically possessed disposition can, depending on what measurement we perform, be manifested in different ways, for the simple reason that there is no preexisting definitely possessed property of having spin in the *z* direction. So there is a legitimate way to defend the (correct) view that dispositionality is as intrinsic as it may be.

9.3.2 Suárez on Dispositions

Suárez construes dispositions in QM in a similar manner, but links his understanding of dispositions to Fine's proof of the unsolvability of the measurement problem (Suárez, 2004b). According to Suárez, dispositions are possessed by quantum systems all the time, even when they are not manifest, and this agrees with the previous point that dispositions are *intrinsic* properties of quantum systems. Importantly, selections are a subclass of measurements, since while the latter are interactions with *all* the properties of a quantum system, selections pick out just *one* of the many intrinsic dispositions of quantum systems.

There is one minor difference between Suárez's view and my definition P1, as in a recent paper he introduces sure-fire dispositions: '*If* object O possesses the deterministic propensity D with manifestation M *then*: were O to be tested (under the appropriate circumstances $C_1, C_2, \ldots$ etc) it would definitely M with probability one.' (2007, p. 429).[12] This would entail that preparing a system in state ψ and measuring it afterwards would count as measuring a sure-fire disposition rather than a categorical property.

Although this difference might seem purely terminological, I prefer to refer to use 'disposition' for properties of entities whose state is not an eigenstate of the observable, i.e., for magnitudes that are not sharply possessed. First of all, my choice helps to focus on 'the collapse' of a state ψ in superposition as a transition from indefiniteness to definiteness of magnitudes, rather than as a transition from propensities to sure-fire dispositions. Furthermore, in bohmian mechanics the particles' positions would count as dispositional properties in Suárez view ('a sure-fire disposition'), and their difference with spin would be specifiable simply in terms of a probabilistic rather than a deterministic descriptions. On the contrary, if we accept the view that we have dispositions relative to observable O whenever we do not have previous values for $O = \sum_n a_n \,|\, v_n\rangle\langle v_n \,|$, the interesting question will of course become whether dispositions are *reducible* to some kind of categorical basis, as it is the case with Bohmian mechanics, or are not so reducible, as it is in the case in other interpretations to be discussed.

There is one last objection that we must discuss before broaching the GRW's ontology from a dispositionalist viewpoint: what sense does it make to claim that a system has a dispositional property when it lacks a precise value for the corresponding quantity? Shouldn't we say that when a physical system lacks a precise value before measurement, not only is there no corresponding categorical property, as it is obvious, but also that there is no dispositional property either? Shouldn't we say that talking of properties, *even if* dispositional ones, is made obsolete by QM's contextuality, and that we should not pour old metaphysical wine in the new barrels provided by mathematical physics? (Daumer et al., 1996).

Well, claiming that a quantum system in a superposition of spin has a disposition for acquiring a precise spin even when if it has no precise spin is at the heart of a dispositional reading of QM. But we should admit that claiming that a system has no precisely possessed property at all before measurement, and that it possesses a disposition to manifest a definite property after a correlation with a larger system, should not be regarded as two *perfectly equivalent* ways of speaking. In the latter description, we are finding a request of explaining something that we still don't understand in detail, namely the existence of a genuine transition from an indefiniteness to a definiteness that is to be regarded as a real physical process, and which should be further studied with the experimental and technical resources of physics.

Accepting the claim that a quantum entity is to be (currently) regarded as a node of dispositions is not a crazy idea, as long as this way of speaking presents some

[12] I thank Suárez for having sent me his manuscript.

advantages relatively to the other, non-property talk. But what kind of advantage can it be, considering that dispositions, typically, don't explain much? Are we back to the *virtus dormitiva* explanation? This is what we will have to inquire in the next section.

9.4 Adding Dispositions and Propensities to GRW

I should make clear from the start that GRW's dynamical reduction models do *not* explicitly rely on dispositions. However, neither do they exclude their existence. In order to see what we could gain by introducing a dispositional language, I will therefore try to summarize the main features of the best known dynamical reduction models by relying on a language introducing irreducible propensities.

According to one of the non-relativistic versions of the dynamical reduction models proposed by GRW (Ghirardi et al., 1986), each non-massless micro-system whose wave function has a certain spatial spread has *an irreducible probabilistic disposition* – a *propensity* – to localize in a region of space given by a diameter of $\sigma = 10^{-5}$cm, in average once every 10^{16} s. The probability of a decay per second is therefore $1/\tau = 1/10^{16}$. These two parameters become two new constants of nature, and specify *to what* and *how often* the localization process occurs.

Obviously, if the system is composed only by one particle, this can remain unlocalized in average for 100 million years (approximately corresponding to 10^{16} s), but since the propensity for a localization is defined with a Poisson distribution such that the probability for the localization of N non-massless particles is N/τ, a system made of $N = 10^{23}$ particles will undergo in average 10^7 localizations per second, and will therefore remain in a dispositional, superposed state for less than 10^{-7} s. Accordingly, in a cubic centimetre, there are approximately $10^7 = 10^{23}/10^{16}$ events of localization, *or 'flashes',* which ensure and explain the localization of the *macroscopic* objects with which we are familiar from our experience.

The localization of a whole system is a consequence of the fact that even if a system is in a dispositional, superposed state, the multiplication of the wave function by a Gaussian localization operator ('the hit') effectively kills the other components of the superposition that are not located close enough to the center of localization.

Analogously, in the relativistic extension of the GRW theory, due to Tumulka (2006a, b), we are given a set of localization events (flashes) and a rule for calculating the probability for the next flash to occur as specified by the wave function.[13] Here is how J. S. Bell summarized this flashy view of the physical universe: 'However, the GRW jumps (which are part of the wave function, not something else) are well localized in ordinary space. Indeed each is centred on a particular spacetime point (*x, t*). So we can propose these events as the basis of the 'local

[13] Interestingly, in this flash model there is no need of postulating a privileged frame for the localization, as it happens with the model in which the mass-density localize. Still, the model suffers from other difficulties on which here I cannot enter.

beables' of the theory. These are the mathematical counterparts in the theory to real events at definite places and times in the real world .... *A piece of matter then is a galaxy of such events*.' (Bell, 1987, 205).

In a different model of the theory, the fundamental entity is a scalar field $\rho = \rho(\boldsymbol{r}, t)$ defined on Newtonian spacetime, with ρ being, at the macroscopic scale, what we call *mass density* of physical objects. In this interpretation, the wave function $\psi(r_1, \dots r_n, t)$ describes the system at a given time, and the square modulus of ψ determines, for each particles i, how much stuff (ρ_i) there is in a given cell: $\rho(r, t) = \sum_i m_i \rho_i(r, t)$. Even though the density of microscopic objects can be in a superposed, dispositional state and therefore enjoy 'the cloudiness of waves', due to the localization mechanisms the mass density of macroscopic objects acquires a precise value in a split second, and the object localizes somewhere via an irreducibly stochastic event.

We should notice that while the wave function leaves in an abstract $3N$ dimensional space, the flashes and the scalar field are both in spacetime, since they are localized wherever the collapse events occur. Important for the purpose of introducing irreducible propensities is the remark that the time and place of the localization processes (their center of collapse), as well as the particular particle or cell that is involved, are chosen at random, and so the localizations themselves are to be regarded as 'spontaneous', or simply *uncaused*. The crucial question at this point is: if this is correct and intended in the model, why introducing dispositions or single-case propensities, *if* the latter are regarded as *causes* of the localizations that are 'inherent' in each microsystem?

First of all, propensities need not be presented as *causes* of the localization process, since we cannot rule out that the collapse be 'spontaneous', or uncaused. Admittedly, there is nothing in the formalism of GRW suggesting this reading, and if the theory remains 'phenomenological' as it is now, the propensity theorist is happy to accept that the real tendency that each single microsystem has to localize is irreducible, but still needed to attribute *single case probabilities* to individual particles. In this hypothesis, it would make sense to say that a universe composed by a single proton would harbour a particle with a disposition to localize once every 100 million years: no reference to ensembles of particles would be possible and therefore no frequencies. In our universe, frequencies would simply be supervenient on, and a manifestation of, such individually possessed propensities.

Frigg and Hoefer (2007), however, have argued that a Humean Best System (HBS) analysis of GRW's probabilities is more plausible than a dispositionalist analysis, and that single case propensities are not needed to defend the view that the probabilities introduced by GRW are as objective and as non-epistemic as it gets. (Frigg and Hoefer, 2007). After all, HBS' *chances* are based on all the local facts in the universe's entire history, and are therefore not to be conflated with subjective degrees of belief as in Bayesian probabilities. Since HBS chances have a factual grounding, they must be regarded as 'objective'. Only, these local facts in the universe history are to be conceived non-dispositionally; probabilities, consequently, are not grounded in modally conceived propensities or powers.

However, it seems to me that the main weakness of a HBS analysis of GRW probabilities depends on its reliance on Lewis' analysis of the nature of laws of nature. If GRW probabilities depend on the probabilistic laws organized within a HBS of the theory, but such laws are, as Lewis has it, supervenient on the local, non-modal facts of the history of the entire universe, how can we make sense of conditional probabilities, which refer to *relations* among state of affairs, and therefore to *universal*? In my view of laws, law-statements are made true by, or more prudently, simply refer to, the dispositions or causal powers possessed by physical systems (Dorato, 2005). From this viewpoint, it is not clear how a HBS reading of the GRW theory can defend an *objectivist* view of probabilities or chances (namely, a non-subjectivist, non-Bayesian approach allowing us to go beyond states of beliefs) without committing itself to some mind-independent *property* or *relations* that microsystems have, and therefore, in a plausible view of properties or relations, *to dispositions or causal powers*.

In Lewis' original idea, what propensity theorists call 'disposition to collapse' really refers instead to *the whole mosaic of local states of affair*, on which collapse laws supervene as axioms or theorems of a single theory combining *simplicity* and *strength*. However, not only are simplicity and strength language-dependent virtues, but they are also intersubjectively shared but *merely epistemic virtues*. Laws in HBS denote nothing but lists or histories of events or occurrent facts, and it is not clear at all whether the Humean mosaic includes or not properties or universals. This alternative generates a dilemma.

If we opt for the former, nominalistic reading of HBS (no property is admitted in the HBS/Lewisian ontology) there are troubles that cannot be overcome. As anticipated before, the probability of the next flash given a set of flashes and the initial wave function is a conditional one, so what we really have is a *relation* between them. Now, how can we claim that this conditional probability describes something in an objective way and is not epistemic if it doesn't refer to such a relation but is simply about a bunch of disconnected, local states of affairs? The question is that if HBS theorists granted that laws in HBS describe relations, they would have thereby overcome nominalism, and therefore one of the main motivations of an HBS' analysis of laws and probabilities.

But perhaps there is nothing inherent in the Lewisian point of view that rules out properties being part of the Humean mosaic, as long as they are conceived as occurrent properties, as opposed to modal ones.[14] Modally conceived properties would infact be dispositions or propensities. And this is the second horn of the dilemma: on a reasonable view of properties, in fact, X is a property of Y if and only if X is a causal power of Y or X is identified by the causal powers of Y (Shoemaker, 1984). And modally conceived properties or propensities have to be readmitted again also by the HBS theorist.

Well, maybe we should not saddle the interpretation of quantum probabilities with complicated metaphysical questions about the identity conditions for

[14] This objection has been voiced by Roman Frigg.

properties. But even if we granted this point, there seems to be another difficulty looming for the HBS reading of the GRW's chances. It is not clear to me how, without asking some help from actual frequentism about actual histories, one is going to distinguish between chancy histories governed by probabilistic laws from 'deterministic histories' governed by sure-fire laws. But since Frigg and Hoefer correctly claim that an appeal to frequentism is a non-starter for GRW, shouldn't they be committed to the existence of propensities in order to make sense of objectivist chances, in the same sense in which Lewis himself is committed to universals in order to add some objectivist constraint to the simplicity and strength of our chosen language? (Lewis, 1983).

Frigg and Hoefer could reply that frequencies *are* part of the Humean mosaic and hence ground probability claims. The difference between HBS and frequentism is that the former position does not assign probabilities *solely* on the basis of frequencies and also takes other epistemic virtues into account (simplicity, strength, etc.). In this way the HBS theorist may drop the notion of a *Kollektiv*, which causes well-known troubles to the von Mises-type frequentist. But this does not mean that the HBS theorist is oblivious of frequencies, or that he needs to resort to propensities.[15]

However, note that virtues like simplicity or strength are *possibly* intersubjectively shared (weakly objective), but are at best a guide to discover mind-independent facts, as they are *epistemic* and language-dependent virtues. In conclusion, the only grip on mind-independent (strongly objective) probabilities that the HBS theorist has is yielded by frequencies: the HBS position then faces a dilemma, since it seems either to collapse on frequentism with all its known problems, or onto *epistemic* views of probability, which are close to subjectivism or bayesianism.

An additional important reason to introduce propensities in GRW should be considered: in the mass density version of the theory, we could try to defend the view that the propensities to localize that each microsystem possesses allow to *explain* the definiteness of the macroworld, in the sense that they allow a *unification of the micro and the macro-world*, characterized by a unique dynamics.[16] And in the flash version of the theory, where *macroscopic* objects *are* collections of 'hits', we could redescribe the ontological assumption of the theory by saying that quantum, non-massless *microscopic* objects whose wave-function has a certain spread in space are a collection of propensities to localize in a small region of space. In both versions, according to GRW, QM is a universal theory, governed by a modified, non-linear Schrödinger's equation. 'Universal' here means that it applies to the micro and macro-world: while single non-massless particles or the microscopic density of stuff may be in superposed states for a long time, despite their propensity to localize, macroscopic bodies are a collection of localization processes.

However, there are two objections to the view that propensities in GRW might explain. One is that all dispositions are in general explanatory empty, the other is that in our particular case the explanatory work is really performed by the actual flashes

[15] This objection was raised by Frigg.

[16] For the theory of explanation as unification, see Friedman (1974), and Kitcher (1976).

or by the localization of mass density, which are *events* or *processes* in spacetime. To the extent that GRW explains by unifying, it is flashes or the localization process of the mass density that ‘unify’, not the propensities to localize, which are unnecessary. Let us quickly analyze these two objections in turn.

The first objection is well-known: do I explain why a piece of glass broke by pointing to its fragility? Well, if I know what fragility refers to (the microscopic structure of certain stuff), I *do* explain why this piece of glass broke by mentioning its fragility, but simply because fragility refers to the structure of glass. However, since the alleged propensity to localize in GRW is *ungrounded* – according to GRW collapses are spontaneous and there is no hidden mechanism for them – how can I claim to explain the localization by adducing an ungrounded disposition? Nevertheless, if I claim that a cloth is impermeable and know nothing about the fabric of the stuff it is made of, there is a sense in which I do explain why I did not get wet, even though for a deeper explanation I need to revert to chemistry. If we agree that explanations have a pragmatic component, and can be regarded as answers to why-questions that depend on the knowledge state of the questioners, why would the piece of information that the coat is impermeable to water fail to provide a *prima facie* explanation for why I did not get wet? If I did not know that the coat was impermeable, coming to know this makes me understand why something occurred. For sure, the kind of information provided by dispositions is *weakly* explanatory, but in some circumstances it can be regarded as providing comprehension.

In the case of the second objection, it must be admitted that the localization process and the propensity to localize are equivalent in terms of unification: we can either describe an object as a swarm of flashes, or depict it as made of particles with a propensity to localize. Equivalently for the mass density version of the theory. The unification is realized in both ways of speaking: if propensities are not indispensable, however, they cannot be ruled out either.

