Reichenbach's Common Cause Principle and Quantum Correlations

PhD Thesis

Iñaki San Pedro García

December 2007

eman ta zabal zazu

Universidad del País Vasco
Euskal Herriko Unibertsitatea

Reichenbach's Common Cause Principle and Quantum Correlations

A Thesis submitted in partial fulfilment of the requirements for the Degree of *Doctor por la Universidad del País Vasco.*

Iñaki San Pedro García

Supervisors: Mauricio Suárez Aller, *Universidad Complutense de Madrid.*
Andoni Ibarra Unzueta, *Universidad del País Vasco, UPV-EHU.*

Iñaki San Pedro, 2007. *Reichenbach's Common Cause Principle and Quantum Correlations*

I hereby declare that this submission is my own original work and that, to the best of my knowledge, it contains no material previously published or written by another person, except where due acknowledgment has been made in the text.

Research towards this thesis has been carried out thanks to a FPI Scholarship of the Spanish Ministry of Science and Education (MEC) associated, initially to the DGICYT Research Project BFF2002-01552: 2002-2005, *Causality, Determinism and Probability in Quantum Mechanics and Relativity Theory*, and most recently to the Research Project *Causation, Propensities and Causal Inference in Quantum Physics* within the DGICYT Research Network HUM2005-01787: 2005-2008 *Classical and Causal Concepts in Science.*

Bilbao, December 2007

A mis padres.

Acknowledgements

This thesis has been made possible thanks to the unconditional support of Mauricio Suárez —both in the personal and as the main scientific supervisor of the work within it— from whom I borrowed many philosophical insights that helped me thinking clearly and critically about many of the issues discussed here. I must thank him as well for showing me the way through in the most difficult stages of my research. Also, for his generosity, his patience and for making things easy, gracias, Mauricio.

Thanks to Andoni Ibarra for his trust and support from the very beginning; for giving me the initial drive and making the whole project possible. Thanks as well for being there with some nice Catalan words whenever we met, spoke or wrote each other.

I would like to thank as well Miklós Rédei for introducing me to the topic and for sharing his valuable insights and ideas on common causes with me; for hosting me in Budapest —making me feel as if I was at home there— and being so generous to me. Also, my thanks to Gábor Hofer-Szabó for helping me with the most technical parts of my research and for his friendship. To both, köszönöm!

I am also very grateful the members of the "Classical and Causal Concepts in Science" Research Network, especially Carl Hoefer and Henrik Zinkernagel, for their support throughout all this time and for making me feel part of the project since the very early stages of my research. To the members of the "Methods of Causal Inference and Representation in Science" Group at the Complutense University, thanks as well for their help and fruitful discussions whenever I joined them in Madrid. And thanks to all at the "Cátedra Sánchez Mazas" —more especially to Thomas Mormann, Hannot, Igor, Luís and Gabriel— for their support and for being so patient and receptive with the always complex issues of quantum mechanics. And to Albert and Chavi, my special thanks for caring and helping me through hard times.

Last but not least, thanks to my family and all that have been close to me throughout, no matter how difficult I made it. Por estar siempre ahí ¡Gracias! And of course to María, que me ha susurrado cada palabra al oído... el lugar, el camino y el destino.

is

Contents

Preface

This thesis is about causation, a fairly complex and often quite elusive concept, which has however a strong intuitive basis. And it is also about quantum mechanics, our empirically most powerful theory, formally precise and of high predictive power, but also a continuous challenge to our intuitions.

Causation has been identified for quite a long time with determinism and constituted one of the cornerstones of what we might call the classical (causal deterministic) mechanistic worldview, the conceptual core of human thought up to the beginning of the 20th century. Such a conceptual picture came to a head with the advancement of quantum mechanics, a then new theory of microscopic phenomena. In particular, the deep philosophical implications of the quantum theory were soon interpreted as to suggest that causal accounts of some physical phenomena were no longer possible. The debate as to whether this is really so is still wide open, however, and this thesis aims to modestly tackle it.

The notion of causation is already a problematic concept on its own at the philosophical level,[1] but surely quantum mechanics has added a new dimension to the problem of its philosophical status. A challenge which is also very closely related to other big philosophical questions, such as whether Nature is deterministic or indeterministic at its most fundamental level, or whether there is such a thing as a reality (that our physical theories describe, more or less accurately). Quantum mechanics (in its orthodox interpretation) is also extremely clear in this respect: the world is fundamentally indeterministic and the theory is not to be interpreted realistically. It is quite remarkable that the attempts at a better understanding of the foundations of quantum mechanics have shown these philosophical problems —big philosophical questions themselves— to be closely related. An outstanding example is the deep relation that quantum mechanics has revealed between questions about the existence of a causal structure in Nature and those regarding the physical

[1] A sign that this is so is the fact that a complete theory of causal interactions has not yet been developed. There is not much consensus even on what such a notion is, nor on how can it be properly characterised.

reality of the observed phenomena. In this respect, the philosophical consequences of the Einstein, Podolsky and Rosen (EPR) paradox were seen to conceal a deep relation between the hidden variables entering Bell's theorem and the (common) causes of Reichenbach's Common Cause Principle.

The relation between causation and Bell's hidden variables was first noted in a remarkable and highly influential argument by van Fraassen in the early 1980's.[2] Van Fraassen's argument was indeed so influential that it established, not only the idea that the hidden variables entering the Bell inequalities are related to the common causes of Reichenbach's Common Cause Principle, but also the view that, as a consequence, there can not exist causal explanations of the EPR correlations. In short, van Fraassen's conclusions about causation in quantum mechanics agree with those of the orthodox interpretation, i.e. that there are certain phenomena (quantum EPR correlations, for instance) which cannot be accounted for in causal terms. As a further consequence (or perhaps a side-effect), the idea that Reichenbach's Common Cause Principle does not constitute a valid general causal inference claim —and that it merely reflects a metaphysical conviction— was reinforced. This has become the received view, and has gained even larger support from the experimental confirmation that quantum mechanics indeed violates the Bell inequalities.

The main aim of my thesis is to revise this standard view and try to shed more light into the issue. I shall claim, in the light of such revision, that contrary to the received view the usual arguments against Reichenbachian common cause explanations of the Einstein, Podolsky and Rosen (EPR) correlations contain unwarranted assumptions. As a result, in my view, these arguments are not valid in general. In particular, I will suggest that the restrictions placed on the postulated common causes —arguably taken to reflect physical locality— seem to be too strong and need to be reconsidered. Thus the question as to whether sensible (and intuitively correct) Reichenbachian common cause explanations of EPR correlations may be provided is still open. This also suggests to view the problem from a different perspective. Instead of asking whether Reichenbachian common cause explanations of EPR correlations exist under the standard conditions (of locality and the like) we may rather pursue the question as to whether quantum mechanical phenomena leaves some conceptual room for intuitively correct and sensible Reichenbachian common cause explanations of EPR phenomena. In other words, we may want to investigate what are the strongest restrictions that one may impose on the idea of Reichenbachian common causes that will allow us to avoid contradictions such as those entailed by Bell's theorem and

[2](van Fraassen, 1982*a*).

yet provide a sensible causal explanation of the EPR correlations. This will turn out to be a second motivation of this thesis, namely to investigate how far into the quantum world can we take a classical concept such as causation. I will pursue this line of thought by providing a highly intuitive causal explanation of the EPR phenomena by means of a Reichenbachian common cause model.

The structure of the thesis is as follows. Chapter 1 provides a very general overview of the enormous conceptual changes that quantum mechanics brought along, specially due to its fundamentally indeterministic character. Attention is paid, in particular, to the impact that the new quantum concepts had on the status of the 'law of causation' —central to the then standard causal deterministic worldview— which was unquestionably thought to underlie any physical theory up to the development of quantum physics.

Chapter 2 introduces the concepts and tools necessary for causal inference in quantum physics. Causal inference faces extraordinary difficulties in quantum mechanics mostly related to indeterminism. This is where probabilistic theories of causation play a crucial role, i.e. in assessing the conditions under which the various tools for causal inference may be reliably applied. The concepts of event and correlation are central to such accounts of causation and hence need to be precisely defined. More specifically, token and type events are distinguished, since they refer to different kinds of causal structure. As for correlations, it turns out that not all correlations arise as a consequence of a deeper causal structure. Thus some necessary (but not sufficient) conditions on correlations are discussed in order to be able to attempt at causal inferences from them. As a result, *purely formal descriptive* correlations are distinguished from *genuinely physical* ones.

Chapter 3 reviews Reichenbach's Common Cause Principle (RCCP) main features —both at an intuitive and a formal level— and its scope of application as a (general) tool of causal inference. In addressing the status of RCCP it is important to clearly distinguish between its metaphysical content and the methodological claim that it carries along. The most prominent counterexamples to RCCP are discussed with this distinction in mind. However, special attention is paid to the the status of RCCP in the context of indeterministic causation. Thus, Cartwright's 'Cheap-but-Dirty' factory example and her generalistion of the fork criterion are discussed in detail. However, in the light of the results by the Budapest School on (Reichenbachian common cause) *completability* it is suggested that Cartwright's arguments are less compelling than originally claimed. The conclusions in this chapter turn out to be of major importance when addressing common cause explanation of quantum correlations in the foregoing chapters.

In Chapter 4 the case of quantum correlations is discussed in detail. The discussion focuses in particular on the so-called Einstein, Podolsky and Rosen (EPR) correlations —whose implications were first considered a challenge for the completeness of the quantum theory— and their significance for the status of Reichenbach's Common Cause Principle (RCCP). Thus, the EPR paradox and thought experiment are explained, along with the idea of hidden variables and the remarkable consequences of Bell's theorem. The second part of the chapter reassesses van Fraassen's highly influential argument showing that Reichenbachian common causes also commit to the Bell inequalities (just as any hidden variable theory does). Such reassessment is first achieved in the light of the distinction in Chapter 3 between the metaphysical and methodological components of RCCP, and secondly by noting some subtleties of the results by the Budapest School, particularly in connection with the ideas of Reichenbachian common cause completability. In particular, the assumption of so-called *common*-common causes suggest that van Fraassen's argument is not conclusive.

In Chapter 5 a recent derivation of the Bell inequalities is discussed, which claims not to involve *common*-common causes. The derivation explicitly requires that the common causes be (Reichenbachian) *individual*-common causes. However, a deeper analysis of the different conditions presupposed in the derivation suggests that they amount to assuming that the postulated Reichenbachian common causes are also *common*-common causes. As a consequence, the derivation cannot be taken to be conclusive either (just as van Fraassen's). It remains an open question whether there exist tenable Reichenbachian common cause explanations of EPR correlations.

I explore in Chapter 6 the possibility of such a (Reichenbachian) *individual*-common cause explanation of EPR phenomena. The key aspect of the common cause model presented is that it explicitly incorporates measurement operations as causal factors. This allows for a reinterpretation of so-called *No-Conspiracy* assumption —which I shall rename *Measurement Independence*—, standard (and necessary) for the derivation of the Bell inequalities. As such, the model is explicitly non-factorizable, which results in the need to revise the ontological status of the postulated common causes. In particular, the model's common causes will need to be interpreted either as non-localised events with local causal powers or, alternatively, as localised events that exert non-local causal influences.

The thesis closes with a summary of the main overall conclusions.

“Chameleons. Such exceptional creatures. The way they change color. Red. Yellow. Lime. Pink. Lavender. And did you know they are very fond of music?”

— Truman Capote, *Music for Chameleons*

Chapter 1

Does Nature Have a Causal Structure?

Whether Nature is causal or not, or in other words, whether the diverse relations between different aspects of our observed reality are a consequence of a deeper causal structure, has been a matter of discussion since the very beginning of philosophical enquiry. A controversy that is obviously and best reflected in the underdetermination of the diverse interpretations of the physical theories we use to describe it. At present this major philosophical question is still a subject of analysis and discussion, particularly after the conceptual changes that quantum mechanics —our most empirically powerful physical theory— introduced in our worldview.

The present chapter aims to provide a general picture, an overview, that is, of the main concepts and related issues that the idea of causation has traditionally involved, and their tension with the new quantum concepts. I will mainly concentrate, however, upon the debate *determinism-indeterminism* —which has been taking place since the early years of the quantum theory— since it illustrates pretty well the impact that quantum mechanics has had on our causal picture of the world.

1.1 Regularities and Determinism

This section provides a very informal (and perhaps unphilosophical) account of how causal concepts arise, but it shall suffice to focus on the issues that the more philosophical discussion forthcoming in the following chapters involve.

1.1.1 Regular Behaviour and Causal Relations

The sense that there is a certain causal structure in Nature may probably have its origins at the very way we perceive the world or, perhaps in a slightly better sense, in the way we think about it. That is, in the way we handle and process our perceptions, and how we interpret them. It is this way of looking and seeing the world that partly shapes our conceptual schema. A fundamental feature of our perceptual capacities is the ability to recognise and classify regular behaviour and patterns in general. It is not surprising for instance that we hear someone saying that "at 11 p.m. it will be already dark" or that the "summer months are hotter than the winter ones". Of course there are places where, depending on the time of the year, it is not dark at 11 p.m. Less controversial seems to be the opinion that the summer months are usually hotter than the winter ones. All in all, almost everyone would agree that the sentences above do express perfectly acceptable observations or empirical facts. They do also reflect some kind of *regular* behaviour, something that happens recurrently. (This is the case at least in my most immediate environment, i.e. I do not recall being in the daylight at 11 p.m. or living a cold summertime).

Our perception of reality is infused by such recurrent observations, which we consider to be regular even if they are not exceptionless —sometimes they do not even occur with high frequency—. Observing and recognising such regularities and patterns is, in fact, fundamental to our understanding of the world since it allows us to leave superfluous information apart and come up with simpler explanations of our observations (including those concerning ourselves). Similarly, we tend to reject the most complicated or conceptually difficult explanations of a given observation.[1] The upshot is that we gain a better (even if partial) understanding of our surrounding world. This is why then we tend to look for recurrent events, regular behaviour in its various forms and (more or less complex) patterns in general. And, as a matter of fact, we have gained (developed), as a consequence, a very strong conceptual background based on this kind of regular associations.

The idea of causation is part of this phenomenological conceptual background. In particular, as a consequence of this way of perceiving and thinking about the world, we ask for the reasons of such regularities and tend to establish links between the different events that constitute them. Those links we usually refer to as causal relations.

The fact that causation was very closely related to such regularities was

[1] This is, by the way, also one of the reasons for which I think we tend to resist sophisticated theories that challenge our most basic perceptions and fundamental understanding of reality, which are sometimes even rooted in our own deepest beliefs.

noted already by Hume. The idea of causation —in different forms or with different meanings, and even referring to different concepts— has been a major philosophical problem already since the Greek philosophers. However, it is Hume's conception of causation, defined in terms of recurrent sequential happening or constant conjunctions, that has shaped to a great extent the approach of modern philosophy towards causation.[2]

Probably one of the most significant features of Hume's account is its very critical attitude towards a metaphysical conception of causation. Hume's critique has been indeed so influential that it can be found to be at the origin of many modern philosophers' causal skepticism. In fact, Hume saw causation as a highly problematic notion. As Salmon points out after quoting Hume's description of causal influence in the introduction to his 1739 "Treatise of Human Nature":[3]

> This [Hume's] discussion is, of course, more notable for factors Hume was unable to find than for those he enumerated. In particular, he could not discover any 'necessary connections' relating causes to effects, or any 'hidden powers' by which the cause 'brings about' the effect.

The passage above identifies probably the greater difficulty that any account of causation faces —even the most actual and sophisticated—, namely that of finding necessary and sufficient conditions for causation. This is a central issue in causal inference and I shall be addressing it in more in detail in Chapter 3 when discussing the status of Reichenbach Common Cause Principle.

1.1.2 Causal Determinism

Another central feature of Hume's concept of causation is determinism. This is not surprising at all since determinism was a fundamental feature of the (mechanistic) worldview ever since the successful development of Newtonian mechanics. In such a worldview, all matter existed in a fixed three dimensional space and evolved deterministically in time (according to Newton's second law).[4] All known physical phenomena could be explained along these

[2]As we will see the main difference between Hume's account and many of the most actual theories of causation stems in the fact that Hume's account assumed causal relations to be deterministic (that is the sense in which 'constant conjunction' is to be understood).

[3](Salmon, 1984, p. 137).

[4]The Newtonian concepts of space and time are necessary concepts assumed a priori in the Kantian sense, i.e. they are not empirical concepts defined by our experiences. Such a view assumes, also a priori, an underlying 'law of causality' by which for any event

terms and cause and effect relations were then identified with matter interactions, mostly in the form of particles deterministically colliding with each other.

This deterministic causal picture mechanics managed to cope with the diverse problems and complexities of the observed reality up to the beginning of the 20th century. It was then that new phenomena were observed and a new conceptual background started to emerge, which gave rise to two new physical theories, Einstein's relativity on the one hand and quantum mechanics on the other. The great impact that Einstein's relativity had on the classical Newtonian worldview is well known. Indeed, Einstein's relativity theories completely changed the Newtonian concepts of space and time by considering them to be a single entity in a first instance, and later suggesting a relation between its geometry and the distribution of matter. However, Einstein's relativity theories left untouched the deterministic character of a 'correct' physical description and thus the tacitly assumed 'law of causality' (see Footnote 4).

