New Vistas on Old Problems

Recent Approaches to the Foundations of Quantum Mechanics

Max Planck Research Library for the History and Development of Knowledge

Proceedings 3

Edition Open Access
2017

New Vistas on Old Problems

Recent Approaches to the Foundations of Quantum Mechanics

Proceedings of an International Symposium, Bern, June 23–24, 2011

Tilman Sauer and Adrian Wüthrich (eds.)

Edition Open Access
2017

Max Planck Research Library for the History and Development of Knowledge
Proceedings 3

Communicated by:
Domenico Giulini

Edited by:
Tilman Sauer and Adrian Wüthrich

ISBN 978-3-945561-20-1
Published 2017 by Edition Open Access,
Max Planck Institute for the History of Science
Reprint of the 2013 edition
Printed and distributed by
PRO BUSINESS digital printing Deutschland GmbH, Berlin
Edition Open Access
http://www.edition-open-access.de

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.d-nb.de.

The Max Planck Research Library for the History and Development of Knowledge comprises four subseries, *Studies*, *Proceedings*, *Sources* and *Textbooks*. They present research results and the relevant sources in a new format, combining the advantages of traditional publications and the digital medium. The volumes are available both as printed books and as online open access publications. They present original scientific work submitted under the scholarly responsibility of members of the Scientific Board and their academic peers.

The volumes of the four subseries and their electronic counterparts are directed at scholars and students of various disciplines, as well as at a broader public interested in how science shapes our world. They provide rapid access to knowledge at low cost. Moreover, by combining print with digital publication, the four series offer a new way of publishing research in flux and of studying historical topics or current issues in relation to primary materials that are otherwise not easily available.

The initiative is supported, for the time being, by research departments of three Max Planck Institutes, the MPI for the History of Science, the Fritz Haber Institute of the MPG, and the MPI for Gravitational Physics (Albert Einstein Institute). This is in line with the *Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities*, launched by the Max Planck Society in 2003.

Each volume of the *Studies* series is dedicated to a key subject in the history and development of knowledge, bringing together perspectives from different fields and combining source-based empirical research with theoretically guided approaches. The studies are typically working group volumes presenting integrative approaches to problems ranging from the globalization of knowledge to the nature of spatial thinking.

Each volume of the *Proceedings* series presents the results of a scientific meeting on current issues and supports, at the same time, further cooperation on these issues by offering an electronic platform with further resources and the possibility for comments and interactions.

Each volume of the *Sources* series typically presents a primary source—relevant for the history and development of knowledge—in facsimile, transcription, or translation. The original sources are complemented by an introduction and by commentaries reflecting original scholarly work. The sources reproduced in this series may be rare books, manuscripts, documents or data that are not readily accessible in libraries and archives.

Each volume of the *Textbooks* series presents concise and synthetic information on a wide range of current research topics, both introductory and advanced. They use the new publication channel to offer students affordable access to high-level scientific and scholarly overviews. The textbooks are prepared and updated by experts in the relevant fields and supplemented by additional online materials.

On the basis of scholarly expertise the publication of the four series brings together traditional books produced by print-on-demand techniques with modern information technology. Based on and extending the functionalities of the existing open access repository European Cultural Heritage Online (ECHO), this initiative aims at a model for an unprecedented, Web-based scientific working environment integrating access to information with interactive features.

Contents

Introduction

Tilman Sauer and Adrian Wüthrich

This volume presents a collection of contributions to the current debate on the interpretation of quantum theory. The collection is neither intended to give a comprehensive overview of a highly active field of philosophical research, nor is it intended to provide a representative collection. The contributions gathered here raise problems, probe alleys of exploration, and try to look at old problems from new or unusual perspectives. We believe that quantum theory by its theoretical structure and empirical validity raises problems that should and can be addressed in a dialogue between physicists and philosophers: we believe that quantum mechanics is in need of interpretation.

The papers of this volume were prepared for a symposium on current interpretational problems of quantum theory held at the University of Bern in June 2011. The symposium was a little bit more specific in its focus. It was announced under the title "Decoherence and No-Signalling." Let us expand a bit on the idea behind the symposium.

As a matter of historical fact, quantum theory has been a subject of interpretational debates ever since its inception. Physicists were puzzled by the quantization of energy that seemed necessary in order to understand black-body radiation, and they were troubled about the status of Niels Bohr's mysteriously successful, yet, axiomatically stipulated quantum postulates. The emergence of modern quantum mechanics is a process of reinterpreting old classical concepts and of trying to come to an understanding of new quantum concepts in a situation that was often explicitly perceived as one of theoretical crisis. Even after the establishment of quantum mechanics in the mid-twenties interpretational questions kept raising their heads. Why does quantum theory often not predict the outcome of a measurement but only give the probability distribution over possible outcomes? Does this reflect our ignorance of some relevant features of the system, do quantum systems not always evolve deterministically? This is just one complex of questions which came up early on in the development of quantum theory.

T. Sauer, A. Wüthrich (eds.), *New Vistas on Old Problems. Recent Approaches to the Foundations of Quantum Mechanics*, Edition Open Access, 2013.

The development of quantum mechanics, from about 1900 to the mid-1930s, was intimately linked to discussions of its consequences for our scientific world view. Albert Einstein, John von Neumann, and Erwin Schrödinger were among those who pointed out in particularly sharp ways the conceptual concerns of many physicists and philosophers with determinism, causality, and observability in quantum mechanical contexts.

In 1932, von Neumann laid down the principles of quantum theory in a concise axiomatic formulation. He also formulated what is, in fact, the common core of most of the aforementioned concerns: quantum mechanics exhibits a "measurement problem." According to quantum mechanics, a "collapse" of the wavefunction unpredictably interrupts the deterministic evolution of a quantum mechanical system in the course of a measurement process. However, quantum mechanics does not provide any criterion of what constitutes a measurement. Quantum mechanics does not tell us to which domain we should restrict the application of the Schrödinger equation. When is the interaction of two systems a "measurement" (which is described by the Schrödinger equation) or an interaction of two quantum systems (which is not)?

In 1935, Schrödinger took the measurement problem to extremes by showing that the unrestricted application of his eponymous equation leads to superpositions of macroscopic objects, such as cats, which should be neither dead nor alive, nor both, nor none of the two. It seems hard to reconcile superposed states of macroscopic objects, such as cats, with our experience, and, insofar, there has to be something which distinguishes the interaction of macroscopically observable objects with quantum systems from the interaction among quantum systems themselves.

Also in 1935, in a joint paper with Nathan Rosen and Boris Podolsky, Einstein challenged the completeness of the quantum mechanical description of composed systems. They considered two sub-systems, say 1 and 2, which, on reasonable grounds, have ceased to interact. They find quantum mechanical descriptions which assign a definite position and momentum to sub-system 2 only if a measurement is performed on sub-system 1. However, they assume, the measurement operations on one sub-system does not influence the physical processes of the other. Sub-system 2 should, therefore, be assigned a definite position and momentum even before a measurement on sub-system 1 has taken place. Because quantum mechanics does not make this assignment, they conclude that it is incomplete.

On the other hand, if the quantum mechanical description is taken to be complete, sub-system 2 is changed from a state with no definite position and momentum to a state with definite position and momentum by the measurement operations on sub-system 1. If the quantum mechanical description is taken to be complete, the collapse of the wave function has to be considered a real physical process.[1]

But the collapse of sub-system 2 is caused *instantaneously* by the measurement of the distant sub-system 1—certainly a "spooky action-at-a-distance." Einstein thus revealed another problematic aspect of the collapse of the wave-function: Not only does quantum mechanics not specify how and why a collapse happens but, also, the collapse of the wave-function is difficult to reconcile with the special theory of relativity and with more general established principles of the separability of composed systems.

In particular through the work of Niels Bohr, an "orthodox" response to all these problems was established as a received view. The so-called Copenhagen interpretation held that those worries by Einstein and others about the interpretational consequences of the theory were unfounded. Although quantum mechanical phenomena continued to attract the attention of philosophers and physicists concerned with foundational issues, several events and developments contributed to the confirmation of this increasingly widely accepted response to quantum mechanics' interpretational problems.

Many physicists regard the theoretical description of decoherence processes as a mere elaboration of the Copenhagen interpretation. The measurement problem is claimed to be overcome by taking into account how superposed quantum mechanical systems disentangle rapidly through the interaction with a many-particle environment. Numerical models and estimates show that decoherence times are much shorter than the resolution of presently feasible measurement techniques.

In a similar vein, the empirical violation of Bell's inequality is taken as a crucial experiment showing the quantum mechanical description may be complete, notwithstanding Einstein's challenge. On this reading, Einstein, together with Podolsky and Rosen, argued that if physical systems interact only locally, the quantum mechanics is an incomplete description of physical systems and their interaction.

[1] The most recent and precise versions of arguments, along these lines, for the reality of the collapse of the wave function have been given by Pusey et al. (2012) and Colbeck and Renner (2012).

However, in 1964, John S. Bell spelled out the notion of locality, which Einstein seems to have had in mind, and showed that any such local theory satisfies an empirically testable inequality. Quantum mechanics, on the other hand, predicted a violation of these inequalities. In the early 1980s, experiments showed Bell's inequality to be violated, a result that vindicated quantum mechanics and, at the same time, ruled out any local alternative. Einstein's reasoning to think quantum mechanics is incomplete was thereby neutralized.

Moreover, it has been proven that the kind of non-locality which the empirical violation of Bell's inequality requires cannot be exploited to send signals faster than light. These no-signalling proofs seem to guarantee a "peaceful coexistence" (Abner Shimony) between quantum mechanics and the special theory of relativity.

Recent philosophical arguments, however, challenge the "new orthodoxy" (Jeffrey Bub) and its appeal to decoherence and no-signalling theorems. As a linear type of evolution, decoherence cannot make superpositions of quantum mechanical states disappear. Therefore, even when complemented by theorems and models of decoherence, quantum mechanics cannot dispense with the need for a collapse of the wave-function, or the need for an explanation of how superposed states are compatible with our experience.

This still generates the essential conundrum of the measurement problem. Similarly, the no-signalling theorems cannot provide what the new orthodoxy requires. The impossibility of sending signals faster than light does not ensure that the core principle of special relativity, Lorentz-invariance, can be satisfied by theories which describe the violation of Bell's inequality. Also, if the non-locality of quantum mechanics is accepted, there is no straightforward reason to dismiss non-local hidden variable theories, such as Bohm's. They are not excluded by the violation of Bell's inequality. If non-locality is accepted, there is no straightforward and sound argument any more why quantum mechanics should be regarded complete. (As mentioned before, there is also no straightforward and sound argument any more why quantum mechanics should be regarded *incomplete*.)

The objective of the symposium was to critically assess whether there still is today a problem of interpretation of quantum mechanics. Taking into account the most recent pertinent developments in philosophy and physics, invited speakers updated an audience interested in both philosophy and physics on the current state of research. Speakers and the audience engaged in a discussion, which challenged the different positions.

Needless to say, that the symposium did not solve any one of the outstanding interpretational problems. What it did was to provide a forum of debate and, in this debate, presented a spectrum of problems, difficulties, approaches, and perspectives. It was meant to be and turned out to be an unlimited and an open-minded debate. Neither did the participants succeed in, or even intend to, exhaust the problems in their various aspects and disguises, nor was any one point discussed until proven valid or invalid. But what the participants did do was to seriously engage in specific aspects of their choosing, laying out difficulties to the best of their understanding and presenting approaches up to a point where they could be taken up by others. The spirit of the symposium, and of this anthology, is to open up new vistas, instead of continuing old debates entrenched in old positions.

The symposium, in this spirit, was first and foremost, a forum to raise concerns. After all, we are still facing opposition—mostly in the physics community—by commentators who flatly deny the need for philosophical reflection. The majority of papers therefore may be presented as expositions of problems. The papers fall roughly into two classes: those which are mainly concerned with how and why a collapse could happen (the measurement problem), and those which are mainly concerned with the problems posed by non-local correlations. We have grouped the contributions accordingly.

1 Measurement of a Quantum System

Alexei Grinbaum's illuminates the foundations of quantum mechanics by focussing on the problem of the concept of an observer. Indicative of the problematic nature of this concept is the "shifty split" between system and observer in foundational accounts of quantum theory. Grinbaum gives us a historical overview of various positions that have been formulated in order to account for the role and function of the observer in quantum mechanics. He argues that common to all those accounts is the notion that observers define what a physical "system" is. More specifically, observers in quantum theory define systems not qua physical constitution, consciousness, or specific experimental setups, but they do so in an information theoretic sense. Taking his clue from this observation, Grinbaum describes observers using the notion of Kolmogorov complexity. Only sufficiently complex systems with a sufficient number of degress of freedom can function as observers of quantum systems, systems that lack the necessary number of degrees of freedom can only be classical observers. Spelling out this proposal in detail, Grinbaum suggests an experimental test for this interpretation. A C-60 fullerene molecule should have enough degrees of freedom vis-a-vis a photon emitting source in order to act as a quantum observer. But it should be able to do so only up to a certain number of photons observed. When saturated after observing too many photons, the fullerene will turn classical, a transition that should be observable, Grinbaum suggests, by measuring the fullerene's heat capacity.

Michael Esfeld and Antonio Vassallo discuss how canonical quantum gravity faces a dilemma. On the one hand, it seems impossible, in canonical quantum gravity, to treat entities which are localized in space-time as primitive and thus solve the measurement problem as does, for example, the alternative quantum theory by David Bohm. On the other hand, there seems to be no viable theoretical proposal, within canonical quantum gravity, as to how (macroscopic) entities, which are localized in space-time, can be reduced to more fundamental entities which are not localized in space-time. For Esfeld and Vassallo, a solution to the dilemma is best sought through a better understanding of the Wheeler-DeWitt equation, which describes the dynamical evolution of the quantum state of space-time.

Jakob Sprickerhof proposes an interpretation of quantum field theory in terms of entities which are localized in space, possibly in widely extended regions, and causally connected through energy-momentum transfer. He constructs his interpretation by modifying the conserved quantity theory of causation developed mainly by Phil Dowe, which is usually held to be incompatible with modern physics. The notoriously difficult to explain EPR correlations are, on Sprickerhof's account, an instance of the measurement problem rather than a problem of superluminal or otherwise non-relativistic causation. According to Sprickerhof, the initial singlet state of a pair of electrons is a spatially extended entity which interacts with the measurement devices; only from the interaction result two separate entities. Accordingly, the EPR correlations do not come about by a causal relation between two space-like separated entities but, rather, by a peculiar process—the measurement—which transforms the one spatially extended entity into two space-like separated entities. To render the EPR correlations less mysterious, Sprickerhof urges, we need to know what happens during a measurement.

Iñaki San Pedro calls into question that some statistical independence conditions, which are crucial in the derivation of Bell-type inequalities, can be justified by appealing to the experimenter's free choice as to which observable be measured in a given run of an experiment. San Pedro identifies questionable implicit assumptions, which are necessary to maintain the usual view that the freedom of experimental choice indeed justifies the statistical independence conditions. Those assumptions involve the temporal ordering of the events in EPR experiments, and, more generally, the temporal ordering of causes and effects. The implicit assumptions also involve the possibility of expressing causal relations in terms of probabilistic relations, and the assumption that the common causes of EPR correlations cannot even *partially* cause the experimenter's measurement choices. If we give up at least one of these implicit assumptions in our causal theory of EPR-type correlations, we can at the same time maintain that experimenters have free will and permit statistical dependencies of measurement choices and common causes for the correlations.

2 Collapse and Non-Locality

Tilman Sauer discusses several lesser known formulations, by Einstein himself, of the EPR paradox. Sauer uses these formulations to bring to the fore what Einstein's particular concern was: Quantum mechanics ascribes two different states to the same physical matters of fact. This seems to be a different concern than that of the EPR paper of 1935, or it is there much less clearly expressed. The last formulation which Sauer discusses, however, confronts us with a puzzle. He either seems not to have mastered the quantum mechanical formalism as applied to the spin properties of particles, or else he was cryptically expressing some idea that needs explication. Was Einstein slowly losing his intellectual faculties, or was he after an innovative solution of the paradox which he had already exposed sufficiently clearly a long time ago?

Simon Friederich localizes the source of both the measurement problem and the problem of quantum non-locality in the ontic conception of a quantum state. These problems can be solved, he suggests, by giving up the notion that quantum states are states that quantum systems are in. Rather, he suggests, we should conceive of quantum states epistemically as expressing the agent's knowledge about quantum systems. Discussing the advantages of such an "interpretation without interpretation", Friederich proceeds to spell out what he calls the "rule perspective" of an epistemic conception of quantum states. Drawing on a distinction due to John Searle between "regulative" and "constitutive" rules, he suggests that the rules by which we assign states to quantum systems are "constitutive" in the sense that they constitute the very meaning of a quantum state. Friederich argues that the rule perspective allows us to be neutral with respect to the broader issue of realism in the foundation of quantum mechanics.

Matthias Egg first reminds us that there are good reasons to believe that, if nothing else, scientific theories tell us the true causal relations which obtain in the world. But Bell-type experiments pose a serious challenge to such a position. In particular, several theorems tell us that Bell correlations cannot be used to send signals. Therefore, these theorems prohibit us to be able to manipulate any cause that would allow us to send a signal in a Bell-type experiment. But then again, according to Egg, it is only through some manipulation of causes that we can obtain knowledge about the specific causal structure, i.e., knowledge beyond the simple fact that there are *some* causal relations. This dilemma leads Egg to the conclusion that the causal realist has two options in the face of Bell-type experiments. The first option is to doubt some of the premises of the no-signaling theorems. The second option is to be realist only about the general claim that there is *some* causal relation between the observed events.

The empirical violation of Bell's inequality forces us to reconsider our most basic foundational concepts. It is, on Adrian Wüthrich's account, an even more profound and consequential fact than the theoretical difficulties associated with the measurement problem. Wüthrich undertakes to prepare the ground for an informed revision of our fundamental tenets when faced with the empirical violation of Bell-type inequalities. He takes issue with the all-too-simple alternative between realism and locality posed by quantum theory's empirical validity. Taking up arguments to the effect that realism itself in some not-too-specific sense entails non-locality, Wüthrich argues that giving up realist convictions may not help us save locality. In order to identify with more precision the consequences that need to be drawn, Wüthrich analyzes the premises of a minimal derivation of a class of Bell-type inequalities. Only with a minimal derivation are we in a position to identify the choices that we have to make. Following this program, Wüthrich gives a more fine-grained logical analysis of the structure of the reduction argument that allows him to qualify necessary assumptions and premises in terms of realist or local spirit.

Even though, historically, much of the interpretational quagmire arose in theoretical reflections on the foundations of quantum physics, those foundational problems are nonetheless real and carry observational consequences. Philip Walther presents results of an experiment that puts to the test the predictions of EPR correlations pushed to a subtle but conceptually significant consequence. We know, Walther reminds us, that quantum mechanical entanglement is real and observable. But this statement usually takes entanglement to be a relation between particles. Can entanglement itself be entangled? To answer this question Walther designed an experiment, in which the usual polarization state of a photon is entangled with the entangled state of a photon pair. Deriving a Clauser-Horne-Shimony-Holt inequality for the correlations between Alice's photon and Bob's entangled photon pair, Walther shows that in this case, too, experiment can decide on the nature of quantum reality. Not surprisingly perhaps, but with profound implications, he found that Bell's inequalities are violated also in the case where the entangled object in question is itself a state of entanglement and that quantum mechanical predictions are empirically confirmed even in this case. Entanglement is real even in the sense that it can itself be entangled.

Acknowledgments

The symposium "Decoherence and No-Signalling: Current Interpretational Problems of Quantum Theory" was funded by the *Mittelbauvereinigung der Universität Bern* (MVUB, University of Bern Intermediate Staff Association) and the *Conférence Universitaire de Suisse Occidentale* (CUSO, Conference of the Universities of Western Switzerland), and the University of Bern's Institute for Philosophy. We appreciate their generous and unbureaucratic financial contributions, and we wish to thank Michael Esfeld for co-organizing the symposium. We also thank the Institute for Philosophy at the University of Bern and its managing director Monika Betzler for providing institutional support and encouragement. A.W. acknowledges financial support from the Swiss National Science Foundation (Grant No. 105211-129730, PI Gerd Graßhoff). We are finally grateful to Jürgen Renn and Peter Damerow and their collaborators of the Edition Open Access at the Max Planck Institute for the History of Science for their enthusiasm and patience in providing a forum of open access publication. One of us (T.S.), a student of Peter Damerow's, would have liked to present this volume to his teacher in appreciation of what he has taught him. Sadly this was no longer possible.

References

Colbeck, R. and Renner, R. (2012). Is a system's wave function in one-to-one correspondence with its elements of reality? *Phys. Rev. Lett.*, 108:150402.

Pusey, M., Barrett, J., and Rudolph, T. (2012). On the reality of the quantum state. *Nature Physics*, 8(6):476–479.

Part I

Measurement of a Quantum System

Quantum Observer and Kolmogorov Complexity

Alexei Grinbaum

Abstract. *Different observers do not have to agree on how they identify a quantum system. We explore a condition based on algorithmic complexity that allows a system to be described as an objective "element of reality." We also suggest an experimental test of the hypothesis that any system, even much smaller than a human being, can be a quantum mechanical observer.*

1 Introduction

Quantum mechanical formalism has an orthodox interpretation that relies on the cut between the observer and the system observed (Dirac, 1930; von Neumann, 1932). This "shifty split" (Bell, 1990) of the world into two parts cannot be removed: the formalism only applies if the observer and the system are demarcated as two separate entities. Physical properties of the system, on one side of the split, do not exist independently of the observer, on the other side of the split, and can only be instantiated during the observation, or 'measurement,' of some dynamical variable of the system chosen by the observer.

The observer is essential for quantum mechanics, but precisely to whom or to what thing does this word refer? There is no consensus. What is extraordinarily difficult, Wheeler (1983) emphasized, is to state sharply and clearly where "the community of observer-participators" begins and where it ends. Quantum mechanics itself says nothing about the physical composition of the observer: whatever is meant by the word has no quantum mechanical description and lies outside physical theory. One cannot infer from a set of quantum mechanical measurements if the observer is a human being, a machine, a stone, a Martian, or the whole Universe.

It is remarkable that although nothing can be said about who the observer is, quantum mechanics gives unequivocal prescriptions for the content of the observer's observation. Differently constituted observers, even if one is a butterfly

T. Sauer, A. Wüthrich (eds.), *New Vistas on Old Problems. Recent Approaches to the Foundations of Quantum Mechanics*, Edition Open Access, 2013.

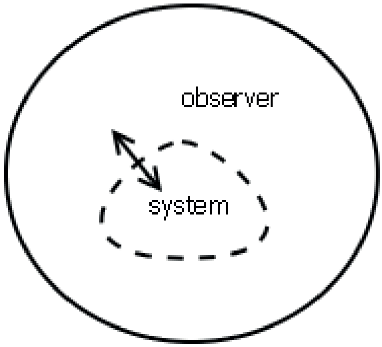

Figure 1: A “shifty split” between the observer and the quantum system.

on Earth and the other a drop of methane on Titan, will obtain the same probabilistic results provided that they manage to measure the same quantum system. How is this possible? Only because the mathematics of quantum mechanics is abstract: it deals with ‘systems’ and ‘observers,’ not with methane or butterflies.

The question of the observer’s exact physical constituency is not dealt with from within quantum mechanics; it is left for philosophy. Different interpretations of quantum mechanics provide different answers. Some say that only conscious human beings can be observers; others state that anything qualifies, even a single electron. Each answer to the question ‘what is an observer’ comes with a corresponding answer to the question ‘can observers disagree’ or ‘what makes them agree.’

As a part of his relative-state interpretation, Everett argued that observers are physical systems with memory, i.e., “parts... whose states are in correspondence with past experience of the observers” (Everett, 1957). We call this a *universal observer* hypothesis: any system with certain information-theoretic properties can serve as quantum mechanical observer, independently of its physical constituency, size, presence or absence of conscious awareness and so forth. In this vein, Rovelli claimed that observers are merely systems whose degrees of freedom are correlated with some property of the observed system: “Any system can play the role of observed system and the role of observing system. […] The fact that observer O has information about system S (has measured S) is expressed by the existence of a correlation […]” (Rovelli, 1996). However, the universal observer hypothesis has remained a controversial statement to this day. For example, Peres claims in the way exactly opposite to Rovelli's, that "the two electrons in the ground state of the helium atom are correlated, but no one would say that each electron 'measures' its partner" (Peres, 1986). Our purpose is to clarify an information-theoretic definition of quantum mechanical observer and to propose a physical test of this hypothesis.

We start by a historical review in Section 2. In Section 3 we give a general definition of observer based on the intuitive feeling that a key component of observation is system identification. Then we apply to it the notion of Kolmogorov complexity, which is the main tool of ensuing analysis. In Section 4 this approach is developed to germinate a definition of quantum and classical systems. In Section 5 we consider a family of observers and require that a system be identified by them in the same way, thereby giving an information-theoretic criterion of an "element of reality." In Section 6 we show that observers can be allowed some disagreement while still maintaining an unambiguous identification of the observed element of reality. Finally, in Section 7 we suggest an experimental test of the universal observer hypothesis.

2 Observer as a Problem

2.1 Observer in the Copenhagen Orthodoxy

Bohr's lecture at Como in 1927 became the foundation of what later came to be known as the Copenhagen interpretation of quantum mechanics. Despite being a universal and orthodox reference in the debate on quantum mechanics, the Copenhagen interpretation has a variety of formulations none of which is the accepted dogma (Howard, 1994; Henderson, 2010). Its main point can be described by the following quote from Bohr: "Only with the help of classical ideas is it possible to ascribe an unambiguous meaning to the results of observation. […] It lies in the nature of physical observation, that all experience must ultimately be expressed in terms of classical concepts." (Bohr, 1934, 94) Two different readings of Bohr's statement exist, related to what exactly is meant by "classical." The first is a straightforward inference that quantum mechanics requires classical mechanics:

> It is in principle impossible to formulate the basic concepts of quantum mechanics without using classical mechanics. (Landau and Lifshitz, 1977, 2)

On the second reading, one maintains that quantum mechanical experiments can only be described by the classical language, which in turn leads us to classical mechanics, but it is the *language* that is a crucial ingredient:

> Bohr went on to say that the terms of discussion of the experimental conditions and of the experimental results are *necessarily* those of 'everyday language,' suitably 'refined' where necessary, so as to take the form of classical dynamics. It was apparently Bohr's belief that this was the only possible language for the *unambiguous communication* of the results of an experiment. (Bohm, 1971, 38)

The choice between classical mechanics and classical language is not always visible and many authors have failed to distinguish between the two formulations. A direct reference to classical mechanics means, for those who support this view, that the world consists of mechanical systems only, whether quantum or classical, and no external observer is necessary. On the contrary, an invocation of the classical language, concepts, or terms—and Bohr himself has always used these words—implies that the formulation of the problem includes somebody who is in command of classical concepts: the experimenter. This experimenter prepares the quantum system and then measures it acting as a quantum mechanical observer. Then observers possess a linguistic faculty because, according to Bohr's dictum, they need to communicate between themselves unambiguously.

The mechanistic point of view and the linguistic point of view differ in the ways they account for the problem of agreement. Mechanistically, one assumes that the properties of classical systems exist objectively, i.e. independently of bringing other systems into consideration. Objectivity then guarantees that any further classical mechanical system will be able to interact with the initial system and acquire the same property, which encodes the measurement result. Mechanistic observers agree because disagreement is not classically observed.

The answer is less straightforward from the linguistic point of view. Observers agree on the results of quantum measurements because they agree on the use of classical concepts. But why do they agree on linguistic usage? How do they establish their agreement? How does it come about? It is a matter of pure convention usually left unanalyzed by the Copenhagen interpretation.

2.2 London and Bauer

John von Neumann's magisterial book on quantum mechanics first appeared in German in 1932 (von Neumann, 1932). It offered a theory of measurement that were to become a standard not only for the Copenhagen interpretation but for quantum mechanics as a physical theory in general. However, von Neumann's musings about the place of the observer were not entirely clear and satisfactory. The formalism worked perfectly, but what was exactly its meaning?

Writing as early as 1939, London and Bauer set the tone of one of the main branches of the post-von Neumann debate on the interpretation of quantum mechanics. They remark that quantum mechanics cannot ascribe properties to the quantum system in itself, only if there is a link to an observer. This observer is for London and Bauer a human person. An objective description of reality is therefore impossible and "it seems that the result of measurement is intimately linked to the consciousness of the person making it" (London and Bauer, 1939, 48).[1] Von Neumann's cut between the observer and the observed system is here pushed to the extreme position, leaving all physical systems on one side, including light photons, the eye and the nerves, and only leaving the observer's 'organ' of awareness, namely consciousness, on the other.

If this were true, why would objectivity be possible at all and why have physicists not yet become solipsists? Why do two physicists agree on what constitutes the object of their observation and on its properties? Because, according to London and Bauer (1939, 49), there exists something like a "community of scientific consciousness, an agreement on what constitutes the object of the investigation." The nature of this scientific consciousness remains however mysterious. London and Bauer only conclude that "it must yet be looked into."

2.3 Wigner

Remarks by London and Bauer were further developed by Eugene Wigner (1961). Consciousness of the observer "enters the theory unavoidably and unalterably" and corresponds to the impression produced by the measured system on the observer. Answering the question of realism, Wigner notes that the wave function "exists" only in the sense that "the information given by the wave function is communicable." Observers agree because they communicate measurement results to one another:

[1] English translation in (Wheeler and Zurek, 1983, 218–259).

> The communicability of information means that if someone else looks at time t and tells us whether he saw a flash, we can look at time $t + 1$ and observe a flash with the same probabilities as if we had seen or not seen the flash at time t ourselves.

The first observer tells us the result of his measurement: communication for Wigner is linguistic. The second observer, in this example of Wigner's, puts himself in the position of the first observer and acts just like him. This is only possible because the second observer has been told the measurement result by the first one.

Do observers actually have to communicate or is it enough to require that they simply *could* communicate the information about the measurement? Here Wigner's position is vague. On the one hand, he states, "If someone else somehow determines the wave function of a system, he can tell me about it [...]," which requires a mere possibility of communication but no sending of actual information. On the other hand, Wigner famously analyzes the following situation labeled 'Wigner's friend':

> It is natural to inquire about the situation if one does not make the observation oneself but lets someone else carry it out. What is the wave function if my friend looked at the place where the flash might show at time t? The answer is that the information available about the *object* cannot be described by a wave function. One could attribute a wave function to the joint system: friend plus object, and this joint system would have a wave function also after the interaction, that is, after my friend has looked. I can then enter into interaction with this joint system by asking my friend whether he saw a flash. [...] The typical change in the wave function occurred only when some information (the *yes* or *no* of my friend) entered *my* consciousness.

Although he calls this situation natural, Wigner is the only one among the founding fathers of quantum theory to have addressed it explicitly. Here Wigner's agreement with his friend is clearly possible thanks to the linguistic communication between them, but this communication itself is not a quantum measurement: whatever the situation, Wigner always knows the question he should put to his friend and fully trusts the answer, always *yes* or *no*. Communication from the friend must actually occur before the wave function could be known by Wigner; it is not enough that this communication be merely possible. As with the linguistic reading of the Copenhagen interpretation, Winger's interpretation involving consciousness of the observer leaves open the exact mechanism, whether a human convention or a physical given, of the agreement between observers.

Wigner also touches on the question of belief and trust in his discussion of the repeatability of experiments in physics. To explore the statistical nature of the predictions of quantum mechanics, it is necessary to be able to produce many quantum systems in the same state; subsequently these systems will be measured. One can never be absolutely sure, Wigner stipulates, that one has produced the same state of the system. We usually "believe that this is the case" and we are "fully convinced of all this" (Wigner, 1976, 267), even if we have not tried to establish experimentally the validity of the repeated preparation of the same state. What is at work here is again a convention shared by all physicists. How do they know that repeated preparations produce the same state if they do not measure each and every specimen in order to verify it? The answer is that they have common experience and a convention on what a 'controlled experiment' amounts to, and their respect of this commonly shared and empirically validated rules enables them to postulate the existence of repeated states even in the situations which had never been tested before. This is how physical theory with its laws and a precise methodology arises by way of abstraction ('elevation,' as Einstein or Poincaré would say (Friedman, 2001, 88)) from the physicist's empirical findings and the heuristics of his work.

2.4 Everett

What is important about the observer? Only his function of an informational agent, not his physical constituency. The need to refer to consciousness exists, insofar as only consciousness can distinguish a mere physical correlation, e.g. of an external system with the observer's eye, from the information actually available to the observer, i.e. such that he can act upon in the future. Other characteristics are irrelevant: say, the observer's age plays no role ("there is little chance of making a big mistake if one does not know [the observer's] age" (London and Bauer, 1939, 43)).

Treating the observer as an informational agent requires that we say precisely what makes different systems possessing information into observers. In other words, what is the nature of a convention shared by all observers, whose workings we notice when apparently unrelated measurements by different observers come out in a consistent way? Brillouin believes that information must be defined with the exclusion of all human element (George, 1953, 360), in which case the convention between observers must necessarily have a physical, as opposed to linguistic or cultural, origin. Watanabe finds such a physical origin in the direction of time shared by all observers: “The past-to-future directions of all observers coincide. This statement has a well-defined physical meaning, for ‘positive time direction’ is a Lorentz-invariant concept” (George, 1953, 387). But both Brillouin’s and Watanabe’s ideas proved to be sterile in the history of physics. Time-directedness, for once, may be a necessary requirement, because, as we shall see later, observers are thermodynamical systems. But it is far from clear if Watanabe’s shared time direction is a sufficient or even a necessary condition.

Hugh Everett (1957) thought differently. Observers are for him systems that can be described in purely physical terms. These systems possess memory, defined as “parts […] whose states are in correspondence with past experience of the observers.” Observers do not have to be human: they could be “automatically functioning machines, possessing sensory apparatus and coupled to recording devices.” Memory records are thus fundamental: it is memory that makes any physical system an observer. Everett continues, “if we are to be able to call the interaction an observation at all, the requirement that the observer’s state change in a manner which is different for each eigenfunction is necessary.” Repeatability relies on the necessary assumption that the eigenstates of the system be unchanged during measurement.

Everett was the first to explicitly consider the problem of several observers. The “interrelationship between several observers” is an act of communication between them, which Everett treats as establishing a correlation between their memory configurations. The consequences are described with remarkable clarity:

> 1. When several observers have separately observed the same quantity in the object system and then communicated the results to one another they find that they are in agreement. This agreement persists even when an observer performs his observation after the result has been communicated to him by another observer who has performed the observation.
>
> 2. Let one observer perform an observation of a quantity A in the object system, then let a second perform an observation of a quantity B in this object system which does not commute with A, and finally let the first observer repeat his observation of A. Then the memory system of the first observer will *not* in general show the same result for both observations. […]

3. Consider the case when the states of two object systems are correlated, but where the two systems do not interact. Let one observer perform a specified observation on the first system, then let another observer perform an observation on the second system, and finally let the first observer repeat his observation. Then it is found that the first observer always gets the same result both times, and the observation by the second observer has no effect whatsoever on the outcome of the first's observations.

2.5 Rovelli

Rovelli (1996) proposed a relational interpretation of quantum mechanics. All physical systems have the capacity to act as observers, i.e., possess information about other physical systems. Information should be seen as an observer-dependent, rather than objective, notion. It is information in Shannon's sense indexed by two parameters: the first related to the observed system about which this information has been obtained, the second related to the observing system that has obtained information about the first system. Trying to remove the second index and to 'liberate' the notion of information from its relational sense is both impossible and meaningless. Objective information independent of the observing system does not exist.

However, the question of intersubjective agreement remains unclear. If each observer acts on his own and information is always defined relative to some observer, why do different observers agree on the measurement result and attribute the same state to the quantum system? For Rovelli, the answer lies in the interaction between observers, which is a quantum mechanical process like any other and, consequently, must be described in the same language of relative information. This information is encoded in the correlations between physical degrees of freedom of the observers. Now imagine three observers in a triangle (see Figure 2), with O_1 observing O_2, O_2 observing O_3, and O_3 observing O_1.

Some of the degrees of freedom of O_1 are correlated with some degrees of freedom of O_2. Similarly, some degrees of freedom of O_2 are correlated with the degrees of freedom of O_3, and the degrees of freedom of O_3 are correlated with those of O_1. These correlations must be self-consistent, i.e., no paradox must emerge from the circular correlation. How is this requirement realized physically? Perhaps it is always the case that when an observer is observed by another system, the degrees of freedom that the other system can 'see' are never the same as the

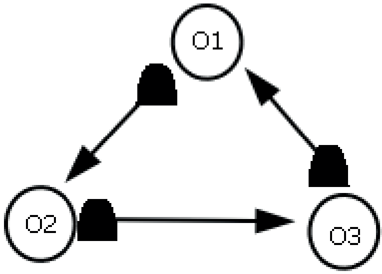

Figure 2: A triangle between three observers.

degrees of freedom in which the observer encodes his own measurement data. Or perhaps some other mechanism is at work, which precludes us from running into a contradiction. This issue is not explored by Rovelli and remains an open question.

3 Observer as a System Identification Algorithm

What characterizes an observer is that it (or he or she) has information about some physical system. This information fully or partially describes the state of the system. The observer then measures the system, obtains further information and updates his description accordingly. The physical processes of measuring the system, updating the information and ascribing a state to the system may happen in many different ways depending on the physical constitution of the observer. The memory of a computer acting as an observer, for instance, is not the same as human memory, and measurement devices vary in their design and functioning. Still one feature unites all observers: whatever they do, they do it to a *system*. In quantum mechanics, defining an observer goes hand in hand with defining a system under observation. An observer without a system is a meaningless nametag, a system without an observer who measures it is a mathematical abstraction. What remains constant throughout measurement is the identification, by the observer, of the quantum system.

Quantum systems aren't like sweets: they don't melt. Take a general thermodynamic system interacting with other systems. Such a system can dissipate, diffuse, or dissolve, and thus stop being a system. If at first a cube of ice gurgling into tepid water is definitely a thermodynamic system, it makes no sense to speak about it being a system after it has dissolved. Quantum systems are not like this. Their states may evolve, but the observer knows how to tell the system he observes from the rest or the environment. An electron in a certain spin state remains an electron after measurement, i.e., it remains a system with a particular set of degrees of freedom. The observer maintains the identity of the system notwithstanding a change in the state of this system that may or may not occur. So, whatever else he might happen to be, the observer is primarily a system identification machine. Different observers having different features (clock hands, eyes, optical memory devices, internal cavities, etc.) all share this central characteristic.

Definition 3.1. An observer is a system identification algorithm (SIA).

Particular observers can be made of flesh or, perhaps, of silicon. ‘Hardware’ and ‘low-level programming’ are different for such observers, yet they all perform the task of system identification. This task can be defined as an algorithm on a universal computer, e.g. the Turing machine: take a tape containing the list of all degrees of freedom, send a Turing machine along this tape so that it puts a mark against the degrees of freedom that belong to the quantum system under consideration. Any concrete SIA may proceed in a very different manner, yet all can be modelled with the help of this abstract construction.

The SIAs whose physical realization may differ share one property that does not depend on the hardware: their algorithmic, or Kolmogorov, complexity. Any SIA can be reconstructed from a binary string of some minimal length (which is a function of this SIA) by a universal machine. As shown by Kolmogorov (1965), this minimal compression length defines the amount of information in the SIA and does not depend (up to a constant) on the realization of the SIA on this or that hardware. The common-lore view of a multitude of individual observers, one hastily printing, another yawning, a third one moving around his DNA strands, should in our opinion be superseded by a view of different SIAs, each with its algorithmic complexity defined via a universal machine.

4 Quantum and Classical Systems

Each quantum system has a certain number of degrees of freedom, which we think about as being independent parameters needed in order to characterize the state of the system. For example, a system with only two states (spin-up and spin-down) has one degree of freedom and can be described by one parameter $\sigma = \pm 1$. If we write down these parameters as a binary string, the Kolmogorov complexity of that string is at least the number of the degrees of freedom of the system, i. e.,

$$K(s) \geq \mathrm{d}_S, \tag{1}$$

where $K(s)$ is the Kolmogorov complexity of the binary string s representing the parameters of the system S, and d_S is the number of the degrees of freedom in S. In what follows the notation $K(s)$ and $K(S)$ will be used interchangeably.

When we say that an observer X observes a quantum system S, it is usually the case that $K(S) \ll K(X)$. In this case the observer will have no trouble keeping track of all the degrees of freedom of the system; in other words, the system will not 'dissolve' or 'melt' in the course of dynamics. However, it is also possible that X identifies a system with $K(S) > K(X)$. For such an observer, the identity of system S cannot be maintained and some degrees of freedom will fall out from the description that X makes of S.