Three final advantages of the propensity talk can be mentioned: if the propensities to localize are metaphysically prior to the localization events, and, contrary to Allori et al. (2008), are ‘metaphysically primitive’, we can presuppose that localization events are something that *occurs* to microsystems in both versions of the theory. And then, by starting with continuants endowed with propensities we might have a better chance to reconstruct a more stable notion of our familiar objects, as bare flashes could not be sufficient[17] (see Frigg and Hoefer, 2007).

Second, if we do not consider the manifestation of the disposition (i.e., the flash itself) as explanatorily ultimate, but leave the room open for a future grounding of the disposition to localize, we may have heuristic reasons to develop GRW, which is still a merely phenomenological theory, into a deeper theory, possibly invoking noise coming from gravitational phenomena (quantum gravity).

[17] This objection is voiced by Frigg and Hoefer (2007), without attempting to counter it.

Third: we do not take into our ontology the configuration space, as Albert and Clifton and Monton did also for GRW: in order to make sense of the role of the wave functions, propensities to localize are enough.[18]

While it must be admitted that none of these three arguments is knock-down, there is little doubt that if we characterize dispositions as we did in the second section, the transition from the indefinite to the definite required by P2 in the present case is illustrated by a theory that is *exact* in the sense of Bell, as it tells us precisely *how often* and *where* the propensity to localize is manifested. It is in this sense that GRW is the best illustration of Heisenberg's idea that QM reintroduces potentiae (which however are to be regarded as real properties) and subsequent transitions to actuality: if we believe that quantum systems before measurements do not have precise values, GRW's postulation of propensities to localize gives us a reason to believe that objects have definite properties when we look at them.

9.5 Dispositions in (some) Non-Collapse Models: Bohr's Interpretation

Despite the fact that trying to figure out what Bohr really thought about QM is a difficult, if not desperate enterprise in the space of a short paper, there seem to be *two* main readings of his approach.

The first comes from a peculiar combination of neopositivist and kantian influences, the second, too often neglected, is based on a dispositionalist reading of his principle of complementarity, to be proposed in the remainder or this section.

The neopositivist strand comes from an application of Einstein's analysis of the notion of simultaneity (1905) within the context of the measurement process of QM. Exactly as, according to Einstein, it is *meaningless* to claim that two events are simultaneous, unless we have specified a particular operational criterion to establish when and in which circumstances two spacelike-related events are to be regarded as 'co-occurring', according to Bohr it is meaningless to attribute a definite property to a quantum system unless we specify a classical measurement context.

Such a first reading of Bohr's understanding of a quantum system before measurement is authoritatively preferred, for example, by Michael Redhead (1987, 49–51). Jan Faye, stressing as he does that Bohr was an entity realist while accepting a form of antirealism about QM as a theory[19] (Faye, 1991), could certainly concur with the view that according to the Danish physicist it is meaningless to attribute before, and independently of, measurement *any kind of properties to quantum systems*.

The Kantian strand of this first reading comes from the possibility of considering the classical language with which we describe the measurement apparatuses as

[18] A similar point has been advocated by Suárez for Bohm's ontology (2007).

[19] This means that the wave function for him was simply a bookkeeping devise good for predictions, but theoretical entities existed in a mind independent fashion

a transcendental condition of possibility to refer to the quantum, noumenal world. Notice that classical apparatuses and quantum entities for Bohr are inseparable, due to the non-divisible nature of the quantum of action that is exchanged between the two. Now, if we really wanted to develop an analogy with Kant's theory of knowledge, we should say that the Kantian categories and the pure forms of intuition are to phenomena of the outer world like the classical apparatuses are to the quantum world. Exactly as the 'phenomena' for Kant are the way in which the noumenal world *appears* to minds endowed with pure forms and categories like ours, the manifestation of the quantum world *via* the choice of a classically describable apparatus must be regarded as an inextricable relation between the noumenal (an Sich) unknowable quantum world and a non-quantum, classical measurement system.

Such a Kantian strand can somehow introduce the second dispositional reading of Bohr that I want to broach now and that I prefer. According to Bohr, two properties are complementary if and only if they are *mutually exclusive* and *jointly exhaustive* (see Murdoch, 1987). I take that this slogan is a central part of Bohr's interpretation of QM. We say that they are *mutually exclusive* because, from the point of view of the classical language, they can be attributed to the same system at the same time only via a contradiction: in classical terms, nothing can be both a particle and wave (if we regard 'having position' and 'being a wave' as referring to *categorical* properties).

However, from a *dispositional* point of view, if we refer to a *quantum* entity, this duality is perfectly legitimate, because we can attribute the *same* particle at the *same* time (i.e., before measurement) a disposition for a particle-like behaviour and a disposition for a wave-like behaviour. Such dispositions are later *selected* by the kind of experiment we wish to perform. The choice of this word 'selection' in not casual, as Suárez bases his dispositionalist approach to QM on the view that measurements chooses or selects a particular, intrinsic disposition of the quantum entity (Suárez, 2004b). This shows, by the way, that his view is *not* at all too distant from Bohr's as I presented it here.

The presence of two dispositions is the reason why in a double-split experiment, complementary properties like the trajectory of the particle (its position) and the interference pattern cannot be simultaneously revealed by the same experiment, given that any apparatus obeys classical physics. Either we know the split through which the particle went, but then we destroy the pattern of interference, or we save the interference, but then we cannot know where the particle went.

On the other hand, if we refer to a quantum system *before* measurement, the complementary properties must be regarded as *jointly exhaustive*, because any attempt at attributing a not-yet measured system only *one* of the two properties would yield an incomplete description of the quantum system. In a word, *an electron is neither a particle nor a wave, but has dispositional features belonging to both concepts*.

Despite lack of direct evidence for the interpretation of Bohr that I am suggesting, I think that it is not absurd to attribute to an entity realist like Bohr the view that microsystems have real tendencies to display well-defined measurement values in a given experimental context, that somehow 'extract' some 'latent aspect' from a mind-independent entity. In this way, Bohr's reading would not differ too much from Heisenberg's.

At this point, it should be obvious why also my second claim seems to be supported. If we attribute a micro-system M a 'real disposition' to show a certain definite value in a measurement context described by a classical apparatus, we explain away certain apparent contradictions of his philosophy, of which he has been accused even by Bell (1987). The problem of a dispositional talk is that in the context of his philosophy it does not improve the physics, as it just amounts to saying that *if we measure a quantum system in a superposition with a particularly prepared physical system, we get a definite result. Since we are not told how, when and why such a definiteness comes into being, the corresponding lack of exactness seems fatal to a realistic project of understanding the physical world.*

Despite the introduction of dispositions, all well-known problems of Bohr's philosophy remain intact, in particular whether the distinction between the classical and the quantum world is *pragmatical* and *contextual*, or is rather physically describable in a precise way.[20] In the former case, Bohr's solution to the measurement problem, even with the addition of intrinsic dispositions for position, momentum, spin, etc, is fine for all practical purposes but is simply an instrumentalistic manoeuvre, covered with a realistic-tasting spice, given by the introduction of dispositions. Of course, there is nothing wrong with instrumentalism per se, but it should be recognized that adding dispositions to a philosophy that, like Bohr's, denies the reality of the collapse, simply adds coherence to his view of QM without increasing our understanding of the physical world.

9.5.1 Suárez's Selective Approach to the Measurement Problem

Unfortunately, it seems to me that the same analysis holds also for Suárez approach to dispositions as selections: 'A selection is an interaction designed to test *a* particular disposition (Fine's 'aspect') of a quantum system. Among the dispositional properties I include those responsible for values of position, momentum, spin and angular momentum. In a *selection*, the pointer position interacts only with the property of the system that is under test' (Suárez, 2004b, 232). In order to represent a given dispositional property, Suárez claims that we can exploit the fact that 'for every property of a quantum system originally in a superposition there is a mixed state which is probabilistically equivalent (for that property) to the superposition' (Suárez, 2004b, 242). Take for instance the two following states, representing respectively the pure state of spin along x, and the mixture $W(x)$:

$$\psi = \left(1/\sqrt{2}\right) \left|\mathrm{up}_x\right\rangle_1 \left|\mathrm{down}_x\right\rangle_2 - \left(1/\sqrt{2}\right) \left|\mathrm{down}_x\right\rangle_1 \left|\mathrm{up}_x\right\rangle_2$$
$$W(x) = \frac{1}{2}P_{[\mathrm{up,\ down}]} + \frac{1}{2}P_{[\mathrm{down,\ up}]}$$

[20] The complaint that Bohr's philosophy relies on an unclear distinction between the classical and the quantum has been notoriously voiced by Bell (1987).

Suárez supposes that the pointer position interacts with only one property of the system $W(O)$, in the example represented by $W(x)$, with O being a particular observable, in our case 'spin along x'. $W(O)$ is not the full state of the system, but simply the state corresponding to its property O.

Now, I think the decisive question to ask is the following: how does a selection of a disposition occur, namely *is the selection a physical proce*ss? If we deny that selections are physical process, the ascription of dispositions to quantum systems is deprived of any interest. By taking this horn of the dilemma, of course we don't have to provide detailed explanations of the selection process, but the ascription of dispositions that are selected by the measurement apparatus looks like a merely *formal trick* to give an account of the transition from pure states to mixtures. It is a solution *by fiat*, so to speak.

On the other hand, by taking the other horn of the dilemma and accepting that selections are *genuine* physical processes,[21] then we need to know more about them, in terms of a more precise, exact *physical* description, of the kind provided by dynamical reduction models. Namely, a description that can, in principle, be tested by experiments, even though the experiments that we can *actually* perform are not capable of testing the theory. If we claim that a measurement 'selects' the appropriate disposition via a genuine physical interaction, we either have the duty to formulate physical hypotheses as to the *when*, *how* and *why*, quantum systems go from a superposition (which is not ignorance-interpretable) to a well-defined value (and then we embrace dynamical reduction models of the GRW type), or we must argue that no such description is possible. But the latter choice is tantamount to give up the hope of explaining what happens in a measurement interaction. Furthermore, if we don't describe the selective process in a more detailed way, we end up treating measurements as special physical interactions and this is certainly unwanted.

In other words, claiming that measurements *are* selections of dispositions without providing physical descriptions of the selecting, genuinely real physical process amounts to sweeping the dust under the carpet. In practice, this would be equivalent to adopting an instrumentalist solution to the measurement problem, which is certainly not in the intention of a proponent of the view that dispositions are real, intrinsic properties of quantum systems.

In a word, we should conclude that selections are an interesting but purely provisional account of measurement interactions. Contrary to the author's intentions, Suárez's alleged solution to the measurement problem is *very similar to* Bohr's, and in order to avoid this trap, he needs to supplement his account with a detailed theory of collapse that can in principle be refuted by experiments: *Suárez's dispositional reading of QM is really committed in some way to the program of dynamical reduction models*.

This conclusion can perhaps be better supported if we conclude by briefly surveying Rovelli's relational account of quantum interactions. According to Rovelli, it is meaningless to attribute an intrinsic, absolute property to a non-correlated system, since 'S has q' is true only for observer/physical system O and may not be true for

[21] This is indeed Suárez's own position (personal communication).

O'. To the extent that 'a variable (of a system *S*) can have a well-determined value *q* for one observer (instrument) (*O*) and at the same time fail to have a determined value for another observer (*O*')' (Laudisa and Rovelli, 2008, Section 2), in this interpretation of QM no sense can be made of any non-dispositional, categorically possessed properties. We could certainly interpret Rovelli's view (and Everettian views, to that effect) as a way to deny the existence of *any* categorically possessed property, and as a way to regard entities as *loci* of purely dispositional properties, whose manifestation is completely dependent on the kind of entity they correlate or interact with.

Notice however, that also this view is hardly explanatory; despite the centrality of the notion of *correlation* or relative state in Rovelli and Everett's view, there is explicitly no intention to offer a clear hypothesis as to *how* and *when* do the correlations occur. Rovelli's view is therefore not different from a form of instrumentalism about the descriptive content of the theory which we have already found in Bohr.

Finally, there is no need of arguing that another important no-collapse view, Bohmian mechanics, renders dispositions wholly dispensable: Bohm's dispositions (the so-called contextual variables) are in fact reducible to positions and context of measurement (Clifton and Pagonis, 1995).

In sum, *if* dispositions have a role in the metaphysical foundations of QM, they must be looked for in GRW kind of theories. Elsewhere, they might contribute to the coherence of an instrumentalist rendering of the theory, but do not help us at all in the interpretive effort that is the task of the philosophy of physics as delineated in the first section of the paper.

Acknowledgement I thank the audience in Madrid for helpful questions and criticism. In addition, I am highly indebted to Roman Frigg, Carl Hoefer and Federico Laudisa for their critical comments to a previous version of this paper, which significantly improved the final result, of which, of course I am the only responsible.

References

Albert D. (1992), *Quantum Mechanics and Experience*, Cambridge: Harvard University Press.

Allori V., Goldstein S., Tumulka R., and Zanghì N. (2008), On the common structure of Bohmian mechanics and the Ghirardi-Rimini-Weber Theory, *The British Journal for the Philosophy of Science*, 59(3), 353–389.

Bell J. S. (1987), *Speakable and Unspeakable in Quantum Mechanics*. Cambridge: Cambridge University Press.

Bell J. S. (1990), Against 'measurement, *Physics World*, 8, 33–40.

Brandom R. (1994), *Making it Explicity*, Harvard, MA: Harvard University Press.

Carnap R. (1936), Testability and Meaning, *Philosophy of Science*, 3(4), 419–471; 4(1) (January 1937): 1–40.

Clifton, R. and Pagonis, C. (1995), Unremarkable contextualism: Dispositions in the Bohm theory, *Foundations of Physics* 25(2), 281–296.

Daumer, M., Dürr D., Goldstein S., and Zanghì N. (1996), Naive realism about operators, *Erkenntnis* 45, 379–397.

Dirac, P.A. (1930), *The Principles of Quantum Mechanics*. Cambridge: Cambridge University Press.

Dorato, M. (2005), *The Software of the Universe*, Aldershot, Hampshire: Ashgate.

Dorato, M. (2007), Dispositions, relational properties, and the quantum world, In M. Kistler, and Gnassounou, B. (eds.), *Dispositions and Causal Powers*, Aldershot, Hampshire: Ashgate, pp. 249–270.

Dorato, M. and Esfeld, M. (2010), GRW as an ontology of dispositions, *Studies in History and Philosophy of Modern Physics* 2010(41), 41–49.

Faye, J. (1991), *Niels Bohr: His Heritage and Legacy*, Dordrecht: Kluwer

Friedman, M. (1974), Explanation and scientific understanding, *Journal of Philosophy*, 71, 5–19.

Frigg, R. and Hoefer C. (2007), Probability in GRW, *Studies in History and Philosophy of Modern Physics*, 38, 371–389.

Ghirardi, G. C., Rimini A., and Weber T. (1986), Unified dynamics for microscopic and macroscopic systems, *Physical Review* D 34, 470–496.

Goldstein, S. (2009), Bohmian mechanics, In E.N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2009 Edition), URL = http://plato.stanford.edu/archives/spr2009/entries/qm-bohm/

Hájek, A. (2010), Interpretations of probability, In E.N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2010 Edition), URL = < http://plato.stanford.edu/archives/spr2010/entries/probability-interpret/>

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

Heisenberg W. (1958), *Physics and Philosophy: The Revolution in Modern Science*, New York, NY: Harper and Row.

Hughes, R. I. G. (1989), *The Structure and Interpretation of Quantum Mechanics*, Cambridge: Harvard University Press.