The conceptual shift introduced by quantum mechanics —the theory of the small— on the other hand challenged the causal deterministic picture throughout, since it assumed Nature to be fundamentally indeterministic. This is not to say, however, that before the rise of quantum mechanics the deterministic classical causal picture —well supported by Newtonian mechanics first and later by Einstein's relativity theories— was unproblematic and free from controversy. It faced, on the one hand the consequences of Hume's skeptical critique at the metaphysical level. But it faced as well more immediate empirically observable difficulties, which were directly related to determinism. In particular, determinism implied for instance that causes *invariably* produced the same effects. Of course it is not difficult to imagine cases in which experience shows this to be false. The most recurrent solution perhaps to such difficulties involving *imperfect regularities* appealed thus to the intuition that indeterminism was a consequence of our lack of information about the system. (If we had had all the information about a physical system, i.e. if we had known the 'complete state of affairs', so it was claimed, we could have predicted with certainty its precise evolution in time.)

Whether this was a tenable position or not —certainly a controversial matter as it stands— it is besides the main point here, but whatever the answer might be, it would not solve the conceptual difficulties regarding indeterminism posed by quantum mechanics, as we will see.

---

there exist another (prior in time) event from which the former follows (according to the dynamical laws of the theory, i.e. Newton's second law).

1.2 Indeterminism and Causal Explanation

1.2.1 Ignorance and Indeterminism

That the indeterminacy implied by the standard interpretation of quantum mechanics was a completely new phenomenon, radically different from classical indeterminism, was already noticed in the early years of the theory.[5] Heisenberg, for instance is very clear about it when commenting on Bohr's attempt —together with Kramers and Slater— to solve the apparent contradiction between the particle and the wave pictures of the theory, which led to the notion of probability wave:[6]

> This concept of the probability wave was something entirely new in theoretical physics since Newton. Probability in mathematics or in statistical mechanics means a statement about our degree of knowledge of the actual situation. [...] The probability wave of Bohr, Kramers, Slater, however, meant more than that; it meant a tendency or something.[...] It introduced something standing in the middle between the idea of an event and the actual event, a strange kind of physical reality just in the middle between possibility and reality.

In fact, classical (Newtonian) mechanics could only account for indeterministic behaviour arising from a lack of knowledge, as noted by Heisenberg in the quote above. This lack of knowledge refers typically to the various boundary conditions (including initial conditions) we may set on the system, and is responsible for the fact that its complete 'state of affairs' cannot be determined.[7] Nor are we able to make an exact prediction about later time evolutions of the system. The typical situation is the flipping of a fair coin. At the moment of flipping a coin, we normally lack a fair amount of precise information about the system, such as as the exact initial position of the coin, the speed with which it is thrown up, the exact point of it in which the impulse is given, its exact shape and mass, etc. And it is this shortage of initial information that prevents us from knowing the exact time evolution

[5] I am here referring mainly to the orthodox interpretation of quantum mechanics, i.e. the Copenhagen interpretation. I shall be doing so in what follows as well. It is worth noting however that there are other versions of quantum mechanics, such as Bohm's, which do not display indeterminism. To the contrary, Bohm's quantum mechanics is explicitly deterministic.

[6] (Heisenberg, 1958, p. 11 of the 2000 reprint).

[7] Usually, the systems featuring this kind of indeterminism may be characterised by fairly complex initial or boundary conditions, as we can see in the example of the fair coin toss.

of the coin. Therefore the outcome of the toss cannot be predicted with certainty.

All we can predict is an 'approximated' result of the toss by assigning probabilities to each of the possible outcomes. It is perhaps worth stressing that the probabilities assigned to the possible outcomes reflect diverse features and characteristics of the system (those we do know). For instance, in the case of a 'fair' coin, symmetry considerations allow us to assign the same probability, i.e. 1/2, to each of the two outcomes.

Events or systems displaying this kind of indeterministic behaviour, such as that of the tossing of a fair coin, are very easy to imagine, for they happen every once and again. The same kind of indeterminism is present, for instance, in experimental error as well. We do not find it conceptually problematic therefore that such events occur. In a sense, we accept this kind of indeterminism in the belief that there is a way of exactly determining the exact outcomes of any such event, even if we have no access to the methods —in the case of experimental error— or information —in the case of boundary conditions— that would allow us to do so. In other words, conceiving indeterminism in such a way assumes an underlying deterministic structure (which we lack knowledge of, however) which also conforms to the (classical) 'law of causality'.

1.2.2 The Fundamental Indeterministic Picture

Quantum mechanical indeterminism is far more challenging. For it suggests that indeterminism is a fundamental feature of reality itself.[8] Quantum indeterminism arises as a consequence of the uncontrollable disturbances at the moment of measurement. In particular, the interaction of the quantum system with the (macroscopic) system that constitutes the measurement apparatus, results in an uncontrollable disturbance that affects both systems equally. As a consequence, measurement outcomes are indeterministic, conforming to Heisenberg's uncertainty relations, and reveal the fundamental indeterminism of the system.[9]

Thus, indeterminism in quantum mechanics does not depend any more

[8] Again, I am referring here to the orthodox interpretation of quantum mechanics.

[9] Quantum indeterminism is thus a consequence of the measurement operations. The formalism itself, however, is fully deterministic. This is pointed out by several authors. Earman (1986) goes as far to claim that quantum mechanics is in some sense even more deterministic than classical mechanics. For the Schrödinger equation governing the evolution of quantum systems is a fully deterministic wave equation with a high stability, in contrast to certain classical systems which are highly sensitive to initial or boundary conditions, i.e. classical chaotic systems.

on the degree of experimental precision achieved, nor on a lack of information about initial boundary conditions. In other words, the quantum kind of indeterminism does not allow to predict precise outcomes of an experiment, even in the ideal case where full knowledge about a system and the corresponding initial and other boundary conditions is achieved. We can see, however, that such an idea of indeterminism does not refer to the unpredictability of the time evolution of the system in the usual sense. For the time evolution of quantum mechanical systems is given by the time-dependent Schrödinger equation, which is essentially a deterministic wave equation. Quantum indeterminism is just fundamental and intrinsic to the quantum systems themselves.

The fundamental indeterminism that quantum theory ascribes to reality is thus fatal for the former causal deterministic picture. In particular, if causation is considered to be deterministic —as in the classical picture—, there is not hope to find causal explanations for the relations between fundamentally and irreducibly indeterministic events. Thus, in quantum mechanics the 'law of causality' is not valid any more.[10] Actually, things get even worse for the classical picture if we endorse the Copenhagen interpretation's reading of Heisenberg's Uncertainty Principle, which takes it that Heisenberg's uncertainty relations place, not only the measurement limits on a quantum system, but also restrictions on the reality of the properties of the system.[11] We will see in more detail in Chapter 4 what the consequences are of such an interpretation.

Given what it has been said so far, it might seem difficult at first glance to provide quantum physics with a causal structure. There seem to be at least two possibilities that can be explored in order to do so. In the first place, we should recall that the above is mainly based on the so-called Copenhagen interpretation of quantum mechanics (see Footnote 5). Although Copenhagen is the most widely accepted formalism, particularly among practising physicists, there are other formalisms describing quantum mechanical phenomena —with exactly equivalent empirical predictive power— that provide however a completely causal deterministic picture. As pointed out, such is the case of Bohmian mechanics. Thus an obvious option in order to interpret quan-

[10]Heisenberg, for instance, is quite harsh on the status, validity and justification of the 'law of causality'. He takes it in particular that the 'law of causality' is a relic of the Cartesian body-mind (*res extensa-res cogito*) separation, which to his mind has been extremely harmful for science and human knowledge (and thus should be conveniently eliminated from scientific practice).

[11]More precisely, Heisenberg's uncertainty relations are seen in this view to impose a restriction on the simultaneous reality of incompatible (or complementary) properties of a system.

tum phenomena causally is to reject the orthodox interpretation and adopt Bohmian mechanics instead.

As a second alternative we might consider modifying our conception of causation by ceasing to regard it as implying (and requiring) determinism any more. This is the case —as well as a partial motivation— of probabilistic accounts of causation, which aim to account for causal influences in terms of probabilistic relations (or infer the former from the later). This is in fact the option I shall be pursuing in this thesis.

Chapter 2

Correlations and Causal Inference

In the preceding chapter I reviewed briefly some of the issues regarding determinism (or indeterminism) and causation. I pointed out that since it is the case that quantum mechanics (in its orthodox interpretation) is a fundamentally probabilistic theory —thus describing an indeterministic world— any proper idea of causation within the theory —if there is such a thing— will need to be laid out in terms of probabilistic causation. Thus, in order to address causal issues in quantum mechanics then we need to define precisely the key concepts that the notion of probabilistic causation is grounded on. In particular, concepts such as that of 'event' or 'correlation' need to be introduced. Furthermore, as we will see, the standard idea of correlation needs to be complemented, somehow, with further requirements in order for causal inferences to be reliable. Only then can the diverse techniques of causal inference be put at work.

2.1 Probabilistic Causation

The central idea of probabilistic causation is that causes raise the probability of their effects. We could perhaps be more precise and say that in probabilistic accounts of causation the probability that a certain (effect) event takes place is *sensitive* to the probability that some other event (its cause) occurs, either positively (increasing it), or negatively (decreasing it) —we will see in Section 2.3 how close this is to the idea of correlation—. In other words, causes are probabilistically (or statistically) relevant to their effects.[1]

[1]In most cases, statistical relevance it implicitly taken to be positive. We can read, for instance in the 'Probabilistic Causation' entry of the Stanford Encyclopedia of Philoso-

We realise already that the appeal to probabilities opens the possibility for causal accounts of indeterministic systems. In fact, indeterminism was one of the motivations for the development of probabilistic theories of causation —and to be able to deal with it to some extent is one of the great advantages of probabilistic causation in relation to regularity accounts—. It is perhaps worth noting that probabilistic theories of causation are insensitive, so to speak, to the source of indeterminism, i.e. to whether the indeterministic character of a system is due to a lack of information or it is fundamental to the system itself. This may be at first sight a completely obvious claim since probability theory, the basis of all probabilistic theories of causation, provides nothing more than a mere formal (mathematical) description of the (empirical) facts and thus 'does not care' at all about what the ultimate origin of the probabilistic behaviour, i.e. of indeterminism, might be.

Now, probabilistic theories of causation are ontologically neutral in this sense, the particular interpretation we might adopt to account for the indeterministic behaviour of a system will have an impact on the event ontology. In particular, if the indeterminism has its origins in lack of knowledge, it seems that a probabilistic cause can only be interpreted as a *partial* cause. For recall that in such a case, determinism would be restored in completion of the causal structure of the system —the information we originally lack, that is—. But the fact that the causal structure needs to be completed just entails that the (probabilistic) cause is a partial cause. In other words, the lack of information entails that there are other causally relevant factors. For exactly the same reasons, this interpretation of indeterminism entails that there is a unique deterministic *total* cause of a given (effect) event, i.e. that which includes all the causal information.

On the other hand, if indeterminism is taken to be a fundamental feature of the system —recall that this was the case in quantum mechanics—, there is not a reason why a probabilistic cause should not be indeed a *total* cause. For, in principle, there is nothing that can stop us thinking that the cause is the only causal relevant factor for its effect. Of course the diagnosis will vary from case to case, since we may know (for whatever reason) whether the

---

phy (Hitchcock, 2007, p. 2) that

> [...] the central idea behind probabilistic theories of causation is that causes *raise the probability* of their effects [...].

However, there are well know cases in which causes act as preventatives of their effects —i.e. causes that lower the probability of their effects— such as in Hesslow's birth control pills example, where the probability of suffering of thrombosis is both raised and lowered by birth control pills consumption (due to the fact that pregnancy is also a positive causal factor of thrombosis). See (Salmon, 1984) for a discussion.

cause is total or partial. In general, though, this information is not available, particularly when it comes to realistic examples. A decision on whether a cause is either total or partial can only be made provisionally in this case —perhaps on other empirical facts or just on intuitive grounds—, until new evidence is found to either confirm or dismiss our initial impressions.

It is remarkable moreover that whether a probabilistic cause is taken to be total or partial seems to make a difference when it comes to the status of certain techniques of causal inference, such as Reichenbach's Common Cause Principle or its generalisation, the Causal Markov Condition.[2] I shall not give a detailed argument justifying such a claim at this stage since I will be discussing Reichenbach's Common Cause Principle in detail in the following chapter. It is enough to point out for now that, as we shall see there, some of the most influential arguments against the validity of Reichenbach's Common Cause Principle for genuinely indeterministic systems —those are arguments along the lines of Cartwright's 'Cheap-but-Dirty/Clean-and-Green' factory example, which also aim to mirror the structure of quantum correlations— rely heavily on the assumption that the postulated common cause is the *only* causally relevant factor of the correlation, i.e. a total cause. It is a common view that such kind of examples show that Reichenbach's Common Cause Principle (and its generalisation, the Causal Markov Condition) fail for genuinely indeterministic systems, while seem to remain valid for deterministic or pseudo-deterministic systems (such as those where there is a lack of causal information). Such arguments suggest therefore that genuinely probabilistic total causes are problematic for causal inference criteria such as Reichenbach's Common Cause Principle. But, as we shall see in the more detailed discussion of Chapter 3, the same conclusion does not follow for probabilistic partial causes, even though they might be genuinely probabilistic causes. Thus, the event ontology assumed in our probabilistic models will be crucial regarding the validity of the causal inference techniques we aim to apply. Before embarking on such discussions at length let us make more precise the concepts and terminology we shall be needing for them.

[2]A related discussion may be found in a recently published paper by Suárez and the author (Suárez and San Pedro, 2007), where the Causal Markov Condition and Redhead's Robustness are shown to be closely related.

2.2 Token Events, Event Types and the Concept of Causation

The concept of event is probably one of the most crucial concepts in any account of causation, whether a probabilistic account or otherwise. But, what do we exactly mean by 'event'? and what is the role of events in causation? These are only two of the many questions regarding the nature of events. It is not my intention to give a precise answer to these and similar questions here. Nor is this needed for my present purposes. It will suffice to provide a general notion of event and make it precise what kind of events probabilistic accounts of causation seem to require.

2.2.1 Token Events and Event Types

Intuitively we may understand the word 'event' to mean 'something that happens' or 'something that takes place'. And it seems part of our intuitions as well that these are the 'things' causation is all about. In particular, we may say that causation is roughly about identifying 'cause-events' and 'effect-events' and unveiling the relation between them —in virtue of which such identification, i.e. 'cause' and 'effect', can be made—. A central issue for a proper understanding of causation is thus to understand precisely what 'events' are: whether there are different kinds of events and how these may be related, if at all; whether different kinds of events imply also different sorts of causal relations, whether causal relations can be sensibly defined only between the same kind of events, or may otherwise be also defined between the different kinds of events, etc. The literature dealing with such issues is extensive and, indeed, the proper characterisation of the notion of event constitutes a philosophical problem by itself. A philosophical problem that we do not need to address in detail here. Thus I shall not attempt to provide precise answers to the questions above. However, defining, even if broadly, the sense in which the term 'event' will be used here seems in order.

One of the most important distinctions in the literature as far as the notion of event is concerned is that between '*token* event' and '*type* event'. Token events are taken to be unique —in the sense that they can only occur once— particular events, located in space and time. As an example, take the sound (of a certain frequency and intensity) of the siren of the Oceana liner at the time of leaving the harbour this evening. Type events, by contrast, are generally understood as events which are not located in space and/or time and which can have several (if not many) instances. We might simply speak, for instance, of the sound of the Oceana siren when leaving the harbour (suppose, perhaps, that the Oceana liner regularly visits the harbour, and makes her siren sound at every departure).

The distinction between type and token events in the terms above, it is usually agreed, is an appropriate and useful distinction at an ontological level. Token events are usually identified with particular facts about a certain object (a physical system, for instance) taking place, while type events are taken to reflect properties of it.[3]

Another controversial matter is whether token events are more fundamental than types. Or even whether tokens may be taken as 'instances' of type events. Again, I shall not discuss these matters here since they are not crucial to what I want to say. Nevertheless, I will endorse throughout this thesis the common view that token events are 'instances' of event types. In other words, I will assume that event types can be reduced to sets of token events. Accordingly, I shall denote token events with bold capital letters (**A**, **B**, etc.), while their corresponding event types will be represented by slanted capital letters (A, B, etc.). Typically, a generic event type A will represent a collection or a set of token events $\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3, \ldots$, i.e. $A = \{\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3, \ldots\}$.

2.2.2 Token-level Causation vs. Type-level Causation

The relation between types and tokens is relevant to causation, because it is generally taken to refer to two different ontologies. This in turn suggests, when it comes to causation, that token events will only be (causally) related to other token events, while event types will typically be either causes or effects of other event types. Take for instance the first time little Oscar managed to shatter a window glass with his high pitched screaming voice. It was at a particular time and place that little Oscar emitted such a powerful sound for the first time; and it was a particular nearby window glass that got completely shattered some (well defined) time after Oscar's shout. This kind of causal relation involves token events only and is therefore usually referred to as *token-level* causation.

On the other hand, *type-level* causal influences do not relate particular events but just kinds of events or properties. Say for instance that little Oscar, having discovered his enormous talent to shatter glass with his voice, decided to make his living from it and joined a circus on tour. As part of every night's performance, little Oscar would emit his high pitched voice to shatter the glass fish tank (full of water) where a beautiful lady struggled to survive, breathlessly. Of course, every time the situation would be slightly different but, after all, little Oscar would manage every night to shatter the

[3]See for instance (Wetzel, 2007).

fish tank and free his lady. And this would in principle happen regardless of where and when a particular performance took place. In this case, it seems, we may just want to say that little Oscar is able to break glass fish tanks with his high pitched screaming voice. It seems then that little Oscar's voice has the capacity to break window glasses or fish tanks. But also, that the glass fish tanks and the windows have the capacity to break when little Oscar's high pitched voice is emitted.