Definition 4.1. System S is called quantum with respect to observer X if $K(S) < K(X)$, meaning that X will be able to maintain a complete list of all its degrees of freedom. Otherwise S is called classical with respect to X.

Suppose that X observes a quantum system S and another observer Y observes both S and X. If $K(Y)$ is greater than both $K(X)$ and $K(S)$, observer Y will identify both systems as quantum systems. In this case Y will typically treat the interaction between X and S as an interaction between two quantum systems. If, however, $K(X)$ and $K(Y)$ are close, $K(X) \gg K(S)$ and $K(Y) \gg K(S)$ but $K(X) \simeq K(Y)$, then Y will see S as a quantum system but the other observer, X, as a classical system. An interaction with a classical system, which we usually call 'observation,' is a process of decoherence that occurs when the Kolmogorov complexity of at least one of the systems involved approaches the Kolmogorov complexity of the external observer. In this case Y cannot maintain a complete description of X interacting with S and must discard some of the degrees of freedom. If we assume that all human observers acting in their SIA capacity have approximately the same Kolmogorov complexity, this situation may provide an explanation of the fact that we never see a human observer (or, say, a cat) as a quantum system.

One welcome consequence of Definitions 3.1 and 4.1 is that Kolmogorov complexity $K(X)$ is not computable. We as human observers do not seem to know the maximum number of the degrees of freedom in a system that we can still keep track of. A photon is certainly a quantum system from our point of view, a simple atom too, a C_{60} perhaps as well, albeit seeing quantum effects with fullerenes is not easy. But we have never seen a quantum system having, say, 10^{23} degrees of freedom. So where does the border run? Is it a number like 6 or 20 or is it $10^n, n > 2$? All we can say is that mathematics shows that human observers cannot compute their own $K(X)$.

5 Elements of Reality

Ever since the EPR paper (Einstein et al., 1935), the question of what is real in the quantum world has been at the forefront of all conceptual discussions about quantum theory. Einstein, Podolsky and Rosen formulated their question with regard to physical *properties*: e.g., is position or momentum real? This is, however, not the only problem of reality that appears when many observers enter the game. Imagine a sequence of observers X_i, $i = 1, 2, ...$, each identifying systems S_n, $n = 1, 2, ...$ System identifications of each S_n do not have to coincide as some observers may have their Kolmogorov complexity $K(X_i)$ below, or close to, $K(S_n)$, and others much bigger than $K(S_n)$. If there is disagreement, is it possible to say that the systems are real, or objects of quantum mechanical investigation, in some sense? We can encode the binary identification string produced by each observer in his SIA capacity as some random variable $\xi_i \in \Omega$, where Ω is the space of such binary identification strings, possibly of infinite length. Index i is the number of the observer, and the values taken by random variable ξ_i bear index n corresponding to "i-th observer having identified system S_n." Adding more observers, and in the limit $i \to \infty$ infinitely many observers, provides us with additional identification strings. Putting them together gives a stochastic process $\{\xi_i\}$, which is an observation process by many observers. If systems S_n are to have a meaning as "elements of reality," it is reasonable to require that no uncertainty be added with the appearance of further observers, i.e., that this stochastic process have entropy rate equal to zero:

$$H(\{\xi_i\}) = 0. \tag{2}$$

We also take this process to be stationary and ergodic so as to justify the use of Shannon entropy.

Let us illustrate the significance of condition (2) on a simplified example. Suppose that $\theta_1, \theta_2, ...$ is a sequence of independent identically distributed random variables taking their values among binary strings of length r with probabilities q_k, $k \leq 2^r$. These θ_k can be seen as identifications, by different SIAs, of different physical systems, i.e., a special case of the ξ_i-type sequences having fixed length and identical distributions. For instance, we may imagine that θ_1 is a binary encoding of the first observer seeing an electron and θ_2 is a binary string corresponding to the second observer having identified a physical system such as an elephant; and so forth. Entropy becomes simply:

$$H = -\sum_k q_k \log q_k. \tag{3}$$

Condition (2) applied to entropy (3) means that all observers output one and the same identification string of length r, i.e., all SIAs are identical. This deterministic system identification, of course, obtains only under the assumption that the string length is fixed for all observers and their random variables are identically distributed, both of which are not plausible in the case of actual quantum mechanical observers. So rather than requiring identical strings we impose condition (2) as a criterion of the system being identified in the same way by all observers, i.e., it becomes a candidate quantum mechanical "object of investigation."

6 Relativity of Observation

Let us explore the consequences of condition (2). Define a binary sequence α_n^i as a concatenation of the system identifications strings of systems S_n by different observers:

$$\alpha_n^i = \overline{(\xi_1)_n}\,\overline{(\xi_2)_n} \dots \overline{(\xi_i)_n}, \tag{4}$$

where index i numbers observers and the upper bar corresponds to "string concatenation" (a detailed definition can be found in Zvonkin and Levin (1970)). Of course, this concatenation is only a logical operation and not a physical process. A theorem by Brudno (1978, 1983) conjectured by (Zvonkin and Levin, 1970) affirms that the Kolmogorov complexities of strings α_n^i converge towards entropy:

$$\lim_{n\to\infty} \lim_{i\to\infty} \frac{K(\alpha_n^i)}{i} = H(\{\xi_i\}). \tag{5}$$

For a fixed i and the observer X_i who observes systems S_n that are quantum in the sense of Definition 4.1, variation of $K(\alpha_n^i)$ in n is bounded by the observer's own complexity in his SIA capacity:

$$K(\alpha_n^i) < K(X_i) \qquad \forall n, \qquad i \text{ fixed.} \tag{6}$$

Hence eqs. (2) and (5) require that

$$\lim_{i\to\infty} \frac{K(\alpha_n^i)}{i} = 0. \tag{7}$$

This entails that the growth of $K(\alpha_n^i)$ in i cannot be faster than logarithmical. Therefore, we have the following:

Proposition 6.1. *An element of reality that may become an object of quantum mechanical investigation can be defined only with respect to a class of not very different observers.*

To give an intuitive illustration, imagine adding a new observer X_{i+1} to a group of observers $X_1, \dots, X_i$ who identify systems S_n. This adds a new identification string that we glue at the end of concatenated string α_n^i consisting of all X_i's identifications of S_n, thus obtaining a new string α_n^{i+1}. The Kolmogorov complexity of α_n^{i+1} does not have to be the same as the Kolmogorov complexity of α_n^i; it can grow, but not too fast, i.e., not faster than the logarithm. Adding a new observation may effectively add some new non-compressible bits, but not too many such bits. If this is so, then $H = 0$ still obtains. Although observers $X_1, \dots, X_i, X_{i+1}$ produce slightly different identification strings, they agree, simply speaking, that an atom is an atom and not something that looks more like an elephant.

The above reasoning applies only to quantum systems S_n in the sense of Definition 4.1. This is because in the case of classical systems different observers may each operate their own coarse-graining, keeping only some degrees of freedom. System identification strings may then differ dramatically and one cannot expect $K(\alpha_n^i)$ to grow moderately.

7 Experimental Test

A previously suggested experimental connection between thermodynamics and theories based on Kolmogorov complexity is based on observing the conse-

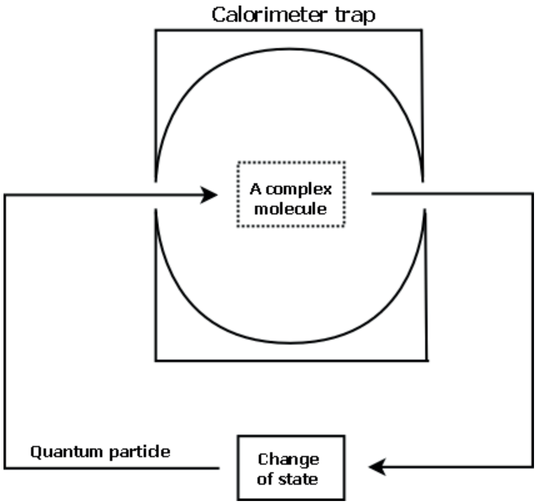

Figure 3: Experiment leading to heat production when physical entropy $\mathcal{S} = H + K$ changes.

quences of a change in the system's state (Zurek, 1989b,a, 1998; Erez et al., 2008). Zurek introduced the notion of physical entropy $\mathcal{S} = H + K$, where H is the thermodynamic entropy and K the Kolmogorov entropy. If the observer with a finite memory has to record the changing states of the quantum system, then there will be a change in $\mathcal{S}$, like the one depicted in Figure 3, and it will lead to heat production that can be observed experimentally.

What we propose here, based on a suggestion by Anton Zeilinger, is a simpler setting that can still serve as a test of the universal observer hypothesis. It does not rely on the measurement of particular states, but on the fact of measurement as such. If a measurement occurs, then the observer has identified the quantum system, and this fact in itself, if repeated, will eventually lead to heat production.

An individual fullerene molecule is placed in a highly sensitive calorimeter and bombarded with photons, which play the role of quantum systems with low $K(S)$ (Figure 4). According to the universal observer hypothesis, the fullerene is a quantum mechanical observer with $K(X) > K(S)$. Thus the absorption of the photon by the fullerene can be described as measurement: the fullerene identifies a quantum system, i.e. the photon, and observes it, obtaining new information.

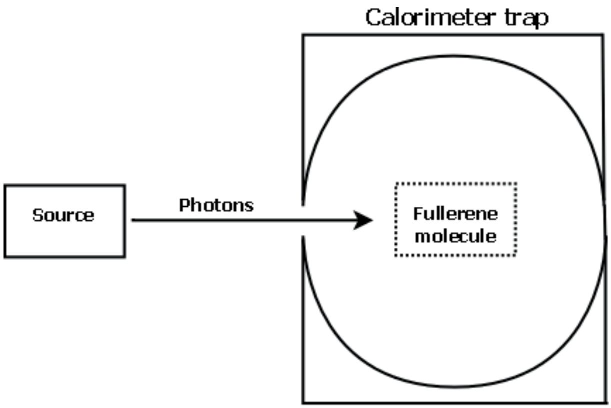

Figure 4: Experiment leading to heat production when observer's memory becomes saturated.

Physically, this process amounts to establishing a correlation between the photon variables, i.e., its energy, and the degrees of freedom of the fullerene. The external observer knows that such a process has occurred but remains unaware of its exact content, so that he is aware that there has been measurement, but doesn't know a precise state of the photon as measured by the fullerene, nor a precise state of the fullerene after measurement.

Informationally speaking, the same process can be described as storing information in the fullerene's memory. If measurement is repeated on several photons, more such information is stored, so that at some point total Kolmogorov complexity will approach $K(X)$. When it reaches $K(X)$, the fullerene will stop identifying incoming photons as quantum systems. Any further physical process will lead to heat production due to the erasure of memory, as prescribed by the Landauer principle that creates information erasure and heat production (Landauer, 1961). Physically, this process will correspond to a change of state of the carbon atoms that make up the fullerene molecule: the calorimeter will register a sudden increase in heat when C_{60} cannot store more information, thereby ending its observer function.

Actual experiments with fullerenes show that this scenario is realistic. A fullerene molecule "contains so many degrees of freedom that conversion of electronic excitation to vibrational excitation is extremely rapid." Thus, the fullerene is a good candidate for a quantum mechanical observer, for “the molecule can store large amounts of excitation for extended periods of time before degradation of the molecule (ionization or fragmentation) is observed” (Lykke and Wurz, 1992). The experiments in which fullerenes are bombarded with photons demonstrate that “the energy of the electronic excitation as a result of absorption of a laser photon by a molecule is rapidly converted into the energy of molecular vibrations, which becomes distributed in a statistical manner between a large number of the degrees of freedom of the molecule [...] The fullerene may absorb up to 10 photons at $\lambda = 308$ nm wavelength before the dissociation of the molecule into smaller carbon compounds” (Eletskii and Smirnov, 1995). We read these results as a suggestion that there should be one order of magnitude difference between $K(S)$ and $K(X)$ and that this allows the fullerene to act as quantum mechanical observer for up to 10 photons at 308 nm wavelength. What needs to be tested experimentally in this setting is heat production: we conjecture that if the same process occurs inside a calorimeter, the latter will register a sudden increase in heat after the fullerene will have observed 10 photons. What we predict here isn’t new physics, but an explanation on a new level, i. e., the level of information, of a physical process, i. e., heat production, which plays a largely overlooked role in the dissociation of fullerenes. We suggest that heat production deserves special attention as a signature of the fullerene’s role as quantum mechanical observer.

As a side remark, imagine that the photon’s polarization state in some basis were fully mixed:

$$\frac{1}{2}(|0\rangle + |1\rangle).$$

While only the energy of the photon matters during absorption, the external observer records von Neumann entropy $H = \log 2$ corresponding to this mixture (the initial state of the fullerene is assumed fully known). After absorption, it is mandatory that this entropy be converted into Shannon entropy of the new fullerene state, corresponding nicely to the uncertainty of the external observer in describing the “statistical manner” of the distribution over a large number of the degrees of freedom. From the internal point of view, we may assume perfect ‘self-knowledge’ of the observer, which puts his Shannon entropy equal to zero. However, his Kolmogorov entropy will increase as a result of recording the measurement information (Zurek, 1989a). Heat produced during the erasure of measurement information is at least equal to Kolmogorov complexity of the string that was stored in observer's memory; but according to quantum mechanics, this heat will not reveal to the external observer any information about the precise photon state observed by the fullerene.

8 Conclusion

The Copenhagen view of quantum mechanics traditionally described quantum systems and observers, epistemologically, as belonging to different categories. On the contrary, the view based on the relativity of observation, as proposed by Everett and later Rovelli, puts all systems on equal grounds and ascribes them only relative states. These two views are not as contradictory as they may seem. Relativity of observation has been understood by some proponents of the Copenhagen school (Hermann, 1935)[2], (Fock, 1971a,b). Information-theoretic treatment of the observer gives a chance to completely overcome the tension. On the one hand, the observer is an SIA and is characterized by its Kolmogorov complexity. On the other hand, quantum mechanics can be reconstructed from information-theoretic axioms and thus seen as a theory of information (Grinbaum, 2007). This puts all systems on equal grounds, in the spirit of Rovelli, while emphasizing the idea of relativity of observation, in the spirit of Fock.

Additionally, information-theoretic treatment of the observer provides a somewhat surprising result developing EPR's notion of "element of reality." One can make sense of a system existing independently of observation, with respect to a class of observers whose Kolmogorov complexities may differ, even if slightly. Equation (7) provides a mathematical criterion for this.

We have analyzed the observer as a system identification algorithm in the context of quantum mechanics. It remains an open question to apply this analysis to quantum field theory, where the task of system identification may look significantly different from the finite-dimensional situation. It also remains an open problem to realize experimentally the setup proposed in Section 7, which may lead to experimental confirmation of the universal observer hypothesis. Putting together this experimental test, which may show that a fullerene can act as an observer for up to 10 photons, and the remark on non-computability of $K(X)$ at the end of Section 4 begs yet another question: is it possible to say that, although $K(X)$ isn't computable in the mathematical sense, physical experiment effectively computes it?

[2] Quoted by Jammer (1974, 207–211).

Acknowledgments

I am grateful to Vasily Ogryzko for stimulating discussions and to Časlav Brukner, Markus Aspelmeyer, Ognyan Oreshkov and Anton Zeilinger for their remarks and hospitality at the Institute for Quantum Optics and Quantum Information of the Austrian Academy of Sciences.

References

Bell, J. (1990). Against measurement. *Physics World*, 3:33–40.

Bohm, D. (1971). On Bohr's views concerning the quantum theory. In Bastin, T., editor, *Quantum Theory and Beyond*, pages 33–40. Cambridge University Press.

Bohr, N. (1934). *Atomic Theory and the Description of Nature*. Cambridge University Press.

Brudno, A. (1978). The complexity of the trajectories of a dynamical system. *Russ. Math. Surveys*, 33:197–198.

Brudno, A. (1983). Entropy and the complexity of the trajectories of a dynamical system. *Trans. Moscow Math. Soc.*, 2:157–161.

Dirac, P. (1930). *The Principles of Quantum Mechanics*. Clarendon Press.

Einstein, A., Podolsky, B., and Rosen, N. (1935). Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.*, 47:777–780.

Eletskii, A. V. and Smirnov, B. M. (1995). Fullerenes and carbon structures. *Physics-Uspekhi*, 38:935–964.

Erez, N., Gordon, G., Nest, M., and Kurizki, G. (2008). Thermodynamic control by frequent quantum measurements. *Nature*, 452:724–727.

Everett, H. (1957). "Relative state" formulation of quantum mechanics. *Rev. Mod. Phys.*, 29:454–462.

Fock, V. (1971a). The principle of relativity with respect to observation in modern physics. *Vestnik AN SSSR*, 4:8–12.

Friedman, M. (2001). *Dynamics of Reason*. CSLI Publications.

George, A., editor (1953). *Louis de Broglie, physicien et penseur*. Albin Michel.

Grinbaum, A. (2007). Reconstruction of quantum theory. *British Journal for the Philosophy of Science*, 58:387–408.

Henderson, J. R. (2010). Classes of Copenhagen interpretations: Mechanisms of collapse as typologically determinative. *Studies In History and Philosophy of Modern Physics*, 41:1–8.

Hermann, G. (1935). Die naturphilosophischen Grundlagen der Quantenmechanik. *Abhandlungen der Fries'schen Schule*, 6:75–152.

Howard, D. (1994). What makes a classical concept classical? Toward a reconstruction of Niels Bohr's philosophy of physics. In Faye, J. and Folse, H., editors, *Niels Bohr and Contemporary Philosophy*, pages 201–229. Kluwer.

Jammer, M. (1974). *The Philosophy of Quantum Mechanics*. John Wiley and Sons.

Kolmogorov, A. (1965). Three approaches to the definition of the concept 'quantity of information'. *Probl. Inform. Transm.*, 1:3–7.

Landau, L. and Lifshitz, E. (1977). *Quantum mechanics*. Pergamon Press.

Landauer, R. (1961). Irreversibility and heat generation in the computing process. *IBM Journal of Research and Development*, 5:183–191.

London, F. and Bauer, E. (1939). *La théorie de l'observation en mécanique quantique*. Hermann, Paris.

Lykke, K. R. and Wurz, P. (1992). Direct detection of neutral products from photodissociated C_{60}. *Journal of Physical Chemistry*, 96:3191–3193.

Peres, A. (1986). When is a quantum measurement? *Am. J. Phys.*, 54:688–692.

Rovelli, C. (1996). Relational quantum mechanics. *Int. J. of Theor. Phys.*, 35:1637–1678.

von Neumann, J. (1932). *Mathematische Grundlagen der Quantenmechanik*. Springer, Berlin.

Wheeler, J. (1983). Law without law. Published in (Wheeler and Zurek, 1983, 182–213).

Wheeler, J. and Zurek, W., editors (1983). *Quantum Theory and Measurement*. Princeton University Press.

Wigner, E. (1961). Remarks on the mind-body question. In Good, I., editor, *The Scientist Speculates*, pages 284–302. Heinemann, London. Reprinted in (Wheeler and Zurek, 1983, 168–181).

Wigner, E. (1976). Interpretation of quantum mechanics. Lectures given in the Physics Department of Princeton University. Published in (Wheeler and Zurek, 1983, 260–314).

Zurek, W. (1989a). Algorithmic randomness and physical entropy. *Phys. Rev. A*, 40:4731–4751.

Zurek, W. (1989b). Thermodynamic cost of computation, algorithmic complexity and the information metric. *Nature*, 341:119–124.

Zurek, W. (1998). Decoherence, chaos, quantum-classical correspondence, and the algorithmic arrow of time. *Physica Scripta*, T76:186–198.

Zvonkin, A. and Levin, L. (1970). The complexity of finite objects and the development of the concepts of information and randomness by means of the theory of algorithms. *Russ. Math. Surveys*, 25:83–124.

From Quantum Gravity to Classical Phenomena

Michael Esfeld and Antonio Vassallo

Abstract. *Quantum gravity is supposed to be the most fundamental theory, including a quantum theory of the metrical field (spacetime). However, it is not clear how a quantum theory of gravity could account for classical phenomena, including notably measurement outcomes. But all the evidence that we have for a physical theory is based on measurement outcomes. We consider this problem in the framework of canonical quantum gravity, pointing out a dilemma: all the available accounts that admit classical phenomena presuppose entities with a well-defined spatio-temporal localization ("local beables" in John Bell's terms) as primitive. But there seems to be no possibility to include such primitives in canonical quantum gravity. However, if one does not do so, it is not clear how entities that are supposed to be ontologically prior to spacetime could give rise to entities that then are spatio-temporally localized.*

Introduction

The research for a theory of quantum gravity (QG), that is, a theoretical framework that extends quantum (field) theory to a theory of gravity, is one of the most long-lived enterprises in modern physics. The term "gravitational quanta" was used for the first time by Léon Rosenfeld (1930), but today—more than 80 years later—there is still no well established physical theory of quantum gravity.

There currently are two main types of approaches to QG. The first one, dubbed *covariant* QG, seeks to find a unification of all the fundamental interactions known in nature by enlarging the standard model of particle physics in order to include gravity, considered as a massless spin-2 field whose quanta are called *gravitons*: (super)string theories are the most notable variant of this type of approach.[1] The other type of approach, called *canonical* QG—the most worked

[1] See, e.g., (Green and Schwarz, 1984; Green et al., 1987).

T. Sauer, A. Wüthrich (eds.), *New Vistas on Old Problems. Recent Approaches to the Foundations of Quantum Mechanics*, Edition Open Access, 2013.

out representative of which is, today, loop quantum gravity (LQG)[2]—focuses on elaborating a formulation of general relativity (GR) suitable of being quantized using a physically well-defined procedure like, for example, Dirac's (1964) procedure; in this case, it seems more appropriate to talk about spacetime geometry rather than the gravitational field as the entity being quantized. For brevity's sake, the paper will deal only with the latter approach.

Although they are work in progress, both covariant and canonical approaches have so far produced many results of physical relevance,[3] showing that at least the leading theories of both types are rather well-developed. Therefore, a philosophical reflection on the foundations of QG is not only a legitimate enterprise, but a necessary step to be taken in order to achieve a better understanding of the conceptual issues involved in the field. Moreover, any theory of quantum gravity has to be empirically adequate, that is, it has to be able to account for the measurement results in quantum physics. In this paper, we shall therefore give an account of canonical QG and consider the challenges in getting from canonical QG to an account of classical phenomena such as measurement outcomes.

1 Canonical Quantum Gravity: A Brief Sketch

The canonical strategy aims to give a quantum description of gravitational phenomena by first formulating the conceptual machinery of GR in a Hamiltonian form and then using the methods of canonical quantization. The "recipe" for canonically quantizing a classical system can be summarized as follows. Given a physical system with n degrees of freedom coordinatized by the configuration variables $\{q_1, \cdots, q_n\}$, its Lagrangian L will satisfy Hamilton's action principle $\delta S = 0$, where S is the usual action defined as:

$$S = \int_{t_1}^{t_2} L dt, \tag{1}$$

[2] See, e.g., (Ashtekar, 1986; Rovelli and Smolin, 1990; Rovelli, 2004).

[3] See, e.g., (Rovelli, 2007).

which leads to the Euler-Lagrange equations[4]:

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{q}_h} - \frac{\partial L}{\partial q_h} = 0. \tag{2}$$

We define the momenta $\{p_1, \cdots, p_n\}$ conjugate to the configuration variables as

$$p_h = \frac{\partial L}{\partial \dot{q}_h}, \tag{3}$$

and we generate the Hamiltonian H of the system by a Legendre transformation:

$$H = \sum_h \dot{q}_h p_h - L. \tag{4}$$

This procedure amounts to switching the system's description from the coordinate system $\{q_1, \cdots, q_n\}$ in the n-dimensional configuration space Q to the coordinate system $\{q_1, \cdots, q_n; p_1, \cdots, p_n\}$ in the $2n$-dimensional phase space Γ. The dynamics of the system is now encoded in the Hamiltonian equations of motion obtained by applying Hamilton's principle to the action (1) where H is given by (4):

$$\begin{aligned} \dot{q}_h &= \frac{\partial H}{\partial p_h} = \{q_h, H\}, \\ \dot{p}_h &= -\frac{\partial H}{\partial q_h} = \{p_h, H\}, \end{aligned} \tag{5}$$

where the binary operation $\{\cdot, \cdot\}$ is the classical Poisson bracket. Such a system can be easily quantized by replacing the phase space Γ with the Hilbert space of complex functions on Q which are square-integrable with respect to Lebesgue measure and by introducing the quantum commutator $[\cdot, \cdot]$ as the quantum mechanical analog of the Poisson bracket. All the classically observable quantities are now self-adjoint operators defined over (a subspace of) the Hilbert space. A function Ψ belonging to this space is the wave function corresponding to a given

[4]Dotted quantities symbolize a derivative with respect to a suitably chosen real parameter. In the case of a classical system the most obvious choice is the usual Newtonian time t. Moreover, from now on, we will always assume that the subscript h ranges from 1 to n.

quantum state of the system whose dynamics is encoded in the Schrödinger equation[5]:

$$i\frac{\partial}{\partial t}\Psi = \hat{H}\Psi. \tag{6}$$

In the case of GR, one would expect that, since the theory has a well-behaved Lagrangian formulation, the above procedure, *mutatis mutandis*, would be carried out without particular problems. However, as we shall see in a moment, constructing a Hamiltonian formulation of GR is far from trivial and can be carried out only for a particular subset of models of the theory.

In the classical case[6], the Hamiltonian description of a system tells us how its dynamical state (q_h, p_h) evolves in absolute time. Moreover, the procedure which took us from (1) to (6), has been carried out only in the very simple case in which all the degrees of freedom of the system are *physical*. It may in fact happen that the solutions of the Euler-Lagrange equations (2) cannot be uniquely determined by specifying a set of initial conditions $\{q_1^0, \cdots, q_n^0; \dot{q}_1^0, \cdots, \dot{q}_n^0\}$, thus being determined only up to an arbitrary function of time.[7] This means that we have the "freedom" to choose this function without altering the physical description of the system. The immediate consequence of this fact is that the transformation (3) turns out to be non-invertible, i.e. there are further relations of the form $\mathfrak{X}_h(p,q) = 0$ called "constraints" between some of the dynamical variables. In this case, things become more complicated because now there is no more a one-to-one correspondece between Lagrangian and Hamiltonian, and the latter is determined only up to a linear combination of the constraints:

$$H' = H + \sum_h c_h \mathfrak{X}_h, \tag{7}$$

[5]Here and throughout the text we set $\hbar = 1$.

[6]The following treatment of constrained Hamiltonian systems is absolutely not rigorous and serves merely heuristic purposes. For an exhaustive treatment of such topics see, e.g., (Henneaux and Teitelboim, 1992).

[7]Just to be a little bit more precise, this happens when $\det||\frac{\partial^2 L}{\partial\dot{q}^i\partial\dot{q}^j}|| = 0$, with $(i, j = 1, \cdots, n)$.

where c_h are unknown coefficients and H is the "unconstrained" Hamiltonian. Hence, the equations of motion (5) now read

$$\begin{aligned} \dot{q}_h &= \{q_h, H\} + c_h\{q_h, \mathfrak{X}_h\}, \\ \dot{p}_h &= \{p_h, H\} + c_h\{p_h, \mathfrak{X}_h\}. \end{aligned} \tag{8}$$

In order to stress the fact that the constraint relations must be handled *after* the evaluation of the Poisson brackets, it is said that they vanish "weakly" and it is written $\mathfrak{X}_h(p,q) \approx 0$.

The constrained Hamiltonian formulation is of paramount importance in the treatment of systems that exhibit gauge invariance. The physical significance of the constraint relations, in fact, is that some degrees of freedom of the system are not physical but just gauge and would be eliminated by solving the constraints ("fixing the gauge"). For this reason, the quantities of most physical relevance for a constrained system - the *observables* - will be the gauge invariant ones, i.e., those quantities that have weakly vanishing Poisson brackets with the constraints. A similar remark can be made in the case of GR, since general relativistic systems exhibit a feature closely related to gauge invariance, i.e. general covariance[8], which simply speaking amounts to the fact that the physical description of a system is independent on the particular coordinatization chosen. Moreover, GR is a background independent theory, and a consequence of this feature is the absence of a primitive notion of time being intended either in a Newtonian or in a Minkowskian sense. Hence, we would expect a Hamiltonian formulation of GR not only to exhibit constraints but also to treat time as a degree of freedom of a gravitational system.[9]

Just to have a schematic idea of what it means to consider time as a degree of freedom of the system, let us consider the toy example of a classical non-relativistic particle.[10] To make things even more simple, let us take the particle to

[8] We will later refer to this feature as "diffeomorphism invariance." Albeit they are *not* the same thing, here we can harmlessly blur the distinction.
[9] The problem of time that we are going to spell out in a moment rests on the fact that this "temporal" degree of freedom turns out to be unphysical.
[10] See, e.g., (Kiefer, 2004, ch. 3, sec. 3.1).

have unit mass and only one spatial degree of freedom $q = x$. If a one dimensional potential $V(x)$ is present, then the Lagrangian of the particle will be

$$L = \frac{1}{2}\left(\frac{dx}{dt}\right)^2 - V(x). \tag{9}$$

In this extremely simple case, the procedure (1)-(6) is straightforward. But what happens if we "parametrize" Newtonian time, i.e., if we consider t as a new degree of freedom? In this case, once we have chosen a suitable real parameter $\tau \in \mathbb{R}$, the new Lagrangian becomes:

$$L' = \frac{dt}{d\tau}L = \frac{1}{2}\left(\frac{dx}{d\tau}\right)^2\frac{d\tau}{dt} - V(x)\frac{dt}{d\tau}. \tag{10}$$

The relation between the two Lagrangians stems from the fact that they must be both compatible with the same action (1), so it must be $Ldt = L'd\tau$. Now let's use equation (3) to calculate the momenta conjugate to x and t. We find

$$p_x = \frac{dx}{d\tau}\frac{d\tau}{dt} = \frac{dx}{dt}, \tag{11a}$$

$$p_t = -\frac{1}{2}\left(\frac{dx}{d\tau}\right)^2\left(\frac{d\tau}{dt}\right)^2 - V(x) = -\left(\frac{1}{2}\left(\frac{dx}{dt}\right)^2 + V(x)\right) = -H. \tag{11b}$$

We immediately notice that the "new" momentum conjugate to x is the same as in the "unparametrized" case, while the time conjugate momentum is nothing but the opposite of the Hamiltonian corresponding to the original Lagrangian L. The "new" Hamiltonian can be quickly calculated from (4) to yield:

$$\begin{aligned} H' = p_x\frac{dx}{d\tau} + p_t\frac{dt}{d\tau} - L' = p_x\frac{dx}{dt}\frac{dt}{d\tau} + p_t\frac{dt}{d\tau} + \frac{dt}{d\tau}(H - p_x\frac{dx}{dt}) = \\ = \frac{dt}{d\tau}(H + p_t) = \frac{dt}{d\tau}\mathfrak{X}_t, \end{aligned} \tag{12}$$

where the Hamiltonian corresponding to the "old" Lagrangian has been introduced through (10) and (4). Equation (12) is slightly more complicated than (7) because it also accounts for the fact that H' is a parametrized version of H but, nonetheless, the fact that $\mathfrak{X}_t$ is a constraint arising from (11b) comes out clearly. Obviously, in this case, the equations of motion cannot be calculated directly from

(8), but must be evaluated by considering (4) and then using the action principle (1)

$$\begin{aligned} S &= \int_{\tau_1}^{\tau_2} \left(p_x \frac{dx}{d\tau} + p_t \frac{dt}{d\tau} - H'\right) d\tau = \\ &= \int_{\tau_1}^{\tau_2} \left(p_x \frac{dx}{d\tau} + p_t \frac{dt}{d\tau} - \mathfrak{X}_t \frac{dt}{d\tau}\right) d\tau \end{aligned} \tag{13}$$

together with the condition $\mathfrak{X}_t \approx 0$. Up to now, we have somewhat followed steps (1)-(5) of the canonical quantization procedure. To "translate" the system into the quantum regime we just substitute the dynamical variables with operators acting on the Hilbert space of wave functions of the system:

$$q \longrightarrow \hat{q}\Psi = q\Psi, \tag{14a}$$

$$p_x \longrightarrow \hat{p}_x\Psi = -i\frac{\partial}{\partial x}\Psi, \tag{14b}$$

$$p_t \longrightarrow \hat{p}_t\Psi = -i\frac{\partial}{\partial t}\Psi. \tag{14c}$$

The remarkable point is that, in case of a constrained system, the only allowed wave functions that encode possible descriptions of the system are those which satisfy the quantum version of the constraints, in this case, taking into account definitions (14), we have:

$$\mathfrak{X}_t \approx 0 \longrightarrow \left(\hat{H} - i\frac{\partial}{\partial t}\right)\Psi = 0, \tag{15}$$

which is, as expected, the usual Schrödinger equation for a single particle. This example may seem rather twisted and artificial but helps us to straightforwardly point out a few important features of the canonical quantization procedure applied to constrained Hamiltonian system. The first is that, while an unconstrained system can be always quite easily parametrized and de-parametrized, i.e., we can always single out and eliminate the "extra" unphysical degrees of freedom, the opposite is not so simple, i.e., given a constrained system, it is in general far from trivial to determine which degrees of freedom are physical and which are not (GR is a clear example of such complexity). The second point is that, for a quantized constrained system, the Hilbert space of possible states of the system does not coincide with the "physical" Hilbert space, i.e., the space of solutions of the dynamical equations of motion (in the previous case, equation (15)). This means that, the more the constraints are complex, the more difficult it will be to sketch how the corresponding physical Hilbert space will look like.[11] However—and this is the third point—even simple constraints not always generate a trivial dynamics. A clear example of this is again (15). Bearing in mind these results, we can now take a closer look at the Hamiltonian formulation of GR.

Historically, the first to accomplish this task—thus completing the work by Dirac (1958)—were Arnowitt, Deser and Misner (Arnowitt et al., 2004), who elaborated the so called ADM formalism. To cut short, we can say that the Hamiltonian formulation of GR must describe how a suitably chosen parametrized physical entity changes for different values of the parameter. A reasonable move to perform before seeking for such physical entity is to put ourselves in the most simple situation possible, i.e., the case in which our general relativistic system consists of nothing but a pure gravitational field.[12] Even in this simplified case, however, there seems to be no natural candidate unless we further restrict ourselves to spacetimes that admit foliations into 3-surfaces Σ_τ: if the 4-manifold has in fact a topology of the type $\mathbb{R} \times \Sigma$, then we can straightforwardly interpret the Hamiltonian formalism of GR as describing how the 3-surfaces of the foliation Σ_τ change by varying the parameter $\tau \in \mathbb{R}$. The most intuitive configuration variables to be adopted in this context are the 3-metrics q_{ab} defined on the 3-surfaces and their conjugate momenta π_{ab} which encode information on how a given surface is embedded in the 4-manifold, i.e., its extrinsic curvature. The ADM formalism follows exactly this line of reasoning but adds a further simplification, i.e., it considers only globally hyperbolic spacetimes, so that the 3-surfaces of the foliation are space-like (Cauchy surfaces)[13] and the parameter τ can be chosen as a global time function.[14] In short, ADM formalism splits the 4-dimensional spacetime into space and time. In this way we can switch the "foliation view" to a more intuitive picture of a single spatial 3-manifold Σ evolving in "time" τ: this is why the ADM formalism is commonly referred to as "geometrodynamics."

[11] This is why we still do not have a completely worked out canonical theory of QG.
[12] To avoid the quicksand of the substantivalism vs relationism debate, we will consider this way of speaking equivalent to saying "empty spacetime" without further elaboration.
[13] Obviously, this choice dramatically cuts out the number of models of GR that can be described by this formalism: all spacetimes that do not admit a global time function are disregarded.
[14] The global hyperbolicity conditions assures the existence of such a function.

The immediate worry of such a splitting of the 4-dimensional GR in a $(3+1)$ theory is that the original 4-diffeomorphism invariance (which is preserved in the Lagrangian formulation) would be broken. Fortunately, this is not the case, since the diffeomorphic invariant character of the theory is now captured[15] in a set of constraints whose basic meaning is that not all points (q_{ab}, π_{ab}) in the phase space represent genuine physical states, i.e., different points related to different but diffeomorphic configurations represent the same physical situation. In general, we speak of a "diffeomorphism constraint," which encodes the 3-diffeomorphism invariance of the 3-manifold Σ, and a "Hamiltonian constraint," which accounts for the fact that the formalism does not depend on the specific parametrization adopted, i.e. on the particular choice of the paramenter $\tau \in \mathbb{R}$. We can collectively refer to these constraints as $\mathfrak{X}_h(q_{ab}, \pi_{ab}) \approx 0$. In the end, then, the full Hamiltonian description for GR will be given by an action principle analog to (13). With all this machinery in place, we can now quantize the theory. We can choose either to solve first the constraints and then to quantize or the other way round: the most common choice is the latter, because it slightly simplifies the calculation, however—at least in principle—both choices lead to the same final results. The immediate consequence of the presence of constraints is that the dynamical evolution of a "gravitational state" Ψ will be generated by a set of equations that resemble (15):

$$\hat{\mathfrak{X}}_h \Psi = 0. \tag{16}$$

Equations (16)—often collectively referred to as Wheeler-DeWitt equation[16]—highlight two major (and interrelated) problems in canonical QG, which are inherited from the classical regime. The first one is the problem of observables. If we take a physically relevant quantity as one which has weakly vanishing Poisson bracket with the constraints (in the classical case) or as an operator that "produces" states annihilated by the constraints[17] (in the quantum case), then the only observables of both theories will be quantities that do not change *in time*. This issue is a direct consequence of the so called "problem of time," i.e. the impossi-

[15]Not *always* captured: there are cases in which the Hamiltonian description of a general relativistic system fails to encode the diffeomorphic invariance of GR. However, we do not need to worry about this issue here.

[16]Some authors call in this way only the Hamiltonian constraint since they interpret it as delivering a genuine dynamics while they consider only the diffeomorphism constraint as generator of the gauge transformations. We will say something more on this kind of interpretations of (16) in Section 3.

[17]More precisely, if $\hat{O}$ is such an operator, then it must be $[\hat{O}, \hat{\mathfrak{X}}_h]\Psi = 0$.

bility to define a classical notion of time neither in GR nor in canonical QG. The Hamiltonian formulation of GR emphasizes this issue by suggesting that any possible notion of time merely refers to a gauge fixing and, hence, is unphysical. An escape route might be to distinguish between quantities compatible only with the diffeomorphism constraints ("observables") and quantities compatible also with the Hamiltonian constraint ("perennials"), as suggested for example by Kuchař (1993). Another possible solution is to discriminate between quantities associated with measurements ("partial observables") and quantities whose value or probability distribution can be predicted by the theory ("complete observables") as suggested by Rovelli (2002). This, of course, partially shifts the problem onto finding a consistent account of measurement in the quantum gravitational context. Tackling these issues is the main task if we want to shed light on a route that leads from the quantum gravitational regime to classical phenomena.

2 The Measurement Problem of Quantum Mechanics

As regards the account of classical phenomena, the very formulation of non-relativistic quantum mechanics poses a problem that is known as the measurement problem. Relativistic quantum mechanics—that is, quantum field theory—faces this problem as well. Quantum gravity being the project of unifying quantum field theory with general relativity theory, it is not to be expected that quantum gravity will solve the measurement problem. Nonetheless, any approach to quantum gravity that is to be empirically adequate has to take a stance on the measurement problem, the question being how to account for measurement outcomes within a quantum theory, including a quantum theory of gravity. Let us therefore go into this problem and consider its consequences for a theory of quantum gravity.

A clear conceptualization of the measurement problem can be found in (Maudlin, 1995, 7):

1A The wave-function of a system is *complete*, i.e., the wave-function specifies (directly or indirectly) all of the physical properties of a system.

1B The wave-function always evolves in accord with a linear dynamical equation (e.g. the Schrödinger equation).

1C Measurements of, e.g., the spin of an electron always (or at least usually) have determinate outcomes, i.e., at the end of the measurement the measuring device is either in a state which indicates spin up (and not down) or spin down (and not up).

The problem is that there can be no formulation of a quantum theory that respects all three of these propositions, because their conjunction is inconsistent: if the wave function yields a complete description of the properties of a system and if it always evolves according to a linear dynamical equation, then it cannot evolve in such a way that it represents a quantum system as having a determinate value of a dynamical property—such as a definite position or a definite value of spin–and a measuring device as indicating such a determinate value.