Kitcher, P., (1976), Explanation, conjunction and unification, *Journal of Philosophy*, 73, 207–212.

Kochen, S. and Specker E. (1967), The problem of hidden variables in quantum mechanics, *Journal of Mathematics and Mechanics*, 18, 1015–1021.

Langton, R. and Lewis D. (1998), Defining 'intrinsic', *Philosophy and Phenomenological Research* 58, 333–345.

Laudisa, F. and Rovelli, C. (2008), Relational quantum mechanics, In E.N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2008 Edition), URL = <http://plato.stanford.edu/archives/fall2008/entries/qm-relational/>.

Lewis, D. (1983), New work for a theory of universals, *Australasian Journal of Philosophy*, 61(4), 343–377.

Margenau, H. (1954), Advantages and disadvantages of various interpretations of the quantum theory, *Physics Today*, 7, 6–13.

Maxwell, N. (1988), Quantum propensiton theory: A testable resolution of the wave/particle dilemma, *British Journal for the Philosophy of Science*, 39, 1–50.

Mumford, S. (1998), *Dispositions*, Oxford: Oxford University Press.

Murdoch, D. (1987), *Niels Bohr's Philosophy of Physics*, Cambridge: Cambridge University Press.

Popper, K. (1982), *Quantum Theory and the Schism in Physics*, volume III of the postscript to the logic of scientific discovery, London: Hutchinson.

Redhead, M. (1987), *Incompleteness, Non-locality and Realism*, Oxford: Clarendon Press.

Sellars, W. (1963), Philosophy and the scientific image of man, *Science, Perception, and Reality*. London: Routledge & Kegan Paul, pp. 1–40.

Shoemaker, S. (1984), *Identity, Cause and Mind*. Cambridge: Cambridge University Press.

Suárez, M. (2004a), On quantum propensities: Two arguments revisited, *Erkenntnis*, 61, 1–16.

Suárez, M. (2004b), Quantum selections, propensities, and the problem of measurement, *British Journal for the Philosophy of Science*, 55, 219–255.

Suárez, M. (2007), Quantum propensities, *Studies in the History and Philosophy of Modern Physics*, 38(2), 418–438. Accessed June 2007.

Tumulka, R. (2006a), A relativistic version of the Ghirardi–Rimini–Weber model, *Journal of Statistical Physics*, 125, 821–840.

Tumulka, R. (2006b), Collapse and relativity, In A. Bassi, Dürr, D., Weber, T., and Zanghì, N. (eds.), *Quantum Mechanics: Are there Quantum Jumps? and On the Present Status of Quantum Mechanics*, American Institute of Physics Conference Proceedings 844, Melville, NY, pp. 340–351.

Van Fraassen B. (1981), A modal interpretation of quantum mechanics, In E. Beltrametti and Van Fraassen, B. (eds.), *Current Issues in Quantum Logic*, New York, NY: Plenum Press, pp. 229–258.

Chapter 10
Is the Quantum World Composed of Propensitons?

Nicholas Maxwell

For well over thirty years I have tried to get across a few simple points about quantum theory – so far with not much success.[1] What I have to say amounts to this. Orthodox quantum theory is unacceptably defective. The defects all arise from the failure to solve the wave/particle problem. A very natural way of solving this problem is to adopt the conjecture that the quantum domain is fundamentally probabilistic. This leads one to a fully micro-realistic, probabilistic version of quantum theory, able to reproduce all the empirical success of orthodox quantum theory, but with as-yet untested predictions that differ from orthodox quantum theory. My message, which admittedly partially overlaps with what others have to say as well, is summed up in a little more detail in the following thirteen sections of this paper.

10.1 Defects of Orthodox Quantum Theory

Orthodox quantum theory (OQT), because it is a theory about *observables*, about *the results of performing measurements on quantum systems*, and not a theory about quantum systems per se, is very seriously defective, to the point of being unacceptable, despite its immense empirical success.

OQT interprets the ψ-function to contain probabilistic information about the outcome of performing measurements[2] on the quantum system (or ensemble of systems) in question. This means that, in order to have physical content, some part of

N. Maxwell (✉)
Department of Science and Technology Studies, University College London,
London WC1E 6BT, UK
e-mail: nicholas.maxwell@ucl.ac.uk

[1] My first published effort goes back to 1972: see Maxwell (1972). See also Maxwell (1973a, b; 1976; 1982; 1985; 1988; 1993b; 1994; 1995; 1998, Chapter 7; 2004).

[2] Throughout, 'measurement' means some process which actually detects quantum systems. A procedure which prepares quantum systems to be in some quantum state is, following Margenau (1958, 1963), here called a 'preparation' rather than a measurement. This distinction between preparation and measurement is crucial for a proper understanding and formulation of quantum theory. See also Popper (1959, 225–226).

M. Suárez (ed.), *Probabilities, Causes and Propensities in Physics*, Synthese Library 347, DOI 10.1007/978-1-4020-9904-5_10, © Springer Science+Business Media B.V. 2011

classical physics must be added to OQT for a treatment of the measuring process. Without the addition of classical physics, OQT can only issue in *conditional* predictions of the form: *if* such and such a measurement is made, the outcome will be such and such, with such and such probability. OQT cannot itself be applied to the measuring process, for then another measuring instrument would be required to measure the first instrument, the second one being described by some appropriate part of classical physics. In general, OQT issues in probabilistic predictions. Schrödinger's time-dependent equation is, however, deterministic. Thus OQT, applied to the quantum system plus measuring apparatus, cannot issue in probabilistic predictions: it would, in effect, predict that the measuring apparatus ends up in a superposition of possible outcomes – until a second measurement is performed with a second measuring apparatus, itself described by classical physics.[3]

It may be objected that all physical theories, even a classical theory such as Newtonian theory (NT), must call upon additional theory to be tested empirically. In testing predictions of NT concerning the position of a planet at such and such a time, optical theory is required to predict the results of telescopic observations made here on earth. But this objection misses the point. NT is perfectly capable of issuing in *physical* predictions without calling upon additional theory, just because it has its own physical ontology. NT, plus initial and boundary conditions formulated in terms of the theory, can issue in the *physical* prediction that such and such a planet is at such and such a place at such and such a time, whether anyone observes the planet or not, without calling upon optical theory or any other theory. This OQT cannot do. It cannot do this because the ψ-function of OQT is interpreted, not as specifying the actual physical states of quantum systems, but rather as containing probabilistic information about the results of performing measurements on the quantum systems in question. OQT, lacking its own quantum ontology, can only issue in predictions about actual physical states of affairs (whether observed or not) if some part of classical physics is employed to describe the measuring instrument.

OQT – the theory with physical content – is thus made up of two conceptually incompatible parts, a purely quantum theoretic part, and some part of classical physics. But this theory, quantum postulates plus classical postulates (QP + CP), suffers from the following seven severe defects, as a direct result of the theory being this *ad hoc* mixture of incompatible quantum and classical postulates.

(1) OQT is *imprecise*, due to the inherent lack of precision of the notion of 'measurement'. How complex and macroscopic must a process be before it becomes a measurement? Does the dissociation of one molecule amount to a measurement? Or must a thousand or a million molecules be dissociated before a measurement has been made? Or must a human being observe the result? No precise answer is forthcoming. (2) OQT is ambiguous, in that if the measuring process is treated as a measurement, the outcome is in general probabilistic, but if this process is

[3] It does not help to employ some special quantum theory of macroscopic phenomena for a treatment of the measuring instrument instead of classical physics: the outcome would still be a severely *ad hoc* theory.

treated quantum mechanically, the outcome is deterministic. OQT is ambiguous concerning the fundamental question as to whether the quantum domain is deterministic or probabilistic. (3) OQT is very seriously *ad hoc*, in that it consists of two incompatible, conceptually clashing parts, QP and CP. OQT only avoids being a straightforward contradiction by specifying, in an arbitrary, *ad hoc* way, that QP applies to the quantum system up to the moment of measurement, and CP applies to the final measurement result. (4) OQT is non-explanatory, in part because it is *ad hoc*, and no *ad hoc* theory is fully explanatory, in part because OQT must presuppose some part of what it should explain, namely classical physics. OQT cannot fully explain how classical phenomena emerge from quantum phenomena because some part of classical physics must be presupposed for measurement. (5) OQT is limited in scope in that it cannot, strictly speaking, be applied to the early universe in conditions which lacked preparation and measurement devices. Strictly speaking, indeed, it can only be applied if physicists are around to make measurements. (6) OQT is limited in scope in that it cannot be applied to the cosmos as a whole, since this would require preparation and measurement devices that are outside the cosmos, which is difficult to arrange. Quantum cosmology, employing OQT, is not possible. (7) For somewhat similar reasons, OQT is such that it resists unification with general relativity. Such a unification would presumably involve attributing some kind of quantum state to spacetime itself (general relativity being a theory of spacetime). But, granted the basic structure of OQT, this would require that preparation and measurement devices exist outside spacetime, again not easy to arrange.

For a fundamental theory of physics, these seven defects are serious indeed.[4]

10.2 Fundamental Defect: Failure to Solve Wave/Particle Problem

The 7 severe defects of OQT just indicated all stem from one fundamental defect: the failure of OQT to solve the wave/particle problem. It is this failure which makes it impossible to interpret the ψ-function of OQT as specifying the actual physical states of quantum systems. As long as no consistent idea is forthcoming as to what kind of entities electrons, protons, atoms and other quantum systems are in physical space from moment to moment, the ψ-function cannot be interpreted as specifying the physical states of actual physical entities in physical space. And the original and fundamental difficulty that lay in the way of developing a consistent idea as to what electrons, atoms etc. are was that no satisfactory solution to the wave/particle problem seemed forthcoming. Electrons and other quantum systems exhibit both wave-like and particle-like properties, as is most apparent in the two-slit experiment, and this seems to present an insuperable obstacle to forming a consistent idea as to what sort of entity these quantum systems can be. Heisenberg decided in effect, when creating matrix mechanics, that no solution to the wave/particle

[4] See Maxwell (1972, 1973b, 1976, 1988 1–8). See also Bell (1987).

problem was forthcoming, and hence the theory would have to be restricted to making predictions about the results of measurement. Schrödinger hoped initially that his wave mechanics could be interpreted to be about wave-like entities in physical space. But any such interpretation was dealt a mortal blow when Born (1926, 1927) interpreted the ψ-function as containing probabilistic information about the results of performing measurements on quantum systems. Wave mechanics given Born's interpretation was able to predict experimental results successfully, whereas the theory given Schrödinger's interpretation, could not. It could not do justice either (a) to the *particle* character of quantum systems, or (b) to the *probabilistic* character of quantum theory, whereas Born's interpretation did justice to both. Bohr repeatedly emphasized that one had to renounce realism about the quantum domain, it being necessary to interpret the new quantum theory of Heisenberg and Schrödinger as being about the results of measurements performed on quantum systems, the measuring process being described by classical physics: see, for example, Bohr (1949).

To the seven defects indicated above we need, then, to add an eighth: OQT fails to solve the quantum wave/particle problem. It fails to be what may be called a 'fully micro-realistic theory' – a theory, that is, which is, in the first instance, exclusively about quantum micro systems, there being nothing in the basic postulates of the theory about measurement at all, even though the theory is, nevertheless, experimentally testable. Or, as John Bell would have put it, OQT is defective because it is about *observables* and not about *beables*: see Bell (1987, Chapter 5).

This eighth defect is the fundamental one. It is from this defect that the other seven stem. Remove this eighth defect, solve the wave/particle problem, develop quantum theory as a fully micro realistic theory exclusively about quantum systems evolving in physical space and time with no reference to measurement or observables whatsoever, and the other seven defects of OQT automatically disappear. An enormous amount of work on what may be called the interpretative problems of quantum theory has, unfortunately, ignored this simple point.[5]

10.3 Probabilism as the Key to the Solution to the Wave/Particle Problem

There is, I suggest, a very obvious possible solution to the quantum wave/particle problem, almost universally overlooked.[6] The denizens of the quantum domain – electrons, atoms, molecules and the rest – are fundamentally *probabilistic* entities,

[5] It is sometimes argued that quantum field theory solves the wave/particle problem. This is not the case at all. Quantum field theory is just as dependent on measurement for its physical interpretation as non-relativistic OQT is.

[6] I do not have space, here, to discuss other approaches to solving the problems of quantum theory, such as Bohmian theory, consistent histories, decoherence, and the many-worlds interpretation. Wallace (2008) provides an excellent survey of these and other approaches.

interacting with one another *probabilistically*, and thus quite unlike anything we have encountered within *deterministic* classical physics. 'Are quantum entities particles or waves?' is the *wrong* question. Instead, we have the following two right questions:

(i) What kind of *unproblematic, fundamentally probabilistic* entities are there, as possibilities?
(ii) Can quantum entities be interpreted to be a variety of such unproblematic, fundamentally probabilistic entities?

We cannot conclude, as a matter of logic, from the probabilistic character of OQT, that quantum theory is telling us that nature herself is probabilistic. This is because, as we saw in Section 10.1 above, OQT is highly *ambivalent* about this crucial issue: see defect (2). It is not clear whether the probabilistic character of OQT reflects probabilism in nature, or whether it is, in some way, the outcome of our measuring interventions. This point is underlined by the fact that there are two interpretations of quantum theory, rivals to the orthodox or Copenhagen interpretation, which hold quantum theory to be fully deterministic – namely the Bohm interpretation, and the many-worlds interpretation.

We can, however, given the probabilistic character of quantum theory, very reasonably conjecture that the quantum domain is fundamentally probabilistic, the laws of this domain, governing the way quantum systems evolve and interact, being probabilistic laws. If this conjecture is correct, it immediately provides us with a very natural route to a resolution of the notorious wave/particle problem. Quantum entities, being fundamentally probabilistic entities, interacting with one another probabilistically, will automatically be quite different from anything encountered within deterministic classical physics. In particular, we should not expect the entities of the quantum domain to be either classical, deterministic *particles*, or *fields*. Quite the contrary, if electrons, atoms, molecules and the rest turned out to be classical particles or fields, it would be a *disaster* for the intelligibility of the quantum domain. The long-standing, traditional effort to understand quantum entities as classical particles or fields has been struggling to solve the wrong problem. The traditional assumption, made by Heisenberg, Born, Bohr, Pauli and the rest, that quantum entities are just too paradoxical, too enigmatic, to be understandable at all (and hence the need to develop OQT as a theory which *evades* the whole problem) is simply based on the failure to take seriously the implications of the thesis that the quantum domain is fundamentally probabilistic.

10.4 Two Kinds of Fundamentally Probabilistic Entity

First, a preliminary, terminological question: what are we going to call hypothetical physical entities that evolve and interact with one another probabilistically? I suggest we call them *propensitons* (Maxwell, 1988, 13).

The two *correct* questions of Section 10.3 then become:

(i) What kinds of propensiton are there, as possibilities?
(ii) Can quantum entities be interpreted to be propensitons of some kind or other? If so, what kind?

As far as (i) is concerned, we can at once distinguish propensitons that evolve in a probabilistic way *continuously* in time, from propensitons that evolve probabilistically *intermittently* in time. Let us call the first *continuous propensitons*, and the second *discrete propensitons*.

A continuous propensiton might be a field-like entity, spread out continuously in space but such that its state at any given instant only determines the state at the next instant *probabilistically*. This remains true for any two states of the propensiton at times t_1 and t_2, however close together t_1 and t_2 may be.