Another matter of debate is whether token-level causation is more fundamental than type-level causation; and whether type-causation can be reduced to token-causation. The debate is of course a follow-up of the debate regarding the ontological priority of tokens over events. Although the debate is important, we do not need to address it here.[4] In this thesis I will often assume that causation is an event-type relation. The main reason is that whatever causation is going on in quantum mechanics, it must be probabilistic.[5] And most commentators agree that probabilistic causation is more suited to event types.[6] But nothing that I will say hinges on that.

So in what follows, the term *event* will refer to *event type*, unless stated otherwise; and expressions involving the word *cause* or *causal relation* must be interpreted in the framework of type-level causation unless otherwise pointed out. Note that this is perfectly consistent with the view that causation is fundamentally a relation between tokens that is then reflected in the probabilistic associations between types.

[4]Just to point out that, some authors take it that type-level causal relations can be reduced to token-level causal relations and are thus intimately related. This is the case, for instance, of (Cartwright, 1989) and (Hausman, 1998). Other authors, to the contrary, consider type and token causation to be completely independent relations and no such reduction can be made. (Sober, 1985) and (Eells, 1991) are perhaps the main proponents of this view.

[5]This is so even aside the question of whether or not quantum probabilities are indeed irreducible.

[6]Two short remarks seem in order that support this view. First, the concept of correlation, which is, as we will see in a moment, at the heart of probabilistic causal inference, is defined only between event types and never between tokens —these can actually be coincident, but never correlated. Second, even if a satisfactory account of correlation between token events could be provided, concepts such as that of Reichenbachian common cause (core to our discussion) may also turn out to be problematic if taken as token events. The point here has to do with the existence of the negation of a token event. It is fairly easy to think of the negation of a type event. However, the negation of a token event seems to be undetermined, since it usually amounts to many possible token events but not a single one. This is not to say that characterisations of probabilistic causality must be exclusively laid in terms of type-level causation. Indeed, accounts of token-level probabilistic causation have also been developed (see for instance Eells (1991, Ch. 6) for one such theory).

2.3 Probabilities and Causal Inference

A fundamental question we must address at this point is how can we know whether a certain statistical dependence or probabilistic relation indeed corresponds to a (probabilistic) causal relation (which on the reductionist picture would in turn allow us to unveil the underlying token causal structure). This is a central question any good account of probabilistic causation should be able to answer. It is here where the different available techniques of causal inference come into play. However, before trying to apply such techniques, we need first to identify unambiguously the probabilistic structures that may potentially reveal causal relations, as it is the case with correlations.

I must just add that I will not be defending any particular account of probabilistic causation here. It shall be enough to describe and discuss the main methods and techniques that have been developed in order to infer causal relations from probabilistic associations, and their range of applicability. Reichenbach's Common Cause Principle will serve as a main guide in this regard, as we shall see in the following chapters.

2.3.1 Sample Data, Probabilities and Correlations

Empirical data provides the main grounds for the inference of causal relations. Typically, data of experiments involving indeterminism —either as lack of complete knowledge about the system, or else due to genuinely stochastic events occurring— may show some patterns and regularities as well, although not in the form of constant conjunctions, as it is the case in regularity accounts of causation. Regularities in such cases only arise in the long run, after many repetitions of the same experimental set up, and can be expressed by means of probabilistic relations. That is, of course, provided that the appropriate techniques of statistical data analysis are applied to the relevant samples (i.e. the observed data). Statistical data analysis is mostly about interpreting the raw data sample so as to infer the process that generates it, i.e. the population, which is at the basis of the probabilistic model. It is common place as part of the statistical treatment of empirical data to extrapolate (large enough) finite sample behaviour to infinite size samples. In this way actual occurrence frequencies of the events that constitute the experiments are used to infer their occurrence probabilities (as their limiting frequencies). And the existing regularities and relations among such occurrences are similarly used to infer the corresponding probabilistic relations, i.e. the correlations. Thus, probabilities, as well as correlations, arise as a result of the embedding of empirical data samples into a model that allows for its interpretation, in the imaginary limiting case of infinite size samples.

The above is a very rough account on how empirical data in the form of occurrence frequencies are related to probabilities, but it shall suffice to stress that the idea of *correlation* arises, at least classically, in the framework of limiting frequencies, i.e. as associated to the idea of probability.[7] Failing to notice this may give rise to unwanted consequences and some misunderstanding, since mere associations in the data sample —perhaps due to the mixing of different populations— may be (wrongly) taken to be correlations.[8] And this is an important point for we shall rely on (genuine) correlations in order to infer causal relations. Indeed, correlations are at the heart of the idea of probabilistic causality.[9]

It will therefore be crucial for our purposes to identify real, *genuine* correlations among all associations in a data sample. We shall attempt to do so by means of the following definition:[10]

Definition 1 *Let $(\mathcal{S}, p)$ be a classical probability measure space with Boolean algebra $\mathcal{S}$ representing the set of random events and with the probability measure p defined on $\mathcal{S}$. If $A,\ B \in \mathcal{S}$ are such that*

$$p(A \wedge B) - p(A) \cdot p(B) > 0, \tag{2.1}$$

then the events A and B are said to be (positively) correlated, and we write $Corr_p(A, B)$.

[7]The quantum case, as we will see in Chapter 4, turns out to be a little bit more complex. This is mainly due to the fact that quantum probabilities need to incorporate typical quantum features, such as the existence of non-commuting observables, which will determine the algebra onto which the quantum probability measure is defined. I shall only speak of classical probabilities for now. This however does not amount to a loss of generality since, as we will see also in Chapter 4, all classical probability results relevant to our purposes have a quantum counterpart. See also the final remarks at the end of this section.

[8]This point has been nicely made by Hoover (2003). Hoover notes, for instance that correlations need to be *inferred* from the associations in the data sample, just like probabilities are inferred from occurrence frequencies. These, as we will see in the next chapter, are the grounds for Hoover's rejection of Sober's 'Venetian Sea Levels and British Bread Prices' putative counterexample to Reichenbach's Common Cause Principle.

[9]This is not to say, of course, that correlation suffices on its own to characterise causation, as will become evident in due course. However, most probabilistic theories of causation take the idea of statistical relevance, which is very closely related to that of correlation, as a starting point. (Statistical relevance relations express the fact that some event type A increases (or decreases) the probability for another event type B to take place.) The difference between the diverse theories are usually the extra assumptions needed to account for spurious correlations, the asymmetry of causal relations, etc.

[10]This definition follows the work by the Budapest School (see Chapter 3 for details). See, for instance, (Hofer-Szabó, Rédei and Szabó, 2002).

This definition makes it suitably clear that correlation is not a property of the raw data sample but a probability dependent notion. In other words, the idea of correlation depends on the model that interprets the data. This is made explicit in the definition above by requiring that the (correlated) events belong in a probability space $(\mathcal{S}, p)$. It is also remarkable that only events that belong in the very *same* probability space may be (genuinely) correlated. This is a crucial remark, since it means that two events A and B that are correlated in a particular probability model $(\mathcal{S}, p)$ may not be so in a different one $(\mathcal{S}', p')$. In fact, in most cases it is quite easy to imagine, for any correlation $Corr_p(A, B)$, an alternative probability space $(\mathcal{S}', p')$, $A, B \in \mathcal{S}'$, where the correlation 'disappears'.

Two further remarks regarding the definition of *genuine* correlation are in order. The first has already been made in passing in Footnote 7 (page 16) and has to do with the fact that the idea of correlation has been defined for the case of classical probability spaces $(\mathcal{S}, p)$. I would like to stress, however, that an appropriate corresponding 'quantum correlation' definition may also be given. However, for the sake of simplicity, I shall only deal with classical probability spaces for now and address the corresponding quantum case in more detail in Chapter 4.

Definition 1 above deserves a further appreciation regarding the 'sign' of so-defined correlations. In this respect Definition 1 can not be said to provide a completely general definition of correlation since it only refers to *positive* correlations, i.e. $Corr(A, B) > 0$. For complete generality, a symmetrical definition must be given of *negative* correlation $Corr(A, B) < 0$. However, this will not turn out to be crucial for my own purposes here. In particular, any further definitions and results involving positive correlations may also be assumed to hold for negative correlations. Therefore, in what follows, reference will be made to positive correlations only, unless explicitly pointed out.

2.3.2 *Purely Formal* and *Genuinely Physical* Correlations

I insisted in the previous section that when it comes to causal inference from probabilistic relations it is crucial to identify what I called *genuine* correlations. I shall recall that for a correlation to be genuine I simply required that it conformed to Definition 1. The key point was, in particular, that the correlated events A and B both belonged to the *same* probability space, i.e. the *same* Boolean algebra $\mathcal{S}$ with the *same* probability measure p.

Identification of *genuine* correlations, although necessary, turns out to be insufficient to warrant adequate causal inference. Whether there exist sufficient conditions for causal inference is indeed a topic of intense debate in the contemporary literature, which I shall not need to address here as such. In the next chapter I will discuss in detail one of the most influential efforts in this regard, however, namely Reichenbach's Common Cause Principle. But before any causal inference techniques are applied we need to analyse further the concept of correlation, in order to establish whether all genuine correlations (as defined in the previous section) are indeed eligible for causal inference. With the foregoing analysis I aim in particular to draw a distinction among *genuine* correlations. I will distinguish in particular between two kinds of *genuine* correlation, those for which we would naturally expect causal explanations, and those that do not seem to have an underlying causal structure. Such a distinction is best appreciated with the aid of a simple example.

Consider the following situation. Suppose that we have a new (very simple) device consisting of a switch, a lamp and a speaker. We expect, in principle, that when the switch is set to position 'ON' the light bulb flashes and a sound is emitted. Suppose however that, although we can manipulate the switch, we do not have direct access to the sound when the switch is set to 'ON', nor do we have direct access to the light. In other words, we cannot directly hear or see to what the device does when the switch is set to 'ON' (say for instance that both the light and the speaker are in a separate and completely isolated room). We can only find out whether the device is working correctly with the aid of two detectors, a photoelectric cell and a sound detector, whose readings are available to us. We may assume for simplicity that both detectors have negligible experimental error.[11]

Now suppose, for the sake of the argument, that in the first test, which I shall call Experiment 1 (EXP1), both detectors, i.e. the photoelectric cell and the sound detector, cannot be set to work simultaneously (perhaps for technical reasons). The first test (EXP1) then consists of a series of runs in which the switch is set to 'ON' (and then back to 'OFF'). Since we cannot use both detectors at once in a single run, we must choose whether to test the sound or the light working correctly every time. Suppose that we decide in a somewhat random way what to measure in each run, and record the corresponding outcomes.

The records may show that, for some reason, in setting the switch to 'ON', if the photoelectric cell was used, the lamp did not always lit up, i.e. in some cases no light was detected. Similarly, the sound detector gave zero output in some of the runs in which it was set to work. These results may

[11] We do not really need to assume this but, as I said, it makes things a bit simpler.

be expressed in terms of probabilities (after the appropriate statistical data analysis if carried out). We might find for instance that the event 'the lamp lites up when setting the switch to the 'ON' position', which I shall refer to simply as 'light', occurs in Experiment 1 (EXP1) with probability P^1_L.[12] On the other hand, we might have observed that the event 'the sound is emitted when setting the switch to the 'ON' position' (i.e. 'sound'), occurs in Experiment 1 (EXP1) with probability P^1_S. That is,

$$p^1(\text{`light'}) = P^1_L,$$
$$p^1(\text{`sound'}) = P^1_S.$$

Moreover, in this experiment we know for sure that an event such as 'the lamp lites up *and* the sound is emitted when setting the switch to the 'ON' position' cannot be observed. (This might be a result of the mentioned technical constraints). Thus,

$$p^1(\text{`light'} \wedge \text{`sound'}) = 0.$$

The above probabilities show that there is a (negative) correlation between the events 'light' and 'sound':

$$Corr^1(\text{`light'}, \text{`sound'}) \neq 0. \tag{2.2}$$

How should we then take such a correlation? Shall we try to explain it in causal terms? The first thing to note is that the above correlation is a *genuine* correlation, as defined in the previous section (see Definition 1).[13] But still, is that correlation intuitively due to some underlying causal structure operating in our device? Certainly not.[14] For the above correlation clearly arises in the context of this particular description of our experimental data, i.e. in this particular model of the data. Furthermore, it does not seem to contain any relevant information as to whether the two correlated events are

[12]It is convenient for reasons that will become evident in a moment to include in the expressions involving probabilities a label (super-index) representing the particular experiment performed (super-index '1' represents EXP1 in this case).

[13]Note in particular that correlations such as (2.2) involve events that belong in the same probability space $(\mathcal{S}, p^1)$.

[14]This answer is of course based on the information we have from the experiment. One could perhaps say that it seems intuitively possible that there is a (causal) relation between'light' and 'sound', but such intuition is simply mistaken in this case. For it responds to some extra (standard) knowledge about the way similar devices work. But, looking only to our recorded experimental data —and not having more information than that given by our measurement devices (however limited they turn out to be)—, we cannot infer any causal structure form correlations such as the above.

also causally, or even physically, related. In particular, the correlation seems to have its origin in the fact that the model used for the description of our experiment includes measurements of two properties that are 'incompatible', in the sense that they cannot be measured simultaneously. (This is simply because two or more 'incompatible' events have zero joint probabilities.) What this means is that the correlation will be completely washed out as soon as the 'incompatibility' disappears. This can be easily accomplished in our case by proposing an alternative model —an alternative probability space $(\mathcal{S}', p^{1'})$, that is— in which there are no incompatible events.

Thus, we see that correlations can sometimes be some sort of unavoidable 'by-product' of some modelling processes. In particular, such correlations do not reflect any specific (physical) property of the system which would in turn suggest that there is an underlying causal structure. This is why I shall refer to them as *purely formally descriptive* or simply *purely formal* correlations, in contrast to *genuinely physical* correlations.

The distinction may be even better seen with the aid of a modified version of our example. Consider the same device as above. There is now a crucial difference however, namely that a method (or another measurement device) has been developed such that both the lamp lighting up and the sound being emitted can be simultaneously detected. Our second experiment (EXP2) then consists in setting the switch to position 'ON' and seeing, by means of our newly developed light-sound detector only, whether the events 'light', 'sound', both 'light' and 'sound', or none of the above are detected. As in the example above we repeat the procedure several times and obtain probabilities corresponding to the four possible events:

$$\begin{aligned} p^2(\text{`light'} \wedge \neg\text{`sound'}) &= P^2_{L\_}, \\ p^2(\neg\text{`light'} \wedge \text{`sound'}) &= P^2_{\_S}, \\ p^2(\text{`light'} \wedge \text{`sound'}) &= P^2_{LS}, \\ p^2(\neg\text{`light'} \wedge \neg\text{`sound'}) &= P^2_{\_\_}. \end{aligned}$$

Again the super-indexes '2' in the above expressions express the fact that the probabilities correspond to Experiment 2 (EXP2). The above probabilities will typically entail that there is a correlation between 'light' and 'sound', as in the case of our first experiment:

$$Corr^2(\text{`light'}, \text{`sound'}) \neq 0. \tag{2.3}$$

There is however a crucial difference between correlations (2.2) and (2.3). For the correlation observed in this second experiment does not involve incompatible events, as it was the case in our first experiment. This is a hint that the correlation between 'light' and 'sound' is a feature of the physical system, and not a product of our measurement limitations. In fact, in this experiment, since the measurement procedure is unique, i.e. we have used a single detector, measurement can be seen to be completely unrelated to the fact that the joint probability of 'light' and 'sound' be different to the product of the individual probabilities, i.e. to the fact that the correlation arises. Therefore, we may conclude that the correlation does indeed reflect some existing (physical) relation between these two properties of the system. Thus, these are the kind of correlations, which I shall call *genuinely physical* correlations, that call for causal explanations. In other words, we shall only attempt to infer causal relations from *genuinely physical* correlations, so-characterised. *Purely formal* correlations, on the other hand, must not be interpreted as causal at all since they need not be in any way the result of the physically underlying features of the data.

2.3.3 Mixing *Purely Formal* and *Genuinely Physical* Correlations

In the two examples above *purely formal* and *genuinely physical* correlations were easily distinguished, partly due to the fact the only one such type of correlation was present in each case (i.e. *purely formal* correlations in EXP1 and *genuinely physical* correlations in EXP2). There are however other cases where it is not so easy to distinguish the two kinds. For instance, more sophisticated versions of the examples above can be proposed in which *purely formal* correlations may be misleadingly taken to be *genuinely physical*. I find of particular interest the cases in which *purely formal* and *genuinely physical* correlations mix-up in the description of a particular experimental set-up.

In order to see how this may happen we may think once more in experiments similar to those in the preceding section. Consider again the 'switch-lamp-speaker' device. Suppose that we decide (for no particular reason though) to use our three detectors for a third experiment (EXP3). As in the first experiment, only one of the three detectors may be put to work in each run (and the choice as to which detector to use in each run is somehow random). We will thus have runs in which only the event 'light' (and ¬'light') can be detected, runs in which it is only the even 'sound' (and ¬'sound') that is detected, and finally runs in which both events (and their negations) may be detected simultaneously, as in EXP2.