The notion of measurement is immaterial to the formulation of this problem. There is no physical definition of what a measurement is: measurement interactions are not a special type of interactions in addition to the strong, the weak, the electromagnetic, and the gravitational interactions, but are simply ordinary physical interactions; and measuring devices are not natural kinds in addition to electrons, protons, the chemical kinds, biological species, etc. Any macroscopic system capable of amplifying the properties of quantum systems can be used as a measuring device. One can therefore replace proposition 1C above with the following, slightly more complicated proposition that does not refer to measurements, but only to positions of macroscopic systems:

1C* The macroscopic systems with which we are familiar—such as, e.g. tables, trees, cats, people, and the like—always (or at least usually) have determinate positions in space, and these systems are composed of microscopic quantum systems.

Consequently, quantum systems, whatever they are, must at least sometimes have positions that are determinate enough so that they can compose macroscopic systems that have determinate positions. But if the wave function specifies all the properties of quantum systems and if the wave function always evolves in accord with a linear dynamical equation, it is impossible that quantum systems have positions that are determinate enough so that they can compose macroscopic systems that have determinate positions, due to the superposition principle and the entanglement of the states of quantum systems.

3 Two Conservative Solutions of the Measurement Problem

The measurement problem shows that if one retains 1C or 1C*—that is, the proposition that macroscopic systems usually have definite positions in space or spacetime—one has to give up either 1A or 1B. Such solutions can be regarded as conservative in the sense that they retain the ordinary presupposition of macroscopic systems having definite positions in space or spacetime so that measurements have definite outcomes.

If one drops 1A and thus maintains that the wave function does not tell the whole story about what there is in the physical world, the only precisely formulated theory that elaborates on this idea is Bohm's quantum mechanics (Bohm, 1952; Bohm and Hiley, 1993). Bohm's theory starts from the trivial fact that macroscopic systems such as measuring devices cannot have a determinate position unless the microscopic systems that compose them also have a rather determinate position. It then adds the—controversial—claim that these microscopic systems cannot acquire a rather determinate position in space and time unless they always have one. In other words, Bohm's theory introduces a determinate value of position for any physical system as an additional variable that is not specified by the wave function. This variable is hidden in the case of microphysical systems in the sense that it is not possible to find out the exact positions of microphysical systems without changing them. On this basis, the quantum probabilities have the same status as the probabilities in statistical mechanics, namely to yield all the knowledge that we can obtain given our ignorance of the exact initial conditions. In short, the ontology of Bohm's theory consists in particles whose positions are correlated with each other and a global law of motion (sometimes referred to as quantum potential or guiding field or pilot wave), spelling out how the positions of the particles taken together develop in time.[18]

It may seem that since Bohm's quantum theory works in terms of particles, it is a non-starter when it comes to quantum field theory and quantum gravity. However, the point of Bohm's theory is to provide an ontology of quantum physics by answering the question of what the formalism tells us about the physical world in terms of it referring to positions of something; that answer is justified by arguing that if the fundamental physical objects, whatever they are, were not characterized by determinate positions, macroscopic objects could not have determinate positions either. The question is whether that latter claim is correct. That claim is

[18] See (Goldstein, 2009, sec. 5 and 15).

not tied to conceiving the fundamental physical objects as enduring particles. Indeed, since (Bell, 1987, ch. 19), there are proposals for a Bohmian quantum field theory around,[19] and there also is a sketch of a Bohmian theory of quantum gravity (Goldstein and Teufel, 2001). The basic idea behind this sketch is to recover a notion of time from (16) as a hidden variable. The starting point[20] is to consider only the diffeomorphism constraint as encoding the gauge freedom of the theory and to take the Hamiltonian constraint as some sort of stationary Schrödinger-like equation which involves a "universal" wave function. Under this framework, we can say that each spacetime point carries three "distinct pieces of physical information" or, less metaphorically speaking, at each point on a 3-surface Σ_t a coordinate system can be found where the 3-metric q_{ab} is represented by a 3×3 matrix in diagonal form with these three pieces of physical information being just the elements on the diagonal. Two of these pieces of information account for the gravitational field (according to the view that gravity is a massless spin-2 field), while the third gives a measure of how much the geometry of Σ_t would change if the point were infinitesimally "pushed" toward a neighbouring 3-surface Σ_{t+dt}. In this sense, this third piece of information generates a notion of "forward in time" which is hidden in the geometry of a 3-surface. Thus, the Wheeler-DeWitt equation accounts for static universal configurations of "all elements of physical reality": what these elements of reality should be and how the theory should single out a wave function from them is still an open question.

The main problem in this context is that Bohmian mechanics is not Lorentz-invariant.[21] Consequently, it breaks the diffeomorphism invariance of general relativity. Thus, in non-relativistic Bohmian mechanics, if one had complete knowledge of the positions of the particles, that knowledge would reveal a preferred foliation of spacetime. However, since one cannot have complete knowledge of the positions of Bohmian particles (given that any measurement changes the positions of the particles), it is also in Bohm's theory not possible to send signals with a superluminal velocity and to know the objective, globally preferred foliation of spacetime.

In standard textbooks from (von Neumann, 1932) on, quantum mechanics is presented in the form of a combination of two radically different dynamics: when no measurement takes place, one uses the Schrödinger equation to calculate the temporal development of the wave function of a quantum system. However, when

[19] See, in particular, (Dürr et al., 2005, 2004).
[20] See, e.g., (Kuchař, 1992).
[21] See (Albert, 1992, 155–161), for a nice illustration why this is so.

a measurement is made, the wave function is supposed to collapse so that it represents the system as having one determinate value of the measured property at the exclusion of all the other ones. Textbook quantum mechanics thus rejects proposition 1B above: the wave function completely describes the properties of physical systems, but under some circumstances—measurements being a case in point—quantum systems change in such a way that they acquire a determinate value of dynamical properties, that change being represented by the collapse of the wave function. Is it possible to make this idea precise so that one specifies when (under what circumstances) and how this change happens? Doing so requires amending the Schrödinger equation. The only precise physical proposal in this sense goes back to Ghirardi, Rimini and Weber (Ghirardi et al., 1986) (GRW). GRW add a stochastic term to the Schrödinger equation such that, in brief, a single microscopic quantum system has a very low objective probability to undergo a spontaneous localization. However, when one considers a macroscopic system that is composed of a huge number of microscopic quantum systems, one of these microscopic systems will immediately undergo a spontaneous localization so that, due to the entanglement, the whole system will be localized. When one couples a quantum system to a macroscopic system, due to the quantum system thus becoming entangled with the huge number of quantum systems making up the macroscopic system, it will also undergo a spontaneous localization very rapidly.

Nonetheless, it remains to be spelled out what exactly in the physical world the GRW dynamics represents, in other words, what the ontology of the GRW theory is. Taking textbook quantum mechanics literally, we have to say that a quantum system such as an electron, when not having a determinate value of position, is smeared out in space. What the GRW dynamics then achieves in improving on the collapse postulate in the textbooks is to describe how this position distribution, which is smeared out in 3-dimensional physical space, develops into rather determinate values. This is indeed the reading of the physical significance of the GRW dynamics that Ghirardi et al. (1995) themselves favor in proposing a mass density ontology: the mass of, say, an electron when it has not a determinate position is literally smeared out in physical space, creating thus a mass density field. However, the mass density ontology, like Bohmian mechanics, is not Lorentz-invariant.

But there is another reading of the GRW dynamics possible. That reading is due to Bell (1987, 205). A good way to access it is via a comparison with Bohmian mechanics: in Bohm's theory, quantum systems *always* have a determinate position, and the determinate value of position is not taken into account in the wave function description. According to what is known as the GRW flash theory, quantum systems have a determinate value of position *only* when the wave function as developing according to the GRW modification of the Schrödinger dynamics indicates such a value (that is, when a spontaneous localization occurs), and these sparse determinate positions are *all* there is in the world. To put it differently, the spontaneous localizations that GRW postulate are conceived as flashes centred around spacetime points, and these flashes are all there is in spacetime. Starting with an initial distribution of flashes, the wave function is a tool to calculate the probabilities for the occurrence of further flashes.

The flash ontology is such that its dynamics can be formulated in a Lorentz-invariant manner, since it abandons the idea of continuous trajectories of anything in spacetime (such as Bohmian particles or field values, or mass densities in Ghirardi's ontology for GRW). Even if one had exact and complete knowledge of the flash distribution, one could not infer from that knowledge an objective foliation of spacetime.[22] More precisely, it is the *only* worked out proposal for an interpretation of what quantum mechanics (or quantum field theory for that matter) tells us about the dynamics of matter in four-dimensional spacetime that has the chance of being Lorentz-invariant (the chance, since the formulation of Tumulka (2006) does not take interacting fields into account).

Both Bohm's theory and the GRW theory—on the mass density version as well as on the flash version—solve the measurement problem by accepting positions of something in spacetime (be it particles, be it field values, be it events such as flashes, be it the density of stuff) as primitive.[23] In other words, they accept what John Bell calls "local beables" as primitive,[24] differing in the local beables that they pose. On this basis, they then can account for definite positions of macroscopic systems in spacetime and thus retain proposition 1C (and 1C*). However, when it comes to quantum gravity, the problem is that the presupposition of accepting positions of something in spacetime as primitive can no longer be taken for granted, since the very concept of spacetime breaks down starting from the ADM formalism. If we take equations (16) as acting on a universal wave function, then the only thing we can talk about are 3-geometries (possibly coupled with matter fields) *as a whole*, which means that we are not dealing anymore with

[22] See (Tumulka, 2006) and (Maudlin, 2011, ch. 10).
[23] See (Allori et al., 2008) for an illuminating comparison of the ontologies of Bohm, GRW mass density and GRW flash.
[24] See (Bell, 1987, ch. 7).

something happening in spacetime but with universal timeless spatial configurations. Thus, in the "extreme" timeless interpretation[25] of (16), we could have, for example, "flashes" in the configuration space (whatever this would mean), but surely not in spacetime. Moreover, even if we could find a suitable interpretation of (16) that permits us to talk about (partial) observables at the Planck scale, still those observables would not be in spacetime but, rather, they would constitute the very fabric of it.[26] Let us therefore look into positions that solve the measurement problem by abandoning 1C.

4 The Everett Interpretation

If one rejects 1C (and thus 1C*), one can regard the wave function as providing a complete description of the properties of physical systems (1A) and one does not have to amend the dynamics (1B). However, one has to replace 1C with an account of how it comes about that it seems to observers that there are determinate values of properties of themselves (their consciousness and their body) as well as their environment. In order to achieve such an account, it is common to draw on decoherence. Although decoherence does not lead to less, but to more entanglement, the quantum system becoming entangled with all the systems in its environment, the wave function of the whole system (quantum system and environment) rapidly develops in such a way that the superposed correlations do not interfere with each other. As far as the formalism of quantum mechanics is concerned, decoherence hence means a development of the wave function (or state vector or density matrix) in a high-dimensional mathematical space such that the interference terms between the superposed correlations vanish. The crucial issue then is to work out an answer to the question of how to get from this development of wave functions in a mathematical space to observers to whom determinate values of dynamical properties appear. Taking simply for granted that such observers somehow emerge out of or supervene on wave functions in a high-dimensional mathematical space evidently does not do the job of a precise physical account.

The only account available in the literature is of the following type: the physical significance of the vanishing of the interference terms between the superposed correlations is that decoherence induces a splitting or a branching of the universe

[25] For a paradigmatic example of such an interpretation see, e.g., (Barbour, 1999). In the next section, we will see in slightly more detail some consequences of this interpretation.
[26] See, e.g., (Rovelli, 2004, ch. 1, sec. 1.2.2).

into many non-interfering branches such that each of the superposed correlations constitutes at least one branch of the universe. Each of these branches that emerge due to decoherence constitutes a quasi-classical world. Thus, there is one branch in which the electron has spin up, the measuring device indicates spin up and the observer is conscious of the measuring device indicating spin up; and there is another branch in which the same electron has spin down, the same measuring device indicates spin down and the same observer is conscious of the measuring device indicating spin down. Since there are many measurements for which there are infinitely many possible outcomes—position measurements are a case in point—, this view is committed to maintaining that decoherence leads to the emergence of infinitely many branches. This position is therefore known as the many worlds interpretation of quantum mechanics, going back to Everett (1957).[27]

However, this proposal leaves a number of questions open. If the idea is that whenever there is decoherence, the whole physical universe develops into many branches, this means that each system in the universe—including its mass, its charge, etc.—is many times copied; but it is unclear how such a physical multiplication of mass and charge could be brought about. Furthermore, it is unclear whether or not the branching concerns spacetime itself. If it did not include spacetime, contradictory predicates would apply to one and the same spacetime region, or even contradictory properties would be instantiated by one and the same spacetime region—such as a measuring device indicating spin up and the same measuring device indicating spin down existing in or being properties of the same unique spacetime region. One can avoid this consequence by conceiving the branching as concerning spacetime itself; but then one would have to develop a physically precise account of how spacetime itself can be many times duplicated whenever there is decoherence so that many superposed spacetimes come into being. Moreover, since the branching is supposed to affect instantaneously the whole of spacetime, it is unclear whether and how the branching could be Lorentz-invariant.[28]

Furthermore, decoherence is a process leading from superposed correlations with interference terms to the vanishing of interference. The account under consideration replies to the question of the physical significance of this process by maintaining that many branches of the universe come into being that do not interfere with each other. But what is the physical significance of the entangled state of the universe prior to the emergence of the branches? Does this state consist in

[27] See (Wallace, 2010) for a concise statement and, in general, the papers in (Saunders et al., 2010).
[28] See (Barrett, 1999, 159–160).

objects being smeared out in spacetime that upon decoherence get split up into all their possible determinate values of position in different branches of the universe? Kiefer (2004, ch. 10, sec. 10.1.2) tries to give a formal account of the appearance of global spacetime variables, such as time itself, from a quantum cosmological context. He summarizes the results as follows:

> The lesson to be drawn is thus that the universe can appear classically only if experienced from within. A hypothetical "outside view" would only see a static quantum world. The most natural interpretation of quantum cosmology is an Everett-type interpretation, since the "wave function of the universe" contains by definition all possible branches. As macroscopic observers, however, we have access only to a tiny part of the cosmological wave function—the robust macroscopic branch which we follow. (*Ibid.*, 318)

Kiefer's approach rests on an approximation technique similar to the so called "Born-Oppenheimer approximation."[29] The basic idea is to decompose the "universal" wave-function Ψ in (16) as follows:

$$\Psi = \Psi_G \times \psi_B, \tag{17}$$

where Ψ_G describes the full gravitational field and ψ_B accounts for the remaining non-gravitational degrees of freedom. Very loosely speaking, this decomposition introduces a picture of the universe where a timeless "global" part generates the dynamical evolution of a "local" part representing the various branches. However, we must be very careful when drawing any conclusion from (17) for the simple reason that it is just an *approximation* and, hence, a light-hearted metaphysical reading of it might mislead us to consider as real features what are just artifacts of the mathematical manipulation of (16).

An even more counter-intuitive—and worrisome—account for the emergence of classical properties is Barbour's (1994a; 1994b) approach. The most important physical entity for him is the "reduced" configuration space Q_0 found by solving the diffeomorphism constraint in (16). Once this space has been found, the Hamiltonian constraint can be interpreted as giving a probability distribution over it. The important point is that there are not many possible probability distributions, but only *one* which is fixed in some way by the structure of the reduced configuration space, which means that there is no Hilbert space of wave

[29] See, e.g., (Sakurai, 1994, 474) for a brief technical treatment.

functions. Each point in Q_0 is a moment or a "now," in the sense that it represents a universal static configuration. This framework calls for an Everett-type interpretation because it involves many "nows," viz. many static pictures of the universe.[30] How can we accommodate in the first place our experience of a time in a such static framework? The answer involves the concept of a "time capsule" which is a "static configuration of part or all the universe containing structures which suggest they are mutually consistent records of processes that took place in a past in accordance with certain laws" (Barbour, 1994b, 2884). A time capsule, then, is a point in Q_0 with an associated peak in the probability distribution. But probability of *what*? It seems that Barbour interprets the wave function as giving the probability for a "now" to be experienced. In his words:

> The timeless wavefunction of the universe concentrates the quantum mechanical probability on static configurations that are time capsules, so that the situations which have the highest probability of being experienced carry within them the appearance of time and history. (Barbour, 1999, 30)

This interpretation reintroduces some sort of link between wave function and observer in a somewhat Copenhagen-like fashion. However, this immensely complicates the matter because, in addition to the strange fact that a mathematical object as the wave function "selects" what elements of Q_0 are to be experienced, we have to give a further account of how the observer "experience" occurs, i. e. how a static "brain configuration" embedded in a universal "now" generates the awareness of change in time. In short, it seems that the cure is much worse than the disease.

Conclusion: A Dilemma

If quantum gravity is to be a fundamental physical theory, it has to include an account of how to get from the entities that are posed as fundamental to classical phenomena such as measurement outcomes, since the evidence for a physical theory consists in measurements. However, as we have argued in this paper, setting out to do so runs into a dilemma: if one endorses a commitment to there really being classical properties (and thus definite measurement outcomes) in the

[30] This way of speaking can be misleading: here the picture *is* the universe, so each "now" represents a distinct universe.

world, accepting proposition 1C (and 1C*) above, then the only worked out accounts available presuppose positions in space or spacetime—of particles, field entities, events such as flashes, or density of stuff—as primitive. However, quantum gravity calls into question such primitives. As we have seen, in the formal process of building a theory of quantum gravity starting from the ADM formulation of GR and ending up with equations (16), the room for accommodating classical properties dramatically shrinks. Already in the classical regime the very notion of spacetime is weakened and eventually disappears in the quantum transition. From this point of view, the "problem of time" in QG is just the tip of the iceberg of a general collapse of the picture given by textbook quantum mechanics.[31] Equations (16), taken at face value, tell us a strange tale of a frozen dynamics of blocks of universal 3-geometries, where the notions of measurement and observable are put in jeopardy: is the wave function of the universe related to measurement outcomes? And what would be the physical meaning of setting up such a measurement? Even the best worked out attempts to recover more familiar notions from this picture (as in LQG, in the first place) cannot do much but end up dealing with quanta of area and volume or other physical entities that are alien to the notion of "position in space at a given time." But if one abandons the commitment to there really being classical properties (and thus definite measurement outcomes) in the world, thus dropping proposition 1C (and 1C*) above, there is no clear strategy available as to how to account for the appearance of a classical world to observers: at the present time, the formally best worked out accounts either involve questionable assumptions (e.g. approximations) or provide partial and unconvincing explanations. However, a literal reading of (16), such as Barbour's, makes it extremely difficult to find an account of the emergence of the classical world. The hope is that from a better understanding of (16) will follow a solution to the dilemma.

References

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

[31] For example, the usual account of unitarity as the fact that the probabilities associated to the possible outcomes of a given experiment must add up to unity *at a fixed time* is obviously problematic in QG.

Allori, V., Goldstein, S., Tumulka, R., and Zanghì, N. (2008). On the common structure of Bohmian mechanics and the Ghirardi-Rimini-Weber theory. *British Journal for the Philosophy of Science*, 59:353–389.

Arnowitt, R., Deser, S., and Misner, C. (2004). The dynamics of general relativity. *arXiv:gr-qc/0405109v1*.

Ashtekar, A. (1986). New variables for classical and quantum gravity. *Physical Review Letters*, 57:2244–2247.

Barbour, J. (1994a). The timelessness of quantum gravity: I. the evidence from the classical theory. *Classical and Quantum Gravity*, 11:2853–2873.

Barbour, J. (1994b). The timelessness of quantum gravity: II. the appearance of dynamics in static configurations. *Classical and Quantum Gravity*, 11:2875–2897.

Barbour, J. (1999). *The End of Time. The Next Revolution in Physics*. Oxford University Press.

Barrett, J. (1999). *The Quantum Mechanics of Minds and Worlds*. Oxford University Press.

Bell, J. (1987). *Speakable and Unspeakable in Quantum Mechanics*. Cambridge University Press.

Bohm, D. (1952). A suggested interpretation of the quantum theory in terms of "hidden" variables. *Physical Review*, 85:166–193.

Bohm, D. and Hiley, B. (1993). *The Undivided Universe. An Ontological Interpretation of Quantum Theory*. Routledge.

Dirac, P. (1958). The theory of gravitation in hamiltonian form. *Proceedings of the Royal Society A*, 246:333–343.

Dirac, P. (1964). *Lectures on Quantum Mechanics*. Belfer Graduate School of Science, Yeshiva University.

Dürr, D., Goldstein, S., Tumulka, R., and Zanghì, N. (2005). Bell-type quantum field theories. *Journal of Physics A*, 38:R1–R43.

Dürr, D., Goldstein, S., and Zanghì, N. (2004). Bohmian mechanics and quantum field theory. *Physical Review Letters*, 93:090402.

Everett, H. (1957). “Relative state” formulation of quantum mechanics. *Reviews of Modern Physics*, 29:454–462.

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

Ghirardi, G., Rimini, A., and Weber, T. (1986). Unified dynamics for microscopic and macroscopic systems. *Physical Review D*, 34:470–491.

Goldstein, S. (2009). Bohmian mechanics. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Spring 2009 edition.

Goldstein, S. and Teufel, S. (2001). Quantum spacetime without observers: Ontological clarity and the conceptual foundations of quantum gravity. In Callender, C. and Hugget, N., editors, *Physics Meets Philosophy at the Planck Scale*, pages 275–289. Cambridge University Press.

Green, M. and Schwarz, J. (1984). Anomaly cancellation in supersymmetric $d = 10$ gauge theory requires $so(32)$. *Physics Letters B*, 149:117–122.

Green, M., Schwarz, J., and Witten, E. (1987). *Superstring Theory*. Cambridge University Press.

Henneaux, M. and Teitelboim, C. (1992). *Quantization of Gauge Systems*. Princeton University Press.

Kiefer, C. (2004). *Quantum Gravity*. Oxford University Press.

Kuchař, K. (1992). Time and interpretations of quantum gravity. In Kunstatter, G., Vincent, D., and Williams, J., editors, 4^{th} *Canadian Conference on General Relativity and Relativistic Astrophysics*, pages 211–314. World Scientific.

Kuchař, K. (1993). Canonical quantum gravity. In Gleiser, R., Kozameh, C., and Moreschi, O., editors, *General Relativity and Gravitation*, pages 119–150. Institute of Physics Publishing, Bristol.

Maudlin, T. (1995). Three measurement problems. *Topoi*, 14:7–15.

Maudlin, T. (2011). *Quantum Non-locality and Relativity*. Wiley-Blackwell, third edition.

Rosenfeld, L. (1930). Über die Gravitationswirkungen des Lichtes. *Zeitschrift für Physik*, 65:589–599.

Rovelli, C. (2002). Partial observables. *Physical Review D*, 65(12):124013.

Rovelli, C. (2004). *Quantum Gravity*. Cambridge University Press.

Rovelli, C. (2007). Quantum gravity. In Butterfield, J. and Earman, J., editors, *Philosophy of Physics*, Handbook of the Philosophy of Science. Elsevier B.V.

Rovelli, C. and Smolin, L. (1990). Loop space representation of quantum general relativity. *Nuclear Physics B*, 331:80–152.

Sakurai, J. (1994). *Modern Quantum Mechanics. Revised Edition*. Addison-Wesley Publishing Company.

Saunders, S., Barrett, J., Kent, A., and Wallace, D., editors (2010). *Many Worlds? Everett, Quantum Theory, and Reality*. Oxford University Press.

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

von Neumann, J. (1932). *Mathematische Grundlagen der Quantenmechanik*. Springer.

Wallace, D. (2010). Decoherence and ontology (or: How I learned to stop worrying and love fapp). In Saunders, S., Barrett, J., Kent, A., and Wallace, D., editors, *Many Worlds? Everett, Quantum Theory, and Reality*, pages 54,72. Oxford University Press.

The Conserved Quantity Theory of Causation and Entangled States

Jakob Sprickerhof

Abstract. *I will argue that processes in Quantum Field Theory can be understood in terms of a new version of the conserved quantity theory of causation. The idea is that causation is the transfer of energy-momentum from cause to effect. This has implications for further topics in the interpretation of quantum physics. I will adopt a proposal due to David Wallace for describing quantum entities as localized in regions of space and show how this gains plausibility by setting it into the context of causation. With this background, I will argue that pre-measurement entangled states are not a structure of two or more related entities, but one spatially extended entity.*

1 Introduction

In this paper, I will investigate two widely discussed topics of philosophy of physics: causation and entanglement. Regarding the first one, I will argue that there is causation in physics and regarding the second, I will argue that entanglement should not be understood as a (causal) relation.

The discussion about causation in physics was opened by Bertrand Russell (1913), but only proliferated from the 1980s on with works amongst others by Philip Kitcher and Wesley Salmon (1989) on explanation and causation and by Nancy Cartwright (1989) on capacities. The conserved quantity theory of causation (CQT) identifies causation as the physical process of the exchange of a conserved quantity between physical entities. The latest version of this theory was formulated by Phil Dowe in 2000, and has mostly been neglected, mainly because it is hardly compatible with modern physics. However, in Sections 2, 3 and 4, I will try to revive the core idea of CQT by showing how it can successfully be applied to Quantum Field Theory (QFT).

T. Sauer, A. Wüthrich (eds.), *New Vistas on Old Problems. Recent Approaches to the Foundations of Quantum Mechanics*, Edition Open Access, 2013.

Entangled states and EPR-correlations were introduced by Einstein, Podolsky and Rosen to demonstrate the incompleteness of quantum mechanics. With the work of John Bell and follow-up experiments, they became an enigma. I see two main problems which are involved here. (1) The problem of localization: How can events be correlated over space-like distances? (2) The measurement problem: How can a superposition of states evolve into distinct states with definite properties? In this paper, I am only concerned with the first problem, which is related to causation. When trying to explain entangled states, the question arises whether there is a causal relation between two entangled objects. To evaluate this, we need *criteria* that tell us what a causal relation is. This motivates Sections 2, 3 and 4. Should it turn out that there is no causal relation involved in entangled states, we need an alternative explanation of the EPR-correlations. This motivates Sections 5 and 6.

The conclusion will be that, before a measurement happened, entanglement between systems can neither be understood as causal relation nor in other structuralistic terms. The alternative picture, that I will draw, is one in which quantum mechanical entities are not (always) localized in a small region of spacetime, but can be extended over a larger region. Accordingly, an entangled state is not a structure formed by two or more entities and a relation, but one extended entity.

2 What Is Causation and How Do We Find It?

Causation is one of the topics in philosophy where there is huge disagreement among scholars. Is there causation or not? If yes, what kind of relation is it? These are just the two most fundamental questions surrounding the subject. Furthermore, there is a vast field of proposals for what causation could be. The range of possibilities reaches from necessary connections over contingent regularities over to manipulation by humans. The ample nature of the concept of causation makes it difficult to even find a starting point for one's investigation.

Phil Dowe introduced a sensible distinction into this discussion, which in my eyes should stand at the beginning of every new proposal for what causation could be; it is the distinction between conceptual and empirical analyses of causation. Conceptual analysis is the task "to elucidate our normal concept of causation," whereas empirical analysis tries "to discover what causation is in the objective world" (Dowe, 2000, 1). The conceptual analysis sets out to clarify the use of the word "causation" in ordinary language. In principle, every competent speaker of a certain language can perform conceptual analysis, without looking into the world. The aim is to explain the concept of causation by expressing it through other concepts, that are better understood, and to spell out the logical consequences of this explication. Empirical analysis, on the other hand, intends to find a process in the world, which can be identified as being a causal process. This is an empirical investigation and thus cannot be accomplished without taking notice of our best sciences; for they are the place to look at, if we want to know how the world is like.

Dowe acknowledges that the distinction is not clear-cut. The ambiguity is made obvious by the asymmetry between both methods of analysis. While conceptual analysis proceeds without taking empirical knowledge into account (though it may have consequences for our way of looking at the world), the empirical analysis is dependent on preliminary conceptual analysis. It is impossible to make plausible that any physical process is a causal relation without having at least a vague understanding of what we mean when we use this concept. If empirical analysis is not backed up by conceptual analysis it could be claimed of anything that it is causation. However, Dowe holds this to be unproblematic. In his eyes, conceptual knowledge about causation is already encoded into science, since scientists are competent speakers of their working language. Hence, we only have to look into science for suitable processes and the conceptual work is already done by the scientists.

Dowe's seemingly 'careless' stance is criticized by Thomas Bontly (2006). He argues that it is unclear how and to which extent the tacit knowledge of scientists about causation is comprised in their scientific theories. Therefore, it is ambiguous whether and how the empirical analysis can be carried out. At least some explicit criteria for what causation is are required before we look into science and see whether there is something that meets them.

It is difficult, if not impossible, to find necessary and sufficient criteria for what causation is. Like Bontly (2006, 192f.), I will present *plausible* criteria, due to several different authors, that a concept of causation has to suffice. This list is not intended to give all and only necessary conditions. Nevertheless, the criteria should be sufficient to identify a physical process as causal and make the choice at least reasonable.

1. *Causation is an (intrinsic) relation between distinct entities* (Menzies, 1996, 98).

   I hold this criterion to be necessary for causation. The conditions under which entities count as distinct need to be specified later. “Intrinsic” means that the causal relation is independent of everything except the entities that are causally related. I have put “intrinsic” in brackets, because Humeans about causation do not agree on this. Leaving it out does not much harm in the context of this paper. Furthermore, it is important to note that the *relation* needs to be specified, in order to accomplish the aim of explaining what causation is. The relation could be for instance spatial connection (one object hitting the other) or a physical force.

2. *One can manipulate the effect by manipulating the cause* (Bontly, 2006, 193).
   This criterion is central for physical practice. Whenever something is changed on the (alleged) cause, something should happen to the (alleged) effect. In the extreme, the effect should disappear when the cause is removed. I take manipulability to be the second necessary criterion for causation. A slightly different criterion, that has nevertheless the same consequences, is expressed by H. D. Mellor (1988, 230): “If an effect is an end, its causes are means to it.”

3. *The cause typically increases the chance of the effect* (Menzies, 1996, 100).
   This leaves the possibility open for chance-lowering causation. However, in general we should observe the effect more often when the cause is present, compared to when it is not. This seems to be neither a necessary nor a sufficient criterion for causation. Nevertheless, if one finds a chance raising relation, this is a good indicator that one has found a causal relation.

4. *Causation is a stable relation between cause and effect.*
   This is a rather vague statement. The interpretation depends on the actual stance one takes towards causation. If one wants to defend a non-Humean theory of causation then “stable” means “necessary.” If on the other hand one wants do defend a Humean theory, then “stable” might mean “regularly.” Additionally, stability is supposed to catch the central meaning of two further criteria by Mellor (1988, 230), namely “causes and effects are evidence for each other,” and “causes explain their effects.”

Some authors hold a fifth criterion to be essential for causation,[1] that I will neglect in the following:

5. *Causes precede their effects in time* (Bontly, 2006, 193).
   In my eyes, this is neither necessary nor sufficient for causation, but something we should leave to physics to decide. If we can identify causation by using the other four criteria then it might be the case that this relation has a fixed order in time. However, it might as well turn out that the order is changed under certain circumstances. The latter, I think, has no influence on whether we found a causal relation or not.

The line of argument now will be first to present Dowe's empirical theory of causation and discuss how it fares in the light of modern physics. The deficiencies of Dowe's theory that show up motivate an updated version of it. Then, in chapter 4, I will show that the first four criteria mentioned above can be applied to QFT and present the updated empirical theory of causation.

3 Phil Dowe's Conserved Quantity Theory

The conserved quantity theory (CQT) is an empirical analysis of causation. Hence, it identifies a process in the world that should fit the criteria mentioned in the last chapter. The idea is that causation is the exchange of a conserved quantity between the cause and the effect. For example in the (classical) Compton effect one photon hits an electron, whereby the momenta of both are changed. If the interaction is free from other influences, then the change of the momenta could only happen by the exchange of momentum between the photon and the electron.

The CQT, in slightly different versions, has already some decades of history; it was first introduced by Jerrold Aronson (1971) then reconsidered by David Fair (1979) and discussed between Phil Dowe (1992, 1995a,b) and Wesley Salmon (1994, 1998). To make a long story short, I will only discuss the latest and most elaborate version of the CQT due to Phil Dowe. It can be put succinctly into three statements:

CQ1. A *causal process* is a worldline of an object that possesses a conserved quantity.

CQ2. A *causal interaction* is an intersection of world lines that involves exchange of a conserved quantity (Dowe, 2000, 90).

[1] For an overview, see (Price and Weslake, 2010).

CQ3. There is a *causal connection* (or thread) between a fact q(a) and a fact q'(b) if and only if there is a set of causal processes and interactions between q(a) and q'(b) such that:

1. any change of object from a to b and any change of conserved quantity from q to q' occur by way of a causal interaction involving the following changes: Δq(a), Δq(b), Δq'(a), and Δq'(b); and
2. for any exchange in (1) involving more than one conserved quantity, the changes in quantities are governed by a single law of nature (Dowe, 2000, 171f.)

A few explanations are needed here. A *worldline* is the trajectory of an object in Minkowski spacetime. An *object* is a member of the set of things, which make up the fundamental ontology of physics. Only objects exist; processes and worldlines are solely means to represent the time evolution of objects and their causal interactions. A *conserved quantity* is every property that is subject to a conservation law in physics. An *intersection* is the meeting of worldlines in a Minkowski spacetime. An *exchange* is the corresponding change of the value of one conserved quantity of at least two objects. Please note that Dowe uses the notion of exchange deliberately to avoid any connotation of transfer of a conserved quantity; where transfer means that exactly the same quantum of a conserved quantity, that is lost by the cause, is acquired by the effect. For us to be using the term transfer instead of exchange, following Dowe, it would be necessary to be able to define identity conditions for amounts of physical quantities, which is not possible (Dowe, 2000, 90f).

The first definition, CQ1, is supposed to capture cases where one object is at the same time the cause and the effect. An example is an object that is moving uniformly in space, where the only cause of the motion is the object's own inertial mass. CQ2 is the definition for simple causal processes, in which cause and effect are directly linked by hitting each other. In cases where cause and effect are not directly linked, because there is another process in between, the definition of a causal connection comes into play. Two examples will help to make things clearer, one example is simple and the other one more complicated:

1. The Compton effect is the fundamental interaction between light and matter. In a (classical) description, the cause is electron a with energy q at time t_1 ($q(a)$ at t_1). The causal interaction is the intersection of the worldlines of electron a and photon b, which changes the energies of both

($\Delta q(a), \Delta q(b)$). The effect at time t_2 is photon b with a smaller energy than at time t_1 ($q(b)$ at t_2).

2. A photon hits an atom and is absorbed. As a result, the atom decays to a different atom with different charge. The cause is photon a with energy q at time t_1 ($q(a)$ at t_1). The first causal interaction is the intersection of photon a with atom b, in which a exchanges energy with b ($\Delta q(a), \Delta q(b)$). The second interaction is the decay of atom b to atom c, in which energy q and charge q' are exchanged ($\Delta q(b), \Delta q(c), \Delta q'(b), \Delta q'(c)$). Following the definition of a causal connection, the effect is atom c with charge q' at time t_2 ($q'(c)$ at t_2) (Dowe, 2000, 172 f.).

It has been criticized that the CQT is circular, because what a conserved quantity is can only be defined by invoking causation: "A conserved quantity is one that remains constant through time in a closed system, but what is a closed system but a system that does not engage in any causal interaction?" (Hitchcock, 1995, 315f). Dowe responds by arguing that for instance energy can be defined in a different way: "energy is conserved [...] on the assumption that there is no net flow into or out of the system" (Dowe, 2000, 95). However, it is unclear whether "net flow" is something other than a causal process (Schaffer, 2001, 810). Nevertheless, this critique is ineffective, since the CQT is not an analytic definition of the concept of causation. If it were, of course, conceptual circularity would be severe. However, the CQT aims at an empirical, rather than conceptual, analysis of causation. It aims to identify processes in the world that can be understood as causal, and this can be done without caring about conceptual circularity.

Unfortunately, there are more problems with Dowe's CQT, especially when trying to fit it to quantum physics. First of all, Dowe's definition of a causal process seems to be obsolete. His only example for a causal process is a case where the inertial mass of an object is the cause for its uniform motion. However, it is clear from the theory of special relativity that uniform motion in one inertial frame is rest in another one and that both frames are on a par with each other, since there is no absolute space or other sort of preferred reference frame. Additionally, CQ1 is in conflict with our intuition that causation involves at least two distinct entities, which can play the roles of cause and effect. Consequently, I do not see any reason why uniform motion should be in need of a causal explanation. Second, the notion of worldlines is highly problematic in quantum physics. The Heisenberg uncertainty principle tells us that an object cannot have a sharp position and a sharp momentum at the same time. Additionally, in general states in quantum physics are not eigenstates but superpositions, so most objects do not have a sharp value of any property. Therefore, it is impossible to define the worldline of an object. This, in turn, makes Dowe's notion of "exchange" opaque, for exchange cannot be defined as the intersection of worldlines. Furthermore, there is no worked out ontology for QFT, so we are ignorant of what the objects of this theory are. This makes it questionable that Dowe's CQT, which is defined in terms of objects, can be applied to QFT.[2]

If the CQT has no answer to these problems, it is certainly a poor theory of causation in physics. In the next chapter, I will take a look into QFT to see whether these problems can be met. It will turn out that according to the four criteria for causation from Section 2 there is causation in QFT. Furthermore, I will introduce a new CQT, adjusted to QFT, that retains from Dowe's theory only the core idea that causation is the exchange of energy.

4 Causal Processes in Quantum Field Theory

Adrian Heathcote (1989) already argued that QFT can be interpreted as describing causal interactions between fundamental objects. This seems immediately plausible; after all, while quantum mechanics can only describe the dynamics of one particle alone or in a potential and interactions only in very simplified models, QFT is the physical theory that broadens quantum mechanics to include interactions. Hence, "all causal influences are the result of forces between objects, all such forces are interactions in the sense of QFT" (Heathcote, 1989, 101f). However, neither did Heathcote bring his theory of causation into precise form, as Dowe did, nor did he show explicitly how the mathematical formalism of the QFT supports his claims. This is what I will do in the following.

I solely rely on the Lagrangian formulation of QFT (LQFT). This probably needs a few words of justification, since most philosophers of physics nowadays discuss the algebraic approach (AQFT). At this place, I can only give a short sketch of the discussion. Roughly, there are two arguments of AQFT proponents against LQFT. (1) LQFT is in a way mathematically ill defined that is in conflict with its claim to describe the fundamental physical world. (2) The rigorous mathematical formalism of AQFT is *eo ipso* superior to study the foundations of

[2] Indeed, there are more problems with the CQT, which I am not going to address here, see (Lupher, 2009) and, for problems that arise in connection with the theory of General Relativity, see (Curiel, 2000; Lam, 2005, 2010).

QFT. The first sort of argument usually criticizes that the renormalization methods in LQFT are a mathematically ill defined and *ad hoc* way to squeeze empirical predictions out of LQFT (Fraser, 2009, 2011). On the contrary, David Wallace (2006, 2011) argues that today it is well understood how divergencies arise due to the failure of LQFT on high energies respectively small distances. Therefore, renormalization methods, that cut off small distances, are “on a sound theoretical footing” (Wallace, 2011, 118). Nonetheless, it has to be admitted that this failure of LQFT could mean that we need a new physical theory for very small distances. Another less problematic justification for the use of renormalization methods is that spacetime might fundamentally be grained or quantized and therefore it makes no sense to try to investigate very small regions of spacetime (Peskin and Schroeder, 1995, 266–268, 798).

As for the second point in favour of AQFT, I do not hold it to be conclusive. The claim that mathematical rigor is important for a scientific theory has to be supported by an argument. How could this argument look like? Presumably, it is tacitly presupposed that nature follows exact mathematical laws (Halvorson, 2007). However, I do not see any other way to support this claim than by physical research. This points to a different perspective on the conflict. From the (meta-)perspective of a scientific realist certainly LQFT is the superior theory. Scientific realism is supported by the no miracles argument, that allows inferring from the success of a theory to the reality of the world that is described by it (Psillos, 1999). Since LQFT is without doubt the most successful physical theory we have, what else could philosophers of physics be realist about if not LQFT? In contrast, AQFT “makes no (non-falsified) empirical predictions whatsoever [and] there is, at present, just no reason to expect that program to succeed” (Wallace, 2011, 120). To end this excursus, I want to add that investigation in AQFT, nevertheless, is a worthwhile program and it will be interesting to see whether an ontological framework for AQFT, should any be found, contradicts the picture drawn by LQFT.