A discrete propensiton is an entity that evolves deterministically until a particular state of affairs arises when, instantaneously, a probabilistic transition occurs, and so on. Discrete propensitons might take the form of spheres which expand steadily and deterministically until – let us suppose – they touch, the condition for the probabilistic transition to occur. The instant two such propensiton spheres touch, each sphere collapses, somewhere within its interior, probabilistically determined, into a tiny sphere of predetermined size. We could modify this slightly by imagining the propensiton sphere is made up of a substance which varies in density in a wave-like way. This determines probabilistically where the tiny sphere is localized, when spheres touch and probabilistic collapse occurs. The tiny sphere, post-probabilistic collapse, is more likely to appear where the pre-collapse substance is dense, and less likely to occur where it is rarefied.

Note that an elementary example of one kind of propensiton – the discrete propensiton – is already beginning to exhibit features somewhat reminiscent of quantum entities!

We can, of course, go on to try to develop further kinds of propensiton. We can seek to introduce *forces* into the propensiton world of possibilities. We can try to design propensitons – continuous or discrete – that are Lorentz invariant. And, germane to our particular concerns here, we can seek to design propensitons that mimic in their behaviour the predictions of OQT – the experimentally confirmed predictions of OQT at least.

The crucial question so far, however, is this: Should we seek to interpret quantum theory as a fully micro realistic theory about continuous or discrete propensitons?

One point deserves to be made straight away. Other things being equal, continuous and discrete propensitons should be treated as, potentially, equally viable, equally intelligible. In particular, the fact that any theory about discrete propensitons will postulate that there are intermittent, instantaneous probabilistic transitions should not be regarded as calling into question the intelligibility of such a theory. There is, from the propensiton perspective, nothing inherently *mysterious* or *inexplicable* about such instantaneous probabilistic transitions. We may hope for a deeper theory that explains such transitions, but we should not be dismayed if this deeper theory should also postulate such instantaneous probabilistic transitions. In particular, to demand that, ultimately, there must be a deterministic explanation for such apparently probabilistic transitions is just to refuse to accept the viability of probabilism at a fundamental level in theoretical physics.

10.5 Guiding Principle: Stay Close to OQT

Ordinarily, in seeking to bring about a theoretical revolution in physics, one should be prepared to develop a radically new kind of theory. But what is being attempted here is rather different. The implication of the argument so far is that the authors of OQT failed to formulate quantum theory properly because they failed to appreciate that probabilism promises to provide a straightforward solution to the apparently insoluble wave/particle paradox, and also failed to appreciate what 'sort of risky game they were playing with reality – reality as something independent of what is experimentally established' (Einstein, 1950, 39). This suggests that, in seeking to develop QT as a fully micro realistic theory about propensitons, we should stick as close as possible to the existing structure of OQT, modifying it just sufficiently to eliminate all reference to observables and measurement from the basic postulates so that the theory becomes fully micro-realistic. And there is another consideration to back up this approach. OQT is an extraordinarily successful theory empirically. Even though fatally defective, it must have got a lot right. This suggests we would be wise, initially at least, to keep as close to the structure of OQT as possible.

If we adopt this approach then, granted we have to choose between the continuous and discrete propensiton, the latter becomes overwhelmingly the better choice. OQT postulates two kinds of evolutions: *deterministic* evolutions in accordance with Schrödinger's time-dependent equation in the absence of measurement, and *probabilistic* evolutions associated with measurement. This mirrors the character of the discrete propensiton as it has been characterized above.

We are led, then, to consider the following idea. The ψ-function is to be interpreted as specifying the actual physical state of discrete propensitons from moment to moment. Schrödinger's equation specifies how these physical states evolve in time as long as no probabilistic transition occurs. Measurement is a *sufficient* condition for a probabilistic transition to occur. Measurement is not, however, a *necessary* condition. It is entirely to be expected, according to this approach, that probabilistic transitions will occur in the absence of measurement. Nothing would be gained if we had to appeal to the imprecise, macroscopic notion of measurement to specify the physical conditions for propensitons to undergo probabilistic transitions: such a propensiton version of QT would reproduce all the defects of OQT. And if the world really is made up of discrete propensitons, and probabilistic transitions occur objectively in nature, it would be very peculiar indeed if these transitions only occurred when physicists make measurements. Propensiton quantum theory (PQT), in order to be a satisfactory, fully micro realistic theory, must specify the conditions for probabilistic transitions to occur in fully micro realistic, quantum mechanical terms.

It is this requirement, incidentally, which ensures that any acceptable, fully micro realistic version of PQT must differ experimentally from OQT. For PQT predicts that probabilistic transitions occur even in the absence of measurement, something which OQT denies. Crucial experiments are in principle possible to decide between OQT and PQT.

Two kinds of problem now face the development of PQT. First, objections may be raised to the possibility of interpreting the ψ-function as specifying the actual physical state of propensiton quantum entities. Second, precise propensiton, quantum mechanical, necessary and sufficient conditions need to be specified for probabilistic transitions to occur. These two kinds of problem are tackled and solved in the next two sections.

10.6 Can the ψ-Function be Interpreted as Specifying the Actual Physical States of Propensitons?

The basic idea is that ψ is to be interpreted as specifying the actual physical state of the propensiton system at any given instant by specifying the value of *probabilistic* properties or *propensities*[7] possessed by the propensitons at the given instant. The notion of propensity is best understood as a probabilistic generalization of the ordinary deterministic notion of dispositional physical property. Physical properties such as mass, charge, rigidity, transparency and so on determine how something changes (or does not change) in certain circumstances. Thus the mass of an object determines how the object will accelerate when subject to a force. Inflammability determines (roughly) that the inflammable object bursts into flames when subject to a naked flame. A propensity is a probabilistic generalization of this deterministic notion of dispositional physical property. Instead of there being just one outcome, there are a number of possible outcomes (possibly infinitely many) and the value of the propensity assigns probabilities to these possible outcomes. An example of a propensity is what may be called the 'bias' of a die – the property of the die which determines the probabilities of the outcomes 1–6 when the die is tossed onto a table. A value of bias assigns a probability to each of the 6 possible outcomes. We can even imagine that the value of the bias of the die itself changes: there is, perhaps, a tiny magnet imbedded in the die and an electromagnet under the table. As the strength of the magnet beneath the table varies, so the value of the bias of the die will change.

[7] Popper introduced the idea of propensities in connection with interpretative problems of QT, see Popper (1957, 1967, 1982) although, as Popper (1982, 130–135) has pointed out, Born, Heisenberg, Dirac, Jeans and Landé have all made remarks in this direction. The version of the propensity idea employed here is, however, in a number of respects, different from and an improvement over, the notion introduced by Popper: see Maxwell (1976, 284–286, 1985, 41–42). It is a probabilistic generalization of the notion of deterministic dispositional or necessitating property introduced in Maxwell (1968): see also Maxwell (1998, 141–155). For a discussion of Popper's contributions to the interpretative problems of quantum theory see Maxwell (2011, Section 6, especially note 19).

Precisely what propensities are attributed to quantum systems by the ψ-function of QT will depend on the precise nature of probabilistic transitions, to be discussed in the next section. But the general idea can be illustrated as follows. Assume that probabilistic transitions are *localizations*. The corresponding propensity attributed to individual quantum systems by ψ would be *position probability density*. As $|\psi|^2$ varies with the passage of time so the value of the propensity, position probability density, varies too.

In order to establish empirically an attribution of a specific value of bias to the die, a number of tests need to be performed (the die needs to be repeatedly tossed) with conditions remaining unchanged. But a specific value of bias is nevertheless a physical property possessed by an individual die. Similarly, ψ attributes specific values of propensities to individual quantum systems; but in order to verify such attributions, a great number of experiments need to be performed, with conditions kept constant, to check up on the probabilistic predictions of the propensity attribution.[8]

The following objections may now be made to the claim that the ψ-function can be interpreted as specifying the actual states of physical systems in physical space at instants of time.

(a) The ψ-function is complex, and hence cannot be used to describe the physical state of an actual physical system.
(b) Given a physical system of N quantum entangled systems, the ψ-function is no longer a function of physical space, but of $3N$ dimensional configuration space. This makes it impossible to interpret such a ψ-function as specifying the physical state of physical systems in physical space.
(c) The ψ-function is highly non-local in character. This, again, makes a realistic interpretation of it impossible.
(d) Interpreting the ψ-function realistically would carry the consequence that when a position measurement is made, and a quantum system that had a state spread throughout a large volume of space, instantaneously collapses into a minute region where the system is detected.

Here, briefly, are my replies.

(a) The complex ψ is equivalent to two interlinked real functions, which can be regarded as specifying the propensity state of quantum systems. In any case, as Penrose (2004, 539) reminds us, complex numbers are used in classical physics, without this creating a problem concerning the reality of what is described. The complex nature of ψ has to do, in part, with the fact that the wave-like character of ψ is not in physical space, except when interference leads to spatio-temporal wave-like variations in the intensity of ψ, and thus in $|\psi(x,t)|^2$ as well.

[8] For a more detailed presentation of these features of PQT see Maxwell (1976, 1982, 1985, 1988).

(b) $\psi(r_1, r_2 \ldots r_N)$ can be regarded as assigning a complex number to any point in 3*N*-dimensional configuration space. Equally, however, we can regard $\psi(r_1, r_2 \ldots r_N)$ as assigning the complex number to N points in 3 dimensional physical space. Suppose $\psi(r_1, r_2 \ldots r_N)$ is the quantum entangled state of N distinct kinds of particle. Then $\psi(r_1, r_2 \ldots r_N)$, in assigning a complex number to a point in configuration space, is to be interpreted as assigning this number to N points in physical space, each point labelled by a different particle. The quantum propensiton state in physical space will be multi-valued at any point in physical space, and also highly non-local, in that its values at any given point cannot be dissociated from values at N–1 other points. If we pick out N distinct points in physical space, there will be, in general, N! points in configuration space which assign different values of ψ to these N physical points, corresponding to the different ways the N particles can be reassigned to these N points. If we pick out just one point in physical space (x_o, y_o, z_o), the ψ-function will in general assign infinitely many different complex numbers to this point (x_o, y_o, z_o), corresponding to different locations of the particles in physical space – there being infinitely many points in configuration space that assign a complex number to this point (x_o, y_o, z_o) in physical space. The N-particle, quantum entangled propensiton is, in physical space, a complicated, non-local, multi-valued object, very different from anything found in classical physics. Its physical nature in 3-dimensional physical space is, nevertheless, precisely specified by the single-valued $\psi(r_1, r_2 \ldots r_N)$ in 3N dimensional configuration space.[9]

(c) As my response to problem (b) indicates, quantum propensitons of the type being considered here, made up of a number of quantum entangled 'particles', are highly non-local in character, in that one cannot specify what exists at one small region of physical space without simultaneously taking into account what exists at other small regions. Propensitons of this type seem strange because they are unfamiliar – but we must not confuse the unfamiliar with the inexplicable or impossible. Non-local features of the ψ-function do not prevent it from specifying the actual physical states of propensitons; propensitons just are, according to the version of PQT being developed here, highly non-local objects, in the sense indicated.

(d) Instantaneous probabilistic collapse is a natural feature of the discrete propensiton. There is nothing inherently impossible or inexplicable about such probabilistic transitions. To suppose otherwise is to be a victim of deterministic prejudice, as we saw in the last paragraph of Section 10.4 above.[10]

[9] This solution to the problem was outlined in Maxwell (1976, 666–667; and 1982, 610). Albert (1996) has proposed that the quantum state of an N-particle entangled system be interpreted to exist physically in 3 N dimensional configuration space. But configuration space is a mathematical fiction, not a physically real arena in which events occur. Albert's proposal is untenable, and in any case unnecessary.

[10] Instantaneous probabilistic collapse is, however, highly problematic the moment one considers developing a Lorentz-invariant version of the theory. This is discussed below, in Section 11.

I conclude that there are no objections to interpreting ψ as specifying the actual physical states of propensitons in physical space.

10.7 Precise Quantum Theoretic Conditions for Probabilistic Transitions to Occur

In order to specify the precise nature of the quantum discrete propensitons under consideration, and at the same time give precision to the version of PQT being developed here, we need now to specify precisely, in quantum theoretic terms (a) the precise quantum conditions for a probabilistic transition to occur in a quantum system, (b) what the possible outcome quantum states are, given that the quantum state at the instant of probabilistic transition is ψ, and (c) how ψ assigns probabilities to the possible outcomes. No reference must be made to observables, measurement, macroscopic system, classically described system or irreversible process.

One possibility is the proposal of Ghirardi, Rimini and Weber (1986) – see also Ghirardi (2002) – according to which the quantum state of a system such as an electron collapses spontaneously, on average after the passage of a long period of time, into a highly localized state. When a measurement is performed on the quantum system, it becomes quantum entangled with millions upon millions of quantum systems that go to make up the measuring apparatus. In a very short time there is a high probability that one of these quantum systems will spontaneously collapse, causing all the other quantum entangled systems, including the electron, to collapse as well. At the micro level, it is almost impossible to detect collapse, but at the macro level, associated with measurement, collapse occurs very rapidly all the time.

Another possibility is the proposal of Penrose (1986, 2004, Chapter 30), according to which collapse occurs when the state of a system evolves into a superposition of two or more states, each state having, associated with it, a sufficiently large mass located at a distinct region of space. The idea is that general relativity imposes a restriction on the extent to which such superpositions can develop, in that it does not permit such superpositions to evolve to such an extent that each state of the superposition has a substantially distinct space-time curvature associated with it.

The possibility that I favour, put forward before either Ghirardi, Rimini and Weber's proposal, or Penrose's proposal, is that probabilistic transitions occur whenever, as a result of inelastic interactions between quantum systems, new 'particles', new bound, stationary or decaying systems, are created (Maxwell, 1972, 1976, 1982, 1988, 1994). A little more precisely:

(I) Whenever, as a result of an inelastic interaction, a system of interacting 'particles' creates new 'particles', bound, stationary or decaying systems, so that the state of the system goes into a superposition of states, each state having associated with it different particles or bound, stationary or decaying systems, then, when the interaction is nearly at an end, spontaneously and probabilistically, entirely in the absence of measurement, the superposition collapses into one or other state.

Two examples of the kind of interactions that are involved here are the following:

$$e^- + \mathrm{H} \to \begin{array}{l} e^- + \mathrm{H} \\ e^- + \mathrm{H}^* \\ e^- + \mathrm{H} + \gamma \\ e^- + e^- + p \end{array}$$

$$e^+ + \mathrm{H} \to \begin{array}{l} e^+ + \mathrm{H} \\ e^+ + e^- + p \\ (e^+/e^-) + p \\ \mathrm{p} + 2\gamma \end{array}$$

(Here e^-, e^+, H, H*, γ, p and (e^+/e^-) stand for electron, positron, hydrogen atom, excited hydrogen atom, photon, proton and bound system of electron and positron, respectively.)

What exactly does it mean to say that the 'interaction is very nearly at an end' in the above postulate? My suggestion, here, is that it means that forces between the 'particles' are very nearly zero, except for forces holding bound or decaying systems together. In order to indicate how this can be formulated precisely, consider the toy interaction:

$$\mathrm{a} + \mathrm{b} + \mathrm{c} \to \begin{array}{ll} \mathrm{a} + \mathrm{b} + \mathrm{c} & (\mathrm{A}) \\ \mathrm{a} + (\mathrm{bc}) & (\mathrm{B}) \end{array}$$

Here, a, b and c are spinless particles, and (bc) is the bound system. Let the state of the entire system be $\Phi(t)$, and let the asymptotic states of the two channels (A) and (B) be $\psi_\mathrm{A}(t)$ and $\psi_\mathrm{B}(t)$ respectively. Asymptotic states associated with inelastic interactions are fictional states towards which, according to OQT, the real state of the system evolves as $\to +\infty$. Each outcome channel has its associated asymptotic state, which evolves as if forces between particles are zero, except where forces hold bound systems together.