Again, the corresponding outcomes are recorded and the following probabilities derived:

$$\begin{aligned} p^3(\text{‘light’} \wedge \neg\text{‘sound’}) &= P^3_{L\_}, \\ p^3(\neg\text{‘light’} \wedge \text{‘sound’}) &= P^3_{\_S}, \\ p^3(\text{‘light’} \wedge \text{‘sound’}) &= P^3_{LS}, \\ p^3(\neg\text{‘light’} \wedge \neg\text{‘sound’}) &= P^3_{\_\_}. \end{aligned}$$

As in the two other examples, the above probabilities will typically entail a correlation between ‘light’ and ‘sound’:

$$Corr^3(\text{‘light’}, \text{‘sound’}) \neq 0. \tag{2.4}$$

Once more, we may ask whether this correlation is to be explained causally or not. We may try to address the issue now by appealing to our distinction between *purely formal* and *genuinely physical* correlations. Thus we shall ask instead whether correlation (2.4) is a *genuinely physical correlation* or not. Intuitively the we may think that it is so, particularly if our experiment allows us measure the joint event ‘light and sound’, i.e. ‘light’ and ‘sound’ can be measured simultaneously. However, the issue does not seem to be as simple as that. Let us see why.

As we can see above, the probabilities found in this third experiment (EXP3) are structurally just the same than those in EXP2, meaning that they correspond to the same observed events. However, they will typically turn out to be numerically different. As a consequence, the induced correlation (2.4) between ‘light’ and ‘sound’, although structurally equivalent to correlation (2.3) in EXP2, will be of a different strength, i.e. will have a different numerical value. That is

$$Corr^3(\text{‘light’}, \text{‘sound’}) \neq Corr^2(\text{‘light’}, \text{‘sound’}).$$

And the reason for this is very simple, namely that in this third experiment, the resulting correlation between ‘light’ and ‘sound’ includes statistical information both about runs in which ‘light’ and ‘sound’ are completely compatible (as in EXP2) and runs in which they are incompatible (as in EXP1). What this means is that $Corr^3(\text{‘light’}, \text{‘sound’})$ contains both *genuinely physical* and *purely formal* (correlational) information. We may then be inclined to say that, since $Corr^3(\text{‘light’}, \text{‘sound’})$ reflects some *genuinely physical* relation between two properties of the system, it seems desirable that it must be explained in causal terms. But if so we must be aware that our causal story may not be such that it fully recovers, so to speak, the correlation. Or in other words, we should not expect to explain the full correlation

$Corr^3$('light', 'sound') in causal terms only. For it contains also information about our particular description of the data (our particular probability model). And, as we saw, this *purely formal* information does not correspond to any physical feature of the system (and thus shall not be accounted for causally).

We must thus keep this distinction in mind —as well as its consequences— when addressing causal inference in the following chapters, specially when discussing correlations in quantum theory, some of which may turn out to be structurally similar $Corr^3$('light', 'sound'), and contain therefore *purely formal* irreducible features.

A brief comment in this regard is in order before closing this chapter. It is well known that quantum mechanics features 'incompatible' observables, in the sense above. In particular, certain physical properties of a system cannot be measured with full precision simultaneously, according to Heisenberg's Uncertainty Principle.[15] We may then wonder whether this 'quantum mechanical' incompatibility is of the same kind as that described in the previous section, and if this is not the case whether it makes a difference as regards the correlations that it may induce.

I think that the answer to the first question is quite straight forward. Quantum mechanical 'incompatibility' is an exclusive feature of quantum systems and is related to measurement operations. However, the measurement limitations that give rise to such 'quantum mechanical' incompatibility are seen (in the orthodox interpretation of the theory) to be inherent to the system, as a consequence of its irreducible indeterministic character. In principle, therefore, there is no possible experiment in which two 'quantum mechanical' incompatible events can be simultaneously measured. In other words, two 'quantum mechanical' incompatible events are experimentally incompatible because they are so at a fundamental physical level. This in clear contrast to the incompatibility described in the previous section.

However, it does not seem to make a difference as regards the physical interpretation of the corresponding induced correlations whether two events are 'incompatible' in either sense. In particular, whenever we are not able to measure two events simultaneously, for whatever reason, be it operational or due to the limits of the physical system itself, *purely formal* correlations will arise. As we will see in Chapter 5 this turns out to be an important issue when discussing a recent derivation of the Bell inequalities.

[15]Incompatibility is the standard terminology in the philosophy of physics literature. Some authors, however, refer to such incompatible observables as *complementary*, a term originally coined by Bohr in the early years of the theory, or *conjugate*, most often found in the physics literature.

Chapter 3

Reichenbach's Common Cause Principle

This chapter is entirely devoted to what it is commonly known as Reichenbach's Common Cause Principle (RCCP). This is perhaps an unfortunate terminology since, as we will see shortly, the expression 'Reichenbach's Common Cause Principle' incorporates two different independent claims.[1] The first claim refers to the existence of common causes (in certain cases) and thus provides a metaphysical guide for causal inference. The second claim, on the other hand, provides a methodological rule for the characterisation of the common causes postulated by the first metaphysical claim.

The clear cut distinction of these two aspects of Reichenbach's Common Cause Principle (RCCP) will be seen to be crucial when assessing its status. In discussing the methodological content of the RCCP I will address its difficulties to stand either as a sufficient or necessary condition for causation. I shall discuss various influential counterexamples in this respect, particularly Cartwright's 'Cheap-but-Dirty/Green-and-Clean' factory example.

The view I shall defend in the remainder of my thesis, however, does not rely on the sufficiency, nor the necessity, of RCCP for common causes. For I want to suggest that we may take advantage of the results that can be obtained by applying RCCP in favourable situations in order to guide our causal inference in more dubious cases, such as the EPR correlations in quantum mechanics. In order to do so we will need to assume the metaphysical part of RCCP to hold, which demands for a careful assessment of its status. I will consider at the end of the chapter Sober's arguments against it and discuss his well known 'Venetian Sea Levels and British Bread Prices'

[1] Such terminology has in fact been a source of some confusion and misunderstanding, specially as regards to its status. This has been lately stressed by Suárez (2007).

example.

3.1 Reichenbach's Common Cause Principle

Reichenbach introduced what has come to be known as Reichenbach's Common Cause Principle (RCCP) in an attempt to characterise the arrow of time. His aim was to *reduce* the direction of time to the asymmetry of causation. Reichenbach assumed that since causes precede their effects in time, causal asymmetry provides a natural characterisation of temporal asymmetry. It is therefore a characterisation of causal asymmetry —by identifying *the* causes and *the* effects— that the principle provides in a first stage.

3.1.1 Correlations and Common Causes

It is not striking at all to observe that the two lamps in my room usually light up and go off simultaneously. Nor would it surprise us on checking a barometer to see that almost every time the reader falls a storm takes place shortly after. Or to imagine two geysers separated at some distance that erupt at irregular intervals but usually almost exactly at the same time.

The above are all very common cases of non-surprising coincidences, examples of which can be found in our every day life. What is remarkable in all those cases is that two apparently causally unrelated events are observed to occur simultaneously with some regularity. Examples as the above are very common and the type of regularities they show led Reichenbach to the idea of common cause. Reichenbach writes in reference to such examples:[2]

> The schema of this reasoning illustrates the rule that the improbable should be explained in terms of causes, [...] The logical schema that governs it may be called the *principle of the common cause*. It can be stated in the form: *If an improbable coincidence has occurred, there must exist a common cause.*

Two remarks are in order regarding the quotation above. First, the 'events' Reichenbach has in mind are not singular space-time localised events, i.e. *token* events, but instances of statistically well defined event *types*.[3] 'Improbable coincidences' therefore may be readily characterised in terms of event type *correlations*, as defined in the previous chapter (see Definition 1

[2](Reichenbach, 1956, p. 157).

[3]I pointed out in the previous chapter that event *types* may be considered to be sets of singular *token* events. This also allows for the corresponding probabilities to be assigned.

in page 16). Not any kind of correlation, however. For the idea of correlation just responds to the objective fact that some sort of 'coincidence' has taken place, but it is insensitive as to whether such 'coincidence' is 'improbable'. By 'improbable' then Reichenbach was referring to some particular sort of correlations, i.e. those representing 'coincidences' which are not to be expected, on the grounds perhaps of a direct causal relation among the correlated events. In this way, Reichenbach postulates the existence of common causes only for those correlations which cannot be traced back to a direct cause link. This will be explicitly stated when formalising Reichenbach's Common Cause Principle in Section 3.1.3.[4]

The second remark has to do with the idea of 'common cause'. The quote above expresses a metaphysical claim, which suggests that 'common causes' are to be found whenever 'improbable coincidences' occur. This claim *per se* says nothing about what probabilistic relations common causes must obey. As such, following Reichenbach and in virtue of its metaphysical content, I shall refer to it as the *Principle of the Common Cause* (PCC).

The Principle of the Common Cause (PCC) on its own says nothing about the nature of the 'common causes' themselves. As pointed out, it cannot be said from the quotation above how these 'common causes' may be identified or characterised in statistical terms. This is accomplished by another criterion introduced by Reichenbach. The *fork* formed by the correlated events A and B together with the common cause C, we are told, must satisfy certain probabilistic relations:[5]

> In order to explain the coincidence of A and B, which has a probability exceeding that of a chance coincidence, we assume that there exists a common cause C. [...] We will now introduce the assumption that the fork ABC satisfies the following relations:

$$\begin{aligned} p(A \wedge B|C) &= p(A|C) \cdot p(B|C), \\ p(A \wedge B|\neg C) &= p(A|\neg C) \cdot p(B|\neg C), \\ p(A|C) &> p(A|\neg C), \\ p(B|C) &> p(B|\neg C). \end{aligned}$$

[4]Moreover we shall later explicitly require that such 'improbable coincidences' are *genuinely physical* correlations in the sense characterised in the previous chapter. That is, correlations which are able to reflect properties inherent to the system itself. Recall that I distinguished *genuinely physical* correlations as opposed to *purely formal*, which I said arise from the manipulation of the statistical data in order for it to be embedded in a probabilistic model, i.e. in the process of modelling the data (see Chapter 2, Section 2.3.2 for details). Reichenbach did not make this requirement explicit but it may be taken to be implicit in at least any defensive version of RCCP.

[5](Reichenbach, 1956, p. 159).

Where $p(X|Y)$ represents the conditional probability of X on Y (with $X = C, \neg C$ and $Y = A, B$), and $X \wedge Y$ represents the joint event "X and Y".[6]

Following Suárez[7] I shall refer to the above four probabilistic relations as *Reichenbach's Criterion for Common Causes* (RCCC) as separate from the statement of the PCC. This criterion also deserves some comments. Note in the first place that the two last expressions are just statistical relevance relations and, as such, are aimed to represent the causal dependence between A and C on the one hand, and B and C on the other. Whether statistical relevance does indeed reflect causal dependence is of course a debatable issue, as we have seen in the previous chapter. However, I pointed out that statistical relevance may be taken as a indicative for causal relations, as it is standard in most theories of probabilistic causation.

As for the first two probabilistic conditions, they express a new restriction on the postulated common cause C. They require, specifically, that if the presence (or the absence) of the common cause is taken into account the correlated events A, B are rendered probabilistically independent. In some sense the common cause C can be said to *screen-off* the correlation $Corr(A, B)$. In fact the first two expressions in the quotation above can be shown to be equivalent to what is commonly known in the literature as the *screening-off* condition, which is probabilistically expressed as $p(A|C \wedge B) = p(A|C)$.[8]

Finally note that Reichenbach explicitly points out that the four statistical relations explain the correlations between A and B in two senses. First, it is noted that the four relations entail that A an B are (positively) correlated, i.e. $Corr(A, B) > 0$ —indeed, this is why Reichenbach refers to the fork ABC as a *conjunctive fork*—. On the other hand, Reichenbach claims that a common cause C satisfying these four relations explains the correlation by rendering A and B statistically independent. In his own words:[9]

> When we say that the common cause C *explains* the frequent coincidence, we refer not only to this derivability of relation (1) $[P(A \wedge B) > P(A) \cdot P(B)]$, but also to the fact that relative to the cause C the events A and B are mutually independent: a *statistical dependence* is here derived from an independence. The common cause is the connecting link which transforms an independence into a dependence.

[6]Note that I do not follow here Reichenbach's original notation but the notation I will be adopting throughout this thesis.

[7](Suárez, 2007).

[8]A simple proof of the equivalence is presented in Appendix A.

[9](Reichenbach, 1956, p. 159). The square brackets in this quotation are mine.

It is remarkable that for Reichenbach the role of his criterion for common causes, and of the screening-off condition in particular, is explanatory. More specifically, what is crucial about the screening-off condition is that common cause explanations that satisfy it seem explanatorily quite powerful. The explanatory power of screening-off common causes, I will argue, may serve as a methodological motivation for relying on Reichenbach's Criterion for Common Causes in the attempt to infer causal information from probabilistic facts, even if it will be patently shown no to hold as a necessary nor a sufficient condition for causation.

In what follows, I will refer to the conjunction of the Principle of the Common Cause (PCC) and Reichenbach's Criterion for Common Causes (RCCC) as *Reichenbach's Common Cause Principle* (RCCP). Thus, whenever Reichenbach's Common Cause Principle (RCCP) is mentioned it will include *both* claims above. Properly distinguishing the two claims is crucial, however, for the assessment of the status of RCCP as a whole, as we will see in the following section.

There is a certain amount of confusion in the literature regarding the terminology and a clarifying note is perhaps in order. In some cases the expression 'Reichenbach's Common Cause Principle' is used to refer just to the four probabilistic relations, i.e. Reichenbach's characterisation of the postulated common causes, which I called Reichenbach's Criterion for Common Causes (RCCC). Both claims are also most often found in the literature under the name 'Principle of the Common Cause' or 'Common Cause Principle'. Since I shall reserve the expression 'Principle of the Common Cause' for the claim in the first quoted passage —on the grounds of its metaphysical significance—, I shall stick to 'Reichenbach's Common Cause Principle', as I said before, when referring to the conjunction of *both* claims.

The main ideas entering RCCP are, obviously the idea of *correlation*, the notion of *screening-off* and finally, as it has been implicitly assumed in both Reichenbach's definition and the discussion above, the fact that common causes must temporally precede both the correlated events.

The concept of correlation was already discussed in the previous chapter (see Section 2.3), and I will focus on the screening-off condition in the following sections. Concerning the temporal precedence of the common cause with respect to the correlated events, I shall only make some brief remarks, since my main interest is to study some aspects of causation that do not essentially depend on temporal asymmetry.

First, it is worth pointing out that this condition is imposed merely on phenomenological observations, such as the ones in the examples at the beginning of this section, and later formalised in terms of the idea of *conjunctive*

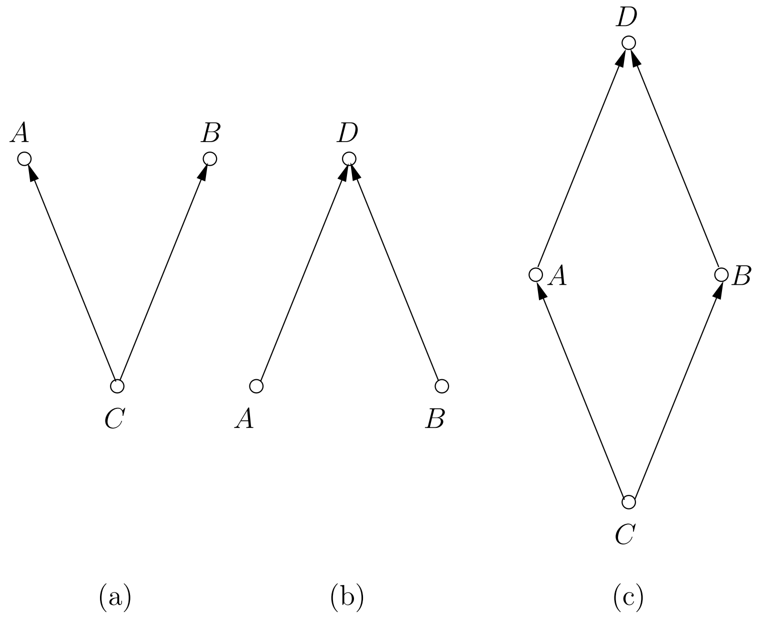


Figure 3.1: Possible conjunctive forks: (a) open to the future, (b) open to the past and (c) closed conjunctive fork.

fork.[10]

As it has been pointed out, events A, B and C that satisfy Reichenbach's Criterion for Common Causes define what it is known as *conjunctive fork* (see Figure 3.1). Reichenbach's intuition was that common causes always appeared in *forks open to the future* or simply *open conjunctive forks* —as in Figure 3.1 (a)—, while a common effect could be represented by a *fork open to the past* or *closed conjunctive fork* (see Figure 3.1 (b)). So a temporal direction was defined in terms of statistical relevance and forwards causation. As I already mentioned, this was Reichenbach's main aim when proposing his common cause principle. Reichenbach's preference for open conjunctive forks, however, is a matter of controversy. In fact, some authors have pointed that there is a tension between the temporal order imposed by the idea of conjunctive fork and the time-symmetry characteristic of microphysics.[11] I shall not worry about such issues here, however, and I shall assume throughout that causes (temporally) precede their effects. This is not to say, however, that the possibility of backwards in time causation is ruled out.[12] In fact, I

[10]The idea of *conjunctive fork* is introduced formally further on, along with the definition of common cause (see Definition 3 in page 33).