After these preliminaries, I will now go into LQFT to find out how far the mathematics can be interpreted in line with the CQT. My treatment of LQFT will be rather informal and, for the sake of brevity, I will only mention the mathematical expressions to which I explicitly refer in order to establish my claims about causation. Usually the calculation of a certain process in LQFT starts with the stipulation of a Lagrangian density. For a typical scattering process, like $e^+e^- \to \mu^+\mu^-$ in quantum electrodynamics, the Lagrangian density is given by (following Greiner and Reinhardt (1996) and Peskin and Schroeder (1995))

$$\mathcal{L}_{QED} = \mathcal{L}_{Dirac} + \mathcal{L}_{e.m.} + \mathcal{L}_{int.} = \bar{\psi}\left(i\partial\!\!\!/ - m\right)\psi - \frac{1}{4}\left(F_{\mu\nu}\right)^2 - e\bar{\psi}\gamma^\mu\psi A_\mu \,.$$

The electrons/positrons for the initial and the muons/anti-muons for the final state are specified by the field ψ ($\bar{\psi} \equiv \psi^\dagger\gamma^0$), the Dirac-matrices γ^μ, and by the mass m. They have Energy E_p, momentum $\vec{p}$, polarized spin 1/2 and electric charge $+1$ (resp. -1). The electromagnetic force is described by the field A^μ (resp. the field-strength tensor $F^{\mu\nu}$) and e, its coupling constant. $\mathcal{L}_{Dirac}$ is the typical Lagrangian for a massive spin-1/2 particle[3], $\mathcal{L}_{e.m.}$ is the Lagrangian for the electromagnetic force and $\mathcal{L}_{int.}$ specifies the coupling of the other two. From $\mathcal{L}_{QED}$ coupled equations of motion for the fields ψ, $\bar{\psi}$ and A^μ can be derived.

What needs to be calculated to obtain the probability of a scattering process, that is the probability of an evolution of a certain initial state to a certain final state, is essentially the overlap of *in* and *out* states:

$$\mathcal{P} = \left|\langle\psi_{out,1}, \psi_{out,2}|\psi_{in,1}, \psi_{in,2}\rangle\right|^2 \,.$$

In and *out* states are usually wavepackets of the form

$$|\phi\rangle = \int \frac{d^3p}{(2\pi)^3}\,\varphi\left(\vec{p}\right) e^{-i(\omega_p t - \vec{p}\cdot\vec{x})}$$

in the time limit $t \to \pm\infty$:[4]

$$\lim_{t\to-\infty} |\phi\rangle = |\psi_{in}\rangle \,,$$
$$\lim_{t\to+\infty} |\phi\rangle = |\psi_{out}\rangle \,.$$

The wavepackets have a momentum that is peaked in momentum space around the definite value $\vec{p}$. They are related via the time limit evolution by a unitary operator

[3] I use the term "particle" in a very loose way; not in the sense of the classical concept of a mass-point, but rather like physicists use it in QFT-textbooks.

[4] From the mathematical point of view, this is an ill-defined expression, because the convergence behaviour of operators in infinite dimensional Hilbert space is not trivial. For more mathematical rigor, see (Greiner and Reinhardt, 1996, ch. 9.2).

$$\psi_{out}(x) = \hat{S}^{-1}\psi_{in}(x)\hat{S},$$

$$\begin{aligned}\langle\psi_{out,1},\psi_{out,2}|\psi_{in,1},\psi_{in,2}\rangle &= \lim_{t_2\to+\infty}\lim_{t_1\to-\infty}\langle\vec{k}_1,\vec{k}_2|\hat{U}(t_2,t_1)|\vec{p}_1,\vec{p}_2\rangle \\ &= \langle\vec{k}_1,\vec{k}_2|\hat{S}|\vec{p}_1,\vec{p}_2\rangle\,,\end{aligned}$$

where S is the so called S-matrix. *In* and *out* states are taken to represent particles which are free, long before and after the interaction. Strictly speaking this is an idealization, since in nature the interaction can never be completely turned off. However, for *in* and *out* states it becomes negligibly small.

In the *canonical quantization* formulation of QFT, the fields become operator fields (sometimes this is called "second-quantization"). For the matter field this means

$$\hat{\psi}(x) = \int \frac{d^3p}{(2\pi)^{3/2}}\sqrt{\frac{m}{E_p}}\sum_s\left(\hat{a}_{ps}u\,(p,s)\,e^{-ip\cdot x} + \hat{a}^{\dagger}_{ps}v\,(p,s)\,e^{+ip\cdot x}\right),$$

where for each particle in the initial state there must be a lowering operator $\hat{a}$, and for each particle in the final state there must be a raising operator $\hat{a}^{\dagger}$. $u\,(p,s)$ and $v\,(p,s)$ are plane wave base functions with momentum p and spin s. Note that, unlike in quantum mechanics, in QFT momentum and energy are classical quantities (c-numbers, in Dirac's terminology). Letting raising and lowering operators act on the vacuum state we create particle states:

$$|p_1s_1,p_2s_2\cdots\rangle = \hat{a}^{\dagger}_{p1s1}\hat{a}^{\dagger}_{p2s2}\cdots|0\rangle\,.$$

Furthermore, the $\hat{S}$ operator has to be expanded into a perturbation series, where the interaction part of the Lagrangian is written in terms of the free asymptotic operator fields and their dynamics. In the leading order of the series, the probability amplitude $\mathcal{P}$ is then given by the invariant S-matrix element $\mathcal{M}$. Essential for obtaining the probability is the evaluation of vacuum expectation values for time ordered operator fields (i.e. n-point Green's functions):

$$G^{(n)}\,(x_1,\ldots,x_n) = \langle 0|T(\hat{\psi}\,(x_1)\cdots\hat{\psi}\,(x_n))|0\rangle\,.$$

For $e^+e^- \rightarrow \mu^+\mu^-$ this can in practice more or less just be read off from the proper Feynman diagram and turns out to be (invoking various symmetry considerations):

$$\mathcal{M} = \bar{v}(p_1 s_1)\,(-ie\gamma^\mu)\,u(p_2 s_2)\left(\frac{-ig_{\mu\nu}}{q^2}\right)\bar{u}(k_2 s_2)\,(-ie\gamma^\mu)\,v(k_1 s_1)\,.$$

Here $u(p)$ and $v(p)$ are spinors that describe the matter-fields, the terms $(-ie\gamma^\mu)$ describe the vertices (i.e. the coupling of the matter-fields) and the middle term describes virtual photons with momentum $q = p_1 + p_2 = k_1 + k_2$. From $\mathcal{M}$ the differential scattering cross-section can be calculated in a straightforward manner:

$$\frac{d\sigma}{d\Omega} \propto |\mathcal{M}|^2\ .$$

The differential cross-section is basically a distribution of energy in space and is measured as the final result of experiments—this is where theory is compared to empirical data. Of course, energy-momentum is conserved over the whole process.

The same result, i.e., the element $\mathcal{M}$ of the S-matrix and the differential cross-section, can be found via the equivalent way of the *path integral* method. Again, what needs to be calculated is the overlapping of an initial with a final state, both of which are eigenstates of the operator field

$$\hat{\psi}(\vec{x}, t)\,|\psi, t\rangle = \psi(\vec{x})\,|\psi, t\rangle\,,$$

and are related by a unitary operator:

$$\langle \psi_{out}, t|\hat{U}(t)|\psi_{in}, t\rangle\ .$$

Only this time, the result is obtained not by calculating vacuum to vacuum transitions with the help of raising and lowering operators, but by evaluating path integrals, $\int \mathcal{D}$, over ‘classical’ (i.e. not operator) fields of the form (following (Greiner and Reinhardt, 1996, 2009))

$$\langle \psi_{out}, t | \psi_{in}, t \rangle = \mathcal{N} \int \mathcal{D}\psi \, exp \left[i \int_{t_{in}}^{t_{out}} d\tau \int d^3x \mathcal{L}(\psi, \dot{\psi}) \right] .$$

The matrix element $\mathcal{M}$ is then given by

$$\mathcal{M} = -ie \int d^4x \, \bar{\psi}(x) \not{A}(x) \psi(x),$$

which leads to the exact same matrix element and differential cross section as the canonical quantization above.[5]

The reality of the in and out going particles is uncontroversial–at least for scientific realists. In contrast, the reality of the virtual or, how I rather like to call them, intermediate particles, from which the factor $\left(-ig_{\mu\nu}/q^2\right)$ in $\mathcal{M}$ stems, is far from clear. Since I will rely on intermediate particles when arguing for causation, I need to provide arguments for their reality.

In canonical as well as in path integral QFT, the Feynman propagator for photons takes on the form:

$$\langle 0 | T(\hat{A}_\mu (x) \cdots \hat{A}_\nu (y)) | 0 \rangle = D_F(x - y) = \int \frac{d^4q}{(2\pi)^4} \frac{-ie^{-iq\cdot(x-y)}}{q^2 + i\epsilon} .$$

In a naive interpretation, this might be the probability amplitude for a point-like photon to travel from point x to point y in spacetime. However, the well known arguments against a point-particle interpretation undermine this naivety. Particles, whatever they may be, do not have classical trajectories and it is questionable in how far they exist locally in spacetime at all. Additionally, there are specific arguments against the realistic interpretation of intermediate particles. They are *prima facie* suspicious, since they are not measurable 'directly'; unlike the initial and final states, they do not appear in bubble chambers or other measurement devices. It is questionable whether they are more than a mathematical part of a perturbation series. However, "if something cannot be 'directly' observed that doesn't mean we cannot have indirect evidence of its existence" (Weingard, 1982, 235). After all, intermediate particles are an indispensible part of a success-

[5] I did not mention decay events so far. Since they are treated in QFT in essentially the same way as scattering events, I assume that everything I say about scattering is true for decays as well.

ful theory, and therefore we have good evidence for their reality. Additionally, photons and other particles that play the roles of intermediate particles do exist as free states and can be 'directly' observed. To be precise, however, it has to be admitted that intermediate particles have more polarization degrees of freedom than their free counterparts. Therefore, not all kinds of intermediate particles exist as free states. Furthermore, intermediate particles do not have to be on mass shell, that is, fulfil the relativistic relation $q^2 = E^2 - m^2$. This does not mean that intermediate particles violate energy conservation. Any fluctuation in energy has to happen on very small timescales, i.e., in accordance with the energy-time uncertainty relation.

The main argument against intermediate particles comes from superpositions.[6] In general, the Feynman propagator cannot be calculated directly, but only in a perturbation series. The final amplitude then is the superposition of the parts of the series and, so the argument goes, therefore there are no discrete particles represented by the parts of the perturbation series. However, this is shared by all kinds of particles in QFT and cannot count as an argument against intermediate particles in particular. Instead, this is just another argument against the literal interpretation of single Feynman diagrams as showing real processes and trajectories. Superpositions do not show that intermediate particles in general do not exist, but that it is only the whole process that has a consistent realistic interpretations and not individual parts of it (Falkenburg, 2007, 237).

Johanna Seibt (2002, 58) challenged philosophers to decide, which QFT should be interpreted: AQFT, LQFT (canonical quantization) or LQFT (path integral formulation). With regard to AQFT, I have already stated my position. Concerning the latter two, I do not see how and why only one formulation could be singled out, because they are fully equivalent. Instead, I see only one possibility, that is, to find an interpretation compatible with both formulations (e.g. not relying on ladder operators). In the following, I highlight three characteristics of the canonical and the path integral formulation, on which I will rely later:

Initial and final states: Initial and final states are defined as not interacting at times long before and after the scattering. In this regard, these states are distinct from one another. They carry a well defined quantity of energy-momentum and have properties that are characteristic for a certain kind of entity (spin, mass, charge).

[6] See (Weingard, 1982, 1988; Teller, 1995, 137).

Local conservation laws: All conservations laws in QFT are local. Thus, either the amount of a certain property in an arbitrary small region of spacetime is constant, or there is a current going through the surface of that region.

Intermediate states: Intermediate states carry the complete energy-momentum of the initial respectively final states. Even though they cannot unambiguously be interpreted as a localized spatial process mediating the energy-momentum they exist in between the initial and final states.

With this background, let me now formulate a new version of the CQT and then discuss how it differs from Dowe's version and why it should still be called causation. The definition is not supposed to be self-contained, but has to be read in light of the foregoing explanations.

C causes E *iff* C is an initial state and energy-momentum is transferred from C by an intermediate state to a final state E.

In comparison to Dowe's CQT, I want to emphasize first of all that initial and final states are not the equivalent to Dowe's objects. For Dowe, talking about objects involves having an ontology, also Dowe's objects have well defined trajectories in Minkowski diagrams. In contrast to this, initial and final states are just what is defined by physics; they are underdetermined concerning ontology. For now, it is not clear whether they are more like classical particles or more like fields or something else and surely they do not have well defined trajectories. This is unproblematic when trying to identify a causal relation. The four criteria in Section 2 are applicable without any information about what kinds of objects we are dealing with. Their formulation is independent of whether cause and effect are point-particles or fields or something else. Therefore, it is possible to identify causal relations in QFT without knowing what kinds of entities initial and final states are.

Second, take the process $e^+e^- \to \mu^+\mu^-$. While Dowe distinguishes between the electron and the positron, or the muon and anti-muon, respectively, and would take either only one as the cause, the effect, or both as separate causes, or effects (see examples one and two in Section 3), I take the electron and positron together as the cause, and the muon and anti-muon together as the effect. The deficiency of Dowe's account can be seen by way of an example. Two electrons e_1 and e_2 with momenta a and b scatter and the momenta change to c and d. Whether after the scattering e_1 has c and e_2 d or *vice versa* cannot be answered in QFT and both possibilities have to be taken into account in the calculation of the scattering amplitude (Greiner and Reinhardt, 2009, ch. 3.3). Therefore, it is not possible to divide the final state in two separate effects and I do not see how and why it should be otherwise for the cause.

Third, I make use of local conservation laws in so far as the causal processes in QFT are continuous processes, though not localized in spacetime. Global conservation of energy-momentum would allow a quantity of energy to cease to exist at one point in space and at the same time come to existence at another distant point in space. Local conservation laws rule out such events, because regions of spacetime can be made arbitrary small and either the amount of a conserved quantity is constant in that region or there is a current into, or out of, that region. This justifies the view that causation is the *transfer* of energy-momentum and not just the *exchange*. I do not, to be clear, defend the position that causal processes in QFT are localized in the sense that everything is moving on thin lines, pictured in Feynman diagrams.

Fourth, it is enough to understand causation only as the transfer of energy-momentum and not of every conserved quantity there is. Energy-momentum is always relevant, it is transferred in every process in QFT and it is the property that is controlled and measured in experiments (most of the time this will happen together with position measurements, in order to measure the distribution of energy in space).

Finally, it is time to ask whether the relation of energy-momentum transfer satisfies the criteria of Section 2, whether it actually is a causal relation. As for the first criterion, the relation is intrinsic, since scattering events, if they are sufficiently isolated in experiments, do not depend on anything else that is going on in the universe. The relata, initial and final states, are distinct by definition, long before and after the scattering, and energy-momentum transfer shows what the relation between the relata exactly is. Moreover, it is almost trivially true that the effect can be manipulated by manipulating the cause; a change in the momentum of the initial state will always change the momentum of the final state. Even though, in quantum physics initial states only produce a final state with a certain probability (the matrix element $\mathcal{M}$), to take the previous example, the probability to observe a muon anti-muon pair in a collider is certainly higher when electrons and positrons collide as compared to when they do not. These probabilities are objective in a sense that they can be reproduced in experiments. Therefore, it is reasonable to think of the relations that are described by QFT as stable relations.

I conclude that QFT can be understood as describing causal processes, i.e., as the transfer of energy-momentum. In the next two sections, I will put causation to work when applying it to the problems of localization and entanglement.

5 The Problem of Localization

States in QFT are usually defined as wavepackets. In experiments, they are focused and localized fairly well by the use of collimators and interactions happen on timescales that are small enough for them to keep their shape. However, no matter how well the collimators work, the wavepacket will not have the form of a delta function, exactly peaked in spacetime and momentum space, but it will have the form of a Gaussian. Since every Gaussian is non-zero everywhere in space, there is a non-vanishing probability for a positive result when measuring an initial or final state everywhere in space. Does it follow that, whatever states are, every state exists everywhere in the universe, whether we measure it or not? This is (one formulation) of the problem of localization in quantum physics.

A promising, but flawed, way to avoid this seemingly counterintuitive conclusion are so called Newton-Wigner states (Newton and Wigner, 1949). These states are exactly localized in coordinate space and have a position operator $\hat{P_x}$ that gives an expectation value of one for a state $|\psi_x\rangle$ localized at spacetime point x , i.e., $\langle\psi_x|\hat{P_x}|\psi_x\rangle = 1$, and an expectation value equal to zero for any point $x' \neq x$, i.e., $\langle\psi_{x'}|\hat{P}_x|\psi_{x'}\rangle = 0$. David Malament (1996) has proven that Newton-Wigner states are inconsistent with special relativity. In the interest of brevity, I will not repeat the proof here but only point out where exactly the contradiction arises. One consequence of special relativity is microcausality. An event at some point x in spacetime is independent of any event at point y that is space-like separated from x. In quantum physics, this is expressed by an equal time commutator relation. If two observables commute

$$\left[\hat{P}(x), \hat{P}(y)\right] = 0 \quad \text{for space-like separated } x,\ y$$

then the outcomes of measurements of these observables are independent of one another.[7] This condition cannot be satisfied by Newton-Wigner states. If the probability for a measurement outcome at x is one, then the probability in a space-like separated point y must be zero. In other words, the statistics of both mea-

[7]This condition seems to be violated by EPR-correlations, but as far as I can see this does not substantially weaken Malament's theorem.

surements are not independent of one another.[8] This dependence could only be established by a signal with superluminal velocity between x and y. This is impossible according to special relativity. Therefore the conclusion of Malament's theorem is that Newton-Wigner states and special relativity are only compatible if the probability for a positive measurement outcome for any observable is zero everywhere in space. This is clearly unacceptable and therefore Newton-Wigner states have to be abandoned.

In general, there is no position operator in QFT. However, since "[a]ll quantum field theories […...] model localization by making observables dependent on position in spacetime" (Halvorson and Clifton, 2002, 18), any observable $\hat{P}(x)$, defined at spacetime point x, is sufficient to refer to an entity localized at x. Here again the problem of localization arises, since this attitude, taken by itself, means that all states in QFT exist at every spacetime point in the universe.

Several authors point to experiments in particle physics, which seemingly show the measurement of small particles, and come to the conclusion that "there remains a particle 'grin'" (Redhead, 1982, 89) "which cannot be dismissed" (French and Krause, 2006, 136).[9] Electrons and photons show up as dots on scintillation screens or photographic plates, α-particles leave tracks in bubble chambers, and so on. It is argued that if the fundamental entities are always measured as localized events, then clearly these entities have to be small particles; this seems to be self-evident and in no need of further explanation. However, on a closer look, this argument turns out to be fallacious. What we are looking for in an ontology are entities that are the cause of observable phenomena. Even if we infer from the observable phenomena that there is something that causes them, we cannot automatically infer that the cause in some ways resembles the phenomena. If we want to find out more about the causes, it is the theory that we have to consult. Halvorson and Clifton (2002, 22) put this argument into a different form when they assert: "In particular, we do not observe particles; rather, there are 'observation events' ." Even though the things that we observe (tracks in bubble chambers etc.) are localized, we do not observe particles but only the effect of a physical interaction between a fundamental entity and the measurement apparatus. Not our observations tell us how the fundamental entities are like, but theory. It might look as though we observe localized particles, however, "[t]hese experiences are illusory!" (Halvorson and Clifton, 2002, 20).[10]

[8] See (Redhead, 1982, 73 f; Saunders, 1994, 89).
[9] See (Bartels, 1999, 182; Kuhlmann, 2010, 83; Teller, 1995, 30).
[10] See (Falkenburg, 2007) for a more thorough discussion of the relevance of experiments for ontology.

In this impasse, David Wallace (2001, 2006) introduced a new conception of how quantum mechanical states could be described as localized in space, which he calls effective localization. It, and the accompanying principle of effective localization (ELP), can be defined as follows (Wallace, 2001, 10):

1. **Effective localisation (qualitative form):** A state $|\psi\rangle$ is *effectively localised* in a spatial region Σ_i iff for any function $\hat{f}$ of field operators $\hat{\phi}, \hat{\pi}$, $\langle\psi|\hat{f}|\psi\rangle - \langle\Omega|\hat{f}|\Omega\rangle$ is negligibly small when $\hat{f}$ is evaluated for field operators outside Σ_i, compared to its values when evaluated for field operators within Σ_i.
2. **The effective localisation principle (ELP) (qualitative form):** A subspace $\mathcal{H}$ of the QFT Hilbert space $\mathcal{H}_\Sigma$ obeys the ELP on scale L iff for any spatial region $\mathcal{S}$ large compared with L, a superposition of states effectively localised in $\mathcal{S}$ is effectivey localised in effectively the same region.

Essentially this means that a state $|\psi\rangle$, in order to be localized in spatial region Σ_i, must have expectation values for a set of operators $\hat{f}_i$ considerably bigger than the vacuum state $|\Omega\rangle$ has for $\hat{f}_i$.

Even though I agree with Wallace, I see two problems. (1) Defining localization in such a way seems to be an *ad hoc* move, only motivated by rescuing some form of localization. Any setting of L is just arbitrary. (2) Wallace is aware that effective localization is only an approximation, but an ontology that tells only what things there are approximately is unacceptable. Is there any way to justify effective localization independently and improve the approximation? I think there is – when effective localization is put into the context of causation.

To see this, two points have to be made explicit. In discussions about scientific realism it is often stressed that we can only reasonably be realistic about unobservable entities that cause observable phenomena.[11] In addition to that, if QFT is understood as describing causal relations between states, the states themselves should be regarded just as what is causally relevant for the process. In this light, the question then is: What is causally relevant for the processes in QFT? Is it a part of the state $|\psi\rangle$ in spatial regions Σ'_i where no experiment, possible today, could find any difference between $|\psi\rangle$ and the vacuum state $|\Omega\rangle$? I do not think so. From the viewpoint of scientific realism and causation we have no reason to believe in something that has no observable effects for our experiments. Therefore,

[11] See (Psillos, 2006) and the contribution by Matthias Egg in this volume.

effective localization is neither *ad hoc* nor an approximation, but a description of entities in whose existence we can have justified belief.

6 Entanglement

In this section, I will explore what follows for the interpretation of entangled states, if causation in QFT and effective localization are taken for granted.

There are two problems involved in entanglement: (1) How can space-like separated events be correlated without having a common cause? (2) How can a superposition evolve into definite states upon measurement? The latter is of course the measurement problem—which I will not treat here. The former question, however, can be investigated to some extent without invoking the measurement problem.

John Bell has shown that EPR-correlations cannot be explained by a common cause, which determines the probabilities of the measurement outcomes of the entangled state. The question as to how space-like correlations are possible, then, can be transferred to whether there is a causal connection or relation of some other sort between entangled particles. A paradigm example for entanglement is the singlet state of two electrons:

$$|\psi\rangle_{\text{singlet}} = \frac{1}{\sqrt{2}}\left(|{\downarrow}\rangle_a\,|{\uparrow}\rangle_b - |{\uparrow}\rangle_a\,|{\downarrow}\rangle_b\right),$$

where $|{\downarrow}\rangle_a$ means that electron a has spin state down etc.

Tim Maudlin (2002, 2007) argues that EPR-correlations can be understood as a causal relation between entangled objects. Of course, he is aware that "[c]orrelation does not imply causation" (Maudlin, 2002, 127). Nonetheless, he believes it to be justified to analyse EPR-correlations in terms of a counterfactual theory of causation: "The *local physical events* A and B are causally implicated with one another if B would not have occurred had A not (and vice versa)" (Maudlin, 2002, 128). Where the notion of "is causally implicated with" is considerably weaker than the usual notion of "is caused by." Here is Maudlin (2002, 128) again: "We do not suppose that it follows from the fact that A is causally implicated with B that A caused B or B caused A." But this makes the relation of causal implication too weak to be still called causation. In fact, I see no difference between causal implication and correlation. No doubt, if we know that B would not have occurred had A not and vice versa, then A and

B are correlated. However, if A is correlated with B means "A↔B," then we can infer from "A is correlated with B" that if B does not occur so does not A, in other words, A had not occurred if B had not (and vice versa).[12] If I understand Maudlin correctly, then causal implication means *nothing more* than a counterfactual, but this makes causal implication and correlation equivalent, since either one follows from the other.

To be fair, this is not the complete position of Maudlin. He is also realistic about natural laws and argues that laws tell us which counterfactuals and causal implications are true:

> Since it is facts about the laws that help us identify the cause […] and since laws are obviously deeply implicated in the evaluation of counterfactuals, I suggest that we stop trying to analyze causation directly in terms of counterfactuals and consider anew how laws play a role in determining causes. (Maudlin, 2007, 148)

Setting aside the problem whether physical theories give a set of laws, Maudlin is silent about which law it is that shows that EPR-correlations are more than mere correlations. As far as I can see, the physical description of EPR-experiments only tells us that if we measure the spin on electron a, then a subsequent measurement of electron b's spin will yield the opposite result, and vice versa—this is correlation and nothing more.

Maybe there are other reasons for holding entanglement and EPR-correlation to be a causal relation. Let us apply the criteria for causation from Section 2. It is clearly true that EPR-correlations are stable relations and if A is correlated with B, then A trivially increases the chance of B. Thus, two criteria out of four are already fulfilled. What about the other two? Can we manipulate one entangled electron by manipulating the other? Before the measurement, from a theoretical as well as an experimental standpoint, this is impossible. First of all, the entangled state $|\psi\rangle_{\text{singlet}}$ cannot be split up into two separate parts that are then manipulated separately. In addition to that, every experimental manipulation on the entangled property (e.g., spin) would destroy the entanglement and therefore cannot count as manipulation. However, maybe measurement is nothing else than a manipulation; after all, the alignment of the magnetic field when measuring the spin of electron a alters the outcome of a measurement of electron b's spin. Fair enough, but now we entered the realm of the measurement problem and it is, to

[12] Arguably, the last step is not purely logical, because there are problems with counterfactuals in formal logic. Nevertheless, in this context of well controlled and isolated experiments it is feasible.

say the least, unclear what happens in a measurement. While decoherence theories may describe measurements as something that alters the measured states, the GRW-flash theory probably does not. In short, before the measurement happens we can certainly not manipulate entangled states and what happens due to the measurement is completely unclear.

The same result is obtained when considering the first criterion, causation as intrinsic relation between distinct entities. Intrinsicness is probably fulfilled, but before the measurement there is no reason to believe in distinct entities. Entangled states, unlike product states, are governed by only one Hamiltonian and therefore there are not two or more things that evolve independently. Also, no physical relation between two parts of an entangled state is measured or appears in the theoretical description. This critique generalizes from causal relations to any relation. It is unclear how entangled states could be described as a relation between distinct entities without any account of how the entities can be defined or what the relation physically is. Again this reasoning is only tenable as long as measurements are neglected.

Is there any alternative to the description of entanglement as a relation? Taking the previous chapter into account, I think there is. If an entangled state (e.g., $|\psi\rangle_{singlet}$) does not consist of two distinct entities, then it can be either cause or effect, but not describe cause and effect at the same time. Furthermore, the discussion of localization has shown that there are in general no point particles, but only entities that exist in spatially extended regions. Applied to entangled states, Wallace's effective localization means that one entangled state is just one entity that is extended in space. Of course, this means that there are entities extended over 20 km and more, but since pointlike entities are no option anyway, this is just another new thing that quantum physics tells us about the world.

To explicate this alternative picture to relationalism, an entangled state should be regarded as one relatum of a causal interaction with a measurement device that has as effect two separated states. Before a measurement happens, the entangled state is just one entity; the measurement destroys the entanglement and leaves two or more distinct entities. This is essentially a different picture (cf. Maudlin) than that of a measurement device that acts on one part of an entangled state, which then causes something in the other part of the entangled state. Though again, this point has no sufficient justification until the measurement problem is solved.

To sum up, an entangled state does not consist of two distinct entities that are measured at space-like distances, but only of a single extended entity that most often is detected in one place (which detection supposedly disrupts the entanglement). However, this does not help much to understand EPR-correlations; it just shifts all the burden over to the measurement problem. What I hope to have shown is that the focus of investigation needs to be transferrred, or else the whole problem has to be reformulated. The question is not how one entangled entity can have causal influence on the other one, but what happens at the measurement. Any discussion whether EPR-correlations are compatible with special relativity misses the point, because there is no reason to believe that some sort of causal or other relation is involved in the formation of entangled states. Rather, the measurement problem needs to be solved, no less and no more, to make entanglement and EPR-correlations less puzzling.

Acknowledgment

I want to thank Michael Esfeld, Tilman Sauer and Adrian Wüthrich for organising the conference and letting me present my unfinished thoughts, as well as the audience and an anonymous reviewer for helpful comments.

References

Aronson, J. L. (1971). On the grammar of 'cause'. *Synthese*, 22:414–430.

Bartels, A. (1999). Objects or events? Towards an ontology for Quantum Field Theory. *Philosophy of Science, Proceedings of the 1998 Biennial Meetings of the Philosophy of Science Association. Part I: Contributed Papers*, 66, Supplement:S170–S184.

Bontly, T. D. (2006). What is an empirical analysis of causation? *Synthese*, 151(2):177–200.

Cartwright, N. (1989). *Nature's Capacities and Their Measurement*. Clarendon Press, Oxford.

Curiel, E. (2000). The constraints General Relativity places on physicalist account of causality. *Theoria*, 15(1):33–58.

Dowe, P. (1992). Wesley Salmon's process theory of causality and the conserved quantity theory. *Philosophy of Science*, 59(2):195–216.

Dowe, P. (1995a). Causality and conserved quantities: A reply to Salmon. *Philosophy of Science*, 62(2):321–333.

Dowe, P. (1995b). What's right and what's wrong with transference theories. *Erkenntnis*, 42:363–374.

Dowe, P. (2000). *Physical Causation*. Cambridge University Press, Cambridge, 1. edition.

Fair, D. (1979). Causation and the flow of energy. *Erkenntnis*, 14:219–250.

Falkenburg, B. (2007). *Particle Metaphysics: A critical account of subatomic reality*. Springer, Berlin.

Fraser, D. (2009). Quantum field theory: Underdetermination, inconsistency, and idealization. *Philosophy of Science*, 76(4):536–567.

Fraser, D. (2011). How to take particle physics seriously: A further defence of axiomatic quantum field theory. *Studies In History and Philosophy of Science Part B: Studies In History and Philosophy of Modern Physics*, 42(2):126–135.

French, S. and Krause, D. (2006). *Identity in Physics: A Historical, Philosophical, and Formal Analysis*. Clarendon Press, Oxford.

Greiner, W. and Reinhardt, J. (1996). *Field Quantization*. Springer, Berlin.

Greiner, W. and Reinhardt, J. (2009). *Quantum Electrodynamics*. Springer, Berlin.

Halvorson, H. (2007). Algebraic quantum field theory. In Butterfield, J. and Earman, J., editors, *Handbook of the Philosophy of Science: Philosophy of Physics, Part A*, pages 731–864. Elsevier, Boston.

Halvorson, H. and Clifton, R. (2002). No place for particles in relativistic quantum theories? *Philosophy of Science*, 69(1):1–28.

Heathcote, A. (1989). A theory of causality: Causality = interaction (as defined by a suitable quantum field theory). *Erkenntnis*, 31:77–108.

Hitchcock, C. R. (1995). Salmon on explanatory relevance. *Philosophy of Science*, 62(2):304–320.

Kitcher, P. and Salmon, W. C., editors (1989). *Scientific Explanation*. University of Minnesota Press, Minneapolis.

Kuhlmann, M. (2010). *The Ultimate Constituents of the Material World. In Search of an Ontology for Fundamental Physics*. Ontos Verlag, Heusenstamm.

Lam, V. (2005). Causation and space-time. *History and Philosophy of the Life Sciences*, 27:465–478.

Lam, V. (2010). Metaphysics of causation and physics of General Relativity. *Humana Mente: Physics and Metaphysics*, 14:61–80.

Lupher, T. (2009). A physical critique of physical causation. *Synthese*, 167(1):67–80.

Malament, D. B. (1996). In defense of dogma: Why there cannot be a relativistic quantum mechanics of (localizable) particles. In Clifton, R., editor, *Perspectives on Quantum Reality : Non-relativistic, Relativistic, and Field-theoretic*, pages 1–10. Kluwer Academic Publishers, Dordrecht.

Maudlin, T. (2002). *Quantum Non-Locality and Relativity: Metaphysical Intimations of Modern Physics*. Blackwell Publishers, Malden (Massachusetts), second edition.

Maudlin, T. (2007). *The Metaphysics within Physics*. Oxford University Press, Oxford.

Mellor, D. H. (1988). On raising the chances of effects: Essays in honor of Wesley C. Salmon. In Fetzer, J. A., editor, *Probability and Causality*, Synthese Library, pages 229–240. D. Reidel, Dodrecht.

Menzies, P. (1996). Probabilistic causation and the pre-emption problem. *Mind*, 105(417):85–117.

Newton, T. D. and Wigner, E. P. (1949). Localized states of elemntary systems. *Reviews of Modern Physics*, 21(3):400–406.

Peskin, M. E. and Schroeder, D. V. (1995). *An Introduction to Quantum Field Theory*. Perseus Books, Reading (Massachusetts).

Price, H. and Weslake, B. (2010). The time-asymmetry of causation. In Beebee, H., Hitchcock, C., and Menzies, P., editors, *The Oxford Handbook of Causation*, pages 414–443. Oxford University Press, Oxford.

Psillos, S. (1999). *Scientific Realism: How Science tracks Truth*. Routledge, London.

Psillos, S. (2006). Ramsey's Ramsey-sentences. In Galavotti, M. C., editor, *Cambridge and Vienna: Frank P. Ramsey and the Vienna Circle*. Springer, Berlin.

Redhead, M. L. G. (1982). Quantum field theory for philosophers. In *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, volume Volume Two: Symposia and Invited Papers, pages 57–99.

Russell, B. (1912-1913). On the notion of cause. *Proceedings of the Aristotelian Society*, 13:1–26.

Salmon, W. C. (1994). *Scientific Explanation and the Causal Structure of the World*. Princeton University Press, Princeton.

Salmon, W. C. (1998). *Causality and Explanation*. Oxford University Press, Oxford.

Saunders, S. (1994). A dissolution of the problem of locality. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 1994:88 – 98.

Schaffer, J. (2001). Review: Physical causation. *British Journal for the Phillosophy of Science*, 52:809–813.

Seibt, J. (2002). 'Quanta,' tropes, or processes: Ontologies for QFT beyond the myth of substance. In Kuhlmann, M., Lyre, H., and Wayne, A., editors, *Ontological Aspects of Quantum Field Theory*, pages 53–93. World Scientific, New Jersey.

Teller, P. (1995). *An Interpretive Introduction to Quantum Field Theory*. Princeton University Press, Princeton, New Jersey.

Wallace, D. M. W. (2001). Emergence of particles from bosonic quantum field theory. http://arxiv.org/abs/quant-ph/0112149v1.

Wallace, D. M. W. (2006). In defence of naiveté: The conceptual status of Lagrangian quantum field theory. *Synthese*, 151(1):33 – 80.

Wallace, D. M. W. (2011). Taking particle physics seriously: A critique of the algebraic approach to quantum field theory. *Studies In History and Philosophy of Science Part B: Studies In History and Philosophy of Modern Physics*, 42(2):116–125.

Weingard, R. (1982). Do virtual particles exist? *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 1982:235–242.

Weingard, R. (1988). Virtual particles and the interpretation of quantum field theory. In Brown, H. R. and Harré, R., editors, *Philosophical Foundations of Quantum Field Theory*, pages 43–58. Clarendon Press, Oxford.

On Free Will and No-Conspiracy

Iñaki San Pedro

Abstract. *In this paper, I challenge the widespread view that* Measurement Independence *adequately represents the requirement that EPR experimenters have free will.* Measurement Independence *is most commonly taken as a necessary condition for free will. A number of implicit assumptions can be identified in this regard, all of which can be challenged on their own grounds. As a result, I conclude that* Measurement Independence-*type conditions are not to be justified by appealing to the preservation of the EPR experimenters' free will.*

1 Introduction

This paper is concerned with a particular aspect of the usual derivations of the Bell inequalities. It is the idea that the inequalities follow partly from the requirement that the EPR experimenters are able to and do make free choices at the moment of setting up the EPR measurement apparatus. In other words, this is the idea that free will is a necessary requirement to be implemented in any physically possible hidden variable theory and therefore necessary for the derivation of Bell's theorem. I do not take this to be a controversial claim. It is less clear however how this requirement for free will is to be actually implemented.

Typical derivations of the Bell inequalities presuppose a common cause—as a hidden variable—onto which several constraints and restrictions are set. Constraints on the postulated common causes are intended to reflect standard requirements of a generic physical system, including temporal order of causal relations or locality considerations. As a result, some version of Bell's *factorizability*—and therefore of a Bell-type inequality—is derived. The strength of such arguments relies on the plausibility of the conditions imposed on the common causes. There is, for instance, an extensive literature regarding the idea of locality, particularly concerning the intuitions leading to the concept of physical locality, the character-

T. Sauer, A. Wüthrich (eds.), *New Vistas on Old Problems. Recent Approaches to the Foundations of Quantum Mechanics*, Edition Open Access, 2013.

isation of the concept itself, its implications and whether it may be appropriately captured and characterised in terms of probabilistic relations.

Less attention has been paid to the requirement that the EPR experimenters do take free independent decisions at the moment of setting up the EPR apparatus for measurement. Roughly, this idea of the EPR experimenters being able to act freely when setting up measurement apparatus is usually taken to entail that the events representing their decisions, and the foregoing corresponding free acts, be causally independent of the hidden variables. This is usually expressed by means of the so-called *No-conspiracy* condition—I shall later refer to this condition, more neutrally, as *Measurement Independence*—, a probabilistic expression which is taken to be necessary for free will.

The aim of this paper is to reassess and ultimately challenge this particular claim. I shall suggest that the fact that the EPR experimenters have free will does not provide a justification for the requirement of *No-conspiracy*. This is not to say that it cannot be justified otherwise. But I shall not pay attention to such issues here since, as pointed out, this paper concerns exclusively the very specific claim connecting free will and *No-conspiracy*.

The paper is divided in two parts. First, I shall motivate that it is indeed commonplace, in the usual arguments for the derivation of the Bell inequalities, to think of free will as being behind the justification of the requirement of *No-conspiracy*. This is done in Sections 2 and 3, where the logical structure of the actual claim I shall later challenge is also made precise. The second part of the paper looks at the various presuppositions involved when invoking *Measurement Independence* as a requirement of free will. In Section 4, I comment on the more general presuppositions so as to be able to later identify more specific (causal) assumptions. These are discussed in detail in Sections 5, 6, and 7, respectively. The paper closes with some brief remarks on some of the implications of the discussion on the previous sections.

2 Free Will in Bell's Theorem

The requirement for free will in itself does not seem to spark off any controversies. In particular, it seems desirable that any theory we propose that aims at a description of nature and that may include or refer to our (human) interaction with it, is to be consistent with the idea of free will; unless, of course, we discard the possibility of free agents from the very start. A more interesting issue concerns the need to represent appropriately the idea of free will within the theory, be it as a piece of mathematical formalism, as some set of background assumptions or presuppositions, etc.

In the context of the derivation of the Bell inequalities the requirement of free will is usually represented by means of a probabilistic expression demanding that the postulated hidden variable must not influence the probabilities of the actual settings of the EPR measurement apparatus. This is the so-called *No-conspiracy* condition:

$$p(m_i|C) = p(m_i), \tag{1}$$

where C stands for the postulated (hidden) common cause and m_i for any of the different possible measurement settings in (both wings of) an EPR experiment.