According to OQT, in connection with the toy interaction above, there are states $\phi_\mathrm{A}(t)$ and $\phi_\mathrm{B}(t)$ such that:

(1) For all t, $\Phi(t) = \mathrm{c_A}\phi_\mathrm{A}(t) + \mathrm{c_B}\phi_\mathrm{B}(t)$, with $|c_\mathrm{A}|^2 + |c_\mathrm{B}|^2 = 1$;
(2) as $t \to +\infty$, $\phi_\mathrm{A}(t) \to \psi_\mathrm{A}(t)$ and $\phi_\mathrm{B}(t) \to \psi_\mathrm{B}(t)$.

According to the version of PQT under consideration here, at the first instant t for which $\phi_\mathrm{A}(t)$ is very nearly the same as the asymptotic state $\psi_\mathrm{A}(t)$, or $\phi_\mathrm{B}(t)$ is very nearly the same as $\psi_\mathrm{B}(t)$, then the state of the system, $\Phi(t)$, collapses spontaneously either into $\phi_\mathrm{A}(t)$ with probability $|c_\mathrm{A}|^2$, or into $\phi_\mathrm{B}(t)$ with probability $|c_\mathrm{B}|^2$. Or, more precisely:

(II) At the first instant for which $|\langle \psi_\mathrm{A}(t)|\, \phi_\mathrm{A}(t)\rangle|^2 > 1 - \varepsilon$ or

$|\langle \psi_B(t)|\,\phi_B(t)\rangle|^2 > 1 - \varepsilon$, the state of the system collapses spontaneously into $\phi_A(t)$ with probability $|c_A|^2$, or into $\phi_B(t)$ with probability $|c_B|^2$, ε being a universal constant, a positive real number very nearly equal to zero.[11]

According to (II), if $\varepsilon = 0$, probabilistic collapse occurs only when $t = +\infty$; (and the corresponding version of PQT becomes equivalent to the many worlds, or Everett, interpretation of quantum theory). As ε is chosen to be closer and closer to 1, so collapse occurs more and more rapidly, for smaller and smaller times t – and, of course, the corresponding versions of PQT become more and more falsifiable experimentally.

The evolutions of the actual state of the system, $\Phi(t)$, and the asymptotic states, $\psi_A(t)$ and $\psi_B(t)$, are governed by the respective channel Hamiltonians, H, H_A and H_B, where:-

$$\mathrm{H} = -\left(\frac{\hbar^2}{2m_a}\nabla_a^2 + \frac{\hbar^2}{2m_b}\nabla_b^2 + \frac{\hbar^2}{2m_c}\nabla_c^2\right) + \mathrm{V}_{ab} + \mathrm{V}_{bc} + \mathrm{V}_{ac}$$

$$H_A = -\left(\frac{\hbar^2}{2m_a}\nabla_a^2 + \frac{\hbar^2}{2m_b}\nabla_b^2 + \frac{\hbar^2}{2m_c}\nabla_c^2\right)$$

$$H_B = = -\left(\frac{\hbar^2}{2m_a}\nabla_a^2 + \frac{\hbar^2}{2m_b}\nabla_b^2 + \frac{\hbar^2}{2m_c}\nabla_c^2\right) + \mathrm{V}_{bc}$$

Here, m_a, m_b, and m_c are the masses of 'particles' a, b and c respectively, and $\hbar = h/2\pi$ where h is Planck's constant.

The condition for probabilistic collapse, formulated above, can readily be generalized to apply to more complicated and realistic inelastic interactions between 'particles'.

According to this fully micro-realistic, fundamentally probabilistic version of quantum theory, the state function, $\Phi(t)$, describes the actual physical state of the quantum system – the propensiton – from moment to moment. The physical (quantum) state of the propensiton evolves in accordance with Schrödinger's time-dependent equation as long as the condition for a probabilistic transition to occur does not obtain. The moment it does obtain, the state jumps instantaneously and probabilistically, in the manner indicated above, into a new state. (All but one of a superposition of states, each with distinct 'particles' associated with them, vanish.) The new state then continues to evolve in accordance Schrödinger's equation until conditions for a new probabilistic transition arise. Quasi-classical objects arise as a result of the occurrence of a sequence of many such probabilistic transitions.

[11] The basic idea of (II) is to be found in Maxwell (1982 and 1988). It was first formulated precisely in Maxwell (1994).

Another approach to specifying the quantum mechanical condition for a probabilistic transition to occur would be to exploit Schrödinger's time-independent equation. Consider again the above toy rearrangement interaction, and let the state of the system

$$\Phi(t) = \mathrm{c}_{\mathrm{A}}(t)\varphi_{\mathrm{A}}(r_{\mathrm{a}}, r_{\mathrm{b}}, r_{\mathrm{c}}, t) + \mathrm{c}_{\mathrm{B}}(t)\varphi_{\mathrm{B}}(r_{\mathrm{a}}, r_{\mathrm{bc}}, t)\phi(r_{\mathrm{bc}}).$$

Here, $\phi(r_{\mathrm{bc}})$ is the stationary state of the bound system (bc) as given by Schrödinger's time-independent equation, r_{a}, r_{b} and r_{c} are the spatial coordinates of a, b and c, and r_{bc} are the coordinates of the centre of mass of (bc). It is assumed that, for any t, $\Phi(t)$ has a unique form when expressed in this way, as long as $|c_{\mathrm{B}}(t)|^2$ is a maximum. The state $\Phi(t)$ jumps into the state $\varphi_{\mathrm{A}}(r_{\mathrm{a}}, r_{\mathrm{b}}, r_{\mathrm{c}}, t)$ with probability $|c_{\mathrm{A}}(t)|^2$ or into the state $\varphi_{\mathrm{B}}(r_{\mathrm{a}}, r_{\mathrm{bc}}, t)\phi(\mathrm{r}_{\mathrm{bc}})$ with probability $|c_{\mathrm{B}}(t)|^2$ when $1/\,|c_{\mathrm{B}}(t)|^2 \int \lfloor j_t| \, dr < \partial$, where j_t is the probability current density at time t into or out of the state $\varphi_{\mathrm{B}}(r_{\mathrm{a}}, r_{\mathrm{bc}}, t)\phi(\mathrm{r}_{\mathrm{bc}})$, the integration being carried out over the relevant configuration space, and $\partial > 0$ is a constant.

But this second proposal is not altogether satisfactory. It is possible that the probability current might be nearly zero only instantaneously, which would not seem to suffice for the probabilistic transition to occur. One could demand that the acceleration of the probability current is nearly zero as well, but the requirement for the probabilistic transition to occur then begins to look somewhat implausibly cumbersome. In what follows I adopt (II), the first condition for probabilistic transitions to occur, and take PQT to refer to that specific version of propensiton quantum theory.

10.8 PQT Recovers all the Empirical Success of OQT

The version of propensiton quantum theory (PQT) just indicated recovers – in principle – all the empirical success of orthodox quantum theory (OQT). In order to see this it is vital to take note of the distinction, already alluded to (see note 1), between *preparation* and m*easurement* (Popper, 1959, 225–226; Margenau, 1958, 1963). A preparation is some physical procedure which has the consequence that if a quantum system exists (or is found) in some predetermined region of space then it will have (or will have had) a definite quantum state. A measurement, by contrast, actually detects a quantum system, and does so in such a way that a value can be assigned to some quantum 'observable' (position, momentum, energy, spin, etc.). A measurement need not be a preparation. Measurements of photons, for example, far from preparing the photons to be in some quantum state, usually *destroy* the photons measured! On the other hand, a preparation is not in itself a measurement, because it does not *detect* what is prepared. It can be converted into a measurement by a subsequent detection.

From the formalism of OQT, one might suppose that the various quantum observables are all on the same level, and have equal status. In fact this is not the case. Position is fundamental, and measurements of all other observables are made up of a combination of preparations and position measurements.[12] PQT, in order to do justice to quantum measurements, need only do justice to *position* measurements.

It might seem, to begin with, that PQT, based on the two postulates (I) and (II), which say nothing about position or localization, cannot predict that unlocalized systems become localized, necessary, it would seem, to predict the outcome of position measurements. PQT does, however, predict that localizations occur. If a highly localized system, S_1, interacts inelastically with a highly unlocalized system, S_2, in such a way that a probabilistic transition occurs, then S_1 will localize S_2. If an atom or nucleus emits a photon or other 'particle' which travels outwards in a spherical shell and which is subsequently absorbed by a localized third system, the localization of the photon or 'particle' will localize the emitting atom or nucleus with which it was quantum entangled.

That PQT recovers (in principle) all the empirical success of OQT is a consequence of the following four points.[13]

First, OQT and PQT use the same dynamical equation, namely Schrödinger's time-dependent equation.

Secondly, whenever a position measurement is made, and a quantum system is detected, this invariably involves an inelastic interaction and the creation of a new 'particle' (bound or stationary system, such as the ionisation of an atom or the dissociation of a molecule, usually millions of these). This means that whenever a position measurement is made, the conditions for probabilistic transitions to occur, according to PQT, are satisfied. PQT will reproduce the predictions of OQT (given that PQT is provided with a specification of the quantum state of the measuring apparatus). As an example of PQT predicting, probabilistically, the result of a position measurement, consider the following. An electron in the form of a spatially spread out wavepacket is directed towards a photographic plate. According to PQT, the electron wavepacket (or propensiton) interacts with billions of silver bromide molecules spread over the photographic plate: these evolve momentarily into superpositions of the dissociated and undissociated states until the condition for probabilistic collapse occurs, and just one silver bromide molecule is dissociated, and all the others remain undissociated. When the plate is developed (a process which merely makes the completed position measurement more visible), it will be discovered that the electron has been detected as a dot in the photographic plate.

---

[12] Popper distinguished preparation and measurement in part in order to make clear that Heisenberg's uncertainty relations prohibit the simultaneous *preparation* of systems in a precise state of position and momentum, but place no restrictions whatsoever on the simultaneous *measurement* of position and measurement. One needs, indeed, to measure position and momentum simultaneously well within the Heisenberg uncertainty relations simply to check up experimentally on the predictions of these relations: see Popper (1959, 223–236).

[13] In fact, from a formal point of view (ignoring questions of interpretation) PQT has exactly the same structure as OQT with just one crucial difference: the generalized Born postulate of OQT is replaced by postulate (II) of Section 10.7. (The generalized Born postulate specifies how probabilistic information about the results of measurement is to be extracted from the ψ-function.)

Thirdly, all other observables of OQT, such as momentum, energy, angular momentum or spin, always involve (i) a preparation procedure which leads to distinct spatial locations being associated with distinct values of the observable to be measured, and (ii) a position measurement in one or other spatial location. This means that PQT can predict the outcome of measurements of all the observables of OQT.

Fourthly, insofar as the predictions of OQT and PQT differ, the difference is extraordinarily difficult to detect, and will not be detectable in any quantum measurement so far performed.

10.9 Crucial Experiments

In principle, however, OQT and PQT yield predictions that differ for experiments that are extraordinarily difficult to perform, and which have not yet, to my knowledge, been performed. Consider the following evolution:

$$\begin{array}{ccccc} & \text{collision} & \text{superposition} & \text{reverse collision} & \\ & & a+b+c & & \\ a+b+c & \longrightarrow & & \longrightarrow & a+b+c \\ & & a+(bc) & & \\ (1) & (2) & (3) & (4) & (5) \end{array}$$

Suppose the experimental arrangement is such that, if the superposition at stage (3) persists, then interference effects will be detected at stage (5). Suppose, now, that at stage (3) the condition for the superposition to collapse into one or other state, according to PQT, obtains. In these circumstances, OQT predicts interference at stage (5), whereas PQT predicts no interference at stage (5), (assuming the above evolution is repeated many times). PQT predicts that in each individual case, at stage (3), the superposition collapses probabilistically into one or other state. Hence there can be no interference.

OQT and PQT make different predictions for decaying systems. Consider a nucleus that decays by emitting an α-particle. OQT predicts that the decaying system goes into a superposition of the decayed and undecayed state until a measurement is performed, and the system is found either not to have decayed or to have decayed. PQT, in appropriate circumstances, predicts a rather different mode of decay. The nucleus goes into a superposition of decayed and undecayed states, which persists for a time until, spontaneously and probabilistically, in accordance with the postulate (II) of Section 10.7, the superposition jumps into the undecayed or decayed state entirely independent of measurement. The decaying system will continue to jump, spontaneously and probabilistically, into the undecayed state until, eventually, it decays.

These two processes of decay are, on the face of it, very different. There is, however, just one circumstance in which these two processes yield the same answer, namely if the rate of decay is exponential. Unfortunately, the rate of decay of decaying systems, according to quantum theory, *is* exponential. It almost looks as if nature is here maliciously concealing the mode of her operations. It turns out, however, that for long times quantum theory predicts departure from exponential decay (Fonda et al., 1978). This provides the means for a crucial experiment. OQT predicts that such long-time departure from exponential decay will, in appropriate circumstances, obtain, while PQT predicts that there will be no such departure. The experiment is, however, very difficult to perform because it requires that the environment does not detect or 'measure' decay products during the decay process. For further suggestions for crucial experiments see Maxwell (1988, 37–38).

There is a sense, it must be admitted, in which PQT is not falsifiable in these crucial experiments. If OQT is corroborated, and PQT seems falsified, the latter can always be salvaged by letting ε, the undetermined constant of PQT, be sufficiently minute. Experiments that confirm OQT only set an upper limit to ε. There is always the possibility, however, that OQT will be refuted and PQT will be confirmed.

It would be interesting to know what limit present experiments place on the upper bound of ε.

10.10 What PQT Achieves

PQT provides a very natural possible solution to the quantum wave/particle dilemma. The theory is fully micro-realistic; it is a theory about 'beables' to use John Bell's term. It makes sense of the mysterious quantum world. There is no reference to observables, to measurement, to macroscopic, quasi-classical or irreversible phenomena or processes, or to the environment, whatsoever. PQT does not suffer from the eight defects, indicated in Sections 10.1 and 10.2, which beset OQT. The theory is restricted, in the first instance, to specifying how quantum micro systems – quantum propensitons – evolve and interact with one another deterministically and probabilistically. But despite eschewing all reference to observables or measurement in its basic postulates, the theory nevertheless in principle recovers all the empirical success of OQT. At the same time it is empirically distinct from OQT for experiments not yet performed, and difficult to perform.

10.11 The Problem of Developing a Relativistic Version of PQT

A major problem does, however, confront PQT: how can this version of quantum theory be made Lorentz invariant? Instantaneous collapse does not seem to accord well with special relativity!

I do not have a solution to this problem. There are, however, a number of points I would like to make in connection with it.

To begin with, it is not the instantaneous, or faster-than-light, character of collapse that violates special relativity. Tachyons – hypothetical particles that travel faster than light (and thus infinitely fast in some reference frame) – do not contradict special relativity. Any such faster-than-light process must, however, be reversible (i.e. such that it can be regarded as travelling in either direction) to be compatible with special relativity. For, given special relativity, in some reference frames the process will travel in one direction, and in others it will travel in the other direction. All these frames are only equally viable if both directions make equal sense physically.

In the case of probabilistic collapse of propensitons, in general the collapse only makes sense if it is instantaneous. Suppose a highly unlocalized system S_1 interacts inelastically and probabilistically with a highly localized system S_2 in such a way that S_1 is localized. In one set of frames, all at rest with respect to each other, the spatial collapse of S_1 is instantaneous. This, given the probabilistic character of the process, makes physical sense. But in other reference frames moving with respect to the first set, S_1 begins to collapse towards S_2 before the probabilistic transition has occurred, anticipating its occurrence, as it were. This hardly makes physical sense. These reference frames are ruled out, on the grounds that they do not make physical sense of what occurs. This clashes with special relativity, which demands that all inertial frames are equally viable.