[11](Price, 1996) is one of the main defenders of such a view.

[12]Assuming that causes temporally precede their effects can be seen somehow as a

am almost certain that some of the foregoing arguments are consistent with a backwards causation interpretation. I will return to the issue in relation to the common cause model for quantum EPR correlations proposed in the final chapter of my thesis.

3.1.2 The Principle of the Common Cause and Reichenbach's Criterion for Common Causes

Although both the Principle of the Common Cause (PCC) and Reichenbach's Criterion for Common Causes (RCCC) are causal claims they have a completely different philosophical significance. On the one hand, the PCC is a metaphysical claim about the existence of common cause events. On the other hand, RCCC is a methodological claim which, although complementary to the PCC, is logically independent from it. In particular, while the PCC tells us about the ontology of the possible causal structure underlying the observed correlation between A and B, Reichenbach's criterion aims to provide the tools for the characterisation of such causal structure. In other words, the four probabilistic relations that define a conjunctive fork ABC are a more or less happy characterisation of the common causes postulated by the first (metaphysical) claim.

The difference between the two claims also becomes clear if we look carefully at the Causal Markov Condition which, as I pointed out in the preceding chapter, is closely related to Reichenbach's Common Cause Principle. The Causal Markov Condition (CMC) is defined, following Hausman and Woodward as[13]:

Causal Markov Condition (CMC): *For all distinct variables X and Y in the variable set $\mathbf{V}$, if X does not cause Y then*

$$p(X|Y \wedge \mathrm{Par}(X)) = p(X|\mathrm{Par}(X)),$$

where $\mathrm{Par}(X)$ *(read* parents of X*) is the set of all direct causes of X in $\mathbf{V}$.*

Now Hausman and Woodward[14] also note that the CMC may be written as two different independent claims:

background premise. However, I shall take such assumption with enough flexibility to allow for it to be updated, perhaps on empirical grounds. Indeed, I believe that whether backwards in time causal influences exist is an empirical matter to be decided by the facts and their most intuitive interpretation.

[13](Hausman and Woodward, 1999).

[14](Hausman and Woodward, 1999, p. 524).

CMC 1: *If A and B in $\mathbf{V} = \{A, B, C, D, \ldots\}$ are probabilistically dependent, then either A causes B, B causes A or A and B are both effects of some **common cause** C in $\mathbf{V}$.*

CMC 2: *If* $\mathrm{Par}(A) \neq \emptyset$*, then conditional on* $\mathrm{Par}(A)$*, A is independent of every other variable in $\mathbf{V}$ except from its effects.*

CMC1 and CMC2 are each generalised versions of the corresponding claims in Reichenbach's Common Cause Principle.[15] For instance, while CMC1 establishes how correlations may be explained —that is, in terms of direct causal interactions of the correlated events or else by postulating a common cause—, it says nothing about how can the postulated common causes be characterised, which is what CMC2 does. In fact, Reichenbach's second claim —which include the four probabilistic relations the fork ABC must satisfy— can be seen to be a particular instance of CMC2 if one considers the set of variables $\mathbf{V}$ to be made up by A, B and C. That is

$$\mathbf{V} = \{A, B, C\}.$$

Now, if the set $\mathrm{Par}(A)$ of all parents of A contains *only* the postulated common cause, i.e. $\mathrm{Par}(A) = \{C\}$, the screening-off conditions required by Reichenbach immediately follow from CMC2.[16]

The distinction above suggests that the status of Reichenbach's Common Cause Principle needs to be addressed separately at the metaphysical and methodological levels. This is what I aim to do in the remaining of this chapter. Before proceeding, however, I shall provide a further formalised version of Reichenbach's proposal.

3.1.3 Formal Version of Reichenbach's Common Cause Principle

A precise definition of the concept of common cause will prove very useful when discussing, not only the status of RCCP in the following sections, but

[15] The sense in which CMC1 is a more general statement than that of Reichenbach (my first quoted Reichenbach's claim in page 26) has to do with the expression 'improbable coincidences'. It is by referring to direct causation that the expression 'improbable coincidence' is better or more precisely defined in CMC1. More in particular, CMC1 defines, so to speak, 'improbable coincidences' as those correlations, i.e. probabilistic dependencies, that cannot be explained by direct causation between the correlated events. As I said before this is not exactly made explicit by Reichenbach.

[16] The first screening-off condition follows by simply noting that B is not an effect of A, nor is B in the set of its parents $\mathrm{Par}(A) = \{C\}$. It follows therefore that, conditional on C, A and B are probabilistically independent, i.e. $p(A \wedge B|C) = p(A|C) \cdot p(B|C)$. The second screening-off condition, i.e. $p(A \wedge B|\neg C) = p(A|\neg C) \cdot p(B|\neg C)$, follows as well in a similar fashion.

also quantum correlations, in Chapters 4, 5 and 6. The formalisation that I present here follows the work by Hofer-Szabó, Rédei and Szabó in late 1990's and early 2000. Following Butterfield I shall refer to this particular philosophical program as the Budapest School.[17]

As a side remark, the results given in this and the following sections (specially in Section 3.4) do not in general hold for all probability spaces. They do hold, however, for the cases that are relevant to us, i.e. for both classical and von Neumann probability spaces.[18] The results will therefore be applicable to both classical and quantum correlations. Bearing this in mind, in what follows I shall only refer for the sake of simplicity to classical probability spaces, unless explicitly stated otherwise.

Recall now the definition of (positive) *correlation* given in the previous chapter[19]:

Definition 2 (Correlation) *Let $(\mathcal{S}, p)$ be a classical probability measure space with Boolean algebra $\mathcal{S}$ representing the set of random events and with the probability measure p defined on $\mathcal{S}$. If $A,\ B \in \mathcal{S}$ are such that*

$$p(A \wedge B) - p(A) \cdot p(B) > 0,$$

then the events A and B are said to be (positively) correlated, and we write $Corr_p(A, B)$.

A *Reichenbachian common cause* of the correlation $Corr_p(A, B)$ may now be unambiguously characterised as follows:[20]

Definition 3 (Reichenbachian Common Cause) *$C \in \mathcal{S}$ is a **Reichenbachian common cause** of the correlation $Corr_p(A, B) \equiv p(A \wedge B) - p(A) \cdot p(B) > 0$ if the following (independent) conditions hold:*

$$p(A \wedge B|C) = p(A|C) \cdot p(B|C) \tag{3.1}$$
$$p(A \wedge B|\neg C) = p(A|\neg C) \cdot p(B|\neg C) \tag{3.2}$$
$$p(A|C) > p(A|\neg C) \tag{3.3}$$
$$p(B|C) > p(B|\neg C) \tag{3.4}$$

where $p(A|B) = p(A \wedge B)/p(B)$ denotes the probability of A conditional on B and it is assumed that none of the probabilities $p(X)$ $(X = A, B, C, \neg C)$, is equal to zero.

[17] Cf. (Butterfield, 2007). See (Hofer-Szabó, Rédei and Szabó, 1999, 2000*b*, 2002) and (Rédei, 2002) for the main results of the Budapest School program.

[18] Von Neumann probability spaces are the appropriate structures for quantum (both relativistic and non-relativistic) probabilities and correlations.

[19] See Definition 1 in page 16.

[20] See for instance (Hofer-Szabó, Rédei and Szabó, 2000*b*, p. 1).

Note that the definition of Reichenbachian common cause above refers to positive correlations. However, the definition can be generalised for negative correlations as well. In what follows, and again for the sake of simplicity, similar definitions and results, will be given for positive correlations only.[21]

One can see that the definition of common cause above is nothing more than a further formalised version of the original criterion offered by Reichenbach for the characterisation of the postulated common causes, which I called Reichenbach's Criterion for Common Causes (RCCC). The definition of Reichenbachian Common Cause may be further formalised at the level of events in the Boolean algebra so that common causes may be viewed as events in the algebra with some imposed restrictions, i.e. the screening-off conditions.[22] Here, the triple (A, B, C) of events in $\mathcal{S}$ is what I referred to as an *conjunctive fork* (see Figure 3.1), provided the conditions in Definition 3 hold.

Note as well that Definition 3 does not make any explicit reference to the metaphysical content of Reichenbach's Common Cause Principle (RCCP). Instead, the PCC is implicitly assumed, i.e. Definition 3 presupposes the existence of common causes —at least one common cause, that is— of the correlation $Corr_p(A, B)$. Thus the existence of Reichenbachian Common Causes is committed to both the principle and Reichenbach's criterion. We may then restate Reichenbach's Common Cause Principle (RCCP) as follows:

Definition 4 (RCCP) *For any two (positively) correlated event types A and B ($Corr_p(A, B) > 0$), if A is not a cause of B and neither B is a cause of A, there exists a common cause C of A and B such that the following independent conditions hold:*

$$p(A \wedge B|C) = p(A|C) \cdot p(B|C) \tag{3.5}$$

$$p(A \wedge B|\neg C) = p(A|\neg C) \cdot p(B|\neg C) \tag{3.6}$$

$$p(A|C) > p(A|\neg C) \tag{3.7}$$

$$p(B|C) > p(B|\neg C) \tag{3.8}$$

where $p(A|B) = p(A \wedge B)/p(B)$ denotes the probability of A conditional on B and it is assumed that none of the probabilities $p(X)$ $(X = A, B, C, \neg C)$ is equal to zero.

At this point an interesting question that one may pose is whether Reichenbachian common causes exist of any given correlation $Corr_p(A, B)$ in $\mathcal{S}$.

[21] This, nevertheless, will not constitute a loss of generality whatsoever since all such definitions, as well as the corresponding results, can be shown to hold as well for negative correlations. See also my remarks following the definition of (positive) correlation (Definition 1 in Section 2.3 of Chapter 2).

[22] See (Hofer-Szabó, Rédei and Szabó, 1999; Rédei, 2002) for a more detailed account of this approach.

In case of a positive answer, we would also like to know under what conditions this might be. To my mind, this question represents somehow a shift the role Reichenbach's Common Cause Principle (RCCP) is taken to have regarding causal inference. The usual approach mainly focuses on the study of RCCP either as a sufficient or as a necessary condition for causation. Examples in this regard are very common and are usually devised to show that both sufficiency and necessity of RCCP are unlikely in most of the cases. I shall discuss some such examples in the following section. The aim to show that the new approach suggested by the question above —as to whether Reichenbachian common causes may be found for any given correlation— introduces a new perspective in the problem of what the role of RCCP is in causal inference.

3.2 On the Philosophical Status of the Screening-off Condition

The distinction introduced in Section 3.1.2 between the metaphysical and methodological claims of Reichenbach's Common Cause Principle (RCCP) allows us to view issues regarding its philosophical status from a better perspective.

Most counterexamples to RCCP found in the literature are aimed at showing that the screening-off condition, the core of Reichenbach's methodological recipe for the characterisation of common causes, does not hold. In such cases it is thus claimed that RCCP does not hold either. However, the status of RCCP may (and perhaps must) be assessed independently at the metaphysical and the methodological level. In particular, independently addressing the metaphysical and the methodological status of RCCP one might find that failure of Reichenbach's Criterion for Common Causes —of the screening-off condition, specifically— does not impugn its metaphysical status, i.e. the status of the Principle of the Common Cause (PCC).

Put it another way, it is consistent to assume the existence of a common cause of an 'improbable coincidence' —and thus endorse the Principle of the Common Cause (PCC)— and then go on to ask whether Reichenbach's Criterion for Common Causes (RCCC) is an adequate characterisation of it. Since the PCC and RCCC are two independent claims, the answer to this question needs not affect the status of the PCC. The converse is also possible. Given a fork ABC that satisfies RCCC —and in particular the screening-off condition—, we may ask whether the screening-off event C is indeed a common cause. But again, since the PCC and RCCC are independent claims, whether the answer to this question is positive or negative need not be informative regarding the status of the PCC. Let us discuss both cases in turn.

3.2.1 Screening-off and Common Causes: Sufficiency

Suppose that a correlation $Corr(A, B)$ is observed such that we are sure that A does not cause B and conversely.[23] Moreover an event C is observed to screen-off the correlation. The question is now whether C is a common cause of both A and B. The simple answer is 'not necessarily'. Indeed, it is very easy to find cases in which a screening-off event C is not a common cause of the two correlated events A and B.

The simplest example is probably that of a common effect. Indeed Reichenbach was already aware that common effects screen-off their (correlated) causes. This is for instance the case in both the diagrams in Figure 3.1(b) and Figure 3.1(c) in page 30. Such cases however are easily ruled out by additionally requiring, as Reichenbach himself did, that the screener-off precedes its effects. Common effects screen-off but do not satisfy this additional requirement.

It is harder to restore the sufficiency of the screening-off condition in situations where the common cause has a deterministic effect D, besides the correlated pair A and B, which occurs prior to them. If such is the case the event D will screen-off the correlation even if it is not causally related to A and B. This situation is depicted in Figure 3.2. This is the case, for instance, when switching on the lights in my room. If I press the switch two lights go on (after some short time). This happens almost invariably. The events representing each of the light bulbs going on are certainly correlated and can be said to be effects of a prior common cause, namely that the switch has been put in the 'ON' position. But suppose that the switch in my room is moreover a 'noisy' switch and *always* emits a 'click' sound when set in the 'ON' position. It is easy to check that although the 'click' is not a common cause of the lighting up of either bulb —nor is a common effect— it will screen-off the correlation.[24]

These are not the only problematic cases, however. More sophisticated examples may be conceived such that a correlation screening-off event C does not seem to have a sensible physical interpretation. It then becomes difficult to say that the screening-off event C is a common cause. I will discuss such

[23] This last condition may be easily motivated by locality arguments, for instance.

[24] In this example that the 'click' screens-off the lighting up of the bulbs becomes almost obvious from the fact that it is a *deterministic* effect of the event "setting the switch to ON". The requirement that the 'click' is a deterministic effect is however not crucial. Only the co-occurrence of A, B and D is.

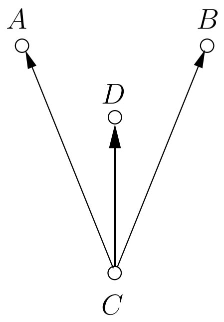


Figure 3.2: A deterministic effect (bold arrow) of the common cause also screens-off the correlation.

cases in more detail in Section 3.4 as a consequence of the so-called common cause completability theorem by Hofer-Szabó et al.

The outcome of this discussion is that *screening-off cannot be regarded as a sufficient condition for common causes*, since it is always possible to find events which are not common causes that nevertheless screen-off certain correlations. It seems now convenient to investigate to what extent examples as the above may affect the status of the Principle of the Common Cause (PCC). Take for instance the example of the light bulbs in my room. As I said, the 'click' that the switch produces when set in the 'ON' position is a screening-off event but not a common cause of the correlation. Does this imply however that there does not exist a common cause of the correlation? I think not. It is thus straightforward to see that although examples may be found which show that screening-off is not a sufficient condition for common causes, we may still claim that a common cause exist. We may claim moreover, not only that a common cause exist, but that such a common cause will screen-off the correlation.[25] This is an important remark which will be crucial in the discussion of common cause completability in Section 3.4.

To sum up, we may say that screening-off cannot be in general taken as a sufficient condition for common causes. However, failure of the screening-off to provide a sufficient criterion for common causes does not affect whether common causes of a certain correlation might indeed exist, leaving intact the status of the Principle of the Common Cause.

[25]In fact, this is precisely the case of the light bulbs example, where 'setting the switch to ON' is a screening-off common cause of both the lights going on.

3.2.2 Screening-off and Common Causes: Necessity

Let us now turn to the status of Reichenbach's Criterion for Common Causes (RCCC) —the status of the screening-off condition in particular— as a necessary condition for common causes.

The failure of screening-off to provide a sufficient condition for common causes seems problematic for any program of causal inference that would want to rely on such a condition. The reason is simply that it seems difficult to identify common causes unless we have such a sufficient condition at our disposal. Once we know that we do not have such a tool at hand the immediate alternative is to proceed negatively, identifying events which are not common causes by means of violations of the screening-off condition. In other words, we may use violations of the screening-off condition to discard the events which are not common causes. This will only work if the screening-off condition is a necessary condition for common causes. Otherwise, violations of the screening-off condition would not provide any useful information about the causal structure. The issue is indeed controversial as several examples have been employed to suggest that there are common causes that violate screening-off.[26]

The most interesting arguments against screening-off as a necessary condition for common causes, for my purposes here at least, are those that involve genuinely probabilistic causes.[27] In particular, counterexamples to Reichenbach's Criterion for Common Causes typically consider correlations between events which occur in tandem —as a result perhaps of some conservation law—, both as an effect of an in principle (patently) obvious common cause. One such example was first proposed by van Fraassen[28] as an argument against Salmon's defence of conjunctive forks, i.e. the screening-off condition, in his theory of causal explanation.[29]

[26]The fate of Reichenbach's Common Cause Principle (RCCP) is tightly related to that of the Causal Markov Condition (CMC). It is usually acknowledged that the CMC holds for deterministic causes but it is controversial whether it does too for genuinely probabilistic ones. The examples I discuss have indeed figured in the intense debate on the status of the CMC.

[27]The main topic of this work, quantum correlations, is indeed taken to constitute a paradigmatic case of genuinely probabilistic causation.