Since in the following sections, I shall argue against this kind of justification of *No-conspiracy*-type conditions —that is, against the view that *No-conspiracy*-type conditions are reasonable, and indeed necessary, conditions to be required in the derivation of the Bell inequalities *because* they reflect the fact that free will is preserved—, I shall first show that, as matter of fact, these ideas are quite widely endorsed by philosophers and physicists alike, including Bell himself. Let us start precisely with Bell's own reflections on the issue:

> [I]t may be that it is not permissible to regard the experimental settings a and b in the analyzers as independent variables, as we did. We supposed them in particular to be independent of the supplementary variables λ, in that a and b could be changed without changing the probability distribution $p(\lambda)$. Now even if we have arranged that a and b are generated by apparently random radioactive devices, housed in separate boxes and thickly shielded, or by Swiss national lottery machines, or by elaborate computer programmes, or by apparently free willed experimental physicists, or by some combination of all of these, we cannot be *sure* that a and b are not significantly influenced by the same factors λ that influence A and B. But this way of arranging quantum mechanical correlations would be even more mind boggling than one in which causal chains go faster than light. Apparently separate parts of the world would be deeply and conspiratorially entangled, and our apparent free will would be entangled with them. (Bell, 1981, C2 57)

Reading the quotation above, one might not be completely convinced that Bell's thoughts as regards probabilistic independence assumptions such as *No-conspiracy* are just thoughts about free will. Indeed, one may note that in the quotation free will is only one among other mechanisms behind the requirement that the experimental settings are regarded as independent variables. So, perhaps, one might argue Bell did not suggest free will was an essential part of the picture, after all. If the EPR measurement apparatus is set exclusively by some computer routine involving random numbers, for instance, with no human action involved at all (not even to run the routine), the argument would go, there would be no reason to appeal to free will.

Despite the reference to mechanisms of this kind however, i.e. random radioactive devices, lottery boxes, etc., it seems clear to me that it was Bell's conviction that the justification of *No-conspiracy*-type conditions by means of random number generators, or other non-human resources, still involves an assumption (hidden or implicit, perhaps) about free will. This is indeed what the final part of the quotation above seems to endorse. Bell concludes there that the deep conspiratorial entanglement in the world as a consequence of the influence of the hidden variable on the measurement settings in turn involves an entanglement as regards our (apparent) free will.

For Bell thus, world conspiracies and the lack of free will seem to go hand in hand. This has also been stressed by several other authors. An example is Huw Price, who provides the following analysis on Bell's thoughts with respect to these issues:

> Bell's Theorem requires the assumption that the properties of a quantum system are independent of the nature of any measurements that might be made on that system in the future—"hidden variables are independent of later measurement settings," to put it in the jargon.
>
> Bell saw that in principle quantum mechanics could be both realist […] and local […], by giving up this *independence assumption*. But he found this solution even less attractive than that of challenging special relativity, for he took it to entail that there could be no free will. (Price, 1996, 231)

Also, in a more recent treatment of the problem, there is a clear sense in which both philosophers and physicists endorse the idea that free will and the kind of independence required by Bell are tightly connected. For instance, Conway and

Kochen's so-called "Free Will Theorem" revolves around the idea that free will is behind such independence of measurement settings, and ultimately behind the fact that there are not world conspiracies of the type described above.[1] As Tumulka (2007) points out in commenting on (Conway and Kochen, 2006):

> […] we should require a physical theory to be non-conspirational, which means here that it can cope with arbitrary choices of the experimenters, as if they had free will (no matter whether or not there exists "genuine" free will). (Tumulka, 2007, 194)

In sum, the claims about probabilistic independence regarding the setting of the EPR measurement apparatus found in the usual arguments for the derivation of the Bell inequalities are made in virtue of us (or perhaps nature, more generally) being capable to act under our (or its) freedom of will.

3 No-Conspiracy and Free Will

I will argue in the following sections that the idea of free will involves, at different levels, a number of causal presuppositions, which I will try to make precise. Causation will then be a central notion in the discussion to follow so it seems convenient to specify further some of the ideas in the previous section in terms of causal notions.

It is not new at all to think of the issues discussed above causally. Van Fraassen (1982) constituted a turning point in this respect, in that he suggested for the first time that the notion of "hidden variable" that appears in Bell's work plays the role of a "cause"—more particularly a "common cause." There is thus in (van Fraassen, 1982) an explicit identification of Bell's "hidden variables" with the notion of "common cause." Therefore the derivation in this context of the Bell inequalities follows by appealing to causal statements. Interestingly enough, van Fraassen (1982) also assumes in his derivation of the Bell inequalities a condition which is equivalent to the independence assumption suggested by Bell himself, which we saw in the previous section. This is the so-called *Hidden Autonomy*. But van Fraassen's *Hidden Autonomy* is different from Bell's original assumption in two respects. First, as pointed out, *Hidden*

[1] It is worth pointing out that the idea of free will in (Conway and Kochen, 2006) does not refer exclusively to humans but is extensible to every particle that could be involved in an EPR experiment, i.e. electrons, photons, etc.

Autonomy is the result of a causal assessment of the EPR scenario and therefore has an explicit causal reading. Second, van Fraassen (1982) does not make any clear reference to the notion of free will, nor to conspiracies, as a motivation of *Hidden Autonomy*. Van Fraassen's justification of *Hidden Autonomy* points rather to the idea that the condition needs to be assumed in order to make sure that the EPR correlations are caused exclusively by the postulated common cause (van Fraassen, 1982, 32).

Despite the fact that van Fraassen (1982) does not make any clear reference to the idea of free will, his *Hidden Autonomy* is, as pointed out, an expression which is equivalent to Bell's independence requirements, or seamlessly to *No-conspiracy*. And, as I have argued above, even if the idea of *No-conspiracy* can be spelled out making no explicit reference at all to (human) experimenters taking free decisions, there is a clear sense in which the notion of free will seems to be behind it. I suggested, more precisely, that requiring *No-conspiracy* is usually justified by appealing, even if not always explicitly, to the notion of free will.

This is the very claim I will be challenging in the remainder of the paper. Before proceeding however, a terminological but in my view important point needs to be made. It has to do with the actual expression used to refer to probabilistic independence conditions such as equation (1), i. e. the expression "no-conspiracy." By making use of this terminology we seem to be tacitly endorsing the view, once more, that violations of such probabilistic independence conditions do indeed entail in some sense or another a conspiracy on the part of nature. Since my aim is to show that this is not so, i. e., that there being free will needs not be expressed by means of a probabilistic independence assumption, I shall refer from now on to expression (1) just as *Measurement Independence*. This is definitively a less prejudiced and more neutral way to refer to such probabilistic independence conditions, the violations of which need not, I will argue, involve any sort of world conspiracy.

To be more precise, what I shall challenge is what I take to be the general agreement that *Measurement Independence* is necessary for *free will*, i.e.

$$\text{Free will} \rightarrow p(m_i|C) = p(m_i). \quad (2)$$

4 What the Idea of Free Will Presupposes

There are a number of presuppositions behind the claim that *Measurement Independence* is necessary for free will but most of them are hardly made explicit in the usual derivations of the Bell inequalities. They can be divided in two classes. We find on the one hand a number of general assumptions, usually in relation to the connection between the experimenters' free decisions and the corresponding actual free acts. In particular, if free will is to be at the origin of the EPR experimenters' decisions to act, it seems a reasonable assumption that there be a robust (one to one) correspondence between the willing of an experimenter to act so and so and the actual act she later commits.[2] Moreover, a "faithful correspondence" of this sort seems to be necessary if we are to make sense at all of free acts—or acts of free will, understood as actual physical events taking place in space and time—and not just free decisions.

On the other hand, there seems to be a general agreement that the notion of free will has some causal import, and that it can therefore be expressed to some extent by means of causal terms. We don't need to review the different proposals to characterise human free will in detail, or the role that causality plays in them. This would take us into a deep metaphysical discussion, away from the purpose of this work. It will be enough, for the sake of the argument, to assume that causation plays in fact a central role when it comes to a description or characterisation of acts of free will. It seems intuitively right to say, for instance, that human free acts are actually free in that they are not *causally* determined or simply influenced by other events we might not even be aware of. Also, it is from a causal perspective that we seem warranted to make claims such as that free will guarantees humans to be able to decide and act freely, or to be able to act differently under changing circumstances, i.e. to revise our decisions to act after reassessment of a situation. These general considerations are, to my mind, rather uncontroversial. It is more intricate, though, how to make more precise and sharp more specific causal assumptions which are behind those considerations.

I would like to pay attention here to three specific presuppositions, all related to some sort of causal view or picture, that the idea of free will, as characterised above, demands. First, if *Measurement Independence*, i.e. expression (1), is to represent some causal statement at all, we need to assume that there is indeed a (faithful) correspondence between causal statements of the interesting kind to us and probabilistic relations.[3] I will refer to such an assumption as the *Cause-*

[2]Whether the (one to one) relation between human free decisions and corresponding free acts needs to be of a causal nature or not is not completely clear. Issues concerning the specific form of such relations will not play any role in the argument here and will therefore be put aside.

[3]There is the issue as to how a proper probabilistic theory of free will would actually look like. We shall not pay attention to such issues here, but it goes without saying that this is a deep and

statistics Link. Second, the specific independence pattern expressed by equation (1) seems to make sense only if a particular event time order as well as a fixed causal order are assumed. This can be made explicit by what I will call the *Time Order* presupposition. Finally, equation (1) is the result of demanding, not only the lack of *some* causal influence between the postulated common cause and the events representing the setting of the experiments (and therefore between the common cause and the experimenters' decisions), but the lack of *all* causal influences between these. This I will refer to as the *No-cause* presupposition.

Note that the three presuppositions above may not be the only assumptions in relation to free will when it comes to *Measurement Independence*. I do take however these presuppositions to be sustaining the intuitive core of *Measurement Independence* as an assumption about free will.

5 Cause-Statistics Link and Causal Explanation

If we endorse the idea that free will can be characterised, perhaps at a basic level, by the presence or lack of certain causal relations we will need to provide a minimal definition at least of what is to be a cause (or, alternatively, what it is for a certain event to be causally influenced). We need, for instance, to be able to tell how a certain event is to causally influence or not the EPR experimenters' free acts (to choose such and such setting for measurement). A common option is to identify, at least to some extent and under certain circumstances, causal dependence (independence) with statistical dependence (independence).

As a first observation thus, and as far as we endorse a probabilistic characterization of causation, we seem to be in need of a robust correspondence between causal relations and probabilistic expressions. In particular, we need to assume that the "translation" of our causal claims into probabilistic expressions are not only sensible but also adequate—at least in the cases we are interested in. This is to say, we need to make sure that the proposed probabilistic relations express unambiguously the actual causal claims they are intended for, and no others. Let us call this the *Cause-statistics Link* assumption.

---

interesting open problem. Miklós Rédei, for instance, has suggested (private conversation) that, if "acts of will" are defined as elements in a Boolean algebra, there could be at least three possible probabilistic conditions that one could claim would express in some sense the idea of us having free will. It would all depend on whether the required independence between the common cause C and the measurement settings m_i is of a logical character, refers to the corresponding probability distributions, or is simply statistical independence.

The *Cause-statistics Link* assumption is not a presupposition about free will *per se*. But, as pointed out, it is needed if we are to make sense of a probabilistic expression (such as *Measurement Independence*) as representing the notion of free will, as far as we take free will to be characterised, if not actually defined, causally.

Now, the *Cause-statistics Link* is a presupposition that can be easily challenged. In fact, there are many counter-examples that show that probabilistic dependence/independence is not necessary for causation, and certainly not sufficient either.[4] Only in some cases and under certain "good" conditions can the *Cause-statistics Link* assumption be considered adequate. So we could conclude that *Measurement Independence*, at least as defined in the context above, is not necessary for free will just by rejecting the idea that causal relations are adequately expressed in terms of probabilistic relations. This move would have however undesired consequences. For instance, if our analysis is motivated to some extent by the desire of explaining the EPR correlations causally—making use for example of the Principle of the Common Cause—we could not afford rejecting the necessity claim (2) on the grounds above. For if common cause explanations are to make sense at all then the *Cause-statistics Link* presupposition needs to be in place. Thus, while rejecting the *Cause-statistics Link* would undermine the claim that *Measurement Independence* is necessary for free will, it would also eliminate any possibility of explaining the EPR correlations in terms of common causes (or any other causal explanation based on the idea that causal relations are captured by probabilistic expressions).

6 Temporal Order of Events and the Direction of Causation

The requirement of *Measurement Independence* in the EPR picture presupposes as well a certain temporal time ordering of the events involved. Common causes, in particular, are assumed to take place *before* measurement operations do—and therefore *before* the corresponding outcomes have been recorded. Let us call that the *Time Order* presupposition.

As a side remark, it needs to be noted that *Measurement Independence* is a relation about *types* of events which are not, strictly speaking, defined as actual spacetime events. Thus, in principle, the notion *before* (*after*) should not apply to them. There is not, to my knowledge, an appropriate and detailed spacetime

[4]The literature on the subject is huge. A classic reference is for instance (Salmon, 1984).

description of event types. A simple way to avoid such problems is to consider type events as constituted by sets or collections of the corresponding tokens. In this view one can then refer to event types spatio-temporally in virtue of them being collections of token events. This should allow us in turn to consider common cause (type) events in *Measurement Independence* as located in the causal past of measurement (type) events.

In any case, the *Time Order* presupposition is very often assumed only implicitly, and with no further justification. The common view seems to be that presupposing this particular time order of events is just as natural—how could it be otherwise?—, so there is really no need for a proper justification. I will suggest however that there are conceivable causal pictures of the EPR experiment in which this time order is altered.

First, note that the presupposed time arrangement that the *Time Order* presupposition demands makes sense only in the context of a further causal assumption, namely that *cause events are in the past of their effects*. In other words, the usual EPR picture involved in the derivation of the Bell inequalities takes it that, if there were some causal story to explain the correlations, causes would take place prior in time to the corresponding effects, i.e. the EPR outcome events. (Again, this seems an intuitively correct, straightforward and innocuous assumption, which may not need further justification.) Thus requiring *Measurement Independence* in the usual derivations of the Bell inequalities involves a combination of an assumption about the *temporal order* of events as well as an assumption about the correct *causal order* to be taken. Both these two presuppositions can be contested independently. This leads to at least three different (causal) pictures, depending on which of the two assumptions above is dropped.

One may want to keep in the first place the temporal arrangement of events initially assumed. That is, one may want to stick to the idea that the postulated common cause is to take place before *both* the measurement operation events (in both wings) and the outcome events. Let me point out that I see no particular reason one must assume this specific time order of events—rejecting it however, would take us to a different causal picture (causal pictures 2 and 3 below). But if we do insist in keeping this specific temporal order the intuition is that, if violations of *Measurement Independence* are not to be tantamount to a world conspiracy (e.g. in the form of a lack of free will), the causal picture should involve backwards in time causation. In particular, the setting of the apparatus for measurement m_i (and perhaps the actual measurement operations) can be thought to be a (future) cause of the postulated common cause C, which would cause in turn the measured

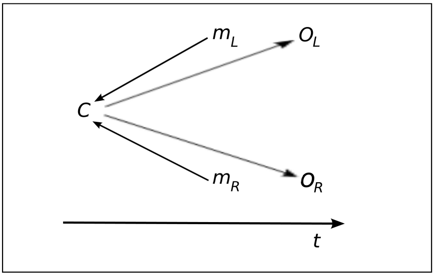

Figure 1: Backwards causation (version 1). Time order of events is preserved but causes propagate backwards in time (causal picture 1).

outcomes O_i (see Figure 1). If the common cause C is located in a sufficiently distant past, this picture turns out to be completely local, hence avoiding the usual conflicts with special relativity. Needless to say that the appeal to backwards causation is taken by many as a highly counterintuitive option. A good argument in favour of such a causal picture however has been made by Price (1994, 1996).

A second causal picture results from rejecting the initially presupposed temporal arrangement of events while keeping the assumption that causes propagate forward in time to cause their effects (San Pedro, 2012). The resulting causal structure is depicted in Figure 2. In this case, the postulated common cause C can be thought to take place sometime in between the actual measurement operations m_i and the occurrence of the observed (correlated) outcomes O_i. That is to say, C is postulated to be in the future of the measurement operations in both wings (and thus after the events representing the experimenters' measurement choices) but in the past of the EPR outcomes. Moreover, measurement is taken to be in this view an explicit causal factor (of both C and the outcomes), hence implying that *Measurement Independence* is violated. As pointed out, the above causal picture retains the most accepted intuition that causes propagate forward in time. This comes at a price nevertheless. Namely, a common cause model built along these lines seems to forcefully involve some sort of (explicit) non-locality.[5]

Finally, a third causal picture would result form the rejection of both the initially presupposed time order of the events involved as well as the intuition that causa-

[5] See (San Pedro, 2012) for details.

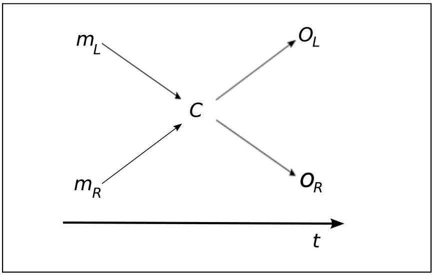

Figure 2: Time order of events is *not* preserved and causes propagate, as usual, forward in time (causal picture 2).

tion propagates forward in time. If backwards causation is again brought into the picture, the time order of events can be easily rearranged such that the postulated common cause C is in the future of the EPR outcome events (and thus, of course, in the future of the events representing the experimenters' choices and/or the actual measurement operations). The common cause may be thought indeed to be situated far enough in the future so as to guarantee that the causal interactions be completely local. (The corresponding causal structure is represented in Figure 3.) Then again, once we consider violations of *Measurement Independence*, the issue of locality seems to be tightly bound to whether or not we allow for backwards causation.[6]

It is not my intention here to discuss how appealing, likely or unlikely any of the above options are.[7] My aim is rather to suggest that in revising the presuppositions of a certain fixed time order of events and/or whether causation propagates forward in time, one can provide sensible causal pictures in which *Measurement Independence* is violated. Violations of *Measurement Independence* do not involve in any of these cases a lack of freedom of will on the part of the EPR experimenters, nor a world conspiracy in the form of an entanglement of "apparently separate parts of the world," to use Bell's terminology. Thus, what

[6]Locality issues are complex and deserve more attention than what we can afford here. See (San Pedro, 2012) for a brief discussion of the implications to the idea of locality due to violations of *measurement independence*.

[7]I point the reader to (Price, 1994, 1996) and to (San Pedro, 2012) for a defense of causal pictures 1 and 2 respectively.

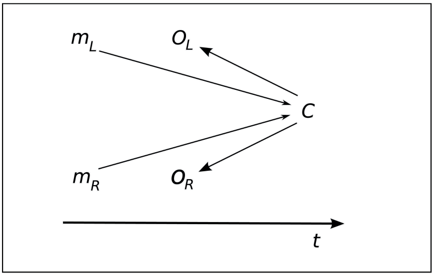

Figure 3: Backwards causation (version 2). Time order of events is *not* preserved and causes propagate backwards in time (causal picture 3).

the above already suggests is that the requirement of *Measurement Independence* in the derivation of the Bell inequalities is *independent* of whether EPR experimenters have free will or not.

7 No Causal Influence at All

In addition to the two assumptions discussed above, for the necessary connection between *Measurement Independence* and free will to stand, it is required that there be no causal influence *at all* from the common cause on the experimenters' free acts when setting the apparatus in such and such direction for measurement.

This *No-cause* presupposition, as we may call it, may turn out to be however too strong a requirement. For demanding no causal influence *at all* seems to suggest either a deterministic causal view as regards the (hidden) common cause events, or at least an idea of cause that exhausts all possible causal factors of a given effect, i.e. a *total* cause. In particular, *No-cause* may be taken to be reasonable in a deterministic context or, alternatively, in the case common causes were thought to be *total* causes of measurement settings. These two are, of course, not the only available options.

In an indeterministic context it is indeed a possibility to conceive the postulated common cause C to be not a total but just a *partial* cause of the measurement setting events m_i. Obviously *Measurement Independence* would not hold in this case. But, would that picture constitute a violation of free will? I don't think so. It is to me very sensible to think that free will would still be preserved even in the case our range of choices, or acts, had been somehow limited. (It seems in fact difficult to think of a situation where we are completely or "unboundedly" free to act.) And this is precisely what seems to be behind the idea of *partial* cause. So in this view free will is again completely compatible with the violation of *Measurement Independence*.

As for deterministic contexts, there is no need to distinguish between *total* and *partial* causes since the presence of *any* cause entails (with probability one) the occurrence of the corresponding effect. In this particular case it does seem intuitively correct to demand no causal influence of any sort if free will is to be preserved. But this may turn out not to be so, after all. In particular, one may want to endorse for instance a *compatibilist* position, and claim that free will is perfectly compatible with a fully deterministic universe.[8] There is no need to revise the *compatibilism-incompatibilism* debate here. I just would like to stress the fact that there are several options available, also in deterministic contexts, where *No-cause* is just too strong an assumption. Under such circumstances then the necessity claim (2) is to be put into question.

In sum, as suggested above, the *No-cause* presupposition may very well be seen to be too strong a condition on the requirement of *Measurement Independence* as a necessary condition for free will. Relaxing it then, opens for the possibility of non-conspiratorial—or free will compatible—violations of *Measurement Independence*.

8 Discussion

I have shown in the discussion above that the commonplace claim by which *Measurement Independence* is taken to be necessary for the whole idea of free will in causal explanations of the EPR correlation is, although apparently correct according to certain intuitions, ultimately mistaken.

The three underlying assumptions I have identified here are all revisable and can be challenged each in its own grounds. As a result, the notion of free will is shown to be compatible with the violation of *Measurement Independence* in different fashions, depending on which of the assumptions is rejected, and with diverse implications in each case. For instance, while it is difficult to make sense of common cause models of EPR if *Cause-statistics Link* is rejected, it seems

[8] Very roughly, *compatibilism* reconciles the idea of free will within deterministic contexts by reducing it somehow to a psychological subjective feature of ours.

plausible to conceive violations of *Measurement Independence* as long as one takes (hidden) common causes to be events that only *partially* cause or influence the EPR experimenters' (partially) free decisions and acts. Most interesting are perhaps the three common cause models that one may conceive in the context of a violation of *Measurement Independence* due to the rejection of either the fixed time order of events usually presupposed in the EPR scenario or the idea that causes propagate forward in time to cause their effects, or both. In discussing them, we saw that whether the models turned out to be local or not depended on which of these two assumptions was dropped. Locality issues, then, can be seen in these three models to be related to considerations about the temporal order of events, or the direction of causation. It would be valuable to know precisely how these are related, but this work needs to be left for further research.

Acknowledgment

Some of the ideas presented in this work were developed during discussion in the context of a broader project, also on the relation between free will and *No-conspiracy*, in collaboration with Miklós Rédei. I thank him for his dedication and generosity. Research supported by the Spanish Ministry of Science and Innovation research project FFI2008-06418-C01-03.

References

Bell, J. (1981). Bertlmann's socks and the nature of reality. *Journal de Physique*, Colloque C2, suppl. au numero 3, Tome 42:41–61.

Conway, J. and Kochen, S. (2006). The free will theorem. *Foundations of Physics*, 36:1441–1473.

Price, H. (1994). A neglected route to realism about quantum mechanics. *Mind*, 103:303–306.

Price, H. (1996). *Time's Arrow and Archimedes' Point*. Oxford University Press.

Salmon, W. (1984). *Scientific Explanation and the Causal Structure of the World*. Princeton University Press.

San Pedro, I. (2012). Causation, measurement relevance and no-conspiracy in EPR. *European Journal for Philosophy of Science*, 2:137–156.

Tumulka, R. (2007). Comment on ‘The free will theorem’. *Foundations of Physics*, 37:186–197.

van Fraassen, B. (1982). The charybdis of realism: Epistemological implications of Bell’s inequality. *Synthese*, 52:25–38.

Part II

Collapse and Non-Locality

(How) Did Einstein Understand the EPR Paradox?

Tilman Sauer

Abstract. *An unpublished formulation by Einstein of the EPR paradox in terms of spin variables raises the question as to his precise understanding of this argument. I review various formulations of the argument in this respect and argue that a core tenet of his understanding was completeness in an ambiguous sense. On the one hand, incompleteness is implied when differences in reality are not captured by the theoretical representation of that reality. But for Einstein quantum mechanical incompleteness also implies a contradiction, i.e. when the same physical state of affairs is described by two formulations that are "different in kind." The critical word here is "different" and it is argued that Einstein intends a notion of "different" that implies empirical non-equivalence. Nevertheless, Einstein elaborates on an example where no-signalling applies, a fact which renders the notion of empirical non-equivalence problematic.*

1 Introduction

A great deal of current philosophical reflections on the foundations of quantum mechanics refers back—directly or indirectly—to the incompleteness argument put forward in 1935 by Albert Einstein, Boris Podolsky, and Nathan Rosen (Einstein et al., 1935). As is often the case with such landmark writings, the assessment of their significance, in our case the assessment of the significance of the EPR argument, changed over time. Niels Bohr (1935) felt challenged to respond to the EPR paper right away with a paper that appeared in the same journal under the same title. In later years, the criticism of the foundations of quantum mechanics associated with Einstein's name was often given short shrift. Einstein had turned, in the eyes of many working physicists, from revolutionary to reactionary, and his later views were considered curious at best. In his 'Subtle is the Lord...,'—still the best and only scientific biography of Einstein that we have—Abraham Pais only devoted a single page to the EPR paper. According to him

T. Sauer, A. Wüthrich (eds.), *New Vistas on Old Problems. Recent Approaches to the Foundations of Quantum Mechanics*, Edition Open Access, 2013.

> it simply concludes that objective reality is incompatible with the assumption that quantum mechanics is complete. This conclusion has not affected subsequent developments in physics, and it is doubtful that it ever will. (Pais, 1982, 456)

In fact, Pais goes on to side with Bohr:

> 'It is only the mutual exclusion of any two experimental procedures, permitting the unambiguous definition of complementary physical quantities which provides room for new physical laws,' Bohr wrote in his rebuttal. He did not believe that the Einstein-Podolsky-Rosen paper called for any change in the interpretation of quantum mechanics. Most physicists (myself included) agree with this opinion. (ibid.)

This was written in 1982, i.e., at around the time when experiments by Aspect et al. (1982) vindicated Bell's suspicion that the EPR argument actually captured something deeper about the conceptual foundation of quantum mechanics (Bell, 1964, 1987). With Bell's famous theorem and with similar other theorems advanced since then it became clear that the notions of causality, reality, and locality that play a central role in the EPR argument do indeed lend themselves to the formulation of precise and testable experimental predictions.[1]

In recent years, many physicists have taken the incompatibility between certain notions of causality, reality, and locality and the empirical data (correctly described by quantum mechanics) less and less as a philosophical stumbling block, that would best be avoided if one does not want to get snarled up in unproductive interpretational subtleties. Instead, more and more physicists came to regard this tension as a productive resource for new ideas about quantum entanglement, quantum computation, quantum cryptography, quantum information, and similar topics. And, at least in their own way of identifying historical tradition and indebtedness, they began to cite, routinely, the original EPR paper. Einstein, the old, stubborn critic of the new quantum mechanics, became a prescient visionary of new revolutionary ideas again.[2]

The question whether Einstein and his attitude toward quantum theory is justly regarded as either stubborn or prescient is not without some interest for us today. When we project back our modern understanding of the EPR argument to any of its original formulations we may find that Einstein's words do not quite fit with

[1] For one such experiment, see the contribution by Philip Walther in this volume.

[2] See (Home and Whitaker, 2007) for a recent reappraisal of Einstein as a critic of quantum theory.

what we would expect he should have said. Depending on our attitude towards Einstein's understanding of physics, we may then either feel challenged to try to make sense of Einstein's way of thinking. We might either hope that we will get some good insights from this historical endeavor that will help us advance our understanding of current philosophical issues. Or else we may find that his words are just incompatible with what we now take as our best understanding of the issue at hand. In the latter case, we might hope to learn something about the conceptual progress that physics has made since the days of Einstein.

2 The EPR Paper

The EPR paper (Einstein et al., 1935) was received by the *Physical Review* on 25 March 1935 and published in its issue of 15 May of that year.[3] It is, of course, a famous paper, indeed one of the most frequently cited works by Einstein, and its text is well-known. Nevertheless, it is difficult to understand, its logic is convoluted, and its technical argument can be regarded as flawed.[4] It has often been observed that it is not the best place to study the EPR argument.

Let me nevertheless remind you of its more well-known features. The abstract summarizes the argument:

> In a complete theory there is an element corresponding to each element of reality. A sufficient condition for the reality of a physical quantity is the possibility of predicting it with certainty, without disturbing the system. In quantum mechanics in the case of two physical quantities described by non-commuting operators, the knowledge of one precludes the knowledge of the other. Then either (1) the description of reality given by the wave function in quantum mechanics is not complete or (2) these two quantities cannot have simultaneous reality. Consideration of the problem of making predictions concerning a system on the basis of measurements made on another system that had previously interacted with it leads to the result that if (1)

[3] For a discussion of the prehistory of that paper, see (Howard, 1990).

[4] For instance, Cooper (1950) pointed out that the original EPR argument depended on the representation theorem for the momentum operator but violated a necessary premise for the applicability of that theorem. This is because the joint wave function was assumed to vanish at some place, which renders the momentum operator Hermitian but no longer self-adjoint, since its domain is restricted to the positive or negative half-line. Einstein rebutted Cooper's argument with a limiting argument. See (Jammer, 1974, 236–238) for a discussion of further contentions of the EPR argument.

> is false then (2) is also false. One is thus led to conclude that the description of reality as given by a wave function is not complete. (Einstein et al., 1935, 777)

It was pointed out many years ago by Arthur Fine (1996) that Einstein was not responsible for the actual composition of the published EPR paper. About that publication, he wrote in a letter to Erwin Schrödinger:

> Dear Schrödinger:
> I was very happy with your long letter, which dealt with my little paper. This one was written, for linguistic reasons, by Podolsky, after many discussions. It did not come out in the end so well quite what I wanted; rather the main point was, so to speak, buried by erudition.[5]

One thing we learn about Einstein's understanding of the EPR argument from this letter is that he regarded the technical details with which the argument was spelled out as irrelevant for the core of the argument. In fact, he thought that the mathematical details actually obscured the main argument in this case.

In Einstein's letter, we learn what the essential point in his understanding was. He talks about a "difficulty," and in response to Schrödinger who had used the term "contradiction," he also used the terms "incompatible" ("unvereinbar") and "contradict" ("widerspricht").

In the following, I want to argue that the contradiction arises with an ambiguity in the concept of completeness as it seemed to be understood by Einstein. According to a *prima facie* reading of the EPR paper, completeness is used in a straightforward sense. There are demonstrable differences, matters of fact, in reality that are not captured in the theory. But on this interpretation it is not easy to see why Einstein considered the argument to be paradoxical. It seems to me that on a second reading the incompatibility or contradiction may rather be located between the fact that there is one and only one real state of affairs but at least two different, i.e. non-equivalent descriptions of this state of affairs. And this is a contradiction under the claim that quantum theory provides a complete description of reality, in the sense that in a complete description, every element of reality corresponds uniquely to one and only one element of the theory.

The problem seen in this way is really not so much one of completeness. It is rather one that is very similar to the problem of overdetermination. If there is a

[5]Translation taken from Don Howard (1985), see also (Howard, 1990); for the original text, see (Meyenn, 2011, Doc. 206).

unique given state of affairs, a complete theory has to provide a unique description of that state of affairs, and if the theory provides different descriptions, then the differences have to be shown to be, at least, empirically equivalent. Ideally, the descriptions should also be logically equivalent but this is, clearly, a stronger requirement. The burden of the EPR argument therefore is to show that there are two different descriptions of the same state of affairs that make empirically different predictions.

Let me provide some textual evidence for this ambiguity in Einstein's use of the concept of completeness.

In his letter to Schrödinger, Einstein repeated the mathematical argument of the EPR paper of expanding the wave function at point B in two different sets of eigenfunctions as Ψ_B and $\underline{\Psi_B}$. He then wrote:

> The essential point now is only the fact, that Ψ_B and $\underline{\Psi_B}$ differ from one another at all. I claim that this being different is incompatible with the hypothesis that the Ψ description is coordinated in a one-to-one way with the physical reality (the real physical state of affairs).[6]

The other premises of the EPR setup are only auxiliary to this conclusion:

> After the interaction the real state of affairs of (AB) consists of the real state of affairs of A and the real state of affairs of B, which two states have nothing whatsoever to do with each other. *The real state of affairs of B now cannot depend on what kind of measurement I perform at A*. ("Separation hypothesis" [...]).[7]

And he concludes the argument by stating again what the contradiction is:

> But then there are two (and in general arbitrarily many) equally valid Ψ_B associated with the same state of affairs of B in contradiction to

[6]"Wesentlich ist nun ausschliesslich, dass Ψ_B und $\underline{\Psi_B}$ überhaupt voneinander verschieden sind. Ich behaupte, dass diese Verschiedenheit mit der Hypothese, dass die Ψ-Beschreibung ein-eindeutig der physikalischen Wirklichkeit (dem wirklichen Zustande) zugeordnet sei, unvereinbar ist." (AEA 22-047; Meyenn, 2011, Doc. 206)

[7]"Nach dem Zusammenstoss besteht der wirkliche Zustand von (AB) nämlich aus dem wirklichen Zustand von A und dem wirklichen Zustand von B, welche beiden Zustände nichts miteinander zu schaffen haben. *Der wirkliche Zustand von B kann nun nicht davon abhängen, was für eine Messung ich an A vornehme.* ("Trennungshypothese" [...])." (AEA 22-047; Meyenn, 2011, Doc. 206), (Einstein's emphasis).

> the hypothesis of a one-to-one or complete description of the real state of affairs.[8]

3 Other Formulations

Einstein makes the same point on his first occasion for discussing the EPR argument in print. This is in his 1936 lecture on "physics and reality." There, he calls the EPR argument "the paradox recently demonstrated by myself and two collaborators."[9] The conclusion is very similar to his explanation in the letter to Schrödinger quoted above. He wrote:

> Since there can be only *one* physical condition of B after the interaction and which can reasonably not be considered as dependent on the particular measurement we perform on the system A separated from B it may be concluded that the Ψ function is *not* unambiguously coördinated with the physical condition. This coördination of several Ψ functions with the same physical condition of system B shows again that the Ψ function cannot be interpreted as a (complete) description of a physical condition of a unit system. (Einstein, 1936a, 376)[10]

In 1948, twelve years later, Einstein gave another formulation of his objection in print. In his contribution to a special issue, edited by Wolfgang Pauli, of the Swiss journal *Dialectica*, he reiterated the EPR-argument with special emphasis on what he saw as the basic assumption of objective reality in physics. He emphasized that the experimenter is perfectly free to choose which observable he wants to measure at the first system S_1, and then he wrote:

[8] "Dann aber gibt es zu demselben Zustande von B zwei (überhaupt bel. viele) gleichberechtigte Ψ_B, was der Hypothese einer ein-eindeutigen bezw. vollständigen Beschreibung der wirklichen Zustände widerspricht." (AEA 22-047; Meyenn, 2011, Doc. 206).

[9] ("[...] eine von mir zusammen mit zwei Mitarbeitern jüngst dargestellte Paradoxie." Jammer (1974, 186) claimed that Einstein never referred to the EPR argument as "paradoxical," a statement that was refuted already by Fine (1996, 47, n. 11).

[10] "Da es nur *einen* physikalischen Zustand von B nach der Wechselwirkung geben kann, welcher vernünftigerweise nicht davon abhängig gemacht werden kann, was für Messungen ich an dem von B getrennten System A vornehme, zeigt dies, dass die Ψ-Funktion dem physikalischen Zustande *nicht* eindeutig zugeordnet ist. Diese Zuordnung mehrerer Ψ-Funktionen zu demselben physikalischen Zustande des Systems B zeigt wieder, dass die Ψ-Funktion nicht als (vollständige) Beschreibung eines physikalischen Zustandes (eines Einzelsystems) gedeutet werden kann." (Einstein, 1936b, 341)

> Depending on this choice we obtain representations of ψ_2 of a different kind, specifically such that depending on the choice of measurement at S_1 different (statistical) predictions result for measurements to be taken at S_2 after the fact.[11]

In his setup of the argument he had earlier distinguished two alternative interpretations. According to one, a particle "really" has a definite location and a definite momentum, and the quantum mechanical description is considered incomplete (Ia). According to the other alternative, a particle has no definite location and no definite momentum before any measurement takes place, and the quantum mechanical description is considered complete (Ib). Einstein then argues:

> From the point of view of the interpretation Ib this means that depending on the choice of the complete measurement at S_1 different real situations are generated, which are described by different ψ_2, $\underline{\psi_2}$, $\underline{\underline{\psi_2}}$ etc.[12]

But, of course, the locality argument is crafted exactly to invalidate the assumption that physically different (as opposed to representationally different) situations can be "generated" by the choice of measurement at the distant wing. Hence we are left again with the situation that different descriptions of different empirical content are coördinated with the same physical state of affairs. The statistical interpretation (in Einstein's understanding) is not a solution for the description of any individual measurement, since it only picks out a subensemble depending on the choice of parameter at one wing, which can make a difference for measurements at the other wing.

Lastly, in his *Autobiographical Notes*, we find the same formulation again. The problem is to have a "different," or "very different" wave function, i.e. one "of a different kind" ("andersartig" or "verschiedenartig") for the same state of affairs:

[11] "Je nach dieser Wahl erhalten wir für ψ_2 eine anders-artige Darstellung, und zwar derart, dass je nach der Wahl der Messung an S_1 verschiedenartige (statistische) Voraussagen über an S_2 nachträglich vorzunehmende Messungen resultieren." (Einstein, 1948, 322)

[12] "Vom Standpunkte der Interpretation Ib bedeutet dies, dass je nach der Wahl der vollständigen Messung an S_1 eine verschiedene reale Situation hinsichtlich S_2 erzeugt wird, die durch verschiedenartige ψ_2, $\underline{\psi_2}$, $\underline{\underline{\psi_2}}$ etc. beschrieben werden." (Einstein, 1948, 322)

> According to the type of measurement which I make of S_1, I get, however, a very different ψ_2 for the second partial system (ψ_2, ${\psi_2}^1$, ...).[13]

And similarly:

> For the same real situation of S_2 it is possible therefore to find, according to one's choice, different types of ψ-function.[14]

Clearly, it is the fact that in quantum theory one obtains several *different* descriptions of the *same* state of affairs was at the core of Einstein's unease about it.

The difference between the different descriptions has to be essential and cannot merely be formal or even notational. The objection cannot be that we may denote the wave function by $\psi(x)$ or $\Psi(x)$ or $\varphi(x)$. Different notations are logically equivalent, and such differences can readily be captured by a distinction between symbolic types and tokens. But the objection also cannot be that we may add a phase factor $e^{i\alpha}$ to the time-independent Schrödinger wave function. The equivalence here is less obvious but it is still a mathematical equivalence, in the sense that the fundamental equation, the Schrödinger equation is invariant under such gauge transformations. But what is the qualitative difference between benign differences in notation or mathematical gauge fixing and fatal differences that render the theory "incomplete?"

As we have seen, in none of the known formulations of the EPR paradox does Einstein give an explicit discussion of what the crucial difference in non-equivalent descriptions might be. Obviously, one would suspect that it would be a difference that render the two descriptions empirically non-equivalent. Einstein seems to suggest that the choice of parameter at one wing entails different predictions about measurement outcomes at the other wing. But it is not clear whether this empirical non-equivalence pertains to the individual measurement or to a statistical ensemble.

Let us review one more formulation of the EPR paradox in this respect. It is another concise, non-technical formulation and it is, in all probability, Einstein's latest formulation of the argument. It is found on the bottom half of a sheet that is

[13] "Je nach der Art der Messung, welche ich an S_1 vornehme, bekomme ich aber ein andersartiges ψ_2 für das zweite Teilsystem (ψ_2, ${\psi_2}^1$, ...)," (Einstein, 1982, 84/85).

[14] "Für denselben Realzustand von S_2 können also (je nach Wahl der Messung an S_1 verschiedenartige ψ-Funktionen gefunden werden." (Einstein, 1982, 84/85)

part of a larger batch of manuscript pages with calculations on general relativity and unified field theory (Sauer, 2007). There is a good chance that it was written down by Einstein after reading David Bohm's (1951) textbook on *Quantum Theory*, in which we find the EPR argument for the first time formulated in terms of spin variables.

Let me quote the formulation in full. In a slightly smoothed English translation it reads (Sauer, 2007, 882):[15]

> Composite system of total spin 0.
>
> 1) The description is assumed to be complete.
> 2) A coupling of distant things is excluded.
>
> If the spin of the subsystem I is measured along the x-axis, it is found to be either 1 or -1 in that direction. It then follows that the spin of the subsystem II equals 0 along the y-direction. But if instead the spin of subsystem I is measured along the y-direction, it follows that the spin of the subsystem II is equal to 1 or -1.
>
> If there is no coupling, then the result of a measurement of the spin of subsystem II may in no way depend on whether a measurement was taken of subsystem I (or on what kind of measurement).
>
> The two assumptions therefore cannot be combined.
>
> If the description is *not* assumed to be complete for the individual system, then that what is being described is not a single system but an ensemble of systems. Then a measurement of subsystem I amounts to the selection of a subensemble of the ensemble of the total system. Then the prediction for a measurement of subsystem II can depend on the choice of the measurement of subsystem I.
>
> The conclusion is valid under the assumption that the assertion of quantum theory is correct, which we can hardly put into doubt.