The only way known to me of reconciling instantaneous collapse and Lorentz invariance is to adopt Gordon Fleming's 'hyperplane dependent' theory: see Fleming (1989). This entails a radical departure from Minkowskian space-time, however, in that it requires that the basic space-time entity is the space-like hyperplane rather than the space-time point. According to the theory, what exists in any small space-time region may depend on what hyperplane it is considered to lie on. Reality is, according to the theory, highly non-local in character, a dramatic departure from special relativity as ordinarily understood.

If we do not adopt Fleming's speculative hyperplane dependent theory, we must just accept, it seems to me, that any version of PQT that postulates instantaneous probabilistic collapse as a real physical phenomenon must be incompatible with special relativity – and general relativity too. Elsewhere I have argued that this does not constitute grounds for rejecting such fundamentally probabilistic versions of quantum theory (Maxwell, 1985, 2006). A successful theory of quantum gravity will almost certainly reveal that both special and general relativity are not quite correct (just as general relativity reveals that Newtonian theory is not correct, and quantum theory reveals that classical physics is not correct). It is conceivable that the inadequacies of special and general relativity lie in their failure to accommodate instantaneous probabilistic collapse. Quantum gravity may require general relativity to be modified so as to accommodate instantaneous probabilistic transitions on spacelike hypersurfaces. Furthermore, elsewhere I have given additional reasons for doubting the spacetime ontology of special and general relativity (Maxwell, 1985, 2006).

It might be thought that if special and general relativity really are inadequate in the way I have just indicated, then this inadequacy would have already revealed itself experimentally. But this need not be correct at all. Fundamentally probabilistic theories which successfully unify special relativity and PQT, and general relativity and PQT, might differ in their predictions from current theories for only very subtle and difficult-to-perform experiments. In particular, a version of PQT that does justice to relativistic effects might only differ experimentally from existing Lorentz invariant quantum electrodynamics for intractable experiments of the type indicated in Section 10.9 above.

In order to develop such a ‘relativistic’ version of PQT, it is necessary, of course, to specify reference frames with respect to which probabilistic collapse is instantaneous. As long as it is possible to specify unambiguously the quantum system within which collapse occurs, these frames might be specified to be those in which the expectation value for the momentum of the system as a whole is zero. It may be, however, that there is a cosmic-wide universal ‘now’ at each instant, probabilistic collapse occurring in such a way as to be instantaneous with respect to this cosmic ‘now’.

10.12 PQT Has Its Roots in Old Quantum Theory

PQT has its roots deep in the history of quantum theory. This is an important point to take into account when it comes to deciding how seriously to take PQT. Far from being a recent, arbitrary, *ad hoc* modification of quantum theory, PQT is, on the contrary, implicit in some of the earliest contributions to the theory, and this ought to count in its favour.

A hint of the basic idea of PQT can even, perhaps, be discerned in Planck’s original creation of quantum theory in 1900. In seeking to derive his law of black body radiation from first principles, Planck was led to postulate that a black body, in equilibrium with light, is made up of harmonic oscillators – atoms or molecules – which absorb and emit light in discrete amounts $E = h\nu$, where E is energy, ν is the frequency of the oscillator, and h is what came to be called ‘Planck’s constant’ (see Jammer, 1966, Chapter 1; Pais, 1982, Chapter 18).

It would have been too much to expect Planck or his contemporaries to have interpreted $E = h\nu$ as a sign that the determinism of classical physics was to yield to probabilism. However, *if* one had been looking for hints of probabilism, this would have been one place to look. $E = h\nu$ is in flagrant contradiction with basic principles of deterministic classical physics. It is not easy to see how the absorption and emission of light, obeying this law, could be a smooth, continuous, deterministic process. It would seem, rather, to have to be an abrupt, discrete and probabilistic process.

One way in which it might be possible to preserve determinism would be to adopt Einstein’s 1905 light quantum hypothesis (see Pais, 1982, Chapter 18). If the energy of light is to be associated with ‘particles’ or photons, scattered at random in the light, and oscillators jump from one energy level to another when they absorb or emit a photon, then it is just about possible to see how determinism might be preserved. Absorption of light is probabilistic, but this is due to the probabilistic distribution of photons in the light: the laws may well be deterministic. (Deterministic emission, however, poses more of a problem.)

In the absence of Einstein's postulate, it is not easy to see how absorption and emission of light can be both deterministic and in accordance with $E = h\nu$.

Planck would not have entertained probabilism for a moment since he sought to derive his law of black body radiation from classical, and therefore deterministic, postulates.

As it happens, grounds did exist, around 1900, independently of Planck's work, for taking probabilism seriously. They arose in connection with radioactivity. In 1900 Rutherford put forward his exponential law of radioactive decay (see Pais, 1986, 120–123). If the instant at which an atom decays is only probabilistically determined, the probability of decay being constant in time, then Rutherford's exponential law follows as an immediate consequence. Probabilism is thus strongly suggested by Rutherford's law. In order to salvage determinism one must suppose that instants of decay are determined either by an appropriately varying environment, or by appropriate variations in the initial states of the decaying atoms. Both possibilities were considered; neither is especially attractive.

Any temptation to interpret the new quantum theory of Planck and Einstein probabilistically would have been considerably reinforced with the advent of Bohr's quantum theory of the hydrogen atom (Jammer, 1966, Chapter 2; Pais, 1986, Chapter 9). According to this theory, the electron in orbit jumps instantaneously from one semi-stable orbit to another, emitting or absorbing light in discrete quantities of energy as it does so, in complete violation of classical physics.

Probabilism and the basic idea of PQT enter the arena quite explicitly, however, with Einstein's theory of spontaneous and stimulated emission of 1916 and 1917 (Pais, 405–412). What Einstein in effect did was to add probabilistic postulates to Bohr's quantum theory of the atom, thereby providing a probabilistic interpretation of the theory. Einstein considered again atoms in equilibrium with radiation, and postulated three probabilistic processes. First, an excited atom has a certain fixed probability per unit time to jump down *spontaneously* to the lower energy state, emitting light. Second, an atom at the lower energy, exposed to radiation, has a certain probability per unit time to undergo *induced absorption*, jumping up to the higher energy level. And third, an excited atom, exposed to radiation, has a certain probability per unit time to undergo *induced emission*, jumping down to the lower energy. For equilibrium, we require that these three processes do not change the overall number of atoms at the two energy levels. From these elementary postulates, Einstein rederived Planck's radiation law.

Einstein's contribution of 1916/1917 can be regarded as providing us with an early version of PQT. It implies that probabilistic transitions occur when an atom jumps from one stationary state to another. This view of the matter receives additional support from the fact that Einstein's postulate for spontaneous emission is entirely in accordance with Rutherford's exponential law of radioactive decay, itself so suggestive of an intrinsically probabilistic occurrence. Einstein himself drew attention to the similarity, and remarked on the probabilistic implications of his contribution at the time. Unfortunately, Einstein's commitment to determinism meant that he failed to support his own contribution *regarded as a probabilistic interpretation of quantum theory*. In a letter to Born in 1920, Einstein declared 'That business about causality causes me a lot of trouble, too. Can the quantum absorption and emission of light ever be understood in the sense of the complete causality requirement, or would a statistical residue remain? I must admit that there I lack the courage of my convictions. But I would be very unhappy to renounce *complete* causality' (Born, 1971, 23). And in 1924 Einstein expressed himself in even stronger terms: 'I find the idea quite intolerable that an electron exposed to radiation should choose *of its own free will*, not only its moment to jump off, but also its direction. In that case, I would rather be a cobbler, or even an employee in a gaming-house, than a physicist' (Born, 1971, 82).[14]

This early, Einsteinian version of PQT (repudiated by its author) would have had to have been modified, of course, once Schrödinger wave mechanics appeared on the scene. One of the great successes of Schrödinger's theory is that it predicts that the frequency of light emitted from an atom is equal to the frequency of the beats that arise because of the different frequencies of the electron in the higher and lower orbit, which in turn means that the atom is in a superposition of the two energy states during the process of emission. Such superpositions of energy levels have, in any case, been detected experimentally. This means we must take the view that such superpositions *exist* but do not persist. They collapse spontaneously and probabilistically when the flow of position probability density between the two states is very nearly zero – or, more precisely, when (II) of Section 10.7 obtains.

10.13 Why Has PQT been Ignored?

Given the important role that the Einsteinian version of PQT played in the history of quantum theory, given the power of PQT to make sense of the quantum domain and solve outstanding problems associated with OQT, and given that PQT may well be experimentally testable, the question naturally rises: Why has PQT been so resoundingly ignored?

The answer is that the physics community has failed to take probabilism seriously. Above all, the author of the first version of PQT abjured probabilism. If we go back to 1926 and to the advent of the new quantum theory of Heisenberg and Schrödinger, we find that those involved split into two camps. On the one hand there was the camp of Einstein, Schrödinger, von Laue and de Broglie, which held that both realism and determinism must be retained whatever the new quantum theory might seem to suggest. On the other hand, there was the camp of Bohr, Heisenberg, Born, Dirac, and most other physicists involved, which held that the new quantum theory necessitated the abandonment of both realism and determinism. These were the lines along which the great quantum battle of the time was drawn. What everyone overlooked was a third option – the only one capable of really making sense of the mysteries of the quantum domain: retain realism but abandon determinism

[14] Subsequently, Einstein came to appreciate that the fundamental objection to OQT is its abandonment of realism rather than determinism: see Born (1971, 168–173), Einstein (1950, 39–40). But Einstein never thought that probabilism might be the key to the solution to the basic problem confronting quantum realism – namely the wave/particle problem. For a discussion of Einstein's attitude towards OQT see Maxwell (1993a, 289–296).

and embrace probabilism instead. It is this third overlooked option that one needs to adopt in order to see the desirability – the *possibility* – of developing Einsteinian PQT so that it comes to provide a viable realistic and probabilistic version of post-Schrödinger quantum theory.

10.14 Conclusions

There are two conclusions.

First, PQT deserves more attention than it has received so far – both the specific version of PQT proposed here, and other, rival versions such as those of GRW and Penrose. There are a host of questions that need answering. What limit do existing experiments place on the upper bound of ε? What experiments are there to test PQT that could realistically be performed? How can PQT be extended to include relativistic quantum theory, QED and other quantum field theories? What are the implications of the probabilism of PQT for quantum gravity? How does the probabilism of PQT relate to the probabilism of theories, such as quantum electroweak theory, that may be regarded as postulating a cosmological episode of probabilistic spontaneous symmetry breaking? What implications does the probabilism of PQT have for views about the nature of time?

Second, in order to solve the problems of quantum theory, what is needed is an end to (usually rather bad) philosophizing about quantum theory, general recognition of the profound defects of OQT, and a return to the customary methods of physics in the search for a better theory: the twin activities of proposing testable conjectures, and subjecting them to experimental tests.

References

Albert, D. Z. (1996), Elementary quantum metaphysics, In J. Cushing, Fine, A., and Goldstein S. (eds.), *Bohmian Mechanics and Quantum Theory*. Dordrecht: Kluwer Academic, pp. 277–284.

Bell, J. S. (1987), *Speakable and unspeakable in quantum mechanics*. Cambridge: Cambridge University Press.

Bohr, N. (1949), Discussion with Einstein on epistemological problems in atomic physics, In P. A. Schilpp, (ed.), *Albert Einstein: Philosopher-Scientist*. La Salle, IL: Open Court, pp. 199–241.

Born, M. (1926), *Seitschrift f*ür *Physik* 37, 863.

Born, M. (1927), *Nature* 119, 354.

Born, M. (1971), *The Born-Einstein Letters*. London: Macmillan.

Einstein, A. (1950), Letter to Schrödinger, In K. Przibram (ed.), *Letters on Wave Mechanics.* Philosophical Library: New York, NY, pp. 39–40.

Fleming, G. (1989), Lorentz-Invariant state reduction and localization. In A. Fine and Forbes, M. (eds.), *Proceedings of the Philosophy of Science Association 1988*. East Lansing, Mich.: Philosophy of Science Association, vol. 2.

Fonda, L., Ghirardi, G. C., and Rimini, A. (1978), Decay theory of unstable systems, *Reports on Progress in Physics* 41, 587–631.

Ghirardi, G. C. (2002), Collapse theories, in stanford encyclopedia of philosophy, available online at http://plato.stanford.edu/achives/ spr2002/ entries/qm-collapse. Accessed 18 Sept 2010.

Ghirardi, G. C., Rimini, A., and Weber, T. (1986), Unified dynamics for microscopic and macroscopic systems, *Physical Review D,* 34, 470–491.

Jammer, M. (1966) *The Conceptual Development of Quantum Theory*, New York : McGraw-Hill.

Margenau, H. (1958), Philosophical problems concerning the meaning of measurement in physics. *Philosophy of Science* 25, 23–33.

Margenau, H. (1963), Measurements and quantum states: Part I. *Philosophy of Science* 30, 1–16.

Maxwell, N. (1968), Can there be Necessary Connections between Successive Events?, *British Journal for the Philosophy of Science* 19, 295–311.

Maxwell, N. (1972), A new look at the quantum mechanical problem of measurement, *American Journal of Physics* 40, 1431–1435.

Maxwell, N. (1973a), Alpha particle emission and the orthodox interpretation of quantum mechanics, *Physics Letters* 43A, 29–30.

Maxwell, N. (1973b), The problem of measurement – real or imaginary? *American Journal of Physics* 41, 1022–1025.

Maxwell, N. (1976), Towards a micro realistic version of quantum mechanics, Parts I and II, *Foundations of Physics* 6, 275–292 and 661–676.

Maxwell, N. (1982), Instead of particles and fields, *Foundations of Physics* 12, 607–631.

Maxwell, N. (1985), Are probabilism and special relativity incompatible? *Philosophy of Science,* 52, 23–43.

Maxwell, N. (1988), Quantum propensiton theory: A testable resolution of the wave/particle dilemma, *British Journal for the Philosophy of Science* 39, 1–50.

Maxwell, N. (1993a), Induction and scientific realism: Einstein, aim-oriented empiricism and the discovery of special and general relativity, *The British Journal for the Philosophy of Science* 44, 275–305.

Maxwell, N. (1993b), Beyond Fapp: Three approaches to improving orthodox quantum theory and an experimental test, In A. van der Merwe, Selleri, F., and Tarozzi, G. (ed.), *Bell's Theorem and the Foundations of Modern Physics*, Singapore: World Scientific, pp. 362–70.

Maxwell, N. (1994), Particle creation as the quantum condition for probabilistic events to occur, *Physics Letters A* 187, 351–355.

Maxwell, N. (1995), A philosopher struggles to understand quantum theory: Particle creation and wavepacket reduction, In M. Ferrero and van der Merwe, A. (ed.), *Fundamental Problems in Quantum Physics*, London: Kluwer Academic, pp. 205–214.

Maxwell, N. (1998), *The Comprehensibility of the Universe: A New Conception of Science*, Oxford: Oxford University Press, (paperback 2003).

Maxwell, N. (2004), Does probabilism solve the great quantum mystery? *Theoria* 19/3(*51*), 321–336.

Maxwell, N. (2006), Special relativity, time, probabilism and ultimate reality, In D. Dieks (ed.), *The Ontology of Spacetime*, Amsterdam: Elsevier, B.V., pp. 229–245.