[28]See (van Fraassen, 1982*b*). This and other arguments by van Fraassen —such as (van Fraassen, 1982*a*)— against common causes are however motivated by a rejection of Reichenbach's Common Cause Principle (RCCP) altogether. Rejection of the metaphysical claim in RCCP follows in van Fraassen as a consequence of the failure of Reichenbach's Criterion for Common Causes (RCCC), i.e. failure of screening-off. See Section 3.4.3 and Table 3.4 for details.

[29]There is no need to reproduce or discuss Salmon's theory of causal explanation here. It will be enough to point out that conjunctive forks were at the heart of his early theory.

The example consists of a particle that collides with an atom. As a result of the collision the atom emits two new particles. Suppose for simplicity that the angle with which each particle is emitted can only take two values, each with probability 1/2. That is to say, Particle 1 may be emitted either at angle θ or at angle θ', each with probability 1/2. Particle 2, on the other hand, may be emitted either at angle $-\theta$ or at angle $-\theta'$, also with probability 1/2 in each case.

Now because of the conservation of linear momentum, if Particle 1 is emitted at angle θ (expressed as 1_θ), Particle 2 must be emitted at angle $-\theta$ (expressed as $2_{-\theta}$), and conversely. More precisely, due to conservation of momentum the corresponding angles at which the particles are emitted are *perfectly correlated*[30]:

$$p(1_\theta \mid 2_{-\theta}) = p(2_{-\theta} \mid 1_\theta) = 1,$$
$$p(1_{\theta'} \mid 2_{-\theta'}) = p(2_{-\theta'} \mid 1_{\theta'}) = 1.$$

A common cause λ may now be postulated such that, if present, the particles are emitted at angles θ and $-\theta$ respectively. Otherwise the particles are emitted at θ' and $-\theta'$. It is finally pointed out in the discussion of the example that if the postulated common cause is deterministic, the joint probabilities of the two particles factorise, i.e. the screening-off condition is satisfied.[31] If the postulated common cause is purely probabilistic, however, screening-off is not longer satisfied. This is easy to see.

Take the deterministic case first. That the postulated common cause be

---

Later modifications included the idea of *interactive fork* in order to account for genuinely probabilistic common causes, as in the examples I am just about to discuss. See (Salmon, 1984) for further details.

[30]Although the expressions below does not intend to provide a formal definition of *perfect correlation* it is easy to see that they correspond the special case of maximal (perfect) correlation in the previous chapter (see the remarks after the definition of correlation in page 16). In what follows, 'prefect correlation' will refer to events conforming to expressions of the kind below.

[31]What it is really going on here is that if screening-off is to hold for perfect correlations, the postulated common cause must then be deterministic. (This is shown, for instance in (van Fraassen, 1982*a*) and, more recently in (Graßhoff, Portman and Wüthrich, 2005). A similar but somehow weaker result is also shown in (Fine, 1982*b*), namely that if there exist a screening-off hidden variable for a perfect correlation then it exists as well a deterministic hidden variable model to explain it, and *vice versa*). In other words, there is not room for indeterminism if the observed correlations are perfect. This is at the heart of a very influential argument that claims that screening-off does not hold for genuinely indeterministic common causes. I will be arguing in Section 3.4.3 that this kind of claim is a bit of an overstatement.

deterministic in the way explained above means that

$$p(1_\theta|\lambda) = p(2_{-\theta}|\lambda) = 1,$$
$$p(1_{\theta'}|\neg\lambda) = p(2_{-\theta'}|\neg\lambda) = 1.$$

Also, for the joint probabilities, we have

$$p(1_\theta \wedge 2_{-\theta}|\lambda) = 1,$$
$$p(1_{\theta'} \wedge 2_{-\theta'}|\neg\lambda) = 1.$$

It is then straightforward to see that the corresponding probabilities factorise (since all probabilities equal one):

$$p(1_\theta \wedge 2_{-\theta}|\lambda) = p(1_\theta|\lambda) \cdot p(2_{-\theta}|\lambda),$$
$$p(1_{\theta'} \wedge 2_{-\theta'}|\neg\lambda) = p(1_{\theta'}|\neg\lambda) \cdot p(2_{-\theta'}|\neg\lambda).$$

Two other screening-off conditions may be written replacing the common cause λ by its negation in the first equation above and the $\neg\lambda$ by the common cause in the second. In that case all are zero probabilities and screening-off is again trivially satisfied. Finally, the 1/2 probabilities for the occurrence of each of the events separately are in this case simply reproduced by assuming that the common cause occurs with probability 1/2, i.e. $p(\lambda) = 1/2$.

The issue turns out to be entirely different if λ is a genuinely probabilistic cause. In this case the occurrences of 1_θ and $2_{-\theta}$ ($1_{\theta'}$ and $2_{-\theta'}$) are still perfectly correlated, exactly as in the deterministic case. Recall as well that the observed probabilities are

$$p(1_\theta) = p(2_{-\theta}) = p(1_\theta \wedge 2_{-\theta}) = 1/2,$$
$$p(1_{\theta'}) = p(2_{-\theta'}) = p(1_{\theta'} \wedge 2_{-\theta'}) = 1/2.$$

But now, since the λ is a genuinely probabilistic cause, the probability for the occurrence of 1_θ ($1_{\theta'}$) given that the common cause is present will be, say, r (r'), different from one, i.e. $p(1_\theta|\lambda) = r$ ($p(1_{\theta'}|\lambda) = r'$). We do not need to know at this point what is the probability $p(\lambda)$ of the common cause.[32] The important point is that due to the restrictions imposed by the conservation of momentum, which entail that the events 1_θ and $2_{-\theta}$ ($1_{\theta'}$ and $2_{-\theta'}$) be perfectly correlated, the following probabilities obtain:

$$p(1_\theta|\lambda) = p(2_{-\theta}|\lambda) = p(1_\theta \wedge 2_{-\theta}|\lambda) = r,$$
$$p(1_{\theta'}|\lambda) = p(2_{-\theta'}|\lambda) = p(1_{\theta'} \wedge 2_{-\theta'}|\lambda) = r'.$$

[32]The result we are aiming to is not dependent of that number. Even in the case the common cause happened to be present in all cases, i.e. $p(\lambda) = 1$, the result above would obtain. The reason will become clear in a moment.

Since the joint probabilities are now equal to the marginal probabilities (and all assumed to be different form 1 or 0) it is easy to check that the corresponding screening-off conditions are now violated:

$$p(1_\theta \wedge 2_{-\theta}|\lambda) \neq p(1_\theta|\lambda) \cdot p(2_{-\theta}|\lambda),$$
$$p(1_{\theta'} \wedge 2_{-\theta'}|\lambda) \neq p(1_{\theta'}|\lambda) \cdot p(2_{-\theta'}|\lambda).$$

Again, two more such violations are obtained by replacing the common cause λ by its negation $\neg\lambda$.

This kind of example then shows that for the case of genuinely indeterministic systems there are plausible common cause explanations of certain correlations —those arising from the conservation of a quantity in particular— which do not fulfil the screening-off conditions (3.1) and (3.2) required by Reichenbach. In other words, screening-off is not a necessary condition on common causes in general.

3.3 Is There the Need for a New Common Cause Criterion?

The conclusion in the previous section poses further interesting questions and leaves us with three possible alternatives, if we are to keep the metaphysical content of Reichenbach's Common Cause Principle, i.e. the Principle of the Common Cause (PCC).[33] First, we may want to impose more restrictions on the idea of common cause so as to find a sufficient condition and thus be able to identify them straight away. A second possibility is to take precisely the opposite direction and weaken our criterion on common causes —Reichenbach's Criterion for Common Causes, that is— in order to obtain a necessary condition. We could proceed negatively in this case by identifying those events which are definitely not common causes. The third alternative would be to keep Reichenbach's Criterion for Common Causes (RCCC), as it stands, and see what can be said about the events which conform to it, i.e. whether they are common causes or not. We will shall import some extra causal information for that. This last option will shape the approach I will be taking in the following chapters but let us see first why that seems the most promising option to take.

[33]Of course this is contrary to the moral van Fraassen wants to draw from his example of the 'bombarded atom' I discussed in the previous section.

3.3.1 Strengthening the Conjunctive Fork Criterion

I take it that Reichenbach's Criterion for Common Causes contains some information about the nature of common causes.[34] Thus, if we are to provide a sufficient condition for common causes we must require that these fulfil a criterion stronger than Reichenbach's Criterion for Common Causes (RCCC), i.e. than screening-off. As I pointed out, if we succeed in finding such a criterion we shall be able to identify directly the right events as common causes.

This option, however, does not seem very appealing, specially if we proceed by just imposing further extra conditions on common causes designed to avoid the specific problems we discussed in Section 3.2.1. This seems completely *ad hoc*. Moreover, I argued at the end of that section that we need not worry much about the screening-off condition not being sufficient for common causes. I pointed out that the fact that there exist screening-off events which cannot be regarded as common causes of a correlation does not have an impact on whether there exists a common cause —and in particular a screening-off common cause— of that correlation. Thus we may conclude, as I suggested in Section 3.2.1, that strengthening conjunctive forks is not really needed for our purposes.

3.3.2 Weakening the Conjunctive Fork Criterion

The second alternative is to weaken Reichenbach's criterion in the hope that a necessary condition for common causes can be provided. This is an option that seems quite reasonable in the light of examples such as van Fraassen's 'bombarded atom' in the previous section. In particular, it seems quite reasonable to ask how the non-screening-off common causes that such examples suggest may be characterised, if at all, in terms of probabilistic relations. Providing such a characterisation would constitute another criterion —a more general criterion than RCCC— for common causes. But once this is achieved we need to ask, once more, whether the new criterion constitutes a necessary condition for common causes (we already know it will not be a sufficient condition[35]). If this is the case, we will then have a tool available for rightfully dismissing certain events as common causes, i.e. those events violating our 'generalised' criterion.

This was the option chosen by Salmon, for instance, who introduced his

[34] This is not a particularly problematic statement in my opinion, and I believe most philosophers would agree with it.

[35] This is quite obvious since weakening the restrictions on common causes will at least allow for the same cases discussed in Section 3.2.1.

interactive forks to that end. We do not need to review Salmon's interactive forks here since they will not play any major role in the foregoing discussion. It will be enough for the purpose of this work to stress that interactive forks constitute an important weakening of the conditions on common causes. The aim was to provide a general necessary condition for these, and thus be able to handle also cases —cases such as the 'bombarded atom' example— where causation is genuinely indeterministic. The crucial point is that Salmon's interactive forks, despite the weak character of the conditions on common causes that they require, do not constitute a necessary condition for common causes either. Thus we seem left to either weaken the criterion further or take the third alternative I pointed out at the beginning of the section. That is, keeping the criterion we already have and see how far can we go with it. This, of course unless we have a good reason to replace our original criterion, even with the same strategy in mind.

As I see it, the only interest we may have in weakening Reichenbach's Criterion for Common Causes would be that the new criterion reaches further, or performs better, when characterising common causes.[36] This is, in fact, the underlying motivation of Cartwright's very influential arguments for the generalisation of Reichenbach's Criterion for Common Causes, i.e. of the conjunctive fork.

Cartwright[37] suggests, on the one hand, that Salmon's interactive forks are too weak a condition on common causes. She thus discards that interactive forks be weaken even further. But Cartwright's reaction is far from keeping Reichenbach's Criterion for Common Causes (RCCC). This is because she thinks that RCCC is clearly not appropriate for the characterisation of common causes, particularly when it comes to cases displaying genuine indeterminism. (This diagnosis is of course a consequence of examples such as van Fraassen's 'bombarded atom'). Cartwright thus proposes to generalise Reichenbach's Criterion for Common Causes, not with the aim to provide a necessary condition for common causes, but with the aim to provide a better criterion for common causes. A criterion which is appropriate for both deterministic and indeterministic scenarios. This is to be achieved also by weakening the conditions on common causes, but *just as much* as to allow for genuinely probabilistic common causes.

It remains to see whether Cartwright's generalisation of RCCC indeed performs better in characterising common causes. If this is so we should then, quite reasonably, take Cartwright's generalisation as the starting point

[36]The sense in which I use the expressions such as 'reaches further' or 'performs better' here is somehow vague at this stage, but it will become more precise in the course of the foregoing discussion.

[37](Cartwright, 1987).

for our inferences about common causes.

3.3.3 Indeterministic Common Causes and the Conjunctive Fork Criterion

Cartwright's generalisation of the conjunctive fork criterion relies on causal modelling techniques, which need not be discussed here. Nor do we need to formulate Cartwright's criterion explicitly since, after all, we just aim to evaluate whether it performs better than RCCC in characterising common causes. It shall thus be enough to discuss and assess Cartwright's arguments suggesting that RCCC is not in general an appropriate criterion for common causes —and specially when it comes to indeterminism— and thus defending that a generalisation of it is in need. These arguments ultimately led her to propose the generalisation of conjunctive forks.[38]

The origin of Cartwright's proposal is van Fraassen's 'bombarded atom' example I presented in the previous section. As I showed there, it is quite straightforward to see that, if a deterministic common cause is postulated, the example does not constitute a problem for the screening-off condition. In fact Cartwright endorses the view that deterministic common causes do in general satisfy screening-off.[39] This would include deterministic instances of Salmon's interactive forks for perfect correlations, i.e. perfect forks.

However, postulating a genuinely probabilistic common cause for the 'bombarded atom' example results in a violation of screening-off. Cartwright puts the blame, not only on the genuinely stochastic character of the com-

[38]It might be worth however pointing out that Cartwright's generalisation of RCCC is the result of a review of the dependencies that should be expected in a three variable causal model when the common cause variable operates under a constraint to produce its effects in pairs. These are mainly considerations about the coefficients that relate the three events in the causal model (derived from the fact that the two effects occur in tandem). In particular, the coefficients relating each of the effects and the common cause are not independent. They happen to be the same since it is assumed that once the common cause operates it does indeed produce both effects at once (although in a genuinely stochastic manner). For the full formal details of Cartwright's generalisation the reader is directed to (Cartwright, 1987), where the criterion was first formulated. See also (Cartwright, 1989) for further details.

[39]This was already shown by Fine (1982*b*). Fine points out moreover that if such is the case then the effects are also determined by each other, i.e. they are perfectly correlated. See Appendix B for a proof. Van Fraassen (1982*a*), and more recently by Graßhoff, Portman and Wüthrich (2005), show a similar result, namely that two-valued screening-off common cause variables of *perfect correlations* are deterministic. A simplified version of the Graßhoff, Portman and Wüthrich (2005) proof is also presented in Appendix B, along with some comments on the relation between Fine (1982*b*)'s results and those by van Fraassen and Graßhoff et al.

mon cause, but also on the fact that the postulated common cause *operates under a constraint* (conservation of momentum):[40]

> But in this case it is not reasonable to expect the probabilities to factor, conditional on the common cause. Momentum is to be conserved, so the cause produces its effects in pairs. [...] Clearly the conjunctive fork criterion is not appropriate here. That is because it is a criterion tailored to cases where the cause operates independently in producing each of its effects: whether one of the effects is produced or not has no bearing on whether the cause will produce the other.

It is important to stress that for Cartwright the violation of screening-off is a consequence of the fact that the common cause is genuinely indeterministic *together* with the fact that the common cause operates under a constraint. Cartwright does not seem to suggest in her discussion that any of these features separately are responsible for the violation of screening-off, albeit she urges us to draw the following lesson from a quite similar example:[41]

> Lesson: where causes act probabilistically, screening-off is not valid. Accordingly, methods for causal inferences that rely on screening-off must be applied with judgement and cannot be relied on universally.

Regardless of the precise meaning of the quote above, let us agree for the sake of the argument that Cartwright's generalisation of the Reichenbach's Criterion for Common Causes (RCCC) responds to the need to account for indeterministic common causes that operate under a constraint.

But, do examples such as van Fraassen's justify a generalisation of RCCC? In other words, is it really the case that RCCC cannot account for indeterministic common causes that operate under a constraint? In order to give an answer to these questions let us first go back to the 'bombarded atom' example which, as I said, is at the origin of Cartwright's arguments in favour of the generalisation of RCCC.[42]

[40](Cartwright, 1987, p. 184).

[41]The example I am referring to is the famous 'Cheap-but-Dirty/Green-and-Clean' factory example, which was first discussed in (Cartwright, 1993). I will go back to this example in Section 3.4 in order to motivate the idea of completability. The quote below is from (Cartwright, 1999*a*, p. 8), where the factory example is also discussed.

[42]Indeed, Cartwright's generalisation of the conjunctive fork is largely motivated by the need to defend the metaphysical content of Reichenbach's Common Cause Principle against van Fraassen's anti-realist claims. See (Cartwright, 1987) for details.

A crucial feature of van Fraassen's example is that the effects (of the atom being bombarded) are *perfectly correlated*.[43] This is also what Cartwright refers to with the expressions 'effects in pairs' or 'tandem effects'. At this point we already see then that the example is quite specific (as for the kind of correlations involved).

The particular case of *perfect correlations* is certainly a very special one. For it can be shown that screening-off common causes —which I also called Reichenbachian Common Causes— of perfect correlations are *deterministic* common causes.[44] To make things clearer we may write such statement as a logical proposition:

$$\mathrm{PCORR} \wedge \mathrm{SO} \rightarrow \mathrm{DC}, \tag{3.9}$$

where PCORR stands for 'perfect correlation', SO for 'screening-off' and DC for 'deterministic common cause'.