The following lines were written at the right margin of the page:

> a) the description by the quantum theory is an incomplete one with respect to the individual system, or
> b) there is an immediate coupling of states of spatially separated things.

[15] For a faithful transcription of the original German manuscript, see (Sauer, 2007, 886).

In view of our preceding discussion, we first observe that Einstein clearly thought the different descriptions of the partial subsystems were *empirically* inequivalent. He argues that the "result of a measurement" would come out differently depending on the choice of parameter in the distant wing. But the explication that Einstein gives before arriving at this conclusion is disturbing. Although less clear from the text, he seems to have in mind a situation of an *individual* measurement. He considers a (now standard) setup for the EPR argument, in which the spin of two quantum particles, each of spin $1/2$ but adding up to a vanishing total spin, are measured along two mutually orthogonal directions at distant wings. But what Einstein asserts does not square with quantum theory, whose assertions he explicitly claims to be "correct:" If Alice measures the x-component of the particle at her wing and finds it to be a definite value (+1 or -1), and if Bob *then* measures the y-component of his particle, he would also find it to be either +1 or -1, and quantum theory does not give a prediction as to which value would be obtained. That is so because a measurement by Alice of the x-component collapses the joint entangled wave function and hence puts the particle at Bob's end into an eigenstate to the x-component. Therefore, measuring the y-component at Bob's end would result in either +1 or -1 with a 50% probability each. Similarly, we read that Einstein is considering the case that Alice is measuring the y-component of her particle's spin, but, apparently, without taking note of the outcome. In that case, Bob's particle will collapse into an eigenstate of the y-component, and, again, he would measure the y-component of his particle to be either +1 or -1. Again, quantum theory cannot predict, which value Bob will actually see, and only predicts that he will see the outcome to be distributed with equal probability 50% between the two possible value. Note that this result does not depend in any way on Alice's taking note of the outcome of her measurement, since, by construction, she could not inform Bob about her measurement result *before* he would actually measure his particle's spin.

The situation is an illustration of a more general no-signalling theorem, which says that the EPR setup is not suited to send information from Alice to Bob faster than with the speed of light, or, in other words, even though the wave collapse occurs instantaneously along the entire space, it does not provide a means to convey significant bits of information.

Let us now read Einstein's argument under the assumption that he was having in mind a statistical reading. Statistically, there is, of course, a significant difference between the two cases. In the first situation, we have no correlation between Alice's and Bob's measurements, in the second situation we have complete anticorrelation. So far so good. But again, this is not what Einstein seems to have had in mind. Another reading seems to fit more natural with the text. According to this reading, Einstein would be thinking about statistical means. But this is problematic, too. We would have, of course, the situation that conditional on Alice's measuring an x-component of $+1$, Bob's y-component would average to a value of 0. The same holds for conditioning Bob's probability on Alice's measuring an x-component of -1. In contrast, Bob's y-component would average to $+1$ (or -1) if conditioned on Alice's measuring the y-component of her particle to be -1 (or $+1$).

But this interpretation does not go well with Einstein's insistence that the outcome of Bob's measurements should "in no way depend on whether a measurement was taken of subsystem I (or on what kind of measurement)." This latter formulation clearly suggests that he was taking the choice of parameter, not the measurement outcome, as the critical experimental intervention that must not have an effect on the measurement outcome at Bob's end. But, as we have seen, no-signalling in the situation at hand tells us that the outcome of Bob's measurement does not depend on the choice of parameter at Alice's experiment, neither for the individual measurement nor for a statistical ensemble.

4 A No-Signalling Theorem by David Bohm

It is even more puzzling that a no-signalling theorem of the kind that we just sketched was discussed and proven explicitly in David Bohm's book, which quite possibly was a source of inspiration for Einstein's new formulation of the EPR paradox (Sauer, 2007). Let me briefly indicate what Bohm had to say on this question. In the penultimate chapter of his textbook on quantum theory (which was still defending Bohr's views of quantum mechanics[16]) Bohm discussed the "quantum theory of the measurement process." The chapter sets out by a general discussion of how to include the measuring apparatus in a quantum mechanical description (very much in the spirit of von Neumann's axiomatic analysis),

[16] In a 1989 interview, Bohm recalled: "First I studied quantum mechanics and relativity, and in doing this I began by more or less accepting the ideas of Niels Bohr. Later I wrote a book called *Quantum Theory*, in which I was really quite strongly in favor of his ideas as I understood them. Well, I became somewhat dissatisfied towards the end of this period, around 1950 when I finished the book. I sent copies of the book to various physicists, including Pauli, Bohr, Einstein. Pauli liked the book. Einstein liked the book, but when I discussed it with him he said he was still not satisfied. Both of us felt that the key question was: 'What is the nature of reality?'." (Bohm, 2004)

then illustrates the general account by a detailed description and analysis of the Stern-Gerlach experiment, and finally comments on the EPR paper. This layout of the chapter allowed Bohm to discuss (apparently for the first time) the EPR argument with spin variables and with a hypothetical experimental setup using Stern-Gerlach apparatus and their analysis. In this context, Bohm very clearly formulated a no-signalling theorem. He wrote about the hypothetical EPR experiment:

> One more significant point arises in connection with this experiment; namely, that the existence of correlations does not imply that the behavior of either atom is affected in any way at all by what happens to the other after the two have ceased to interact. (Bohm, 1951, 618–619)

He also gave a proof of this statement. Define basic wave functions ψ_c and ψ_d as $\psi_c = u_+(1)u_-(2)$ and $\psi_d = u_-(1)u_+(2)$, where u_+ and u_- denote "the one-particle spin wave functions representing, respectively, a spin $\hbar/2$ and $-\hbar/2$, and the argument (1) or (2) refers, respectively, to the particle which has this spin." Bohm had earlier in the chapter argued that a measurement of the z-component of the spin at one wing would generate uncontrollable phase factors $e^{-\alpha_c}$ or $e^{-\alpha_d}$ to the spin wave functions, such that the total spin of the joint system is no longer defined. He could therefore now proceed to give a brief proof of his no-signalling theorem which states more precisely that the expectation value of any function g of the spin $\sigma = (\sigma_x, \sigma_y, \sigma_z)$ of particle 2 does not depend on whether or not a measurement of some spin component of particle 1 was done. Here is how Bohm phrased his proof:

> To prove this statement, we first evaluate the mean of any function $g(\sigma_2)$ of the spin variables of particle No. 2 alone. With the wave function before a measurement took place, we obtain
>
> $$\overline{g}_0(\sigma_2) = \frac{1}{2}(\psi_c^* - \psi_d^*)g(\sigma_2)(\psi_c - \psi_d) = \frac{1}{2}\left[\psi_c^* g(\sigma_2)\psi_c + \psi_d^* g(\sigma_2)\psi_d\right]$$

> (By virtue of the orthogonality of ψ_c and $g(\sigma_2)\psi_d$.) After the spin of the first particle is measured, the average of $g(\sigma_2)$ becomes
>
> $$\overline{g}_f(\sigma_2) = \frac{1}{2}(\psi_c^* e^{-i\alpha_c} - \psi_d^* e^{-i\alpha_d}) g(\sigma_2)(\psi_c e^{i\alpha_c} - \psi_d e^{i\alpha_d})$$
> $$= \frac{1}{2}\left[\psi_c^* g(\sigma_2)\psi_c + \psi_d^* g(\sigma_2)\psi_d\right]$$
>
> This is the same as what was obtained without a measurement of the spin variables of particle No. 1. The behavior of the two spins is, however, correlated despite the fact that each behaves in a way that does not depend on what actually happens to the other after interaction has ceased. (Bohm, 1951, 619)

Einstein, I believe, must have read and known this passage, which is part of a longer chapter that also discusses his EPR paper. Yet, he does not seem to acknowledge in his manuscript notes the fact that the EPR setup for spins as discussed by Bohm obeys a no-signalling theorem. Such a theorem, it seems, would undermine his argument that the EPR paradox is a contradiction in the sense that quantum theory here gives rise to two empirically different descriptions of the same physical state of affairs.

5 Concluding Remarks

What do we make of this? It appears that Einstein either did not understand what quantum theory actually predicts in the EPR situation that he was considering, or, at least, that he did not bother to spell out the theory's prediction carefully in so many details. But this conclusion is at odds with the fact that Einstein, for all we know otherwise, had an excellent understanding of quantum theory as well as of statistical physics and its underlying concepts. On the other hand, history has shown that Einstein's intuition behind the EPR setup was well borne out by the subsequent development of quantum physics. Should the conclusion then be that Einstein's general realist philosophical perspective let him anticipate difficulties of quantum theory[17] even though he would not spell them out in his later years as clearly as he would have done when he was still young?

[17] For Einstein's (and Paul Ehrenfest's) prescient anticipation, as it were, of the quantum measurement problem in their discussion of the difficulties in interpreting the result of the Stern-Gerlach experiment, see (Unna and Sauer, forthcoming).

References

Aspect, A., Dalibard, J., and Roger, G. (1982). Experimental test of Bell's inequalities using time-varying analyzers. *Phys. Rev. Lett.*, 49:1804–1807.

Bell, J. (1964). On the Einstein-Podolsky-Rosen paradox. *Physics*, 1:195–200.

Bell, J. (1987). *Speakable and Unspeakable in Quantum Mechanics*. Cambridge University Press.

Bohm, D. (1951). *Quantum Theory*. Prentice-Hall.

Bohm, D. (2004). *On Creativity*. Routledge Classics.

Bohr, N. (1935). Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.*, 48:696–702.

Cooper, J. (1950). The paradox of separated systems in quantum theory. *Proceedings of the Cambridge Philosophical Society*, 46:620–625.

Einstein, A. (1936a). Physics und reality. *Journal of the Franklin Institute*, 221:349–382.

Einstein, A. (1936b). Physik und Realität. *Journal of the Franklin Institute*, 221:313–347.

Einstein, A. (1948). Quanten-Mechanik und Wirklichkeit. *Dialectica*, 2:320–324.

Einstein, A. (1982). Autobiographisches/Autobiographical Notes. In Schilpp, P. A., editor, *Albert Einstein: Philosopher-Scientist*, pages 1–96. Open Court.

Einstein, A., Podolsky, B., and Rosen, N. (1935). Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.*, 47:777–780.

Fine, A. (1996). *The Shaky Game. Einstein, Realism and the Quantum Theory (2n ed.)*. University of Chicago Press.

Home, D. and Whitaker, A. (2007). *Einstein's Struggles with Quantum Theory. A Reappraisal*. Springer.

Howard, D. (1985). Einstein on locality and separability. *Studies in History and Philosophy of Science*, 16:171–201.

Howard, D. (1990). 'Nicht sein kann, was nicht sein darf,' or the prehistory of EPR: Einstein's early worries about the quantum mechanics of composite systems. In Miller, A., editor, *Sixty-two Years of Uncertainty: Historical, Philosophical, Physics Inquiries into the Foundations of Quantum Physics*, pages 61–111. Plenum Press.

Jammer, M. (1974). *The Philosophy of Quantum Mechanics*. John Wiley and Sons.

Meyenn, K., editor (2011). *Eine Entdeckung von ganz außerordentlicher Tragweite. Schrödingers Briefwechsel zur Wellenmechanik und zum Katzenparadoxon*. Springer.

Pais, A. (1982). *'Subtle is the Lord': The Science and the Life of Albert Einstein*. Oxford University Press.

Sauer, T. (2007). An Einstein manuscript on the EPR paradox for spin observables. *Studies in the History and Philosophy of Modern Physics*, 38:879–887.

Unna, I. and Sauer, T. (forthcoming). Einstein, Ehrenfest, and the quantum measurement problem. *Annalen der Physik (Berlin)*.

Quantum Theory as a Method: the Rule Perspective

Simon Friederich

Abstract. *The paper presents a "therapeutic" account of quantum theory, the "Rule Perspective," which attempts to dissolve the notorious paradoxes of measurement and non-locality by reflection of the nature of quantum states. The Rule Perspective is based on the epistemic conception of quantum states—the view that quantum states are not descriptions of quantum systems but rather reflect the assigning agents' epistemic relations to the systems. The main attractions of this conception of quantum states are outlined before it is spelled out in detail in form of the Rule Perspective. The paper closes with an assessment of the status of quantum probabilities in the light of the considerations presented before.*

1 Introduction

By most accounts, the measurement problem and the problem of quantum "non-locality," that is, the tension between quantum theory and relativity theory, are the two most important challenges in the foundations of quantum mechanics. Possible ways to react to these problems range from changing the dynamics (as in GRW theory) to adding determinate particle and field configurations (as in pilot wave approaches) to adopting a non-standard metaphysical picture in which the universe continuously splits up in an immense number of branches (the Everett interpretation). A completely different approach to the solution of the two problems is to try to *dissolve* them by showing that they arise from misunderstandings of the notions in terms of which quantum theory is formulated and disappear as soon as these misunderstandings are removed. Such a perspective on the quantum mechanical formalism is offered by the epistemic conception of states—the view that quantum states do not describe the properties of quantum systems but rather reflect the state-assigning agents' epistemic relations to the systems. The main attraction of this idea is that it offers a reading of the quantum mechanical for-

T. Sauer, A. Wüthrich (eds.), *New Vistas on Old Problems. Recent Approaches to the Foundations of Quantum Mechanics*, Edition Open Access, 2013.

malism that avoids both the measurement problem and the problem of quantum non-locality in a very elegant way.[1]

There are two very different types of accounts that are based on the epistemic conception of states. Those of the first type are hidden variable models where the quantum state expresses incomplete information about the configuration of hidden variables that obtains in that any configuration of hidden variables (which one might call an "ontic" state) is compatible with several quantum states, which one might therefore characterize as "epistemic."[2] In this paper, I will only talk about accounts of a second, very different type. Accounts of this second type try to dissolve the paradoxes of measurement and non-locality without presenting a theory of "ontic" states of quantum systems at all. They can be characterized as "interpretation[s] without interpretation"[3] in the sense that, according to them, quantum theory is fine as it stands without any additional technical vocabulary such as hidden variables, branching worlds, dynamics of collapse, or whatever else. Such a perspective on quantum theory can be called "therapeutic" in that it holds the promise to "cure" us from what is seen as unfounded worries about foundational issues like the measurement problem on the basis of conceptual clarification alone.

The question of whether this promise can be fulfilled will be discussed in this paper, beginning in Section 2, where I briefly review the dissolutions of the paradoxes of measurement and non-locality offered by the epistemic conception of states. Sections 3 and 4 present a more specific account, which I propose to call the "Rule Perspective," that fleshes out the basic idea of the epistemic conception of states in more detail. Subsequently, Section 5 assesses the status of quantum probabilities in the light of the considerations presented before. The paper closes in Section 6 with a short remark on why the Rule Perspective, despite being based on a reading of quantum states as non-descriptive, is not a form of instrumentalism in that it *presupposes* rather than denies that physical states of affairs are describable in objective (yet non-quantum) terms.

---

[1] For studies defending versions of the epistemic conception of states and views in a similar spirit, see (Fuchs and Peres, 2000; Mermin, 2003; Caves et al., 2002a,b; Fuchs, 2002; Pitowsky, 2003; Schack, 2003; Bub, 2007; Caves et al., 2007; Spekkens, 2007; Fuchs, 2010; Healey, forthcoming: Friederich, 2011).

[2] See (Spekkens, 2007; Harrigan and Spekkens, 2010) for illuminating discussions of accounts of this type, both from a systematic and from a historical perspective.

[3] See (Fuchs and Peres, 2000).

2 Dissolving the Paradoxes

The measurement problem arises from the fact that if quantum states are seen as states that quantum systems "are in," evolving in time always according to the Schrödinger equation, measurements rarely have outcomes.[4] In quantum theoretical practice, the problem is solved "by brute force," namely, by invoking von Neumann's notorious projection postulate, which claims that the state of the measured system "collapses" to an eigenstate of the measured observable whenever a measurement is performed. Measurement collapse is commonly criticized on two grounds, first, that it is in stark contrast to the smooth time-evolution according to the Schrödinger equation in that it is abrupt and unphysical and, second, that we are given no clear criteria for distinguishing between situations where time-evolution follows the Schrödinger equation on the one hand and situations where collapse occurs on the other.[5]

The dissolution of the measurement problem in the epistemic conception of states has two different aspects: First, a conceptual presupposition for formulating the measurement problem is rejected by denying that quantum states describe the properties of quantum systems in the sense that the very idea of quantum states as states that quantum systems "are in" is not accepted. This makes it impossible to argue, as in standard expositions of the measurement problem, that according to the law of quantum mechanical time-evolution the measured observable cannot have a determinate value in the state the measured system is in. Second, the epistemic conception of states offers a very natural justification of measurement collapse by means of which the measurement problem is avoided in quantum mechanical practice. If one accepts the idea that quantum states in general depend on the assigning agent's epistemic situations with respect to the systems that the states are assigned to, the collapse of the wave function appears completely natural in that it merely reflects a sudden and discontinuous change in the epistemic situation of the assigning agent, not a mysterious discontinuity in the time-evolution of the properties of the system itself. The epistemic conception of states thus removes the inconsistency of the standard, ontic, perspective on quantum states with the fact that measurements evidently do have outcomes. It does not somehow *explain* the emergence of determinate outcomes but attempts

[4]At least if one assumes, as usual, the so-called eigenstate-eigenvalue link which says that for a system in a state ψ an observable A has a definite value a if and only if it is an eigenstate of the operator corresponding to A with eigenvalue a.

[5]See, for instance, (Ruetsche, 2002, 209) for a lucid account of these two criticisms.

to cause the felt need for a dynamical explanation of these to vanish. I shall briefly return to how this is done in the Rule Perspective at the end of Section 5.

To see the dissolution of the "paradox of quantum non-locality," that is, the difficulty of reconciling quantum mechanical time-evolution in the presence of collapse with the requirement of Lorentz covariance as imposed from relativity theory,[6] it is useful to consider as a specific example a two-particle system in an EPR-Bohm setup where two systems A and B are prepared in such a way that those knowing about the preparation procedure assign an entangled state, for instance, the state $\frac{1}{\sqrt{2}}(|+\rangle_A|-\rangle_B - |-\rangle_A|+\rangle_B)$, for the combined spin degrees of freedom. The two systems A and B are brought far apart and an agent, Alice, located at the first system, measures its spin in a certain direction. Having registered the result, she assigns two no longer entangled states to A and B, which depend both on the choice of observable measured and on the measured result. Another agent, Bob, located at the second system, may also perform a spin measurement (in the same or in a different direction of spin) and proceed to assign a pair of no longer entangled states to the two systems in an analogous way. Now the intriguing challenge for the ontic conception of quantum states as states quantum systems "are in" is to specify at which time which system is in which state and, in order to preserve compatibility with relativity theory, to do so in a Lorentz covariant manner. The difficulty is most pressing for cases where the measurements carried out by Alice and Bob occur in space-like separated regions, perhaps even in such a way that each of them precedes the other in its own rest frame.[7] In that case there is clearly no non-arbitrary answer to the question of which measurement occurs first and triggers the abrupt change of state of the other. Existing attempts to overcome this problem make quantum mechanical time-evolution dependent on foliations of spacetime into sets of parallel hyperplanes, but so far no such approach has found widespread acceptance.[8]

If one adopts the epistemic conception of quantum states, the problem of reconciling quantum theory and relativity theory disappears and the sudden change of the state Alice assigns to the second system appears very natural: Alice knows

[6] See (Maudlin, 2011, ch. 7) for a very useful and up-to-date exposition of this difficulty.

[7] See (Zbinden, 2001) for a discussion of experiments carried out in such a setup.

[8] See (Fleming, 1988) for a hyperplane-dependent formulation of state reduction and (Myrvold, 2002) for a defense of that approach. For criticism, see (Maudlin, 2011, ch. 7), which comes to a rather general negative verdict as to whether relativity theory and standard quantum theory can be consistently combined at all, based, however, on the presupposition that the ontic conception of quantum states is correct.

about the preparation procedure for the combined two-particle system, and it is not surprising that the result of her measurement of the first system may affect her epistemic condition with respect to the other. By interpreting the state not as a description of the system itself but as reflecting her epistemic situation we need not assume that her measurement of the first system has a physical effect on the second. Predictions for the results of measurement that are derived on the basis of entangled states may still be baffling and unexpected, but no conflict with the principles of relativity theory in form of superluminal effects on physical quantities does arise. Even though quantum mechanics is non-local in the sense of violating Bell-type inequalities, it does not involve any superluminal propagation of objective properties.

3 Quantum Bayesianism and Its Problems

In the previous sketch of the dissolution of the paradoxes of measurement and non-locality the question of *in which sense* quantum states may be regarded as reflecting the epistemic conditions of the assigning agents was not answered. Sometimes the epistemic conception of states is read as the claim that a quantum state represents our knowledge of the probabilities ascribed to the values of observables determined from it via the Born rule. Marchildon, for instance, identifies that view with the epistemic conception of states when he claims that "[i]n the epistemic view [of states], the state vector (or wave function or density matrix) does not represent the objective state of a microscopic system [...], but rather our knowledge of the probabilities of outcomes of future measurements."[9] However, as has been convincingly argued by Fuchs,[10] the notion of knowledge of quantum probabilities is incompatible with the epistemic conception of states, so the latter should evidently not be identified with the view described by Marchildon.

The reason why the notion of knowledge of quantum probabilities is incompatible with the epistemic conception of states has to do with the "factivity" of knowledge, that is, the conceptual feature of the notion of knowledge that knowing that q is possible only if q is indeed the case. As stressed above in the discussion of the dissolution of the measurement problem, the epistemic conception of states does not acknowledge the existence of a *true* state of a quantum system, a state it "is in." According to this perspective, different agents having different

[9] See (Marchildon, 2004, 1454).

[10] See (Fuchs, 2002), fn. 9 and sec. 7, in particular. See also (Timpson, 2008, sec. 2.3).

knowledge of the values of observables of a quantum system may legitimately assign different quantum states to it. The idea that quantum states reflect the assigning agents' knowledge of quantum probabilities, however, is incompatible with the assumption that different agents may legitimately assign different states to one and the same system. For if indeed probabilities were the objects of the assigning agents' knowledge, an agent might know the probability p of a certain measurement outcome E to occur, which would mean that, due to the factivity of "knowledge," p would be the one and only correct, the *true* probability for E to occur. Since this holds for any possible measurement outcome E to which the state assigned to the system ascribes some probability p, the probabilities obtained from this state would be the true ones and any other assignment of probabilities that differs from an assignment of these would simply be wrong. This conclusion would be incompatible with the claim that the states assigned by different observers to the same quantum system may legitimately be different. Therefore, the epistemic conception of states cannot be spelled out by saying that quantum states reflect knowledge about probabilities.

A viable (yet, as we shall see, too radical) option for adherents of the epistemic conception of states is to say that quantum states reflect the assigning agents' subjective *degrees of belief* about possible measurement outcomes. An account, which is based on this idea, has been worked out in great detail by Fuchs, Caves, and Schack and is now widely known as *quantum Bayesianism*. According to Fuchs, quantum states reflect our beliefs about what the results of "our interventions into nature"[11] might be. The probabilities encoded in quantum states, from this perspective, are not the objects of the beliefs reflected in these states, but they measure the degrees to which the agents assigning the states believe that the measurement outcomes will occur.

Since degrees of belief may differ from agent to agent without any of them necessarily making any kind of mistake, quantum Bayesianism does not encounter the same problems as the view that quantum states represent our knowledge of probabilities. It does, however, have a drawback in that it goes extremely far in characterizing elements of the quantum mechanical formalism as subjective in order to be consistent as an epistemic account of states. The most radical feature of quantum Bayesianism, arguably, is its denial of the fact that, for any given measurement setup, the question of *which* observable is measured in that setup might have a determinate answer. Fuchs argues as follows for this view:

[11] This is how Fuchs describes it, see (Fuchs, 2002, 7).

> Take, as an example, a device that supposedly performs a standard von Neumann measurement $\{\Pi_d\}$, the measurement of which is accompanied by the standard collapse postulate. Then when a click d is found, the posterior quantum state will be $\rho_d = \Pi_d$ regardless of the initial state ρ. If this state-change rule is an objective feature of the device or its interaction with the system—i. e., it has nothing to do with the observer's subjective judgement—then the final state must be an objective feature of the quantum system. (Fuchs, 2002, 39)

Fuchs' main point seems to be that if a set $\{\Pi_d\}$ of projection operators is objectively associated to a given experimental setup, registering a "click d" means that the state to be assigned after measurement is Π_d independently of which state ρ has been assigned to the system before. The post-measurement state seems to be fixed and we seem to have ended up with a *true* post-measurement state Π_d—a result which is incompatible with the epistemic conception of states. Therefore, it seems that the question of which observable is measured in which experimental setup can have no determinate answer according to the epistemic conception of states.

This conclusion, however, is extremely difficult to swallow. If sound, one could legitimately regard it as a *reductio* of the idea that the foundational problems of quantum theory can be dissolved by the epistemic conception of states. In actual quantum mechanical practice experimentalists agree almost always on which observable is measured by which device, and quantum mechanics could hardly be as empirically successful as it is if this were not the case. Furthermore, if measurement could never be regarded as measurement of a determinate observable in any given context, it would not make any sense to ask for any measured value to which observable it belongs. Knowledge of the values of observables would be excluded as a matter of principle, for one could never decide which observable some given value is a value of. Quantum mechanical practice, however, clearly seems to presuppose that we often do have knowledge of the values of at least some observables, and even if one adopts the radical position that microscopic observables (however one actually defines them) do never possess determinate values, this option is not available for the macroscopic systems to which we have more direct access but treat them quantum mechanically as many-particle systems by the methods of quantum statistical mechanics (e.g. when computing heat capacities, magnetic susceptibilities, and the like). At least approximate knowledge of macroscopic quantities, such as volume, particle number, temperature, pressure, and macroscopic magnetization is usually presupposed, and denying that such knowledge is possible seems not a promising option.

By claiming that the question of which observable is measured in which setup has no determinate answer quantum Bayesians consciously reject not only the notion of a quantum state a quantum system *is in*, but also the notion of a state assignment being performed *correctly*. While quantum Bayesians have successfully shown that the notion of a quantum state a quantum system is in can be avoided in a class of cases where it appears to be absolutely essential (in so-called "quantum state tomography"),[12] they have not been able to establish an analogous argument for the dispensability of the notion of a state assignment being performed correctly. This notion, however, seems to play an essential role in quantum mechanical practice, for instance in the case of systems being prepared by a (so-called) *state preparation device*, where any state assignment that deviates from a highly specific one is counted as wrong by all competent experimentalists. State preparation can be described as a form of measurement in that only systems exhibiting values of an observable lying within a certain interval are allowed to exit the device on the "prepared states" path, so accepting the notion of a state assignment being performed correctly is equivalent to allowing the question of which observable is measured in which setup to have a determinate answer. In the following section, I present an account which fleshes out the notion of a state assignment being performed correctly without invoking the notion of a state a quantum system is in.

4 Constitutive Rules and State Assignment

The most promising strategy to make sense of the notion of a state assignment being performed correctly without relying on the notion of a state a quantum system is in is to argue that to assign correctly means to assign in accordance with certain rules governing state assignment.[13] From the perspective of the epistemic conception of states one will have to think of these rules as determining the state an agent has to assign to the system depending on what she knows of the values of its observables. Examples of the rules according to which state assignment is

[12] See (Caves et al., 2002b).
[13] See (Friederich, 2011), sec. 4 and 5, for the slightly more detailed original version of the considerations presented in this section.

performed are unitary time-evolution in accordance with the Schrödinger equation (which applies whenever no new information about measurement data comes in), Lüders' rule (a generalized version of the von Neumann projection postulate) for updating one's assignment of a quantum state in the light of new data, and the principle of entropy maximization, which is used in contexts where a state should be assigned to a system where none was assigned before. To understand the peculiar status which is ascribed to these rules in the epistemic account of states proposed here, it is useful to compare their role in the present account to their role in the standard—ontic—conception of quantum states as descriptions of quantum systems. In this context, a terminological distinction proposed by John Searle (1969) in his theory of speech acts is very useful for clarifying the differing roles of the rules of state assignment in ontic accounts of quantum states and in the epistemic account of states proposed here.

Searle introduces the distinction between the two types of rules as follows:

> I want to clarify a distinction between two different sorts of rules, which I shall call *regulative* and *constitutive* rules. I am fairly confident about the distinction, but do not find it easy to clarify. As a start, we might say that regulative rules regulate antecedently or independently existing forms of behavior; for example, many rules of etiquette regulate inter-personal relationships which exist independently of the rules. But constitutive rules do not merely regulate, they create or define new forms of behavior. The rules of football or chess, for example, do not merely regulate playing football or chess, but as it were they create the very possibility of playing such games. The activities of playing football or chess are constituted by acting in accordance with (at least a large subset of) the appropriate rules. Regulative rules regulate pre-existing activity, an activity whose existence is logically independent of the rules. Constitutive rules constitute (and also regulate) an activity the existence of which is logically dependent on the rules. (Searle, 1969, 33f)

The standard ontic conception of quantum states as states quantum systems are in conceives of the rules governing state assignment as regulative rules. This can be seen by noting that whenever an agent assigns a quantum state to a quantum system what she aims at, according to the ontic conception of states, is to assign the state in which the system really is (or at least some reasonable approximation to it). This goal, however, can be specified without relying in any way on the rules which the agent follows in order to achieve it. Consequently, from the perspective of ontic accounts of quantum states, the notion of a state assignment being performed correctly is "logically independent of the rules"[14] according to which it is done. In these accounts, the role of the rules of state assignment is that of an instrument or a guide to determine the state the system really is in (or some reasonable approximation to it). From this perspective, state assignment can be characterized "antecedently [to] or independently" of the rules according to which it is performed. These rules are therefore conceived of as regulative rules in ontic accounts of quantum states.

In the epistemic account of quantum states proposed here, in contrast, the rules of state assignment play an entirely different role: They can be neither a guide nor an instrument for determining the state the system is in, for the notion of such a state is rejected. The basic idea, instead, is that to assign in accordance with the rules of state assignment is what it *means* to assign correctly, so the notion of a state assignment being performed correctly is itself *defined* in terms of these rules. It is therefore, as Searle writes, "logically dependent on the rules" according to which it is done, so these rules should be conceived of as *constitutive* rules in an epistemic account of states that preserves the notion of a state assignment being performed correctly without accepting the notion of a state a quantum system is in.

Having introduced the basic idea of the "Rule Perspective" as an epistemic account of quantum states that conceives of the rules of state assignment as constitutive rules, we can now come back to Fuchs' argument that there can be no determinate answer to the question of which observable is measured in which experimental setup. According to Fuchs, if the observable measured were an objective feature of the device, the measured result would impose objective constraints on the state to be assigned to the system after measurement, which he regards as in conflict with the basic idea of the epistemic conception of states that there is no agent-independent *true* state of the system. As I shall argue now, however, the epistemic account of states presented here is perfectly compatible with the view that the question of which observable is measured in an experimental setup has a determinate answer.

To see this, assume that the observable that is measured in a given experimental setup is an objective feature of the measuring device and that, in accordance

[14]Phrases within quotation marks in this and the following paragraphs are all taken from the passage from (Searle, 1969) just cited.

with the Rule Perspective, the rules of state assignment require an agent having performed a measurement with that setup to assign some specific quantum state to the measured system in order to assign correctly. Does it follow from these assumptions that the state she has to assign to the system after measurement must be regarded as the state it really is in, in conflict with the epistemic conception of states? Clearly not: All that follows is that those agents who have obtained information about the registered result must update the states they assign to the system in accordance with Lüders' rule, taking into account the registered result. With regard to the case considered by Fuchs this means that those knowing that the "click d" has been registered must update their states to Π_d after measurement in order to assign correctly. To arrive at the further conclusion that Π_d is the *true* state of the system one would have to demonstrate that assigning any other state than Π_d would amount to making a mistake, whatever one knows of the values of observables of the system. One would have to show, in other words, that assigning a state that is different from Π_d would be wrong not only for those who know that the "click d" has been registered, but also for those who don't.

This point is enforced by noting that there can be agents assigning states to the system who might not have had a chance to register the "click d." Registering it may have been physically impossible for them, for the process resulting in the "click d" may be situated completely outside their present backward light cone. If we adhere to the discipline that the states assigned by these agents reflect their epistemic relations to the system, it makes no sense to hold that they ought to assign the state Π_d as well because, given their epistemic relations to the system, the rules of state assignment advise them to assign differently. According to the epistemic conception of states, their state assignments have to be adequate to their epistemic condition with respect to the system, so they would not only not be obligated to assign Π_d, it would even be *wrong* for them.

In the following section I discuss what ramifications the perspective on the rules of state assignment as constitutive rules has for the interpretation of quantum probabilities.

5 The Interpretation of Quantum Probabilities

Having considered the status of the rules of state assignment in the epistemic account of quantum states, which I have called the "Rule Perspective," I now turn to the interpretation of probabilities derived from quantum states via the Born

Rule. While quantum Bayesianism describes them as subjective degrees of belief in accordance with the personalist Bayesian conception of probability, the Rule Perspective ascribes to them a more objective character which is arguably better in agreement with their actual role in quantum theoretical practice.

The most important sense in which quantum probabilities remain *subjective* in the Rule Perspective is that, for the same observable and the same quantum system, they may differ from agent to agent without one of them making a mistake. Inasmuch as one regards *any* quantities exhibiting an agent dependence as subjective one will therefore conclude that the Rule Perspective classifies quantum probabilities as subjective. However, quantum probabilities as conceived by the Rule Perspective can be seen objective in other, no less important regards, for instance in that the question of which probability *should be assigned* by an agent to the value of an observable is regarded as having a determinate, objective answer whenever the epistemic situation of that agent is sufficiently specified. Depending on whether a state assignment is performed correctly, the probabilities computed from the state via the Born Rule are either correct or incorrect in an objective way.[15]

Quantum Bayesianism stresses the non-descriptive, normative character of quantum theory by arguing that "[i]t is a users [sic] manual that any agent can pick up and use to help make wiser decisions in this world of inherent uncertainty" (Fuchs, 2010, 8) and by claiming that the Born Rule imposes norms on how to form our beliefs as regards "the potential consequences of our experimental interventions into nature" (Fuchs, 2002, 7). The Rule Perspective agrees, but it adds that quantum theory not only provides us with norms of how to "make wiser decisions," *given* the quantum states we have assigned to quantum systems, but also with norms of how to assign these states to the systems in the first place. The normative character which quantum theory has according to both quantum Bayesianism and the Rule Perspective is obscured by the formulation that quantum states are "states of belief,"[16] which is sometimes found in the writings of quantum Bayesians. This formulation is misleading because it invites the reading that quantum states, instead of being descriptions of physical *objects*, are descriptions of agents and their beliefs. This would imply that an assignment of a quantum state to a quantum system by an agent would be adequate if and only if the probabilities derived from that state corresponded

[15] See (Healey, forthcoming, sec. 2), for a more detailed investigation of in which sense quantum probabilities can be regarded as objective and agent-dependent at the same time.

[16] See (Fuchs, 2002, 7; Schack, 2003; Fuchs, 2010, 18).

exactly to the agent's degrees of belief, for only under this condition can the state be said to correctly describe the agent's system of beliefs. It is clear, however, that if quantum Bayesianism takes seriously its own characterization of quantum theory as a normative "manual" to "make wiser decisions," it need not regard quantum states as descriptions of anything, neither of the systems themselves nor of the assigning agents' degrees of belief. Much more naturally, it regards them as *prescriptions* for forming beliefs and for acting in the light of available information. According to this perspective, quantum states are not literally *states* of anything, neither of objects nor of agents.[17]

Having discussed the question in which sense quantum probabilities are subjective and in which sense they are objective, I now turn to the question of what quantum probabilities are probabilities *of*. To answer this question, the notion of a *non-quantum magnitude claim*—"NQMC" in what follows—is very useful, which has recently been introduced by Richard Healey in the context of his "pragmatist approach" (Healey, forthcoming) to quantum theory, which is in many respects similar in spirit to the Rule Perspective. An NQMC is a statement of the form "The value of observable A of system s lies in the set of possible values Δ." Healey refers to these statements as "non-quantum" since "NQMCs were frequently and correctly made before the development of quantum theory and continue to be made after its widespread acceptance, which is why [he calls] them non-quantum" (Healey, forthcoming, 25). It is perhaps possible to identify NQMCs with what, for Heisenberg, were descriptions in terms of "classical concepts,"[18] but Healey objects against this use of "classical" to avoid the misleading impression that an NQMC "carries with it the full content of classical physics." Another plausible reason for not calling NQMCs "classical" is that some of them use *non-classical* concepts such as spin. According to the Rule

[17]It seems likely that this observation can be used to answer a criticism brought forward against quantum Bayesianism by Timpson. According to this criticism, a quantum Bayesian is committed to the systematic endorsement of pragmatically paradoxical sentences of the form "I am certain that p (e. g., that the outcome will be spin-up in the z-direction) but it is not certain that p" (Timpson, 2008, 604), for instance when assigning a pure state, which necessarily ascribes probability 1 to at least one possible value of an observable. Timpson offers the first half of this sentence ("I am certain that p") as a quantum Bayesian translation of a state assignment ("I assign $|\uparrow_z\rangle$"), but the quantum Bayesian might reject this translation by claiming that the Born Rule probabilities derived from the state she assigns are not measures of her *actual* degrees of belief but rather prescriptions for what degrees of belief she *should* have, given certain presuppositions.

[18]See, for instance, (Heisenberg, 1958, 30).

Perspective, NQMCs are descriptively used in quantum mechanical practice, in contrast to quantum states.

As I shall argue now, the notion of an NQMC is useful in answering the question what the probabilities derived from the Born Rule are probabilities of. The most straightforward reading of the Born Rule

$$prob_{\rho}(A \in \Delta) = Tr(\rho \Pi^{A}_{\Delta}), \tag{1}$$

where Π^{A}_{Δ} denotes the projection on the span of eigenvectors of A with eigenvalues lying in Δ, is that it ascribes a probability to a statement of the form "The value of A lies in Δ," that is, to an NQMC. This reading, however, is in need of further qualification in that, as the famous no-go results due to Gleason, Bell, Kochen, Specker, and others suggest, not all NQMCs can simultaneously have determinate truth values. A common reaction to this problem is to restrict the interpretation of the Born Rule to measurement outcomes, saying that the probabilities derived from it are to be understood as conditional on measurement of A, and to deny that quantum theory has any empirical significance outside measurement contexts. However, as Healey notes,[19] this solution is unsatisfying not only from a hardcore realist but even from a more pragmatically-oriented point of view in that ascribing a probability to a NQMC can in some cases be legitimate with respect to situations where no measurements are performed at all. As an example of a situation where this is the case, consider a double-slit experimental setup, where electrons passing through a double-slit are coupled to a heat bath of scattering photons. In this case, different from that of electrons not coupled to photons, no wave-type interference pattern can be observed on a screen behind the double-slit and the probabilities for electrons on the screen can be computed from their probabilities passing through the individual slits. In that sense, one can treat the NQMCs "The electron is in the volume interval Δ_1" (implicating that it passes through the first slit) and "The electron is in the volume interval Δ_2" (implicating that it passes through the second) as having determinate truth values for each electron. One can thus consistently interpret the Born Rule as ascribing probabilities to these NQMCs even if it is not experimentally determined for the electron through which slit it actually passes.

[19] See (Healey, forthcoming, sec. 3). The reasoning given in the text is strongly based on the discussion of recent diffraction experiments presented there in great detail.

Generalising this observation, Healey proposes that quantum theory "license[s] claims about the real value only of a dynamical variable represented by an operator that is diagonal in a preferred basis" (Healey, forthcoming, 16). His proposal, in effect, is that ascribing a probability to an NQMC involving reference to an observable A is legitimate for an agent whenever the density operator ρ she assigns to the system is diagonal in a preferred Hilbert space basis, typically selected by environment-induced decoherence. In that case the *reduced* density operator assigned to the system by agents taking into account its coupling to the environment and performing the trace over the environmental degrees of freedom becomes (at least approximately) diagonal in an environment-selected basis.[20] As in the special case of the double-slit setup with photons mentioned before, it is unproblematic in these contexts to assume that NQMCs about observables having this basis in their spectral decomposition have determinate truth values. It is therefore natural for the Rule Perspective to hold that the Born Rule defines probabilities just for those NQMCs which refer to observables whose spectral decomposition makes the density matrix assigned by an agent uniquely diagonal. Arguably, the Rule Perspective should say that these NQMCs are what quantum probabilities are probabilities *of*.