Maxwell, N. 2011, Popper's paradoxical pursuit of natural philosophy, In J. Shearmur, and G. Stokes (eds.), *Cambridge Companion to Popper*, Cambridge: Cambridge University Press.

Pais, A. (1982), *Subtle is the Lord. . .* , Oxford: Oxford University Press.

Pais, A. (1986), *Inward Bound*, Oxford: Oxford University Press.

Penrose, R. (1986), Gravity and state reduction, In R. Penrose and Isham, C. J.(eds.), *Quantum Concepts of Space and Time*, Oxford: Oxford University Press, pp. 129–146.

R. Penrose (2004), *The Road to Reality*, London: Jonathan Cape.

Popper, K. (1957), The propensity interpretation of the calculus of probability and the quantum theory, In S. Körner (ed.), *Observation and Interpretation*, London: Butterworth, pp. 65–70.

Popper, K. (1959), *The Logic of Scientific Discovery*. London: Hutchinson (first published in 1934).

Popper, K. (1967), Quantum Mechanics without "The Observer", In M. Bunge (ed.), *Quantum Theory and Reality*, Berlin: Springer, pp. 7–44.

Popper, K. (1982), *Quantum Theory and the Schism in Physics*, London: Hutchinson.

Wallace, D. (2008), The philosophy of quantum theory, In D. Rickles (ed.), *Ashgate Companion to the New Philosophy of Physics*, London: Ashgate.

Chapter 11
Derivative Dispositions and Multiple Generative Levels

Ian J. Thompson

11.1 Introduction

Recently, much philosophical work has emphasized the importance of *dispositions* for realistic analyses of causal processes in both physics and psychology. This is partly because of the attractiveness of the thesis of *dispositional essentialism*, which holds that all existing things have irreducible causal powers, and such views are advocated in (Bird, 2004; Cartwright, 1983; Chakravartty, 2003; Elder, 1994; Ellis, 2000, 2001; Ellis and Lierse, 1994; Fetzer, 1977; Harré and Madden, 1975; McKitrick, 2003; Molnar, 2004; Mumford, 1995, 1998; Shoemaker, 1984; Swoyer, 1982 and Thompson, 1988). The thesis opposes the views of (Ryle, 1949: ch. 5) who sees dispositions as merely 'inference tickets' or 'promises', and (Armstrong, 1969) who sees them as derived from universal laws combined with non-dispositional properties. (Mumford, 2005) articulates a common aspect of dispositional essentialism, to imagine how the concept of universal laws could be rather *replaced* by talk of specific objects and their dispositions.

Recent critics of dispositional essentialism have pointed, for example, at Least Action Principles (Katzav, 2004), and Gauge Invariance Principles (Psillos, 2006), both of which principles appear to be independent laws that do not follow the pattern of aggregations with dispositions of the constituents. It might therefore appear that we have to move our understanding beyond that of simple dispositions. Related complexities are described in the works of (Krause, 2005) and (Stachel, 2005); who consider the difficulties arising from the identity of indistinguishable particles in quantum mechanics.

Certainly in physics and elsewhere, there are a great number of dispositional ideas such as *force*, *potential*, *propensity* that we should try to understand more systematically. There are other ideas of *energy*, *probability* and *virtual fields* that could well be linked with concepts of dispositions. Maybe a more sophisticated theory of dispositions will allow us to make headway in understanding least action principles

I.J. Thompson (✉)
Current address: Lawrence Livermore Laboratory, L-414, Livermore, CA 94551, USA
e-mail: IJT@ianthompson.org

M. Suárez (ed.), *Probabilities, Causes and Propensities in Physics*, Synthese Library 347, DOI 10.1007/978-1-4020-9904-5_11, © Springer Science+Business Media B.V. 2011

and gauge invariance within the framework of dispositional essentialism. I therefore continue the analysis of *kinds* of dispositions using suggestions from physics, to consider the possibility of what I will call *derivative dispositions*, and later consider whether these together may form a structure of what may be called *multiple generative levels*. This paper therefore consists of proposals for what those concepts might mean, and analyses of examples in physics and psychology that appear to need such concepts for their understanding. We will need to distinguish cases whereby new dispositions come about from rearrangement of parts, from possible cases where they are 'derived' or 'generated' in some more original way.

11.2 Beyond Simple Dispositions

11.2.1 Changing Dispositions

Most examples of dispositions in philosophical discussions are those, like fragility, solubility, radioactive instability, whose effects (if manifested) are *events*. If a glass exercises its fragility, it breaks. If salt shows its solubility, it dissolves, and the manifestation of radioactive instability would be a decay event detected say with a geiger counter. However, physicists want to know not merely *that* these *events* occur, but also *how the dispositions themselves may change* after the manifestation event. In the cases here, the fragility of the parts or the stability of the nuclei may change as a result of the manifestation events, and it is an important part of physics to describe the new (changed) dispositions as accurately as possible. Such descriptions are part of more comprehensive dynamical theories, as distinct from descriptive accounts of events.

Sometimes, *new* dispositions may be ascribable after an event that could not have been ascribed before the event. The fragments of a broken glass may be able to refract light in a way that the intact glass could not, for example. The dissolved salt may be to pass through a membrane, in contrast to the dispositions of the initial salt crystals. The fragments arising from a nuclear decay may possibly decay by emitting electrons in a way the parent nucleus could not.

In general, it often appears that new dispositions may be truthfully ascribed as the result of a prior disposition's operation. When this happens, as in the above examples, it at least appears that new dispositions come into existence as the manifestation of previous dispositions. Since now one disposition leads to another, some philosophical analysis is called for.

11.2.2 Rearrangement Dispositions

The existence of some of these new dispositions may perhaps be successfully explained as the rearrangement of the internal structures of the objects under discussion, when these are composite objects. The refraction by pieces of broken glass, in contrast to the original smooth glass, has obvious explanations in terms of the shapes of the new fragments. Salt's diffusion through a membrane, once dissolved, is presumably because of the greater mobility of salt ions in solution compared with the crystal form.

Science is largely successful in explaining such dynamical evolutions of empirical dispositions of natural objects. It bases the explanations in terms of changes in their structural shapes and arrangements of their parts, along with the fixed underlying dispositions or propensities of these parts. It is from the dispositions of these parts that, according the structure, all their observed dispositions and causal properties may be explained.

The existence of new dispositions by rearrangement of the parts of an object may be taken as non-controversial within existing philosophical frameworks. It appears that typical philosophical analyses can readily accommodate the way the derivative dispositions of an *aggregate* are explained in terms of recombinations of the dispositions of its parts.

11.2.3 Derivative Dispositions

However, it appears that not all dynamical changes of dispositions occur by rearrangements of parts, and those that are *not* rearrangements are what in this paper I want to call *derivative dispositions*. There are some cases, to be listed below, where new dispositions come into existence, without there being any known parts whose rearrangement could explain the changes. The next section gives some examples from physics and psychology of what appear to be such derivative dispositions, and this is followed by a more general analysis of how these might work.

11.3 Examples of Derivative Dispositions

11.3.1 Energy and Force

If we look at physics, and at what physics regards as part of its central understanding, one extremely important idea is *energy*. Physics talks about kinetic energy as energy to do with motion, and potential energy as to do with what *would happen* if the circumstances were right. More specifically, if we look at definitions of force and energy which are commonly used to introduce these concepts, we find definitions like

- *force:* the tendency F to accelerate a mass m with acceleration F/m.
- *energy:* the capacity E to do work, which is the action of a force F over a distance d,
- *potential energy field:* the field potential $V(x)$ to exert a force $F = -dV/dx$ if a test particle is present.

Furthermore, we may see a pattern here:

- *potential energy field:* the *disposition* to generate a *force*, and
- *force:* the *disposition* to accelerate a mass, and
- *acceleration:* the final result.

We cannot simply identify for example ‘force’ and ‘acceleration’, because, as (Cartwright, 1983 points out, force is not *identical* to the product *ma*: it is only the *net force* at a point which can have that effect. An individual force is only by itself a tendency which may or may not be manifested. It is a disposition, as is energy generically, as well as potential energy. It is generally acknowledged that ‘force’ is a disposition: my new point is that it cannot be reduced either to ‘acceleration’ or ‘energy’.

I therefore take these as an example of two successive *derivative dispositions*, where the effect of one disposition operating is the generation of another. An electrostatic field potential is a disposition, for example, the manifestation of which – when a charge is present – is not itself motion, but which is the presence now of a derivative disposition, namely a force. The manifestation of a force – when acting on a mass – may or may not occur as motion, as that depends on what other forces are also operating on the mass. The production of a force by a field potential does not appear to be something that occurs by means of the rearrangements of microscopic parts, but appears to be more fundamental, and almost sui generis. It appears that *field potentials*, *force* and *action* form a set of multiple generative levels, and this situation is clearly in need of philosophical inspection.

Admittedly, many physicists and philosophers often manifest here a tendency to say that *only* potential energy is ‘real’, or conversely perhaps that ‘only forces are real’, or even that ‘only motion is real’, and that in each case the other physical quantities are only ‘calculational devices’ for predicting whichever is declared to be real. Please for a while apply a contrary tendency to resist this conclusion, at least to the end of the paper. In Section 11.5 I will be explicitly evaluating such reductionist strategies, along with a discussion of the comparative roles of mathematical laws and dispositional properties within a possible dispositional essentialism.

11.3.2 Sequences, or Levels?

We normally *think* of energy, force and acceleration as the sequential stages of a process. However, in nature, there is still energy even after a force has been produced, and forces continue to play their roles both during and after accelerations. This means energy does not finish when force begins, and force does not finish when acceleration begins, but, in a more complicated structure, all three continue to exist even while producing their respective derivative dispositions. The best way I can find to think better of this more complicated structure is that of a set of ‘multiple generative levels’. In this way we can think of a ‘level of energy’ which persists even while it produces forces, and is ‘level of forces’ even as they produce accelerations.

Admittedly the idea of a 'level' is a spatial metaphor for what is not itself spatial, but the metaphor still serves to illustrate my argument.

11.3.3 Hamiltonians, Wave Functions and Measurements

In quantum physics, energy (the total of the kinetic and potential energies) is represented by the Hamiltonian operator $\hat{H}$. This operator enters into the Schrödinger wave equation $\hat{H}\ \Psi(x,t) = i\hbar\partial\Psi(x,t)/\partial t$, which governs all quantum wave forms $\Psi(x,t)$. It thus generates all time evolution, and hence all fields of probabilities $|\Psi(x,t)|^2$ for measurement outcomes. The principal dynamics in quantum physics are specified by knowing what the initial state is, and what the Hamiltonian operator is. These remarks apply to quantum mechanics as it is practised, by using Born's statistical interpretation and then naively saying that the quantum state changes after a measurement to one of the eigenstates of the measurement operator. (This is the much discussed 'reduction of the wave packet', which we may agree at least appears to occur.)

We may therefore consider quantum physics in the following 'realistic' way. We have the Hamiltonian which is to do with total energy, which is somehow 'active' since it is an *operator* which operates on the wave function and changes it. The Schrödinger equation is the rule for how the Hamiltonian operator produces the wave function, which is a probabilistic disposition (a propensity) for action. This wave function (in fact its squared modulus) gives a probability for different of macroscopic outcomes of experiments, and the wave function changes according to the specific outcome.

Such is the structure of quantum physics as it is practised, and we may observe a sequence of derivative dispositions in operation:

- *Hamiltonian operator:* the fixed disposition to generate the wave function by evolving it in time,
- *wave function:* the probabilistic disposition (a 'propensity wave') for selecting measurement outcomes, and
- *measurement outcome:* the final result.

It appears again that we have multiple generative levels, with the set of *Hamiltonian*, *wave function* and *selection event*. Note here also that the final result is the weakest kind of 'minimal' disposition, which influences merely by selection, because it *is* a selection. It appears as the last of a sequence of derivative dispositions, as a kind of 'bottom line' if we want to include it within the framework of multiple generative levels.

Admittedly again, reductionist tendencies may be applied. Most commonly, it may be denied that there are distinct measurement outcomes in any ontological sense, and that they may only be approximately defined within a coarse-grained 'decoherent history'. Advocates of the Many Worlds Interpretation, or of Decoherence theories, take this view. Others such as Bohr take the opposite view:

he holds that only the measurement outcome is real, and that the Hamiltonian and wave function are calculational devices and nothing real. These views in tension will be discussed in Section 11.5.

11.3.4 Virtual and Actual Processes

Taking a broader view of contemporary physics and its frontiers, we may further say that the 'Hamiltonians, wave functions and measurements' of above describe just the dispositions for a class of 'actual processes'. The Hamiltonian is the operator for the total energy, containing both kinetic and potential energy terms. However, we know from Quantum Field Theory (QFT) that, for example, the Coulomb potential is composed 'in some way' by the exchange of virtual photons. Similarly, we also know from QFT that the mass in the kinetic energy part is not a 'bare mass', but is a 'dressed mass' arising (in some way) also from many virtual processes. This again suggests the theme of my paper: that the Hamiltonian is not a 'simple disposition', but in fact is *itself derivative* from some prior 'generative level'. In this case the needed generative level could be called that of 'virtual processes', in contrast to that of 'actual processes'.

The class of virtual processes, as described by QFT, have many properties that are opposite to those of *actual* processes of measurement outcomes. Virtual events are at points (not selections between macroscopic alternatives), are interactions (not selections), are continuous (not discrete), are deterministic (not probabilistic), and have intrinsic group structures (e.g. gauge invariance, renormalisation) as distinct from the branching tree structure of actual outcomes. All these contrasts (which I do not have the space to expound here) suggest that virtual processes should be distinguished from actual events. The guiding principles have different forms: virtual processes are most commonly described by a Lagrangian subject to a variational principle in a Fock space of variable particle numbers, whereas actual processes, as discussed above, deal with the energies of specific observable objects leading to definite measurement outcomes.

It is more certain that virtual processes form a simultaneous 'level' in addition to the 'level' of Hamiltonians, propensities and measurements. This is because virtual processes are clearly occurring perpetually and simultaneously with Hamiltonian evolution, as they are necessary to continually 'prepare and form' the 'dressed' masses and potentials in the Hamiltonian. Dressed masses and potentials persist during Hamiltonian evolution. In atoms and molecules, virtual processes such as photon exchanges to generate the Coulomb potentials exist continuously as a kind of background for observable processes.

11.3.5 Pregeometry and the Generation of Spacetime

Field theories such as QFT still use a geometric background of spacetime, and there is currently much speculative work in quantum gravity research to determine how this spacetime might arise. Wheeler started interest in 'pregeometry': the attempt to formulate theories of causal processes which do not presuppose a differentiable manifold for spacetime. Rather, his aim was to encourage speculation as to how spacetime might *arise*. Most commonly, the task has been taken as showing how spacetime may turn out to be a 'statistical approximation' in some limit of large numbers of hypothetical pregeometric processes. Proposals have involved spinors by (Penrose, 1987); 'loop quantum gravity' as described for example in (Rovelli, 1998); and 'causal sets' according to (Brightwell et al., 2003).

If some pregeometry could be identified, I would speculate that a good way of seeing this would be as a distinct pregeometric level within a structure of derivative dispositions. That is, instead of spacetime being a statistical approximation (in the way thermodynamics is a statistical approximation to molecular gas theories), it could be better imagined that spacetime is an aspect of derivative dispositions that have been generated by 'prior' pregeometric dispositions. This is admittedly speculative, but it does follow the pattern of some current research, so I use it as an example of how the philosophical analysis of dispositions may yet interact fruitfully with modern physics. This appears to be useful particularly since the very aim of 'deriving spacetime' has itself been called into question by (Meschini et al., 2005).