But, since we are worried here about genuinely indeterministic common causes, we must consider the negation of the expression above instead. That is:

$$\begin{aligned} \neg\mathrm{DC} &\rightarrow \neg(\mathrm{PCORR} \wedge \mathrm{SO}) \\ &\rightarrow (\mathrm{PCORR} \wedge \neg\mathrm{SO}) \vee (\neg\mathrm{PCORR} \wedge \neg\mathrm{SO}) \vee (\neg\mathrm{PCORR} \wedge \mathrm{SO}). \end{aligned} \tag{3.10}$$

We immediately see from this expression that postulating indeterministic common causes conceivably conveys three possible situations: (i) that the (indeterministic) common cause *does not* satisfy screening-off and its effects *are* perfectly correlated; (ii) that the (indeterministic) common cause *does not* satisfy screening-off and its effects *are not* perfectly correlated; and finally (iii) that the (indeterministic) common cause *does* satisfy screening-off and its effects *are not* perfectly correlated.

The situation described in van Fraassen's example includes a purely indeterministic common cause of a perfect correlation. This will thus correspond to the first case enumerated, i.e. the effects of the (indeterministic) common cause *are* perfectly correlated and the common cause *does not* satisfy screening-off.[45] It is hence not surprising at all that the screening-off is vi-

[43]See the remarks following the Definition 1 in page 16, as well as the remarks in Footnote 30.

[44]This is surely true for a two-valued common cause variable V, i.e. $q \in \{q_1, q_2\}$ such that $Vq_1 = C$ and $Vq_2 = \neg C$. See my remarks in Footnote 39 and Appendix B for details.

[45]Thus can be seen better if we take the conjunction of expression (3.10) and PCORR, i.e. 'Eq. (3.10)' $\wedge$ PCORR. For this yields,

$$\neg\mathrm{DC} \wedge \mathrm{PCORR} \rightarrow \mathrm{PCORR} \wedge \neg\mathrm{SO},$$

which is just the case of the 'bombarded atom' example.

olated there. Thus Cartwright is completely right when she claims that "in this case it is not reasonable to expect the probabilities to factor, conditional on the common cause. Momentum is to be conserved, so the cause produces its effects in pairs."

That claim is in the first part of my first quote of her above. The second part of the quote, however, seems to bear a stronger claim, which I do not completely share. In particular, we are told that the conjunctive fork criterion is not appropriate in this example "because it is a criterion tailored to cases where the cause operates independently in producing each of its effects: whether one of the effects is produced or not has no bearing on whether the cause will produce the other." But is this really so? What we can say from the expression above is that as long as the correlations are not perfect, the conjunctive fork criterion, i.e. screening-off, can perfectly hold for indeterministic common causes.[46] To make the point sharper, it all comes down to whether the correlations we want to explain are perfect or not. And, does a common cause that operates under a constraint —such as conservation of momentum— produce perfect correlations in all cases? Certainly not, as we can see with the help of a very simple example.

Take, for instance, van Fraassen's 'bombarded atom' example and slightly modify it such that the atom splits in three fractions (instead of two) after being bombarded, which move away at angles θ_1, θ_2 and θ_3, each with a given probability. As in the original example, the values that θ_1, θ_2 and θ_3 may take are also bound by momentum conservation.[47] If such were the case the correlations $Corr(\theta_1, \theta_2)$, $Corr(\theta_1, \theta_3)$ and $Corr(\theta_2, \theta_3)$ between the three angles would not be perfect any more. That is to say

$$\begin{aligned} p(\theta_1|\theta_2) \neq 1 \neq p(\theta_2|\theta_1), \\ p(\theta_1|\theta_3) \neq 1 \neq p(\theta_3|\theta_1), \\ p(\theta_2|\theta_3) \neq 1 \neq p(\theta_3|\theta_2). \end{aligned}$$

What this shows is that although the constraint has some impact as to whether correlations arise —and on the strength of such correlations as well— it is not sufficient on its own for the common cause to produce its 'effects in pairs', i.e. to generate perfect correlations. Whether perfect correlations

[46] Again, this can be seen clearer if we take the conjunction of expression (3.10) and $\neg$ PCORR, i.e. 'Eq. (3.10)' $\wedge \neg$ PCORR. This yields,

$$\neg \mathrm{DC} \wedge \neg \mathrm{PCORR} \rightarrow (\neg \mathrm{PCORR} \wedge \neg \mathrm{SO}) \vee (\neg \mathrm{PCORR} \wedge \mathrm{SO}),$$

which is just the claim above.

[47] Conservation of momentum would imply that the sum of the three angles be 2π, or in other words, that the sum of the vectors representing the three trajectories be zero.

arise due to a common cause operating under a constraint is dependent on the number of effects that the common cause 'generates', so to speak.

Hence the second part of my first quotation of Cartwright above seems a bit of an overstatement about screening-off and genuinely indeterministic common causes operating under a constraint. Since, while such a claim is true for for common causes that produce their 'effects in pairs', it does not seem correct in the remaining cases, even if the common cause operates under a constraint. Moreover I do not see so straightforward the fact that perfect correlations arise uniquely in the context of the constraints Cartwright is thinking about (such as those imposed by the conservation of a quantity). But this is just a side point, in any case.

The upshot of this discussion is that examples such as van Fraassen's 'bombarded atom' and the like that involve genuine indeterminism only show that, as a general rule, screening-off is violated if the common cause produces perfectly correlated effects —whether this is due to the common cause operating under a constraint or not does not seem crucial—. For all other cases, we may still keep Reichenbach's Criterion for Common Causes all the same.

3.3.4 Screening-off and Indeterministic Common Causes are not Incompatible

In fact, expression (3.10) above confirms that indeterministic common causes may satisfy screening-off, only never for the case of perfect correlations. A possible reading of expression (3.10) —perhaps the most natural and hence the one I will be following here— is as follows:

> Non-perfect correlations may have different explanations in terms of indeterministic common causes, some involving screening-off common causes, and non-screening-off common causes some others.

In other words, for a non-perfect correlation we may very well happen to find two (or more) common cause explanations: one in terms of Reichenbachian Common Causes —which screen-off the correlation—, and the other in terms of non-screening-off events. In fact, in the following section (and the remaining of this work) I will be arguing that there are fairly good chances that a Reichenbachian Common Cause explanation may be provided for any correlation. This argument will also elude the problems that we have seen Reichenbach's Criterion for Common Causes has when it comes to indeterminism with the particular case of perfect correlations. So my results will be claimed to hold all the same whether it is perfect correlations or non-perfect

Common Causes

	Deterministic	Indeterministic
PCORR	**RCCP**	¬ **RCCP**
non-PCORR		¬ **RCCP** or **RCCP**

Table 3.1: The status of Reichenbach's Common Cause Principle (RCCP) for *perfect correlations* (PCORR) and standard *non-perfect correlations* (non-PCORR) as regards to deterministic and genuinely probabilistic (indeterministic) common causes.

ones that we are dealing with. This will justify that there is really no need for a generalisation of Reichenbach's Criterion for Common Causes.

To sum up, we may distinguish between at least two different kinds, so to speak, of common causes. We have, on the one hand, *Reichenbachian* common causes (RCC), which screen-off their corresponding correlations.[48] These may be either deterministic or indeterministic common causes. On the other hand, indeterministic common causes that operate under a constrain to produce perfectly correlated effects may also be considered. Those are *non-Reichenbachian* common causes —since they do not fulfil Reichenbach's Criterion for Common Causes (the screening-off condition more in particular)—. These common causes will typically conform to restrictions weaker than screening-off (see Figure 3.3).[49] These results are also shown in Table 3.1.

I shall stress once more that no generalisation of the conjunctive fork criterion turns out to be a necessary condition on common causes. Necessity may be achieved only by importing extra causal information about the system at hand. The only difference between the different criteria thus stems on the

[48] See Definition 3 (page 33) for a complete definition of *Reichenbachian* common cause.

[49] Other different types of common cause may perhaps be distinguished among non-Reichenbachian common causes. Those conforming to Salmon's interactive forks are different than those characterised by the generalisation of the conjunctive fork advanced by Cartwright (1987). I will not consider this possibility further however, for, as I shall argue in a moment, regardless of whether Reichenbach's Common Cause Principle (and his criterion for common cause in particular) is sufficient or necessary for common causes, it shall still be a reliable tool for causal inference under the right circumstances.

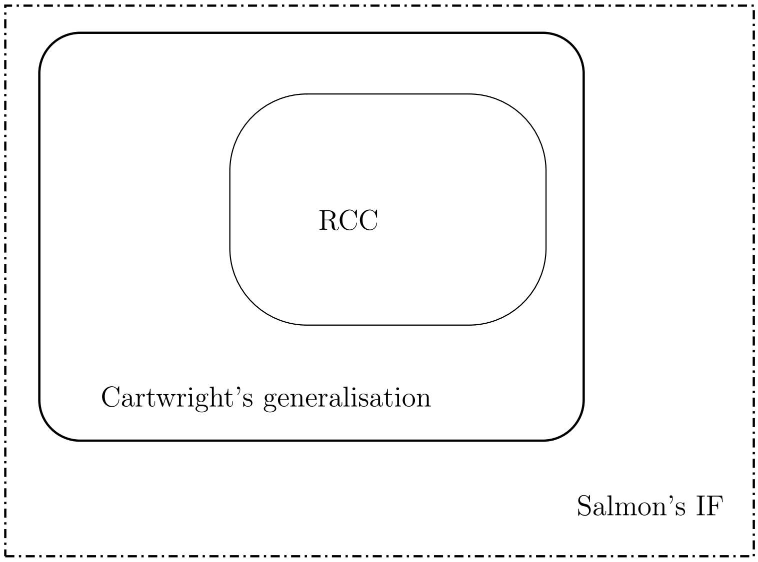


Figure 3.3: *Reichenbachian* common causes (RCC), Cartwright's *generalisation* and Salmon's *Interactive Forks* (IF).

amount of such extra information needed in order to be able to make reliable causal inferences. What this suggests is that although one criterion may be in some sense considered more general than the others, all of them shall be taken at an equal footing. In particular there seems not to be a philosophical reason to favour one over the others. It is only by importing prior causal information that an appropriate criterion for the characterisation of common causes may be provided. This attitude is in line with Cartwright's general philosophy and it is usually expressed under the famous slogan 'no causes in, no causes out'.

3.4 Common Cause Completability

We have seen then that Reichenbach's Criterion for Common Causes (RCCC) —and more particularly screening-off— is neither a sufficient nor a necessary condition for common causes. Moreover, there is little hope that we find either stronger or weaker characterisations of common causes that would constitute respectively sufficient or necessary conditions for them, unless we are prepared to import some extra causal information. My proposal is thus that we keep RCCC and see how far we can go with it. In other words, I shall adopt RCCC as a methodological default rule for causal inference.[50] This was the third option I mentioned at the beginning of the preceding section.

I hope that in doing that we may succeed in using Reichenbach's criterion as a methodological guide in domains where we lack —or are unsure about— our causal intuitions, as it is the case of quantum physics. In this section I shall thus attempt to give sufficient reasons to consider Reichenbach's Common Cause Principle as a useful tool for the study of causal relations, also in the quantum case.

3.4.1 Intuitive Motivation for the Existence of Reichenbachian Common Causes

Let us then investigate the background causal information that is needed in order for Reichenbach's criterion, i.e. RCCC, to be reliable for causal inference.

We have already learned that for some correlations there might be different causal explanations, some in terms of screening-off common causes, some others not. But we did not say, really, whether some of these correlations may only admit non-screening-off common causes —indeed perfect correlations in indeterministic contexts seemed good candidates—. We might then want to ask whether this is so, i.e. whether there are correlations that cannot be explained at all in terms of screening-off common causes. Put it differently, are there screening-off common cause explanations available to us for any given correlation? Or, yet in other words, is it the case that every time a correlation is observed, a common cause C is also 'found' such that the correlation is explained in virtue of its being screened-off by C?

We may attempt to answer these questions by going back to potentially problematic examples, such as van Fraassen's 'bombarded atom' discussed earlier in this chapter. Another such example is the notorious 'Cheap-But-Dirty/Clean-And-Green' factory example, due to Cartwright.[51]

The 'Cheap-But-Dirty/Clean-And-Green' factory example has been widely discussed in the literature, particularly in the context of discussions of the Causal Markov Condition. What is interesting about this example is that the causal structure is presupposed. Hence there is no doubt as to whether the common cause is precisely what we are told it is. Moreover, the example assumes that all causal connections are represented, which amounts to saying that the system is causally complete.[52] What I present here is a simplified

[50] I follow Williamson (2005) in this, who also defends this view.

[51] Cf. (Cartwright, 1993).

[52] Completeness however is a strong assumption, often unwarranted.

version of the example which shall be enough to display the main features of Cartwright's original version.[53]

Suppose that a chemical factory (the 'Cheap-But-Dirty' factory) produces a chemical A through a genuinely probabilistic process. In such a process, the probability of actually getting the product A is 80%. Suppose moreover that whenever the chemical A is produced, another (pollutant) chemical B is also produced as a by-product. The probability for the pollutant B to be produced is then 80% as well. The production process of the factory thus clearly correlates the production of the chemical A with the production of the pollutant B.

Environmental concerns arise when the pollutant is detected. Everything points to the process for the production of the chemical A in the 'Cheap-But-Dirty' factory as responsible of the production of the pollutant B as well. In other words, the process seems to be the 'common cause' C of both the chemical A and the pollutant B. The factory management, however, would not concede that, and defend their innocence with an argument that relies on the screening-off condition. If the process C for the production of the chemical A in the 'Cheap-But-Dirty' factory were to be the cause of the production of the pollutant B as well, then conditional on the process, the probability that the chemical and the pollutant are produced together would factorise, i.e. $p(A \wedge B|C) = p(A|C) \cdot p(B|C)$.

But what is really going on, we are told, is that even if the process is running, i.e. even if the common cause C is present, the cause only 'fires' 80% of the times. There is no reason why C should screen-off the correlation in this case. This is best seen if we look at the probability space $(\mathcal{S}, p)$, over formed by the Boolean algebra $\mathcal{S}$ containing the three events and all their possible conjunctions, and the probability measure p that assigns probabilities to each of the elements of $\mathcal{S}$. That is

$$\mathcal{S} = \{A, B, C, A \wedge B, A \wedge C, B \wedge C, A \wedge B \wedge C\},$$

and the probability measure p assigns the following probabilities:

$$\begin{aligned}
&p(A) = p(B) = 0.8, \\
&p(C) = 1, \\
&p(A \wedge B) = 0.8, \\
&p(A \wedge C) = p(B \wedge C) = 0.8, \\
&p(A \wedge B \wedge C) = 0.8.
\end{aligned}$$

[53]The detailed version may be found in (Cartwright, 1993, 1999a,b).

The above assumes first that $(\mathcal{S}, p)$ contains *all* causal influences, i.e. is causally complete, as required in the example. We may also assume for the sake of simplicity that the common cause C is always present. Thus, since C only produces its effects 80% of the time, the probabilities for $A \wedge C$ and $B \wedge C$ are both 0.8. If, in addition, we assume (again for simplicity) that in the absence of the cause none of the chemicals A or B are produced, we get that the probability that A will occur is 0.8, and similarly for B. The same applies to $A \wedge B \wedge C$. The actual numbers are not so important. The important fact is that the model reproduces the example's features. In particular, we have that

$$\begin{aligned} p(A \wedge B|C) &= 0.8, \\ p(A|C) &= 0.8, \\ p(B|C) &= 0.8. \end{aligned}$$

It becomes now clear that screening-off is not satisfied:

$$p(A \wedge B|C) \neq p(A|C) \cdot p(B|C).$$

However, C is by construction the common cause of both A and B. So we can conclude that ***the** common cause C does not screen-off the correlation.* I have emphasised the article 'the' in the sentence above to stress that C is *the only* possible candidate for common cause in the probability space $(\mathcal{S}, p)$. Recall that this is ensured by the assumption that the probability space $(\mathcal{S}, p)$ is causally complete, i.e. that $(\mathcal{S}, p)$ contains all possible causes of A and B.

As we have seen, Cartwright's natural move at this point is to weaken the criterion that characterises our common causes within $(\mathcal{S}, p)$. However, there is another possible move, perhaps as natural as Cartwright's own, that also follows the intuition that a common cause explanation may be given. In a sense, the alternative is 'just a matter' of rewriting the italicised sentence above as "***a** common cause C does not screen-off the correlation*". What this suggests is that there might exist, under the right conditions, another event C' which is a common cause of the correlation and which screens it off.

The 'just a matter' however is not such a simple matter. It first requires us to drop the assumption that the probability space $(\mathcal{S}, p)$ contains all causal influences. In other words, we need to assume now that $(\mathcal{S}, p)$ may be causally incomplete. I have already pointed out, completeness is quite a strong assumption. There are few practical cases, if at all, in which completeness may be warranted. For, in general, it seems difficult to know whether a cause is in fact a total cause of its effect, or just a partial cause.

These considerations partly underlie the intuition that more detailed causal explanations are always possible. We may say perhaps that the background causal information that was initially assumed in Cartwright's example may not turn out to be the right causal information.

3.4.2 Extensibility and Common Cause Completability

The motivation for the search of new causal explanations may very well be that different explanations are required for different purposes. In some cases a given causal explanation might just not be good enough. In this sense, recall that in Reichenbach's opinion, screening-off common causes provide quite powerful explanations. We may then follow Reichenbach's intuitions at this stage and look for alternative causal explanations in terms of (Reichenbachian) common causes.