According to this line of thought, whether an agent is entitled to treat an NQMC as having a determinate truth value depends on the state she assigns to the system and, therefore, on her epistemic situation. In contrast to this agent dependence, however, as Healey notes, "the content of an NQMC about a system s does not depend on agent situation ... [in that] ... it is independent of the physical as well as the epistemic state of any agent (human, conscious, or neither) that may make or evaluate it" (Healey, forthcoming, 25). Thus, while quantum theory itself is non-descriptive according to the Rule Perspective insofar as quantum states are not regarded as descriptions of physical states of affairs, it nevertheless functions as a way of *organizing* descriptive claims. It does so, first, by giving a criterion of under which conditions these claims can be treated as having determinate truth values and, second, by providing a method of computing probabilities for these claims to be true. In particular, it licenses NQMCs about macroscopic (pointer) observables for which the density matrices we assign to them when we take into account decoherence effects are typically (at least approximately) diagonal. If we want to use the quantum mechanical formalism correctly, we therefore have

[20] See (Schlosshauer, 2005) for a helpful introduction to decoherence and clarification of its relevance for foundational issues.

to treat statements about measurement devices as having determinate truth values and in that sense must assume that measurements do have outcomes. This line of thought may dispel the felt need of accounting for determinate measurement outcomes in terms of a dynamical explanation and in that sense completes the dissolution of the measurement problem.

6 Reality Presupposed

In the previous section quantum theory was characterized from the point of view of the Rule Perspective as a method of organizing descriptive "non-quantum" claims and attributing probabilities to them. This makes it clear that the Rule Perspective, far from denying that objective descriptions of physical states of affairs can be given, is based on the *presupposition* that objective descriptions exist in the form of NQMCs. The potential accusation against the Rule Perspective that it supposedly relies on an implausible and unattractive anti-realism or instrumentalism completely misses the mark: The Rule Perspective, as we see, cannot even coherently be formulated without making the realist assumption that physical states of affairs exist and that they can be described in terms of NQMCs.

If at all, the Rule Perspective is non-realist in the sense that it reads the most genuinely *quantum* conceptual resource—quantum states—as non-descriptive. As I have argued in (Friederich, 2011), however, this does not exclude an interpretation of the "structure and internal functioning" (Timpson, 2008, 582) of the quantum theoretical formalism as a whole to reflect objective features of physical reality itself. Furthermore, quantum theory as a method of organizing descriptive claims in the sense discussed in the previous section can still be seen as having been *discovered* rather than freely created or invented by the human mind [21] Conceiving of quantum states as tools of organizing descriptive claims depending on one's epistemic situation rather than as descriptions of physical states of affairs themselves does not mean to deny that physical states of affairs do exist. It just means that the conceptual resources needed to describe them must be non-quantum.

[21] See (Friederich, 2011, sec. 6) for more detailed considerations on this option.

References

Bub, J. (2007). Quantum probabilities as degrees of belief. *Studies in History and Philosophy of Modern Physics*, 38:232–254.

Caves, C. M., Fuchs, C. A., and Schack, R. (2002a). Quantum probabilities as Bayesian probabilities. *Physical Review A*, 65:022305.

Caves, C. M., Fuchs, C. A., and Schack, R. (2002b). Unknown quantum states: The quantum de Finetti representation. *Journal of Mathematical Physics*, 44:4537–4559.

Caves, C. M., Fuchs, C. A., and Schack, R. (2007). Subjective probability and quantum certainty. *Studies in History and Philosophy of Modern Physics*, 38:255–274.

Fleming, G. N. (1988). Lorentz invariant state reduction, and localization. In Fine, A. and Forbes, M., editors, *PSA 1988, Vol. 2*, pages 112–126. East Lansing, MI: Philosophy of Science Association.

Friederich, S. (2011). How to spell out the epistemic conception of quantum states. *Studies in History and Philosophy of Modern Physics*, 42:149–157. Preprint: arXiv:1101.1975.

Fuchs, C. A. (2002). Quantum mechanics as quantum information (and only a little more). In Khrennikov, A., editor, *Quantum Theory: Reconsideration of Foundations*, pages 463–543. Växjö University Press. arXiv:quant-ph/0205039.

Fuchs, C. A. (2010). QBism, the perimeter of quantum Bayesianism. arXiv:1003.5209.

Fuchs, C. A. and Peres, A. (2000). Quantum theory needs no 'interpretation'. *Physics Today*, 53:70–71.

Harrigan, N. and Spekkens, R. W. (2010). Einstein, incompleteness, and the epistemic view of quantum states. *Foundations of Physics*, 40:125–157.

Healey, R. (forthcoming). Quantum theory: a pragmatist approach. *British Journal for the Philosophy of Science*. Preprint: arXiv:1008.3896.

Heisenberg, W. (1958). *Physics and Philosophy: The Revolution in Modern Science*. George Allen & Unwin. repr. 2007 by Harper, New York, page numbers referring to this edition.

Marchildon, L. (2004). Why should we interpret quantum mechanics? *Foundations of Physics*, 34:1453–1466.

Maudlin, T. (2011). *Quantum Theory and Relativity Theory: Metaphysical Intimations of Modern Physics*. Wiley-Blackwell, 3rd edition.

Mermin, N. D. (2003). Copenhagen computation. *Studies in History and Philosophy of Modern Physics*, 34:511–522.

Myrvold, W. C. (2002). On peaceful coexistence: is the collapse postulate incompatible with relativity? *Studies in History and Philosophy of Modern Physics*, 33:435–466.

Pitowsky, I. (2003). Betting on the outcomes of measurements: A Bayesian theory of quantum probability? *Studies in History and Philosophy of Modern Physics*, 34:395–414.

Ruetsche, L. (2002). Interpreting quantum theories. In Machamer, P. and Silberstein, M., editors, *The Blackwell Guide to the Philosophy of Science*, pages 199–226. Blackwell.

Schack, R. (2003). Quantum theory from four of Hardy's axioms. *Foundations of Physics*, 33:1461–1468.

Schlosshauer, M. (2005). Decoherence, the measurement problem, and interpretations of quantum mechanics. *Reviews of Modern Physics*, 76:1267–1305.

Searle, J. (1969). *Speech Acts*. Cambridge University Press.

Spekkens, R. W. (2007). Evidence for the epistemic view of quantum states: A toy theory. *Physical Review A*, 75:032110.

Timpson, C. G. (2008). Quantum Bayesianism: A study. *Studies in History and Philosophy of Modern Physics*, 39:579–609.

Zbinden, H. (2001). Experimental test of nonlocal quantum correlations in relativistic configurations. *Physical Review A*, 63:022111.

Causal Realism in the Context of Bell-type Experiments

Matthias Egg

Abstract. *After introducing the main idea of causal realism and discussing one of the key motivations for this position, I will review an argument by Tim Maudlin to the effect that there is superluminal causation in Bell/EPR-type experiments. I will then compare the concepts of causation used by causal realism and by Maudlin, in particular with respect to the importance they attach to practical controllability or manipulability of the causes. In conclusion, I will sketch how the causal realist can react to the impact of the EPR case.*

1 Introduction: Why Causal Realism?

A central question in the debate on scientific realism concerns the validity of inferences to the best explanation (IBE). Most strands of realism, whether they admit it or not, rely on this form of inference in one way or another. Accordingly, IBE has been a major target of antirealist criticism. The most basic line of criticism is that IBE commits the fallacy of *affirming the consequent*: if x explains y and y is true, it does not follow that x is true. Realists have, roughly speaking, responded to this charge in two ways. The first one, which I will not discuss here, consists in claiming that the theoretical virtues which mark out one explanation as *the best* can serve as an argument for the truth of this explanation (Psillos, 1999, ch. 8). The second realist response focuses on a specific class of scientific explanations, namely causal explanations, and claims that the particular character of the causal relation allows us to infer the *explanans* from the *explanandum*. If we explain some observed phenomenon y by saying that it was brought about by a cause x, it is legitimate to conclude that x really occurred. This is the kind of scientific realism that I call *causal realism*.

But how are we to understand this relation of *bringing about* between x and y? More precisely: how is causality supposed to do the job the realist wants it

T. Sauer, A. Wüthrich (eds.), *New Vistas on Old Problems. Recent Approaches to the Foundations of Quantum Mechanics*, Edition Open Access, 2013.

to do? As a first approximation, one might think that if x causes y, then x is a necessary condition for y, so that, whenever we observe y, we can be sure that x occurred as well. The inadequacy of this proposal becomes obvious once we consider how causal reasoning works in actual science. If, for example, particle physicists want to test whether some process x really occurs, they typically try to detect the products of x. More precisely, they calculate what kind of signal the process x should produce in their detector, and then they look for this signal (let us denote it by y) in the experimental data. Now typically, the mere occurrence of y by no means establishes the reality of x, because there are usually some alternative ways in which a signal of type y could have been produced. This is what physicists call "background." A case for x will only be made if there is a part in the counting rate for y that cannot be attributed to background. It will then in general be false to say that whenever y occurs, x has occurred as well. But it will be true that at least in some cases (though we may not know in which ones), y would not have occurred if x had not occurred. The truth of this counterfactual statement for at least some tokens of the event type y seems to be an essential part of what it means for x to cause y (regardless of the difficulties that a general account of causation in terms of counterfactuals may face). And the truth of this counterfactual establishes realism with respect to x.

Having thus tied causal realism to the truth of certain counterfactual statements, I might seem to have blurred the distinction between causal and theoretical realism. For it is not unique to causal explanations that they support certain counterfactuals. The same is true for theoretical explanations; they contain laws, and an essential aspect of lawhood is the property of supporting counterfactuals. But these counterfactuals are not of the right sort to lend the same kind of support to theoretical realism that causal statements lend to causal realism. Here is why: laws support counterfactual claims concerning their instances. For example, Boyle's law supports claims like "if I reduced the volume of this gas at constant temperature, its pressure would increase." But what theoretical realism aspires to establish by means of IBE is not the truth of a singular statement but of the law itself. In order to achieve this, a claim of the following form would be needed: "If law L did not hold, phenomenon y would not occur." This is actually not just a counter*factual*, but a counter*legal* statement, and the mere fact that L explains y implies nothing about the truth of such a statement. What it implies is that L is part of a sufficient condition for the occurrence of y, in the sense that L, conjoined with some initial conditions, allows us to derive a statement describing y. But using L in a theoretical explanation involves no speculation about what would happen if L did not hold. By contrast, as we have seen above, citing an entity x in a causal explanation of y essentially involves a claim about what would (or would not) have happened had x not occurred. In other words: the counterfactual statements that give rise to causal realism are an integral part of causal explanations. As a consequence, causal explanations are more closely linked to realism than theoretical explanations, and this is one motivation for causal realism.

To end this introduction, I briefly mention two other motivations for causal realism. One is that causal realism seems to be a promising strategy against a recent objection to scientific realism, introduced by Kyle Stanford and known as the *problem of unconceived alternatives*.[1] The other is that causal realism can be profitably combined with ontic structural realism, spelling out the latter as a metaphysics of causal structures (Esfeld, 2009).

2 The Argument for Superluminal Causation

When we ask whether there is superluminal causation in Bell/EPR-type experiments, a discussion of what exactly we mean by causation seems inevitable. I will touch on one aspect of this discussion in the next section, but first I will review an argument in favor of superluminal causation, which claims to involve only the most uncontroversial application of the notion of causation. Furthermore, it claims to hold independently of which particular solution to the quantum measurement problem one happens to prefer. The argument is from chapter 5 of Tim Maudlin's book *Quantum Non-Locality and Relativity*, an updated third edition of which has just recently appeared (Maudlin, 2011).

Maudlin starts by specifying a sufficient condition for a causal implication between two events: "The local physical events A and B are causally implicated with one another if B would not have occurred had A not (or vice versa)" (Maudlin, 2011, 117). This condition bears some resemblance to the counterfactual claim discussed in Section 1, but there is an important difference in the focus of inquiry: in Section 1, we started from an observable event y and asked about the reality of its unobservable cause x. Here, A and B are both observable events (typically the outcomes of measurements) and the question is whether there is a causal link between them.

[1]Chakravartty (2008, sec. 4) in response to Stanford (2006).

Obviously, the fact that A and B are causally implicated with one another in this sense does not yet imply that either A caused B or vice versa. If two television sets are tuned to the same program, it is correct to say that a certain picture would not have appeared on the first screen, had it not appeared on the second one. But it is not the case that the appearance of the picture on one of the screens caused the appearance on the other. Instead, there is a common cause for the two events, namely the signal sent out by the broadcasting company.

In the context of EPR-experiments, it is very natural to think that the observed correlations are due to a common cause, since these experiments typically involve particles coming from a single source, detected at different locations. Since the particles do not travel faster than the speed of light, the event of their emission at the source lies in the backward light cones of both the detection events. Therefore, if the emission could serve as a common cause explanation for the correlations, there would be no need for superluminal causation. So in order to argue for superluminal causation, we do not only need to show that two space-like separated events A and B are causally implicated with one another, but also that the implication cannot be traced to an event situated in the backward light cones of A and B. This is captured by Maudlin's *sufficient condition for superluminal influences:*

(SI) [G]iven a pair of space-like separated events A and B, if A would not have occurred had B not occurred even though everything in A's past light cone was the same then there must be superluminal influences. (Maudlin, 2011, 118)

It is obvious from the context that by "influences" Maudlin hear means "causal influences." Notice that the claim is not that there is a direct causal influence from either A to B or B to A. The causal connection between A and B may be due to a common cause C, but the condition (SI) states that C must lie outside A's backward light cone. But this is to say that there is superluminal causation between C and A. So whether we opt for direct causation between A and B or for some common cause, in either case there is superluminal causation.

How do we evaluate a counterfactual statement like the one in (SI), in order to decide whether (SI) is actually fulfilled in the context of EPR-experiments? Maudlin's answer is that "if we have gotten the laws of nature right, then we can know about at least some unrealized possibilities. Given a set of laws we may be able to evaluate counterfactuals, and thereby to discern some causal connections" (Maudlin, 2011, 120). At this point, a contradiction with causal realism might seem to arise, since, as argued in the introduction, causal realism maintains that our knowledge of the laws of nature is significantly less secure than our knowledge of causes. But if the evaluation of a causal claim like (SI) depends on a knowledge of certain laws of nature, then it seems that, contrary to what the causal realist believes, laws are epistemically prior to causes. However, this seeming contradiction can be resolved by distinguishing *fundamental* from *phenomenological* laws. It is only the former that arouse the causal realist's suspicions, because their acceptance depends crucially on their explanatory virtues, and, as discussed above, L's explaining y does not imply L's truth. By contrast, phenomenological laws derive their support from the simple fact that they accurately describe what is observed in experiments. The causal realist can endorse laws of this type wholeheartedly, and nothing more is required here. Consider, for example, the first part of (SI), the claim that "A would not have occurred had B not occurred." No deep theory is needed to justify this claim. Once we accept that there is a systematic correlation between the measurement outcomes in the left and the right wings of an EPR experiment ("systematic" in the sense that it can be expressed by a phenomenological law), we may infer that in at least some cases, the left outcome would have been different, had the right outcome been different.[2]

A somewhat more detailed treatment is needed to assess the second part of (SI), namely the claim that even if we held fixed everything in the past light cone of A (or B), the correlation between A and B would remain intact. But even here, the evaluation of the counterfactual does not depend on any specific theory. It only has to take into account that the measurement process which gives rise to the events A and B can be either deterministic or (irreducibly) stochastic. The two cases require two different treatments, but the result will be the same.

In the *deterministic* case, the assumption that a common cause located in the intersection of the backward light cones of A and B is responsible for the correlation implies a Bell-type inequality. The experimentally well-confirmed violations of such inequalities rule out any local-deterministic common cause model

[2] Maudlin formulates his argument in terms of perfect correlations, and in this case it is always true that the left outcome would have been different had the right outcome been different. This is of course highly idealized. Although I do think that experiments can in principle provide warrant for even idealized phenomenological laws, I will not argue for this here. It seems to me that Maudlin's argument goes through even with imperfect correlations, as long as they are assured to be non-accidental (and only the most radical sceptic will doubt that this latter fact can be established experimentally). Nevertheless, I will below assume perfect correlations whenever this simplifies the argument.

for the EPR correlations. In other words, assuming a deterministic measurement process, the correlation between the events A and B cannot be attributed to any causal factor located in the intersection of their past light cones. Thus (SI) is essentially[3] satisfied in this case. That the same is true for *indeterministic* models can most easily be seen in the case of a perfect correlation between A and B. If the measurement process that leads to A is truly stochastic, it could have come out differently even if its complete backward light cone remained unchanged. But had A come out differently, so would B (and vice versa), as required by (SI).

Since (SI) holds for deterministic as well as indeterministic models, it follows that the existence of superluminal causation is established independently of any specific approach to the measurement problem. Maudlin concludes: “Reliable violations of Bell’s inequality need not allow superluminal signaling but they do require superluminal causation” (Maudlin, 2011, 141). This contrast brings us to the topic of the next section, namely signaling and how it relates to the concept of causation.

3 Causation, Manipulability, and Signaling

One reaction to the verdict of the previous section is to ask whether Maudlin has rigged the game by helping himself to too weak a notion of causation. If there is a causal relation between A and B, should it not at least in principle be possible to bring about a variation in B by manipulating A? And if so, should it not be possible to use that manipulation to send a signal from A to B, thereby violating some no-signaling theorem? Maudlin discusses this question in a section entitled “But is it causation?” (Maudlin, 2011, 135–141) There he argues that the exploitability or controllability of the causal relation should not be part of the concept of causation and that no-signaling should therefore not be taken to imply no-causing: “In general if one adds control of one variable to a counterfactual-supporting connection one gets signaling, but the addition is strictly irrelevant to the existence of the causal connection” (Maudlin, 2011, 137).

No matter whether or not one agrees with this characterization, one might ask at this point if the issue is relevant at all. It certainly is interesting to learn about these non-local dependencies, but does it make any difference whether we call

[3] There is a small argumentative gap here, because (SI) requires holding fixed the *entire* past light cone of A and not just the part that overlaps with the past light cone of B. See (Maudlin, 2011, 122) for an argument closing this gap.

them causal or not? Well, from the perspective of causal realism, it does make a difference which structures are causal and which are not, because this affects the decision on how far the realist commitment should extend. Furthermore, practical manipulability has played an important role in some of the arguments for causal realism. Consider, for example, Ian Hacking's famous pronouncement about electrons: "So far as I'm concerned, if you can spray them then they are real." (Hacking, 1983, 23) It thus seems worthwhile to consider in more detail how the lack of manipulability affects realism in the context of EPR experiments.

As we saw in the last section, the argument for superluminal causation (in Maudlin's sense) is independent of any specific choice of theoretical model and it is backed by strong experimental evidence. So far, the story is perfectly acceptable to the causal realist. But what exactly does this commit him to? To the reality of superluminal influences, of course, but what kind of influences and between which relata? Maudlin's argument can only be as general as it is by refusing to answer these questions. For illustration, let us look at two possible ways to account for EPR correlations.[4]

The most obvious option is to postulate a direct superluminal influence from A to B (or vice versa). Apart from being faster than light, such influences would be unusual in yet other ways: their strength does not seem to diminish when the distance between A and B is increased[5] and the influence is discriminating in that it only affects particles that have previously interacted with each other.

The second option tries to avoid such action at a distance by denying that there are two entities, one in each wing of the experiment, influencing each other across a space-like interval. Rather there is one single quantum structure that brings about events A and B. But of course we still have a superluminal influence: event A, for example, is caused by the whole, non-separable quantum structure which spans both wings of the experiment, so A is influenced by something which is not confined to its past light cone.

I will not enter into a discussion about which of these models is preferable or less objectionable. The point is that they are both compatible with what experiments tell us about EPR-like arrangements. The situation is similar to one that appears frequently in empirical research based on statistics: we observe a corre-

---

[4]For more examples, see (Suárez, 2007).

[5]A referee has helpfully pointed out that this is actually not so unusual; the strong nuclear force, as described by quantum chromodynamics, even *increases* with increasing distance between the interacting particles. This is true, but such behavior is restricted to subatomic distances. By contrast, EPR correlations have been shown to extend over distances of several kilometers.

lation between two variables, but we do not know whether this correlation is due to a direct causal influence from one variable to the other or to a common cause of the two variables. This analogy allows us to see why causal realism places such emphasis on the practical manipulability of alleged causes. For this is precisely what enables us in many cases to discern the causal structure that holds between the variables. Consider once again the example of the two TV sets from the previous section. The fact that we cannot, by manipulating the image on the first screen, influence what appears on the second screen, strongly favors the common cause hypothesis over the hypothesis that one of the images causes the other.

The controllability of a causal factor, therefore, has an important epistemic significance over and above the obvious fact that a controllable causal factor may open the way for technical applications. In the absence of controllability, we might have strong empirical evidence for the existence of some causal relations (as the argument in the previous section shows), but we do not have this type of evidence for claims about the precise causal structure of the situation.

4 Conclusion: Towards a Bell-informed Causal Realism

Although the result of the last section justifies causal realism's insistence on controllability/manipulability, it seems to have a rather devastating implication for this type of realism in the context of Bell-type experiments. If causal realism can only get off the ground when manipulation of the relevant factors is possible and if the no-signaling theorem prohibits such manipulation in the EPR cases, then it seems that causal realism is simply irrelevant in this context. I will now sketch two possible strategies to avoid this conclusion.

The first strategy starts from the observation that the no-signaling theorems are theoretical results, so they are only as well confirmed as the theories from which they are derived. Of course, standard quantum mechanics is extremely well confirmed, but by itself, it does not provide us with a satisfactory account of what happens in a measurement. Since this is precisely one of the ingredients of the EPR puzzle, it is at least possible that a theory which solves the measurement problem will yield predictions on signaling that differ from those of standard quantum mechanics. One might object that this is not a very plausible scenario, given that there are presently no empirical indications for a violation of no-signaling. But the fact that something has not been possible up to now does not imply that it should be impossible in principle. The history of science contains many examples of clever experimenters who gained access to domains that were previously thought to be inaccessible in principle.

But even if future inquiry should lead to progress along these lines, some ambiguities in the causal structure of EPR correlations will probably remain unresolved. In the previous section I merely discussed a very coarse-grained distinction between a common cause and a direct cause model, which, if we presuppose controllability of the measurement events, would be comparatively easy to resolve experimentally. However, each of these generic options is compatible with a number of more fine-grained models concerning the precise character of the causal influence, and experiments may fail to distinguish between these models. This point reflects a general dilemma for causal realism: on the one hand, the causal realist wishes to limit his commitments to claims for which there is direct experimental evidence, on the other hand, such evidence may not be available for detailed claims about the causes with which causal realism is concerned.

The second strategy for an improved causal realism seeks to deal with this dilemma in a constructive manner, not really seeing it as a dilemma at all, but as a hint on how to differentiate between different grades of commitment within one's realism. On this view, the two horns of the dilemma correspond to *two kinds of warrant* (called *causal* and *theoretical* by Suárez (2008) and Egg (forthcoming)), which can be ascribed to scientific claims. Claims that are causally warranted form the hard core of the realist's commitment, because they are as secure as any empirical claim can be. In particular, they can be defended against any kind of antirealist criticism (though not against radical skepticism, of course). The price to pay for this security is a lack of specificity. We saw an example of this in the discussion of superluminal influences: there is no reasonable doubt that there are such influences, in other words, we have causal warrant for their existence, but this commitment does not include any details about their precise nature. However, causal realism does not advocate complete agnosticism with regard to these details. In the absence of direct experimental evidence, causal realism can draw on the resources of standard scientific realism and evaluate the theoretical warrant for the different models that are compatible with experience, in order to arrive at a more detailed picture than what the evaluation of causal warrant by itself would yield.

The work of formulating such a detailed picture for the EPR experiments remains to be done. But whatever the precise outcome will be, the approach of causal realism will have the advantage of transparently delimiting the parts of the picture to which we should be strongly committed as opposed to the more speculative parts. The fascinating thing about EPR is that even the causally warranted core part of the picture contains the potential for a serious clash with common sense and special relativity.

References

Chakravartty, A. (2008). What you don't know can't hurt you: realism and the unconceived. *Philosophical Studies*, 137:149–158.

Egg, M. (forthcoming). Causal warrant for realism about particle physics. *Journal for General Philosophy of Science*.

Esfeld, M. (2009). The modal nature of structures in ontic structural realism. *International Studies in the Philosophy of Science*, 23(2):179–194.

Hacking, I. (1983). *Representing and Intervening*. Cambridge University Press, Cambridge.

Maudlin, T. (2011). *Quantum Non-Locality and Relativity*. Wiley-Blackwell, Chichester, 3rd edition.

Psillos, S. (1999). *Scientific Realism: How Science tracks Truth*. Routledge, London.

Stanford, P. K. (2006). *Exceeding Our Grasp: Science, History, and the Problem of Unconceived Alternatives*. Oxford University Press, New York.

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

Suárez, M. (2008). Experimental realism reconsidered: How inference to the most likely cause might be sound. In Hartmann, S., Hoefer, C., and Bovens, L., editors, *Nancy Cartwright's Philosophy of Science*, pages 137–163. Routledge, New York.

Locality, Causality, and Realism in the Derivation of Bell's Inequality

Adrian Wüthrich

Abstract. *For several years, a significant disagreement has persisted between (mainly) philosophers, and (mainly) physicists concerning the consequences of the empirical violation of Bell's inequality. On the one hand, it is claimed that empirical violations of Bell's inequality show us that the world is "non-local," which might give rise to a serious conflict with the special theory of relativity. On the other hand, several authors maintain that this conflict can be avoided by giving up a "realistic" interpretation of quantum mechanics. I use a recent derivation of a Bell-type inequality from Reichenbach's principle of the common cause to explicate the different notions of "non-locality" and "realism" involved in the debate, and to assess the adequacy of the different claims. My two main conclusions are: First, if "realism" is understood as the existence of sufficient conditions for the measurement outcomes in the experimental setups under consideration, giving it up will not allow us to maintain locality. Second, however, I will argue that there is, in fact, a plausible notion of "realism" which we could reject in order to save locality. Instead of challenging the special theory of relativity, this option impugns common notions of causality.*

1 From a Dilemma to a "Monolemma"

The empirical violation of Bell's inequality forces us to deny at least one of the premises from which it is derived.[1] There is, however, a significant disagreement as to what assumptions about the physical world or our theories of it are at stake.

On the one hand, people (mainly physicists) maintain that it is an open question whether we should admit that there are non-local interactions in the world, and

[1] The inequality is named after John Bell, who derived it in 1964. The volume by Cushing and McMullin (1989) includes some of the first influential discussions of the philosophical implications of the inequality.

T. Sauer, A. Wüthrich (eds.), *New Vistas on Old Problems. Recent Approaches to the Foundations of Quantum Mechanics*, Edition Open Access, 2013.

thus challenge the special theory of relativity,[2] or whether we should, rather, give up a certain notion of realism:

> The experimental violation of mathematical relations known as Bell's inequalities sounded the death knell of Einstein's idea of "local realism," in quantum mechanics. But which concept, locality or realism, is the problem? (Aspect, 2007, 866)

On the other hand, (mainly) philosophers urge that this stance is untenable because even if we dispense with our cherished notions of realism we have to accept non-local interactions in the world. Their argument is the following.

One can derive Bell's inequality from essentially the following two assumptions:

- The outcome of a measurement at one wing in an Einstein-Podolsky-Rosen (EPR) experiment[3] does not causally depend on the outcome or the measurement operations at the other wing.
- The outcomes of the measurements in an EPR experiment are co-determined by some events which are themselves causally independent of the measurement operations.

The first of these two assumptions comprises locality conditions, for which one can make strong supportive arguments based on the theory of special relativity. The second assumption postulates something akin to what Einstein called, in the EPR paper, "elements of reality." The existence or not of elements of reality, and the properties that they instantiate should be, moreover, independent from the actions of the observer, according to the second assumption.

Let us, in line with this "zeroth" characterization, abbreviate the two set of assumptions by "Locality_0" and "Realism_0," respectively. Then, the derivation of Bell's inequality can be represented as the following entailment. Let us call it, for the purposes of the present discussion, "Bell's theorem":

$$(\text{Locality}_0 \ \& \ \text{Realism}_0) \vDash \text{Bell's Inequality}.$$

Because of the empirical violations of the consequence of this entailment, i. e., of Bell's inequality, at least one of the two sets of premises has to be false. If the

[2] See (Maudlin, 2002) for a detailed discussion of the extent to which non-local interactions are incompatible with the theory of special relativity.

[3] (Einstein et al., 1935); see also (Bohm, 1951).

derivation is *minimal* in the sense that dropping one of the premisses will lead to an invalid argument, then the argument gives no reason to give up more than one of the two problematic sets of premises.

However, according to several authors, the derivation from the above two premises is not minimal, because, according to them, Realism is a necessary condition for Locality, i. e.

$$\text{Locality}_0 \vDash \text{Realism}_0.$$

They take it that Einstein, Podolsky, and Rosen have established that claim:

> Many believe that because Bell starts by assuming the world conforms to what is called local realism, he therefore proved that *either* locality or realism is violated. Thus, the world could be local if it violates "realism." But this idea overlooks—or misunderstands—that the original "EPR" argument of Albert Einstein, Boris Podolsky and Nathan Rosen rules out the possibility of quantum locality without the realism Bell uses. (Albert and Galchen, 2009, 31)

If Albert and Galchen are right and Realism is, indeed, a necessary condition for Locality, "Bell's theorem" gives rise to a derivation of Bell's inequality from the assumption of Locality alone:

$$\text{Locality}_0 \vDash \text{Bell's Inequality.}$$

This is also the conclusion that Maudlin draws:

> So experiments verifying the violation of Bell's inequality would doom locality *tout court*. (Maudlin, 2002, 20)

Thus, according to Albert, Galchen, and Maudlin, we cannot avoid the challenges, which a violation of Locality most likely poses to the theory of special relativity, by giving-up Realism alone.

2 Details of the Derivation of Bell's Inequality

In order to assess the adequacy of the divergent claims by, on the one hand, the physicist Aspect and, on the other hand, Albert, Galchen, and Maudlin, I will sketch, in the following, a more detailed and explicit derivation of a Bell-type inequality, which is based on work by Graßhoff et al. (2005). This derivation is minimal with respect to an important class of other derivations. For example, the authors do not require different correlations to have the same common cause. I will subdivide the derivation into two steps so as to be able to compare it more readily with Maudlin's, and Albert and Galchen's arguments to the effect that Bell's inequality derives from locality conditions alone.

I will set aside some auxiliary assumptions. This is, of course, always a pragmatic instead of a principled decision and depends on where one wants to put the focus of the investigation. Most of the auxiliary assumptions are made explicit by Graßhoff et al. (2005), but even there the authors did, for instance, not address, in any detailed manner, the question as to how to interpret the probabilities which are used in the derivation. I will bracket out this question here, too.

The first step of the derivation proceeds from three principal assumptions to an intermediate conclusion. The assumptions involve perfect correlations, causal independence of correlated outcomes, and a variant of Reichenbach's (1956) principle of the common cause:

PCORR: Upon parallel measurements, the outcomes in an EPR experiment are perfectly anti-correlated.

C-OI: The measurement outcomes at one wing are not causally relevant for the outcomes at the other wing.

SCR: If two types of events A and B are correlated and neither A is causally relevant for B nor vice-versa then there exists a third type of event C which statistically "screens-off" A from B:

$$p(AB|C) = p(A|C)p(B|C).$$

From these three assumptions follows (given the auxiliary assumptions characterizing the experimental setup and a suitable interpretation of the probabilities) that there are sufficient conditions for the correlated effects.[4] More precisely, for instance, setting up both the apparatus on the left and on the right to measure the spin of the particle relative to direction no. 1, L_1R_1, together with a certain type of screening-off event, C_{11}^{+-}, is a sufficient condition for the outcome of the mea-

[4]See, e.g., (Graßhoff et al., 2005) for details.

surement at the left wing being $+$, and also for the outcome at the right wing being $-$. In a similar way, we arrive at the following four conditional statements:[5]

$$L_1 R_1 C_{11}^{+-} \rightarrow L_1^+, \qquad L_2 R_2 \neg C_{22}^{+-} \rightarrow R_2^+,$$
$$L_2 R_2 C_{22}^{+-} \rightarrow L_2^+, \qquad L_3 R_3 \neg C_{33}^{+-} \rightarrow R_3^+.$$

Since we assume that

C-PI: the measurement operations on the particle at the distant wing of the experimental setup are not causally relevant to the outcome of measurements on the particle in question,

the above conditionals should be true even without the statements, in the antecedent, that the measurement apparatus at the distant wing is set up in a particular way. That is, for instance, we can discard L_2 from the sufficient condition for R_2^+, or R_1 from the sufficient condition for L_1^+. This manipulation of the antecedent, is an application of the regularity theory of causality as developed by Mackie (1974), and elaborated by, e.g., Baumgartner (2008), and Graßhoff and May (2001). According to these theories of causality, if, for instance,

$$L_1 R_1 C_{11}^{+-} \rightarrow L_1^+$$

would cease to be valid when L_1 is discarded, L_1 would, in fact, be causally relevant for L_1^+ given certain additional conditions.[6]

Although I here performed the transition from the conditionals which contain both measurement settings to the conditionals which contain only one of them using a theory of causal regularities, the essential assumption involved in the transition is again screening-off. For, under suitable interpretations of the probabilities,

[5] By the symbols L_1, L_2, and L_3, I denote, respectively, the statements that the measurement apparatus in the left wing of the experiment is set to measure the spin in the first, second, or third direction. L_1^+, L_2^+, and L_3^+ denote that the outcome of a spin measurement in the left wing of the experiment in the first, second, or third direction has the outcome "spin up." R_1, R_1^+ etc. denote the corresponding statements for the right wing of the experiment. C_{11}^{+-}, C_{22}^{+-}, and $\neg C_{33}^{+-}$ denote, respectively, the statements that the screening-off event for the correlation between L_1^+ and R_1^-, L_2^+ and R_2^-, or L_3^+ and R_3^- are instantiated in a given run of the experiment. $\neg C_{22}^{+-}$ and $\neg C_{33}^{+-}$ mean, respectively, that C_{22}^{+-} or C_{33}^{+-} is not instantiated.

[6] For details, see, e.g., (Baumgartner, 2008, 342).

the above manipulation of the conditionals is tantamount to reading C-PI as the claim that, e.g., $L_1C_{11}^{+-}$ screens off L_1^+ from R_1, i. e.,

$$p(L_1^+|R_1L_1C_{11}^{+-}) = p(L_1^+|L_1C_{11}^{+-})$$

or, equivalently,

$$p(L_1^+R_1|L_1C_{11}^{+-}) = p(L_1^+|L_1C_{11}^{+-})p(R_1|L_1C_{11}^{+-}).$$

Thus, if we do not take the distant measurement operations to be causally relevant for the nearby outcome and assume an interpretation of causal irrelevance which either implicitly or explicitly is a screening-off condition, we end up with the following sufficient conditions for the outcomes:

SUFF:

$$L_1C_{11}^{+-} \to L_1^+, \qquad R_2\neg C_{22}^{+-} \to R_2^+,$$
$$L_2C_{22}^{+-} \to L_2^+, \qquad R_3\neg C_{33}^{+-} \to R_3^+.$$

We can graphically represent these relations by way of *causal graphs* (see Figure 1).[7]

From these conditionals, we can derive the relative frequencies of the outcomes in terms of the relative frequencies of the screener-offs and the relative frequencies of the measurement operations. If we add the condition that the screener-offs are statistically independent from the measurement operations, we can even express the frequencies of the measurement outcomes in terms of only the screener-offs. That is, if we assume

NO-CONS: The conjunction of the measurement operations and the conjunction of the screener-offs are statistically independent from each other:[8]

$$p(C_{11}^{+-}\neg C_{22}^{+-}|L_1R_2) = p(C_{11}^{+-}\neg C_{22}^{+-}), \text{ etc.,}$$

[7] Alternative causes for the outcomes are not represented and not discussed here.

[8] I use $p(X|Y)$ to denote the conditional probability of an instantiation of events of type X, given that an event of type Y is instantiated.

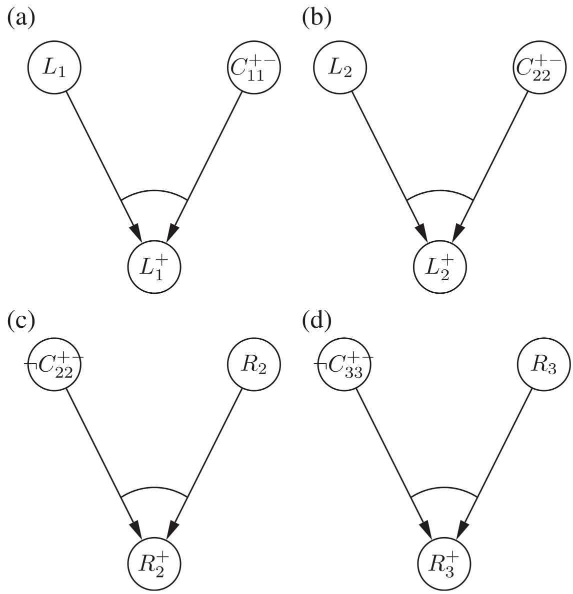

Figure 1: Four graphical representations of complex sufficient conditions, following Graßhoff et al. (2005). Cf. causal graphs in the sense of (Baumgartner, 2006, 73–79).

we have

$$p(L_1^+ R_2^+ | L_1 R_2) = p(C_{11}^{+-} \neg C_{22}^{+-})$$
$$= p(C_{11}^{+-} \neg C_{22}^{+-} C_{33}^{+-}) + p(C_{11}^{+-} \neg C_{22}^{+-} \neg C_{33}^{+-}),$$

$$p(L_2^+ R_3^+ | L_2 R_3) = p(C_{22}^{+-} \neg C_{33}^{+-})$$
$$= p(C_{11}^{+-} C_{22}^{+-} \neg C_{33}^{+-}) + p(\neg C_{11}^{+-} C_{22}^{+-} \neg C_{33}^{+-}), \text{ and}$$

$$p(L_1^+ R_3^+ | L_1 R_3) = p(C_{11}^{+-} \neg C_{33}^{+-})$$
$$= p(C_{11}^{+-} C_{22}^{+-} \neg C_{33}^{+-}) + p(C_{11}^{+-} \neg C_{22}^{+-} \neg C_{33}^{+-}).$$

Because the right-hand side of the last equation contains only terms which also occur in the right-hand side of the previous two equations, the relative frequency which is expressed by the last equation cannot be larger than the sum of the other two frequencies, i. e.,

BELL:
$$p(L_1^+ R_3^+ | L_1 R_3) \leq p(L_1^+ R_2^+ | L_1 R_2) + p(L_2^+ R_3^+ | L_2 R_3).$$

This is a version of Bell's inequality, which is violated significantly in the experiments which measure the relative frequencies to which the inequality refers.[9] Therefore, we must conclude that at least one of the assumptions from which the inequality has been derived must be false.

3 Locality and Reality Conditions

According to my sketch of the detailed derivation of Bell's inequality, we need the assumptions of (i) perfect correlation, causal independence of the outcomes on each other (ii) and on the distant measurement operations (iii), (iv) the principle of the common cause, and (v) the statistical independence of the common causes on the measurement operations. That is, the following entailment holds:

PCORR & C-OI & C-PI & SCR & NO-CONS ⊨ BELL

[9] In this form, the inequality is discussed, e.g., by van Fraassen (1982), and has probably first been derived by Wigner (1970). The first significant empirical violation of Bell's inequality has been observed by Aspect et al. (1982).