11.3.6 Psychology

There are many examples of apparent derivative dispositions in everyday life, in psychology, in particular in cognitive processes. Such dispositions are involved whenever the accomplishment of a given disposition requires the operation of successive steps of kinds different from the overall step. The original disposition on its operation therefore generates the 'derived dispositions' for the intermediate steps, which are means to the original end. An original 'disposition to learn', for example, can generate the derived 'disposition to read books', which can generate further 'dispositions to search for books'. These dispositions would then generate dispositions to move one's body, which in turn lead ultimately to one's limbs having (physical) dispositions to move. These successively generated dispositions are all *derived* from the original disposition to learn, according to the specific situations.

Another example of sequential and derivative dispositions is the ability to learn. To say that someone is easy to teach, or that they are musical, for example, does not mean that there is any specific action that they are capable of doing. Rather, it means that they well disposed to learn new skills (whether of a musical or of a general kind), and that it is these new skills which are the dispositions that lead to specific actions.

In this I follow (Broad, 1925): that there are 'levels' of causal influence. We might allow that particular dispositions or intentions are best regarded not as the most fundamental causes, but as 'intermediate stages' in the operation of more persistent 'desires' and 'motivations'. The intention to find a book, for example, could be the product or derivative of some more persistent 'desire for reading', and need only be produced in the appropriate circumstances. Broad would say that the derived dispositions were the *realisation* of the underlying dispositions. These are called 'levels' rather than simply 'sequences' because the underlying motivation still exists during the production of later levels: it operates simultaneously with the derivative dispositions. It is not the case that 'desire for reading' ceases during the act of reading, for it is rather then at its strongest and in fulfilment.

11.4 Analytical Scheme

11.4.1 Generative Sequences

The first general idea is that 'multiple generative levels' are a sequence $\{A \rightarrow B \rightarrow C \rightarrow ..\}$ in which A 'generates' or 'produces' new forms of B using the present form of B as a precondition. We say that B derives from A as its manifestation. Then B generates C in the same way. This sequence may perhaps continue until an end Z, say, where the only activity is the 'selection' described below.

This rough scheme does not tell us, however, how A, B, etc might be *changed* as a result of their operation. Presumably this occurs often, as for example in naive quantum theory, when a wave function is changed after it generates a particular measurement outcome. We want to consider the philosophy for a general scheme which might explain the (apparently mysterious) logic of the 'reduction of the wave packet'. In order to formulate such a general scheme, let us extract some guidelines from our example derivative dispositions listed previously. To do this, we will need to first distinguish the concepts of principal from instrumental and occasional causes.

11.4.2 Principal, Instrumental and Occasional Causes

(Davidson, 1967) argues that causality is a two-place relation between individual events. Thus causal relations are certainly not just implications from the description of the first event to that of the second event, but are something more real. The reality of causality, however, does not thereby automatically include such components as dispositions and propensities, although (Steiner, 1986) wants to extend Davidson's ideas in this direction. In the present paper, I want to allow *both* dispositions *and* previous events to be causes, although in different senses.

Thus I recommend that distinctions ought to be made between all of the following:

- the 'Principal Cause': that disposition which operates,
- the 'Occasional Cause': that circumstance that selects which dispositions operate,
- the 'Instrumental Cause': the origin of the occasional cause, so is another cause by means of which the Principal Cause operates.

The overall pattern is therefore that 'Principal causes operate according to occasional causes, which arise from instrumental causes'.

All three kinds of causes appear to be necessary for any event in nature, for example, when a stone is let fall: the principal cause is the earth's gravitational attraction, the occasional cause is our *act* of letting go, and instrumental cause is the muscle movements in our finger releasing the stone. Its hitting the ground is thus caused by our letting go, but only as an instrumental and then occasional cause. Many common uses of 'cause' (including that of (Davidson, 1967)) refer to occasional causes rather than principal causes, as it is only in this 'occasional' sense that *events* can be said to be causes. Previous events cannot be efficacious causes, (Emmet, 1984) points out, in the sense of 'producing' or 'giving rise to' their effects, since events per se are not themselves powers, but clearly they do make some difference whether they happen or not. This is because events are the changes in powers, but change itself is not a power but the property of powers. The instrumental cause *is* a genuine causal contributor, and may be said to 'set the stage', by making suitable conditions (namely, the occasional cause) for selecting the operation of the principal cause.

I acknowledge that using the phrase 'occasional cause' brings in perhaps an unnecessary amount of philosophical debate, but I see essentially the same questions occurring here as there. We need some generic concept to refer to the circumstances, conditions, or occasions that must obtain in order for a disposition to manifest itself.

11.4.3 Causal Sequences in Physics

Consider now a electron of fixed charge and mass moving in an electrostatic potential, according to classical electrostatics. At a given place x, the derivative of the potential $V(x)$ gives the force, and the force gives acceleration which in turn changes the velocity of electron, and it moves to a new place. In our framework of derivative dispositions, we see that the potential is a disposition which generates another, namely the force. It does so, moreover, according to the place of the electron. The electrostatic potential is therefore the principal cause of the force, and the place of the electron is the occasional cause. A place or any other spatiotemporal property *by itself* is never an efficacious cause, but it can be said to be the circumstance by means of which the potential generates the force. In general, when we include magnetism and radiation, such properties will include velocities and accelerations. Perhaps these properties will themselves require further dispositional analyses as in (Lange, 2005).

Note that we never have forces causing potentials to exist where they did not before, nor can places. Let us generalise by surmising a set of generative levels {Potential $\rightarrow$ Force $\rightarrow$ Places}, such that the principal causation is always in the direction of the arrow, and the only 'backward' causation is by selection with an occasional cause. The only feedback 'back up the sequence' is therefore with the conditional aspect of certain occasions. The specific operation of prior dispositions does not happen continually or indiscriminately, so needs to be selected, and thus there is an essential role for 'particular occasions' as preconditions.

Consider secondly the quantum mechanical evolution of a system from time t_0 that is subject to measurement selections at various later times t_1, t_2, etc. The quantum mechanical story is as follows. The initial quantum state $\Psi(t_0)$ is evolved according to the Schrödinger equation by the Hamiltonian $\hat{H}$ for $t < t_1$. Consider the measurement for operator $\hat{\mathrm{A}}$ occurring at $t = t_1$, the operator having an eigen-expansion $\hat{\mathrm{A}}u_\lambda = a_\lambda u_\lambda$. In practical quantum mechanics, the quantum state changes to $\Psi(t_1^+) = u_\lambda$ if the result of the measurement is the eigenvalue a_λ, which occurs with probability $p_\lambda = |< u_\lambda| \Psi(t_1) >|^2$. The new state $\Psi(t_1^+)$ is then evolved similarly for $t < t_2$, the time of the next measurement.

Seen in terms of derivative dispositions, the Hamiltonian is the disposition to evolve an initial state $\Psi(t_0)$ to new times t, generating $\Psi(t) = \exp(-i\hat{H}t/\hbar)\Psi(t_0)$. The new $\Psi(t)$ are themselves another disposition, namely a propensity to produce measurement outcomes with the various probabilities $p_\lambda = | < u_\lambda|\Psi(t) > |^2$. The final results are the discrete selection events at the times of measurement. These discrete events have themselves only the minimal causal powers as they influence the future evolutions of the wave function. In that sense, they are in our scheme just the 'occasional causes' according to which other dispositions may operate. The principal dispositions are first the Hamiltonian operator that starts the whole process, and then the wave functions considered as fields of propensity for different selection events.

Summarising the quantum mechanical case, we see that here again, the principal causes act 'forwards' down a set of multiple generative levels, yet whose range of actions at any time is selected from all those presently possible, as constrained by past events. Those events thereby become occasional causes. Because the wave functions before a measurement event are the cause of that event, those wave functions are thereby the instrumental cause of the new wave functions after the measurement.

11.4.4 Conditional Forward Causation

From our examples, we may generalise that all the principal causation is 'down' the sequence of multiple generative levels {A $\rightarrow$ B $\rightarrow$ ...}, and that the only effect back up the sequence is the way principal causes depend on previous events or occasions to select their range of operation. Let us adopt as universal this asymmetric relationship between multiple generative levels: *that dispositions act forwards in a way conditional on certain things already existing at the later levels*. We regard this as a simple initial hypothesis, and will have to observe whether all dispositions taken as existing in nature can be interpreted as following this pattern.

We may therefore surmise that A, the first in the sequence, is the 'deepest underlying principle', 'source', or 'power' that is fixed through all the subsequent changes to B, C, etc. Conditional Forward Causation, the pattern we saw from physics, would imply that changes to B, for example, come from *subsequent* operations of A, and not from C, D,.. acting in 'reverse' up the chain. We would surmise, rather, that the *subsequent* operations of A are now conditioned on the results in B, C, D, etc. The operations of A are therefore the *principal causes*, whereas the dependence of those operations on the previous state of B is via *instrumental causation*, and the dependence on the results in C, D,... is via *occasional causation*. I would like to suggest that this is a universal pattern for the operation of a class of dispositions in nature, namely those that do not follow from the rearrangement of parts of an aggregate object.

11.5 Reductionism and Dispositional Essentialism

In all the apparent examples of multiple generative levels given here, many physicists and philosophers of physics will want to assert the particular 'reality' of one of the levels, and say that the prior levels are 'merely calculational devices' for the behaviour of their chosen real level.

For example, some assert in electromagnetic theory that only the field tensors (incorporating the electric and magnetic vector fields) are 'real', and that the vector potential (incorporating the electrostatic potential) is a calculational device with no reality. To this end, they note the gauge uncertainties in the vector potential, which for electrostatics is the arbitrariness in setting the level of zero potential energy. Against this, many have noticed that the scattering of electrons in the Bohm-Aharonov experiment is most succinctly explained in terms of the vector potential, not the field tensor. It turns out that, strictly speaking, it is loop integrals of the vector potential which carry physical significance. I conclude that there are non-trivial physical and philosophical questions about the relative 'reality' of potentials and forces which require not immediate preferences but considered responses.

We also saw how reductionist tendencies may be manifest in quantum theories. 'Decoherent history' accounts of quantum mechanics want to keep the wave function according to the Schrödinger equation, and deny that macroscopic outcomes occur in a reality, and only allow them to be approximate appearances. The founders of quantum theory such as Bohr and Wheeler, however, took the opposite view, that an electron is only 'real' when it is being observed – when it makes the flash of light at a particular place – not while it is travelling. In their opposite view, the Hamiltonian and wave function are calculational devices and nothing real, having only mathematical reality as portrayed by the mathematical name 'wave function'.

The views which make prior or later levels into 'mere' calculational devices can be critiqued from the point of view of dispositional essentialism. This view encourages us in general to *not* invoke arbitrarily mathematical rules for the laws of nature, but, as (Mumford, 2005) suggests, *replace* the role of laws by that of the dispositional properties of particular objects. The question of simplicity, to be answered in order to apply Occam's criterion, is therefore whether it is simpler to have multiple kinds of objects existing (even within multiple generative levels) each with simple dispositions, or simpler to have fewer kinds of existing objects, but with more complicated laws governing their operation. The discussion in the literature about interpreting the Bohm-Aharonov effect is trying to answer precisely this question, once it had been established that different approaches were both adequate in explaining the phenomenon.

In the present paper, I have shown many more apparent examples of multiple generative levels, each composed of derivative dispositions. The questions of simplicity, and adequacy, will have to be examined in all of these cases as well. Nevertheless, I believe that the concepts introduced here enable us to take a more comprehensive and universal view of physical dispositions (such as those of potentials and forces, or of Hamiltonians and wave functions) that otherwise appear to be *ad hoc* when taken individually.

References

Armstrong, D. M. (1969), Dispositions are causes, *Analysis* 30, 23–26.

Bird, A. J. (2004), The dispositionalist conception of laws, *Foundations of Science* 152, 1–18.

Brightwell, G Dowker, F., García, R. S., Henson, J., and Sorkin, R. D. (2003), Observables in causal set cosmology, *Physical Review D* 67, 084031.

Broad, C. D. (1925), *Mind and Its Place in Nature*. London: Routledge and Kegan Paul.

Cartwright, N. (1983), *How the Laws of Physics Lie*. Oxford: Clarendon Press.

Chakravartty, A. (2003), The dispositional essentialist view of properties and law, *International Journal of Philosophical Studies* 11, 393–413.

Davidson, D. (1967), Causal relations, *Journal of Philosophy* 64, 691–703.

Elder, C. L. (1994), Laws, natures, and contingent necessities, *Philosphy and Phenomenological Research* 54, 649–67.

Ellis, B. (2000), Causal laws and singular causation, *Philosphy and Phenomenological Research* 61, 329–351.

Ellis, B. (2001), *Scientific Essentialism*, Cambridge: Cambridge University Press.

Ellis, B. and Lierse, C. (1994). Dispositional essentialism, *Australian Journal of Philosophy* 72, 27–45.

Emmet, D. (1984), *The Effectiveness of Causes*, New York: State University of New York Press.

Fetzer. J. H. (1977), A world of dispositions, *Synthese* 34, 397–421.

Katzav, J. (2004), Dispositions and the principle of least action, *Analysis* 64, 206–214.

Krause, D. (2005), Separability and Non-Individuality: Is it possible to conciliate (at least a form of) Einstein's realism with quantum mechanics? PhilSci-Archive, URL: http://philsci-archive.pitt.edu/archive/00002431/. Accessed 5 Jan 2007.

Harre, R. and Madden, E. H. (1975). *Causal Powers*, Oxford: Oxford University Press.

Lange, M. (2005), How can instantaneous velocity fulfill its causal role? *Philosophical Review* 114, 433–468.

McKitrick, J. (2003), The bare metaphysical possibility of bare dispositions, *Philosphy and Phenomenological Research* 66, 349–369.

Meschini, D., Lehto, M., Piilonen, J. (2005). Geometry, pregeometry and beyond, *Studies in History and Philosophy of Modern Physics* 36, 435–464.

Molnar, G. (2004), *Powers*, Oxford: Oxford University Press.

Mumford, S. (1995), Ellis and Lierse on dispositional essentialism, *Australian Journal of Philosophy* 73, 606–12.

Mumford, S. (1998), *Dispositions*, Oxford: Oxford University Press.

Mumford, S. (2005), Laws and Lawlessness, *Synthese* 144, 397–413.

Penrose, R. (1987), *Spinors and Space-Time: Volume 1, Two-Spinor Calculus and Relativistic Fields*, Cambridge: Cambridge University Press.

Psillos, S. (2006), What do Powers do when they are not Manifested? *Philosphy and Phenomenological Research* 72, 137–156.

Rovelli, C. (1998), Loop Quantum Gravity, *Living Reviews of Relativity* 1: 1. URL: http://www.livingreviews.org/lrr-1998-1.

Ryle, G. (1949), *The Concept of Mind*, Chicago: University of Chicago Press.

Shoemaker, S. (1984), *Identity, Cause, and Mind*, Oxford: Oxford University Press, (2nd ed., 2003).

Stachel, J. (2005), Structure, individuality and quantum gravity, In D. P. Rickles, French, S. R. D., and Saatsi, J (eds.) *Structural Foundations of Quantum Gravity*, Oxford: Oxford University Press, pp. 53–82.

Steiner, M. (1986), Events and Causality, *Journal of Philosophy* 83, 249–264.

Swoyer, C. (1982), The Nature of Natural Laws, *Australian Journal of Philosophy* 60, 203–223.

Thompson, I. J. (1988), Real dispositions in the physical world, *British Journal Philosophical Science* 39, 67–79.

Name Index

Subject Index