We performed the derivation in two steps, the first being the derivation of sufficient conditions for the measurement outcomes (SUFF), from the perfect correlations between causally independent outcomes (PCORR, C-OI), using a variant of the principle of the common cause (SCR):

$$\text{PCORR \& C-OI \& C-PI \& SCR} \vDash \text{SUFF}$$

Using the intermediate conclusion SUFF, we proceeded, in a second step, to the derivation of Bell's inequality (BELL) from the causal independence of the outcomes on the distant measurement operations (C-PI), and the statistical independence of the screener-offs and the measurements operations:

$$\text{C-PI \& NO-CONS \& SUFF} \vDash \text{BELL}$$

This entailment shows us that we cannot maintain SUFF without giving up at least one of C-PI or NO-CONS. If SUFF, i. e. the existence of sufficient conditions for the outcomes, is our notion of "realism" any "realistic" account of the quantum correlations in question will have to feature a causal dependence of the outcomes on the distant measurement operations or a statistical dependence between the measurement operations and the "elements of reality," understood as the screener-offs of the perfect correlations:

$$\text{SUFF} \vDash (\neg\text{C-PI} \vee \neg\text{NO-CONS})$$

However, the first step of our derivation shows us that even if we are willing to give up "realism" in the sense of sufficient conditions for the outcomes, we still have to give up either C-OI, C-PI or SCR, or deny that the empirical evidence lends itself to the description as perfect correlations of the outcomes with parallel measurement settings.

$$\neg\text{SUFF} \vDash (\neg\text{C-OI} \vee \neg\text{C-PI} \vee \neg\text{SCR} \vee \neg\text{PCORR})$$

Therefore, giving up realism (understood as SUFF) does not help us as much as it might seem at first sight in avoiding the possible conflict between quantum mechanical phenomena, such as the EPR experiment, and the special theory of relativity. When considering the advantages of giving up "realism," we should not only take into account the costs of maintaining it but also the costs that we nonetheless have to carry if we do give it up.

Understood along these lines, Maudlin's and Albert's and Galchen's critical remarks are to the point. However, the way they phrase the critique has several unattractive consequences.

If they maintain that the derivation of Bell's inequality is a reductio argument against locality conditions alone, they must subsume all of C-PI, NO-CONS, C-OI, SCR, and PCORR under their notion of locality. Also, if they regard the first step of the derivation as a variant of EPR's argument, which derives Realism from Locality, and the second step, "Bell's theorem" in their terms, as a derivation of Bell's inequality from Locality and Realism, the notion of locality in the first and the second step of the derivation is not necessarily the same. In the first step, "Locality" must entail the causal independence of the measurement outcomes and the version of the principle of the common cause:[10]

$$\text{LOC}_{\text{EPR}} \vDash \text{C-OI \& C-PI \& SCR}$$

In the second step, "Locality" must entail the causal independence of the outcomes on the distant measurement operations and the statistical independence of the measurement operations and the screener-offs:

$$\text{LOC}_{\text{BELL}} \vDash \text{C-PI \& NO-CONS.}$$

In both steps, they must identify "Realism" with SUFF:

$$\text{Realism} = \text{SUFF.}$$

These consequences are unattractive mainly because assumptions such as NO-CONS or SCR have to be classified as "locality" conditions. Yet, NO-CONS, to begin with, is a condition relating event types which are instantiated by *time-like* separated events, whereas I take the usual referents of locality conditions to be event types which are instantiated by *space-like* separated events.[11]

[10] I do not include PCORR in LOC_{EPR} because it reflects idealized empirical data and is not meant by anyone to qualify as a locality or reality condition. For the present purposes, it is best seen as an auxiliary assumption.

[11] Space-like separation of two events means that they happen far apart from each other and very soon one after the other or even simultaneously; in fact, the spatial distance divided by the temporal distance is larger than the speed of light, for space-like separated events. For time-like separated events, the spatial distance divided by the temporal distance is smaller than the speed of light. According to the theory of special relativity, the temporal order of space-like separated events depends on the frame of reference. Therefore, one is usually reluctant to admit causal dependence between two space-like separated events; in some frames of reference, the cause would happen after its effect.

Also SCR is not a particular statement about space-like separated events. Rather, it applies to any two types of events and says that there cannot be a correlation between them when they do not cause each other or when there is no screener-off of the correlation.

Of course, from a logical point of view, one can define whatever one pleases as long as the definitions are consistent. But in order to relate the discussion about the consequences of the empirical violation of Bell's inequality to the more general concepts of "locality" and "realism" we are well advised to link the premises of the derivation as closely as we possibly can to the meaning these terms have in their typical use. And on this measure, there is a much better alternative than the broad notion of locality which includes virtually any premise of Bell's inequality.

The alternative is to identify NO-CONS and SCR with "Realism," i. e.,

$$\text{Realism} = \text{NO-CONS \& SCR},$$

and C-OI and C-PI with "Locality":

$$\text{Locality} = \text{C-OI \& C-PI}.$$

Like this, on the one hand, the locality conditions state particular cases of the prohibition that an event type be causally relevant for another event type when their coinciding instances are space-like separated. On the other hand, the realism conditions spell out the intuition "that there has to be something" accounting for our observations, and that this "something" is there, independent of our eventual measurement operations.

4 Thinking About Local, "Non-Realistic" Models

The previous analysis opens up the possibility, *contra* Maudlin, Albert, and Galchen, of maintaining locality and giving up realism instead. Local, "non-realistic" models would satisfy both C-PI and C-OI, but not both NO-CONS and SCR.

Models which violate only NO-CONS will have to address delicate, though perhaps not unresolvable, issues involving backward in time causation,[12] or a hitherto unknown class of causal relations between the measurement operations and the events which are associated with the production of quantum mechanical singlet states.

[12] See, e.g., (Price, 1996).

Models which violate SCR are the other local but non-realistic option. These models will have to allow correlations between events which do not cause each other and which do not have a screener-off. However, whether such correlations can exist, is a question which has to do with our notion of *causality*, and the answer to this question is logically independent from the requirement that space-like separated events do not cause each other.

Acknowledgments

Research for this article was supported by the Swiss National Science Foundation under the grant no. 105211-129730 (principal investigator Gerd Graßhoff). A generous postdoctoral fellowship (Sept. 2011 – Apr. 2012) from the University of Pittsburgh's Center for Philosophy of Science gave me a particularly encouraging opportunity to write up these results.

References

Albert, D. Z. and Galchen, R. (2009). A Quantum Threat to Special Relativity. *Scientific American*, 3:26–33.

Aspect, A. (2007). To Be or Not to Be Local. *Nature*, 446:866–867.

Aspect, A., Dalibard, J., and Roger, G. (1982). Experimental Test of Bell's Inequalities Using Time-Varying Analyzers. *Physical Review Letters*, 49(25):1804–1807.

Baumgartner, M. (2006). *Complex Causal Structures. Extensions of a Regularity Theory of Causation*. PhD thesis, University of Bern.

Baumgartner, M. (2008). Regularity Theories Reassessed. *Philosophia*, 36:327–354.

Bell, J. (1964). On the Einstein-Podolsky-Rosen paradox. *Physics*, 1:195–200.

Bohm, D. (1951). *Quantum Theory*. Prentice Hall, New York.

Cushing, J. and McMullin, E. (1989). *Philosophical Consequences of Quantum Theory: Reflections on Bell's Theorem*. University of Notre Dame Press.

Einstein, A., Podolsky, B., and Rosen, N. (1935). Can Quantum-Mechanical Description of Physical Reality Be Considered Complete? *Physical Review*, 47(10):777–780.

Graßhoff, G. and May, M. (2001). Causal Regularities. In Spohn, W., Ledwig, M., and Esfeld, M., editors, *Current Issues in Causation*, pages 85–114. Mentis.

Graßhoff, G., Portmann, S., and Wüthrich, A. (2005). Minimal Assumption Derivation of a Bell-type Inequality. *The British Journal for the Philosophy of Science*, 56(4):663–680.

Mackie, J. L. (1974). *The Cement of the Universe: A Study of Causation*. Clarendon Press, Oxford.

Maudlin, T. (2002). *Quantum Non-Locality and Relativity: Metaphysical Intimations of Modern Physics*. Wiley-Blackwell, 2nd edition.

Price, H. (1996). *Time's Arrow and Archimedes' Point: New Directions for the Physics of Time*. Oxford University Press.

Reichenbach, H. (1956). *The Direction of Time*. Dover Publications.

van Fraassen, B. C. (1982). The Charybdis of Realism: Epistemological Implications of Bell's Inequality. *Synthese*, 52(1):25–38. Reprinted in revised form in Cushing and McMullin (1989).

Wigner, E. P. (1970). On Hidden Variables and Quantum Mechanical Probabilities. *American Journal of Physics*, 38(8):1005–1009.

Entanglement as an Element-of-Reality

Philip Walther

Abstract. *Entanglement—according to Schrödinger (1935) the essential property of quantum mechanics—teaches us that the properties of individual quantum systems cannot be considered to be (local) elements of physical reality before and independent of observation. Yet it is a widespread point of view that the way the observations on, say, two particles are correlated, i.e. the specific type of their entanglement, can still be considered as a property of the physical world. Here I discuss a previous experiment (Walther et al., 2006) showing that this is explicitly not the case. The correlations between a single particle property, the polarization state of a photon, and a joint property of two particles, the entangled state of a photon pair in a three-photon entangled state, have been measured. It is shown that the correlations between these properties can obey a cosine relation in direct analogy with the polarization correlations in one of the triplet Bell states (Bell, 1964). The cosine correlations between the polarization and entangled state measurements are too strong for any local-realistic explanation and are experimentally exploited to violate a Clauser-Horne-Shimony-Holt (CHSH) Bell inequality (Bell, 1964; Clauser et al., 1969). Thus, entanglement itself can be an entangled property leading to the notion of entangled entanglement.*

1 Introduction

In general, quantum mechanics only makes probabilistic predictions for individual events. Can one go beyond quantum mechanics in this respect? More than seventy years ago, in 1935, Einstein, Podolsky and Rosen (EPR) argued that quantum theory could not possibly be complete (Einstein et al., 1935). They showed that one could infer perfectly complementary properties, like position and momentum of an individual particle, by performing a corresponding measurement on the distant particle that is quantum-mechanically entangled with the first one. Based firmly on plausible assumptions about locality, realism, and theoretical

T. Sauer, A. Wüthrich (eds.), *New Vistas on Old Problems. Recent Approaches to the Foundations of Quantum Mechanics*, Edition Open Access, 2013.

completeness, they further argued that quantum states cannot be a complete description of physical reality, but rather give only a statistical one of an ensemble of intrinsically different quantum systems. While at the time, Bohr (1935) famously argued against EPRís conclusions, in particular against their notion of "reality" as assuming the systems have intrinsic properties independently of whether they are observed or not, it was not until almost 30 years later that the EPR program could be formulated in terms of an experimentally-testable prediction. I am, of course, referring to the landmark discovery of John Bell (1964) that EPRís premises of locality and realism put measurable limits on the strength of correlations between outcomes of remote measurements on a pair of systems. These limits are known as Bell inequalities and quantum mechanics does not satisfy them.

Since Bell's initial discovery, a large volume of theoretical and experimental work has been devoted to this subject. Experimental violations of Bell inequalities have been demonstrated using pairs of polarization-entangled photons (Freedman and Clauser, 1972; Fry and Thompson, 1976; Aspect et al., 1982; Ou and Mandel, 1988; Shih and Alley, 1988; Weihs et al., 1998), even under strict Einstein locality requirement, using other photonic degrees of freedom such as energy-time (Tapster et al., 1994; Tittel et al., 1998) and angular momentum (Vaziri et al., 2002), trapped ions (Rowe et al., 2001), and even neutron systems (Hasegawa et al., 2004). Multiphoton entanglement experiments have been performed demonstrating all-versus-nothing arguments against local realism (Pan et al., 2000) by exploiting so-called Greenberger-Horne-Zeilinger (GHZ) states (Greenberger et al., 1989), where single measurement outcomes can be incompatible with local realistic models. Aside from outstanding loopholes, which have not all been closed simultaneously in a single experiment (Weihs et al., 1998; Rowe et al., 2001), these experiments all but rule out the possibility of local realistic theories. However, common to all previous Bell experiments, regardless of the implementation, is that the measured degrees of freedom corresponded to properties of individual systems. Entanglement itself, as a property of the composite systems, was usually considered an objective property.

The experiment discussed in the following, however, demonstrated the first example of a Bell-inequality violation where an entangled state itself qualifies as an EPR element of reality. Specifically, a measurement of the single particle at Alice's side defines the relational property between the two other particles, without defining their single-particle properties. Therefore, only the joint state of the two qubits at Bob's side is an element of reality. The correlations between the polarization state of one photon and the entangled state of another two are experimentally demonstrated to violate the Bell inequality. This shows that entanglement itself can be entangled. The notion that entanglement itself can be an entangled property was originally proposed in the context of (Zeilinger et al., 1992; Krenn and Zeilinger, 1996).

2 An Experiment on Entangled Entanglement

In Figure 1, a schematic for the experiment is shown in which three photons are prepared in an entangled state

$$|\Phi^-\rangle_{1,2,3} = \frac{1}{\sqrt{2}}\left(|H\rangle_1 |\phi^-\rangle_{2,3} - |V\rangle_1 |\psi^+\rangle_{2,3}\right), \quad (1)$$

where the subscripts label different photons, the kets $|H\rangle_1$ and $|V\rangle_2$ represent states of horizontal and vertical polarization, respectively, of photon 1 and $|\psi^+\rangle_{2,3} = 1/\sqrt{2}\left(|H\rangle_2 |V\rangle_3 + |V\rangle_2 |H\rangle_3\right)$ and $|\phi^-\rangle_{1,2} = 1/\sqrt{2}$ $\left(|H\rangle_1 |H\rangle_2 + |V\rangle_1 |V\rangle_2\right)$ represent two (out of four possible) so-called Bell-states (maximally entangled states) of photons 2 and 3. Since the entangled state of photons 2 and 3 is entangled with the polarization state of photon 1, the state in Eq. (1) can be referred to as entangled entanglement. Photon 1 is moving freely in one direction to Alice, while the photons 2 and 3 are moving into the opposite direction to Bob. Aliceís photon 1 is now subjected to a polarization measurement along the axis θ_1. For simplicity, the settings are restricted to the linear polarization measurement, i.e., θ_1 lies within the x-y plane of the Poincaré sphere. If the polarization is found to be parallel to the axis θ_1 (outcome +1), the photon will be projected onto the state $|H'\rangle_1 = \cos\theta_1 |H\rangle_1 + \sin\theta_1 |V\rangle_1$, or when to be found perpendicular (outcome -1), it will be projected onto the state $|V'\rangle_1 = -\sin\theta_1 |H\rangle_1 + \cos\theta_1 |V\rangle_1$. Photons 2 and 3 at Bobís side are subjected to a specific joint measurement that can also only result in two different outcomes. In relation to the experiment, photons 2 and 3 are labelled as B and D, respectively, due to being emitted into the spatial mode B and D (Figure 2). Bob's measurement setting is denoted by the angle θ_2. The measurement will project the two photons onto either the state $|\phi^{-'}\rangle_{B,D} = \cos\theta_2 |\phi^-\rangle_{B,D} + \sin\theta_2 |\psi^+\rangle_{B,D}$ (outcome +1) or $|\psi^{+'}\rangle_{B,D} = -\sin\theta_2 |\phi^-\rangle_{B,D} + \cos\theta_2 |\psi^+\rangle_{B,D}$ (outcome -1). The outcome +1 will be identified by joint registration of photons 2 & 3 at the pairs of detectors, (1 and 2) or (3 and 4), while the outcome -1 will be identified by firing of pairs of detectors (1 and 3) or (2 and 4). When Alice and Bob choose the orientations θ_1 and θ_2 of their measurement apparatuses the initial state transforms to

$$|\Phi^-\rangle_{1,B,D} = \cos(\theta_1 + \theta_2)\frac{1}{\sqrt{2}}\left(|H\rangle_1 |\phi^-\rangle_{B,D} - |V\rangle_1 |\psi^+\rangle_{B,D}\right) + \sin(\theta_1 + \theta_2)\frac{1}{\sqrt{2}}\left(|V\rangle_1 |\phi^-\rangle_{B,D} - |H\rangle_1 |\psi^+\rangle_{B,D}\right). \qquad (2)$$

The quantum state in Eq. (1) has the remarkable property that it is the same for any choice of local settings θ_1 and θ_2 such that $\theta_1 = -\theta_2$, i.e., it is invariant under this set of locally unitary transformations. This entails *perfect correlations*: if polarization along θ_1 is found to be +1 (-1) for photon 1, then with certainty the result of the measurement for setting θ_2 will be found to be +1 (-1) for photons 2 and 3, and vice versa. Because of the perfect correlations, the result of measuring any entangled state $\cos\theta_2 |\phi^-\rangle_{B,D} + \sin\theta_2 |\psi^+\rangle_{B,D}$ or $-\sin\theta_2 |\phi^-\rangle_{B,D} + \cos\theta_2 |\psi^+\rangle_{B,D}$ can be predicted with certainty by previously choosing to measure the polarization of photon 1 along the axis $\theta_1 = -\theta_2$. By locality (in EPR's words):

> Since at the time of measurement the two systems no longer interact, no real change can take place in the second system in consequence of anything that may be done to the first system,

the measurement performed on photon 1 (photons 2 and 3) can cause no real change in photons 2 and 3 (photon 1). Thus, by the premise about reality (in EPR's words):

> If, without in any way disturbing a system, we can predict with certainty (i.e. with probability equal to unity) the value of a physical quantity, then there exists an element of physical reality corresponding to this physical reality,

the entangled states of photons 2 and 3 are elements of reality for any θ_2 (and similarly for photon 1 and its polarization along θ_1). Remarkably, the individual properties of either photon 2 or 3 are not well-defined, as individual detection events at detectors 1, 2, 3, and 4 are random and cannot be predicted by previously choosing to measure a property of photon 1. Therefore, the EPR elements of reality for entangled states of photons 2 and 3 may exist even without existence of these elements for their individual properties.

In the following, I will demonstrate that the conjunction of EPR's propositions, which lead to the establishment of entangled states as elements of reality, is in conflict with the quantum-mechanical prediction. This incompatibility will be shown by deriving CHSH Bell inequality [4] for correlations between individual properties of photon 1 and joint properties of photons 2 and 3 from EPR premises and experimental demonstration of their violation by quantum mechanical predictions.

While any Bell state can be converted into any other Bell state by only single-qubit rotations on one of its constituents (Mattle et al., 1996), the argument is constructed by using a specific subset of two of the Bell states, $|\psi^+\rangle_{2,3}$ and $|\phi^-\rangle_{2,3}$, since they are coherently mixed through the polarization rotation introduced by a half-wave plate (HWP), which makes such an experiment feasible. Using only this HWP, projective measurements onto maximally entangled states of the form $\cos\theta_2\,|\phi^-\rangle_{2,3} + \sin\theta_2\,|\psi^+\rangle_{2,3}$ at Bobís side can be controlled. For consistency throughout this paper, the angle θ has been adopted to mean the rotation of a polarization in real space. Thus the same polarization rotation on the sphere is $2\theta_2$ and that rotation is induced by an HWP which is itself rotated by only $\theta_2/2$.

The experimental setup is explicitly explained in (Kwiat et al., 1995): The three-photon state is created using a pulsed ultraviolet laser (pulse duration 200 fs, repetition rate 76 MHz), which makes two passes through a type-II phase-matched β-barium borate (BBO) nonlinear crystal (Mattle et al., 1996), in such a way that it emits highly polarization-entangled photon pairs into the modes A & B and C & D (Figure 2). Transverse and longitudinal walk-off effects are compensated using an HWP and an extra BBO crystal in each of modes A through D. By additionally rotating the polarization of one photon in each pair with additional HWPs and tilting the compensation crystals, any of the four Bell states can be produced in the forward and backward direction. The source is aligned to produce the Bell state, $|\phi^+\rangle$, on each pass of the pump. Photons are detected using fibre-coupled single-photon counting modules and spectrally and spatially filtered using 3nm bandwidth filters and single-mode optical fibres. While classically correlated states cannot be correlated at the same time in complementary bases, the quality of entanglement is confirmed by the measured visibilities of each generated photon pair, which exceeded 95% in the H/V basis and 94% in the complementary $|\pm\rangle = 1/\sqrt{2}(|H\rangle \pm |V\rangle)$ basis.

Bell pairs contain only two-particle entanglement. To entangle them further, one photon from each pair needs to be superimposed: those in modes A and C, on a polarizing beamsplitter (PBS1). Provided those photons overlap at the beamsplitter and emerge from different output ports, a four-photon GHZ state is generated (Mattle et al., 1996) $|\Psi\rangle = 1/\sqrt{2}\left(|H\rangle_B\,|H\rangle_D\,|H\rangle_1\,|H\rangle_T + |V\rangle_B\,|V\rangle_D\,|V\rangle_1\,|V\rangle_T\right)$. The PBS is an optical device that transmits horizontally-polarized photons and reflects vertically-polarized photons. The PBS implements a two-qubit parity check: if two photons enter the PBS from the two different input ports, then they must have the same polarization in the H/V basis in order to pass to the two different output ports. Then, rotations incurred in quarter-wave plates (QWP) and the subsequent projection of the trigger photon in mode T onto $|H\rangle_T$ reduces the four-particle GHZ state to the desired three-photon entangled state $|\Phi^-\rangle_{1,B,D} = \frac{1}{\sqrt{2}}\left(|H\rangle_1\,|\phi^-\rangle_{B,D} - |V\rangle_1\,|\psi^+\rangle_{B,D}\right)$.

The polarization of single photons can easily be measured by using linear polarizers. As is common in Bell experiments, the angle, θ_1, defines the state on which the linear polarizers projects. In this work, for Bobís measurement, a Bell-state analyzer based on a PBS (Pan and Zeilinger, 1998) is used. By performing a check that the parity of the photons is even, the PBS acts as a $|\phi^{\pm}\rangle$-subspace filter. The two Bell states in this subspace, $|\phi^+\rangle$ and $|\phi^-\rangle$, have opposite correlations in the $|\pm\rangle$ basis and can easily be distinguished using a pair of linear polarizers. By orienting those linear polarizers so that one is along the $|+\rangle$ direction and the other along the $|-\rangle$ direction, a projective measurement onto $|\phi^-\rangle$ is completed. Since an HWP in mode B can interconvert $|\phi^-\rangle$ and $|\psi^+\rangle$ in a controllable way, Alice can choose her projective measurement before her PBS is set to an angle $\theta_2/2$. This is directly analogous to the projections onto the polarization state.

Correlation measurements were carried out by rotating Aliceís polarizer angle, θ_1, in 30^o steps while Bob's HWP was kept fixed at $\theta_2/2 = 0^o$ or 22.5^o. Four-fold coincidence counts at each setting were measured for 1800 seconds. These data are shown in Figure 3. The count rates follow the expected relation $N(\theta_1, \theta_2) \propto \cos^2(\theta_1 + \theta_2)$ in analogy with the expected rates from the standard two-particle Bell experiment. The experimentally obtained data have visibilities of $(78 \pm 2)\%$ in the H/V-basis and $(83 \pm 2)\%$ in the $|\pm\rangle$ basis. Both of these visibilities surpass the crucial limit of $\sim 71\%$ which, in the presence of white noise, is the threshold for demonstrating a violation of the CHSH-Bell inequality. Thus, for the proper choices of measurement settings it is expected that the entangled

entangled state should be able to demonstrate a conflict with local realism using Aliceís polarization state and Bobís maximally-entangled state.

For the state, $|\Phi^-\rangle_{1,B,D}$, the expectation value for the correlations between a polarization measurement at Bob and a maximally-entangled state measurement at Alice is $E(\theta_1, \theta_2) = \cos[2(\theta_1 + \theta_2)]$. The correlation can be expressed in terms of experimentally-measurable counting rates using the relation

$$E(\theta_1,\theta_2) = \frac{N(\theta_1,\theta_2) + N(\theta_1 + \frac{\pi}{2},\theta_2 + \frac{\pi}{2}) - N(\theta_1,\theta_2 + \frac{\pi}{2}) - N(\theta_1 + \frac{\pi}{2},\theta_2)}{N(\theta_1,\theta_2) + N(\theta_1,\theta_2 + \frac{\pi}{2}) + N(\theta_1 + \frac{\pi}{2},\theta_2) - N(\theta_1 + \frac{\pi}{2},\theta_2 + \frac{\pi}{2})} \quad (3)$$

where $N(\theta_1, \theta_2)$ is the number of coincidence detection events between Alice and Bob with respect to their set of analyzer angles θ_1 and θ_2. These correlations can be combined to give the CHSH-Bell parameter, $S = |-E(\theta_1, \theta_2) + E(\tilde{\theta}_1, \theta_2) + E(\theta_1, \tilde{\theta}_2) + E(\tilde{\theta}_1, \tilde{\theta}_2)|$, where $S \leq 2$ for all local realistic theories. For the settings $\{\theta_1, \tilde{\theta}_1, \theta_2, \tilde{\theta}_2\} = \{0^o, 45^o, 22.5^o, 67.5^o\}$, the correlations calculated from quantum mechanics for our state yields $S = 2\sqrt{2}$. This value violates the CHSH Bell inequality and is therefore incompatible with the assumptions of local realism (Fry and Thompson, 1976).

In the experiment, four-fold coincidence counts at each measurement setting were accumulated for 1800 seconds. Each four-fold coincidence signalled 1) the successful creation of two pairs, 2) the successful entangling operation at PBS1, 3) the reduction of the state to the three photon state onto the requisite state, $|\Phi^-\rangle_{1,B,D} = \frac{1}{\sqrt{2}}\left(|H\rangle_1 |\phi^-\rangle_{B,D} - |V\rangle_1 |\psi^+\rangle_{B,D}\right)$.

As is shown in Eq. 3, each correlation is a function of four such data points. The counting rates are shown in Figure 4 for the 16 required measurement settings. These counting rates allow us to calculate the four correlations $E(\theta_1, \theta_2) = 0.69 \pm 0.05$, $E(\theta_1, \tilde{\theta}_2) = -0.61 \pm 0.04$, $E(\tilde{\theta}_1, \theta_2) = -0.58 \pm 0.04$ and $E(\tilde{\theta}_1, \tilde{\theta}_2) = -0.60 \pm 0.04$. Furthermore, those correlations give the experimental Bell parameter, $S = 2.48 \pm 0.09$. This Bell parameter violates the CHSH inequality by 5.6 standard deviations.

3 Conclusion

This year, the Bell inequality turned 47. Since their inception, Bellís inequalities have been the subject of immense theoretical and experimental interest. Initially, this effort was focused on purely foundational issues, but more recently, this work has grown into the burgeoning field of quantum information. Even with all of this attention to this topic, Bell tests have been considered only using single particle properties. The experimental work discussed here is the first Bell test where this restrictive constraint has been lifted.

This result also shows that the naive realistic view of "particles" being physical entities that can be entangled is too simplistic and narrow as no single particle properties are entangled in the present experiment. Therefore from an information-related point of view it only makes sense to speak about measurement events (detector "clicks") whose statistical correlations may violate limitations imposed by local realism and thus be entangled.

References

Aspect, A., Dalibard, J., and Roger, G. (1982). Experimental test of Bell's inequalities using time-varying analyzers. *Phys. Rev. Lett.*, 49:1804–1807.

Bell, J. (1964). On the Einstein-Podolsky-Rosen paradox. *Physics*, 1:195–200.

Bohr, N. (1935). Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.*, 48:696–702.

Clauser, J. F., Horne, M. A., Shimony, A., and Holt, R. A. (1969). Proposed experiment to test local hidden-variable theories. *Phys. Rev. Lett.*, 23:880–884.

Einstein, A., Podolsky, B., and Rosen, N. (1935). Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.*, 47:777–780.

Freedman, S. and Clauser, J. (1972). Experimental test of local hidden-variable theories. *Phys. Rev. Lett.*, 28:938–941.

Fry, E. S. and Thompson, R. C. (1976). Experimental test of local hidden-variable theories. *Phys. Rev. Lett.*, 37:465–468.

Greenberger, D., Horne, M., and Zeilinger, A. (1989). Going beyond Bell's theorem. In Kafatos, M., editor, *Bell's Theorem, Quantum Theory, and Conceptions of the Universe*, pages 69–72. Kluwer.

Hasegawa, Y., Loidl, R., Badurek, G., Baron, M., and Rauch, H. (2004). Violation of a Bell-like inequality in single-neutron interferometry. *Nature*, 425:45–48.

Krenn, G. and Zeilinger, A. (1996). Entangled entanglement. *Phys. Rev.*, A54:1793–1797.

Kwiat, P., Mattle, K., Weinfurter, H., Zeilinger, A., Sergienko, A. V., and Shih, Y. (1995). New high-intensity source of polarization-entangled photon pairs. *Phys.Rev.Lett.*, 75:4337–4342.

Mattle, K., Weinfurter, H., Kwiat, P., and Zeilinger, A. (1996). Dense coding in experimental quantum communication. *Phys. Rev. Lett.*, 76:4656–4659.

Ou, Z. Y. and Mandel, L. (1988). Violation of Bell's inequality and classical probability in a two-photon correlation experiment. *Phys. Rev. Lett.*, 61:50–53.

Pan, J.-W., Bouwmeester, D., Daniell, M., Weinfurter, H., and Zeilinger, A. (2000). Experimental test of quantum nonlocality in three-photon Greenberger-Horne-Zeilinger entanglement. *Nature*, 403:515–519.

Pan, J.-W. and Zeilinger, A. (1998). A Greenberger-Horne-Zeilinger-state analyzer. *Phys. Rev.*, A57:2208–2211.

Rowe, M. A., Kielpinski, D., Meyer, V., Sackett, C., Itano, W., Monroe, C., and Wineland, D. (2001). Experimental violation of a Bell's inequality with sufficient detection. *Nature*, 409:791–794.

Schrödinger, E. (1935). Die gegenwärtige Situation in der Quantenmechanik. *Die Naturwissenschaften*, 23:807–812, 823–828, 844–849.

Shih, Y. H. and Alley, C. O. (1988). New type of Einstein-Podolsky-Rosen-Bohm experiment using pairs of light quanta produced by optical parametric down conversion. *Phys. Rev. Lett.*, 61:2921–2924.

Tapster, P. R., Rarity, J. G., and Owens, P. C. (1994). Violation of Bell's inequality over 4 km of optical fiber. *Phys. Rev. Lett.*, 73:1923–1926.

Tittel, W., Brendel, J., Zbinden, H., and Gisin, N. (1998). Violation of Bell inequalities by photons more than 10 km apart. *Phys. Rev. Lett.*, 81:3563–3566.

Vaziri, A., Weihs, G., and Zeilinger, A. (2002). Experimental two-photon three-dimensional quantum entanglement. *Phys. Rev. Lett.*, 89:240401–240405.

Walther, P., Resch, K., Brukner, C., and Zeilinger, A. (2006). Experimental entangled entanglement. *Phys. Rev. Lett*, 97:020501–020505.

Weihs, G., Jennewein, T., Simon, C., Weinfurter, H., and Zeilinger, A. (1998). Violation of Bell's inequality under strict Einstein locality conditions. *Phys. Rev. Lett.*, 81:5039–5043.

Zeilinger, A., Horne, M., and Greenberger, D. (1992). Higher-order quantum entanglement. In Han, D., Kim, Y., and Zachary, W., editors, *Squeezed States and Quantum Uncertainty*, NASA Conference Publication 3135. National Aeronautics and Space Administration, College Park.

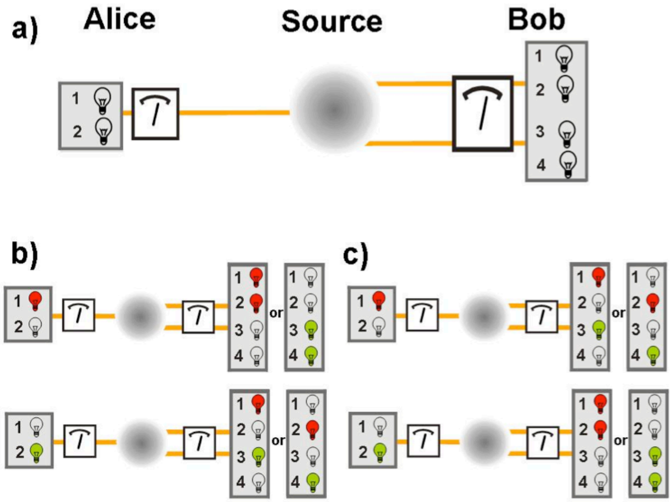

Figure 1: Schematic for the Bell experiment based on an entangled entangled state. a) A source emits three entangled photons in such a way that one photon is received by Alice and the two other photons by Bob. Alice controls an analyzer that makes measurements of the polarization of her photon. When the photonís polarization is measured to be parallel to orientation, θ_1, of the analyzer, the measurement outcome is +1 (red light bulb) or -1 (green light bulb) when perpendicular. In contrast, Bob makes projective measurements onto a two-particle entangled state, where again the orientation of the apparatus is defined by the angle, θ_2. Bobís outcomes are defined as +1, when detectors 1 & 2 (red light bulbs) or 3 & 4 (green light bulbs) are firing, or -1 when detectors 1 & 3 or 2 & 4 are firing. b) When Alice and Bob measure with the same measurement settings, i.e. $\theta_1 = -\theta_2$, they observe perfect correlations, which appear in four possible configurations, given by +1. However, when they measure in a different basis, i.e. $\theta_1 \neq \theta_2$, they will also observe four possible anti-correlations c), given by -1. The correlation measurements with different measurement settings form the basis of a test of local realism using entangled entanglement.

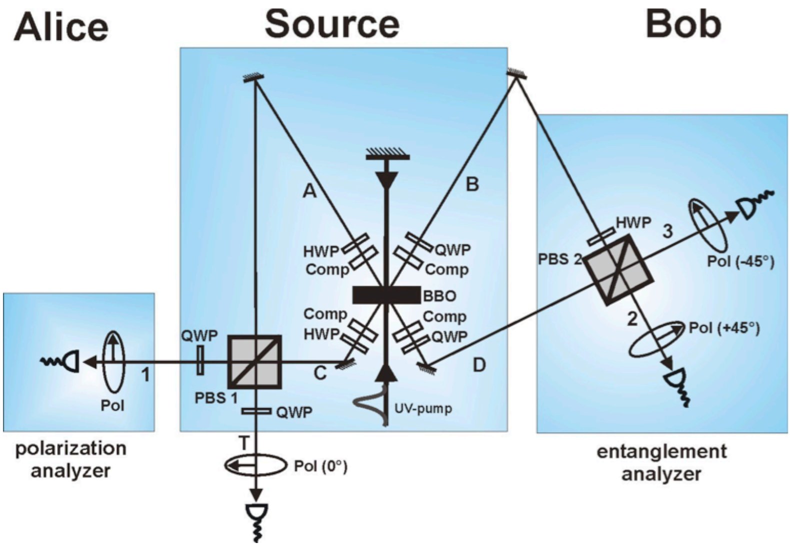

Figure 2: Setup for the experimental realization. A spontaneous parametric down-conversion source emits polarization-entangled photons in the Bell state, $|\phi^+\rangle$, into both the forward pair of modes A & B and backward pair of modes C & D. After superimposing the modes A & C at the polarizing beamsplitter PBS1, passing each mode through a quarter-wave plate (QWP), and projecting the trigger qubit T onto the state $|H\rangle_t$ generates the entangled entangled state $|\Phi^-\rangle_{1,B,D} = \frac{1}{\sqrt{2}}\left(|H\rangle_1 |\phi^-\rangle_{B,D} - |V\rangle_1 |\psi^+\rangle_{B,D}\right)$. The photon in mode 1 belongs to Alice, who uses a linear polarizer for her single-particle polarization measurements, determined by the angle, θ_1, of her polarizer. The photons in mode B and D belong to Bob, who uses a modified Bell state analyzer to make projections onto a coherent superposition of $|\phi^-\rangle_{B,D}$ and $|\psi^+\rangle_{B,D}$, where the mixing angle, θ_2, is determined by the angle, $\theta/2$, of the half-wave plate (HWP) in mode B.

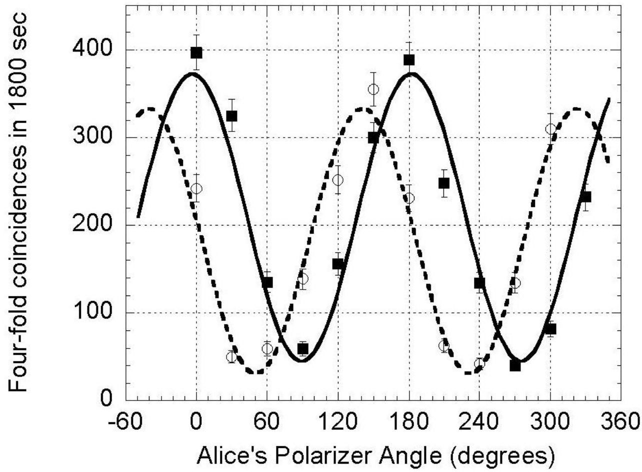

Figure 3: Measured coincidence fringes between Alice and Bob for the entangled entangled state. Bobís half-wave plate was initially set to 0∞, so that he made fixed projective measurements onto the state $|\phi^{-}\rangle_{B,D}$. The total number of four-fold coincidence counts measured in 1800 seconds as a function of the angle of Alice's polarizer is shown as solid squares. Fitting the curve to a sinusoid (solid line) yields a visibility of $(78 \pm 2)\%$. Bob then changed his measurement setting to project onto the state $\frac{1}{\sqrt{2}}\left(|\phi^{-}\rangle_{B,D} + |\psi^{+}\rangle_{B,D}\right)$, and the procedure was repeated. The data for these settings are shown as open circles. The sinusoidal fit (dotted line) yields a visibility of $(83 \pm 2)\%$.

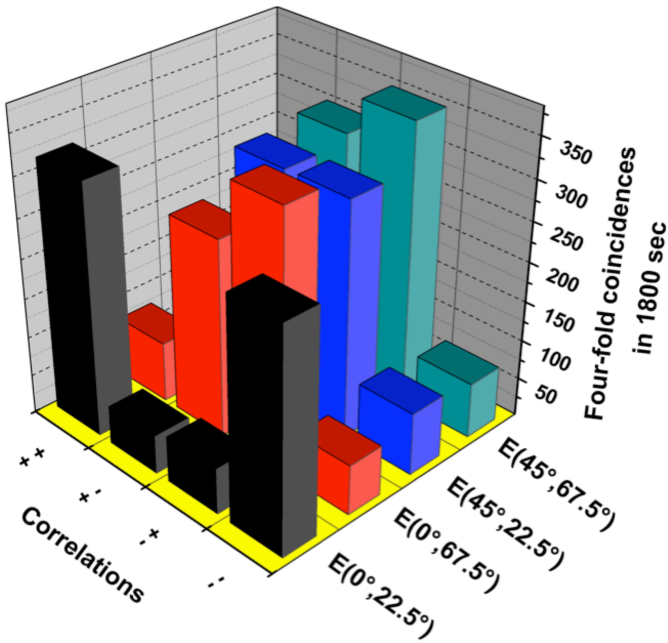

Figure 4: Experimental results obtained by measuring correlations for violating a CHSH Bell inequality. The Bell inequality is comprised of 4 correlations, in this case between the polarization state measured by Alice and the entangled states measured by Bob. Each of these correlations in turn can be extracted from 4 coincidence counting rates. The requisite coincidence measurements for the 16 different measurement settings are shown. Each measurement was performed for 1800 seconds. For measurement settings, $\{\theta_1, \theta_2\}$, the axis labels ++, +–, –+, – – refer to the actual settings of $\{\theta_1, \theta_2\}$, $\{\theta_1, \theta_2 + \pi/2\}$, $\{\theta_1 + \pi/2, \theta_2\}$, and $\{\theta_1 + \pi/2, \theta_2 + \pi/2\}$ respectively. These data can be combined to give the Bell parameter $S = 2.48 \pm 0.09$.

Contributors

Matthias Egg	Department of Philosophy University of Lausanne Switzerland Matthias.Egg@unil.ch
Michael Esfeld	Department of Philosophy University of Lausanne Switzerland Michael-Andreas.Esfeld@unil.ch
Simon Friederich	Philosophisches Seminar University of Göttingen Germany email@simonfriederich.eu
Alexei Grinbaum	CEA-Saclay/ LARSIM Gif-sur-Yvette France alexei.grinbaum@cea.fr
Iñaki San Pedro	Department of Logic and Philosophy of Science Complutense University Madrid Spain inaki.sanpedro@filos.ucm.es
Tilman Sauer	Einstein Papers Project California Institute of Technology USA, and Institute for Theoretical Physics University of Bern Switzerland tilman@caltech.edu

Jakob Sprickerhof
Department of Philosophy
University of Leeds
UK
phjs@leeds.ac.uk

Antonio Vassallo
Department of Philosophy
University of Lausanne
Switzerland, and
Institute of Philosophy
University of Warsaw
Poland
antonio.vassallo@unil.ch

Philip Walther
Faculty of Physics
University of Vienna
Austria
philip.walther@univie.ac.at

Adrian Wüthrich
Institute for Philosophy
University of Bern
Switzerland, and
History of Science
Technical University Berlin
Germany
awuethr@philo.unibe