Leaving aside the worries about completeness, the search for such an event, however, will require moreover that the original probability space be enlarged, so to speak. Such intuitions can be readily formalised in order to gain precision and make the issue clearer.

In cases such as the above example, where the probability space $(\mathcal{S}, p)$ does not contain a screening-off common cause of the correlation $Corr(A, B)$, I will say that the probability space $(\mathcal{S}, p)$ is *Reichenbachian common cause incomplete*.[54]

Definition 5 (RCCI Probability Space) $(\mathcal{S}, p)$ *is called a* ***Reichenbachian common cause incomplete*** *probability space if it contains a pair of correlated events A and B but does not contain a Reichenbachian common cause C of the correlation Corr(A,B).*

As we have seen it is not difficult to find examples of such common cause incomplete probability spaces.[55] What is more, Reichenbachian common cause incomplete probability spaces are quite likely to be found describing usual experimental data. The question is then how are we to deal with such (Reichenbachian common cause incomplete) causal structures?

As pointed out in the previous section, one seems to have two alternatives when facing a Reichenbachian common cause incomplete probability space.

[54] This is just a little bit different to the original claim by (Hofer-Szabó, Rédei and Szabó, 1999) which refer to such a probability space as simply 'common cause incomplete'. I have introduced the term Reichenbachian to account for the possibility of the existence of non-Reichenbachian common causes, as in the above examples.

[55] Indeed, van Fraassen's and Cartwright's examples reviewed in the preceding sections are both instances of that.

We may opt, following Cartwright, for a weakening of the common cause criterion. Alternatively, we may embark on a search for a screening-off common cause. We already discussed the first option. Let us now discuss the second option.

First, we are searching new events (i.e. the Reichenbachian common causes) which are not contained in the original probability space $(\mathcal{S}, p)$, so we need to characterise precisely how this is to be done. This is achieved by the definition of *extension* of a probability space:

Definition 6 (Extension) *The probability space $(\mathcal{S}', p')$ is called an* ***extension*** *of $(\mathcal{S}, p)$ if there exist a Boolean algebra embedding h of $\mathcal{S}$ into $\mathcal{S}'$ such that $p(X) = p'(h(X))$, for all $X \in \mathcal{S}$.*

Extensibility, as expressed above, allows then for the enlargement of the original probability space so that new events are included. Moreover, Definition 6 ensures that the extension operation be consistent with the old event structure $(\mathcal{S}, p)$. Note that in extending a probability space $(\mathcal{S}, p)$ into $(\mathcal{S}', p')$ we are not only enlarging, so to speak, the set of events $\mathcal{S}$ —this is the role of the embedding h in Definition 6— but also the probability measure p is changed. Thus the embedding h needs to be defined such that the initial probabilities and correlations are maintained under the new probability measure p'. In other words, correlations should stay invariant under the extension operation, that is $Corr(A, B) \equiv Corr_p(A, B) \equiv Corr_{p'}(A, B)$. Thus, ensuring that the extension operation is consistent in this sense is crucial for our purposes. Otherwise, it would be completely meaningless to ask whether a Reichenbachian common cause of the *original* correlations exist in the extended space.

The definition of extension above provides the tools for enlarging any given probability space. It does not tell however how many such extensions exist of a given probability space $(\mathcal{S}, p)$. Nor does it tell us in what circumstances such extended probability space $(\mathcal{S}', p)$ will contain the Reichenbachian common causes we are looking for. In this respect, we are not even sure at this stage whether there exists one such extension —among all admissible extension operations of a given probability space $(\mathcal{S}, p)$— so that Reichenbachian common causes are to be included in it. In order to address these issues we need to introduce the idea of Reichenbachian common cause completability (RCC Completability):

Definition 7 (RCC Completability) *Let $Corr(A_i, B_i) > 0$ $(i = 1, 2, \ldots, n)$ be a set of correlations in $(\mathcal{S}, p)$ such that none of them possess a common cause in $(\mathcal{S}, p)$. The probability space $(\mathcal{S}, p)$ is called* ***Reichenbachian common cause completable*** *with respect to the set $Corr(A_i, B_i)$ if there exists an*

extension *$(\mathcal{S}', p')$ of $(\mathcal{S}, p)$ such that $Corr(A_i, B_i)$ has a Reichenbachian common cause C_i in $(\mathcal{S}', p')$ for every $i = 1, 2, \ldots, n$.*

Completability, as defined above, is hence the key for successfully searching Reichenbachian common causes of any given correlation. The question is now whether any incomplete probability space $(\mathcal{S}, p)$ —incomplete as defined in Definition 5, that is— can be extended such that is (Reichenbachian) common cause completable. In other words, when is a probability space $(\mathcal{S}, p)$ Reichenbachian common cause completable?

The notion of completability provides further methodological grounds for Reichenbach's Criterion for Common Causes to be applicable, and is also related to the idea of hidden variables, which I will introduce in the next chapter in relation to quantum EPR correlations.

Going back to our main purpose, Hofer-Szabó et al. show that an extension $(\mathcal{S}', p')$ may *always* be found for a Reichenbachian common cause incomplete probability space such that it contains (Reichenbachian) common causes for all the original correlations.[56] In other words, as Hofer-Szabó et al. themselves express in the form of a proposition:[57]

> Every classical probability space $(\mathcal{S}, p)$ is common cause completable with respect to any *finite* set of correlated events.

Let me emphasise again that although all definitions and propositions presented here refer to classical probability spaces, similar work can be developed for von Neumann spaces as well, giving completely equivalent results to those presented here for the quantum mechanical case.[58]

It needs to be noted as well that the results presented here are relative to a specific set of correlations. The obvious consequence is that, even if Reichenbachian common cause completeness is achieved for that particular set of correlations —by extending the original probability space—, the new extended probability space may be incomplete with respect to another set of correlations, which we must expect to arise after the extension operation is performed. We may then ask whether there exists an extension of $(\mathcal{S}, p)$

[56] See (Hofer-Szabó, Rédei and Szabó, 1999) for the details of the proof.

[57] (Hofer-Szabó, Rédei and Szabó, 1999, p. 384).

[58] I shall partly rely on such results for the discussion of quantum correlations in the foregoing chapters. On the other hand, this issue aside, it is remarkable that the results above refer only to a *finite* set of correlated events ($i = 1, 2, \ldots, n$). The question whether equivalent results may be obtained for an *infinite* set of correlated events is still open. However, I shall not worry about such issues here since my discussion of quantum EPR correlations in the following chapters will always involve finite sets of correlated events. In such cases, the quantum equivalent to the above results will provide the grounds for the claim that Reichenbachian common causes of EPR correlations may always be found.

such that the resulting probability space $(\mathcal{S}', p')$ is Reichenbachian common cause complete not only with respect to the original correlations in $(\mathcal{S}, p)$ but also with respect to all other correlations that might have arisen as a result of the extension operation. Such an extension is said to be a Reichenbachian common causally closed probability space.[59]

Whether Reichenbachian common cause closed probability spaces exist or not, however, will not greatly concern us here.[60] For the main purposes of this work it will be enough to ensure that Reichenbachian common cause *complete* spaces exist for a given set of finite correlations.

3.4.3 Indeterministic Reichenbachian Common Causes

The results in the previous section suggest that a Reichenbachian common cause explanation may be provided of any given correlation, as long as we are prepared to drop an assumption of completeness about the causal structure of the original probability space where the correlation is observed. If so then we seem to have at our disposal a powerful tool for causal inference. In particular, we can now test how Reichenbachian common cause explanations fare in any situation where correlated (and non-directly causally related) events are found. But of course, as I said in the previous section, a Reichenbachian common cause explanation may not be the only possible common cause explanation. We may then want to compare how the different criteria for common causes fare in a particular case.

I partly addressed such issues in Section 3.3. We can now see that, just like Salmon and Cartwright, proponents of the Budapest School, would agree as regards the validity of Reichenbach's Criterion for Common Causes in deterministic contexts. But their reasons for agreeing would be different. For, while Salmon and Cartwright would consider the case to be a particular instance of interactive forks and *generalised* forks respectively, proponents of the Budapest School would take it that no extension is needed for such a deterministic common cause.

Indeterministic common causes however are a different issue, as we have seen. Let us consider again the 'Cheap-but-Dirty/Green-and-Clean' factory example. Recall that the motivation for such an example was to show that

[59] (Gyenis and Rédei, 2004).

[60] Without going into detail, I just would like to point out that, in order to be able to give an answer to this question, one needs to characterise and make precise the causal dependence-independence relations between the events within the probability space. This is a very interesting question, since it tackles issues such as the degree of causal closeness of our modern physical theories or other commonly used theoretical structures.

Reichenbach's Criterion for Common Causes did not hold for genuinely probabilistic causes. For such cases, Cartwright suggested to weaken Reichenbach's criterion for common causes. An alternative that I advanced was to reject the completeness assumption *a priory*. This option seems now well supported by the extensibility results in the previous section. For what they show is that there is a way of extending the original probability space such that a Reichenbachian common cause may be found. Now what is clear is that in the particular case of the factory example such Reichenbachian common cause must be deterministic since the production of the chemical A and that of the by-product B are perfectly correlated. But this seems not to be a problem from the point of view of common cause completability. Thus, the extensibility results allow us to avoid as well the problems discussed in Section 3.3 concerning perfect correlations.

I also argued in Section 3.3 that when correlations are not perfect we simply do not know whether Reichenbach's Criterion for Common Causes is appropriate. But it turns out that we now know a little bit more. We know in particular that a Reichenbachian common cause may be theoretically postulated for any given correlation. The extensibility results moreover are applicable to both deterministic and indeterministic Reichenbachian common causes. Thus, there is no reason why non-perfect correlations may not be explained in terms of genuinely probabilistic (indeterministic) common causes. This is in fact what the extensibility results suggest and it is in accordance with my conclusions in Section 3.3.

There is however an issue that might make my arguments in favour of Reichenbachian common cause explanations less compelling, even in the indeterministic case. For problems may arise at an interpretative level when referring to extensibility and common cause completability, which might cast doubts about the physical applicability of these results. More in particular, since the extensibility theorem and common cause completability are purely formal mathematical results, it may be claimed that their implications are valid at the formal algebraic level only. That is, some may claim that perhaps Reichenbachian common causes postulated by extending the original probability space may in some cases lack physical interpretation. This suggests that what I have been calling 'common cause completability' —following the original terminology of the Budapest School— is merely 'screening-off completability'. In other words, what the extensibility theorems just show is that it is always possible to extend the original probability space such that a *screening-off* event is found for the correlation. And, again, we will need to include further extra (causal) facts in our explanation if we are to confirm that a screening-off event found in that fashion is indeed a common cause. This is an important and fair criticism but the following observation may

Genuinely Prob. Causation (Indeterminism)			
	RCCC	¬ RCCC	
		General. Fork	Interact. Fork
PCC	**Budapest School**	**Cartwright**	**Salmon**
¬ PCC		**van Fraassen**	

Table 3.2: The positions of van Fraassen, Salmon, Cartwright and the Budapest School towards both the Principle of the Common Cause (PCC) and Reichenbach's Criterion for Common Causes (RCCC) as regards to genuinely probabilistic (indeterministic) causation.

help to mitigate it.

The extensibility theorem presented above does not say anything as to whether there is a unique extension of the original probability space. On the contrary, it seems plausible to think that there can be in general several such extensions and therefore different putative Reichenbachian common causes of a given correlation. The physical interpretation of a Reichenbachian common cause resulting from an extension of a given probability space will thus depend on the choice of extension. The question to ask then is not simply whether there exists an extension of the original probability space such that it contains a Reichenbachian common cause of the correlation we wish to explain. Rather, the question to ask is whether there exist an extension of the original probability space such that it contains a *physically interpretable* Reichenbachian common cause of the correlation we wish to explain. I will assume in what follows that the answer to this question is positive in general, although I understand that this is a far from trivial assumption.

Table 3.2 contains the views discussed so far as regards indeterministic common causes. The main differences between Salmon, Cartwright and the Budapest School regarding the status of Reichenbach's Criterion for Common Causes (RCCC) are displayed. The Budapest School maintains that the PCC and RCCC both hold in genuinely indeterministic contexts, including quantum mechanics. This is very much in opposition to van Fraassen. On the other hand, Cartwright and Salmon both accept the validity of the PCC but only if RCCC is suitably expanded. They differ, however, on the expansion they defend since Salmon rejects screening-off altogether and defends interactive forks, while Cartwright defends a criterion weaker than Reichenbach's but more restrictive than Salmon's interactive forks. To sum up, with the exception of van Fraassen, all the other three proposals discussed —Salmon's, Cartwright's and the Budapest School— accept the metaphysical part of Reichenbach's Common Cause Principle, i.e. the Principle of the Common Cause (PCC).

As a conclusion we then may say that while Reichenbach's Criterion for Common Causes, or better Reichenbach's Common Cause Principle as a whole, is not a necessary condition for the existence of common causes (nor a sufficient one), there are good reasons to think that Reichenbachian common causes may provide good causal explanations of any given correlation. Those will not be, however, the only common cause explanations available. Indeed this view endorses that causal explanations are not unique, but depend on the objective particular purposes of the explanation. It will be interesting to see how well Reichenbachian common cause explanations fare in relation to other possible causal explanations.

3.5 Venetian Sea Levels and British Bread Prices

I pointed out in the previous section that, apart from van Fraassen, the main differences among the three positions discussed with respect to the status of Reichenbach's Common cause Principle (RCCP) are to be found at the methodological level, as regards to the characterisation of the postulated common causes. More specifically, Salmon, Cartwright and the Budapest School all retain the metaphysical content of RCCP, i.e. the Principle of the Common Cause (PCC). That the PCC holds is also crucial for the proposal I aim to develop here. In fact, appealing to extensibility and Reichenbachian common cause completability seems completely hopeless if the PCC did not to hold. Thus, the most harmful counterexamples to RCCP are probably those that attempt to show that there is something wrong generally with the PCC. Sober's well known 'Venetian Sea Levels and British Bread Prices' example[61] and van Fraassen's critique, specially regarding quantum correlations,[62] are probably the two most influential such approaches. In this section I shall concentrate on Sober's argument, and leave van Fraassen's for

[61](Sober, 1987, 2001).

[62]See (van Fraassen, 1982*a*) for instance.

the next chapter.

3.5.1 A Counterexample to the Principle of the Common Cause

Sober's 'Venetian Sea Levels and British Bread Prices' example is devised to refute the metaphysical content of Reichenbach's Common Cause Principle (RCCP). The example, however, has some methodological consequences since, as we will see in a moment, it suggests that there exist certain correlations —e.g. between sea levels in Venice and bread prices in Britain— which cannot be accounted for in terms of common causes. Sober thus urges us to reject Reichenbach's Common Cause Principle in favour of the Likelihood Principle. I shall nevertheless concentrate only on the metaphysical implications of the argument.

Let us review Sober's example in detail. It is the case that the sea level in Venice (VSL) and the cost of bread in Britain (BBP) have been (monotonically) increasing during a given period of time. Table 3.3 displays such trends in the values of Venetian sea levels and British Bread prices in accordance with the actual example. In relation with such data, Sober observes that 'higher than average values' of Venetian sea levels British bread prices are correlated:[63]

> As I claimed initially, higher than average bread prices *are* correlated with higher than average sea levels.

That this is so is clear from the data in data in Table 3.3. If we express by '$\mathrm{VSL}_i > \langle \mathrm{VSL} \rangle$' the fact that 'the Venetian sea level in year i is higher than average', we can easily check that the probability of observing a 'higher than average' Venetian sea level in year i is

$$p\left(\mathrm{VSL}_i > \langle \mathrm{VLS} \rangle\right) = 1/2.$$

Similarly, for British bread prices we have that

$$p\left(\mathrm{BBP}_i > \langle \mathrm{BBP} \rangle\right) = 1/2.$$

[63] Cf. (Sober, 2001, p. 334). Sober's appeal to 'higher than average values' rather the values themselves (of either Venetian sea levels or British bread prices) is mainly motivated by some critiques to an earlier version of the counterexample (Sober, 1987). We do not need to review such arguments here since they will not play any important role in the foregoing discussion. The important point is that Sober's later formulation of the counterexample —involving 'higher than average' correlated values— stands.

Venetian Sea Levels and British Bread Prices

Year (i)	1	2	3	4	5	6	7	8	$\langle \text{Year} \rangle = 4.5$
VSL	22	23	24	25	28	29	30	31	$\langle \text{VSL} \rangle = 26.5$
BBP	4	5	6	10	14	15	19	20	$\langle \text{BBP} \rangle = 11.625$

Table 3.3: Sober's Venetian sea levels and British bread prices data (Sober, 2001, p. 334).

We can also calculate the joint probability of both:

$$p\left[(\text{VSL}_i > \langle \text{VSL} \rangle) \wedge (\text{BBP}_i > \langle \text{BBP} \rangle)\right] = 1/2.$$

These three probabilities entail that:

$$p\left[(\text{VSL}_i > \langle \text{VSL} \rangle) \wedge (\text{BBP}_i > \langle \text{BBP} \rangle)\right] \\ -p\left(\text{VSL}_i > \langle \text{VLS} \rangle\right) \cdot p\left(\text{BBP}_i > \langle \text{BBP} \rangle\right) > 0. \quad (3.11)$$

Thus, everything seems to confirm that in fact 'higher than average values' of Venetian sea levels and British bread prices are (positively) correlated. So, the question is how this correlation is to be explained away, i.e. whether it may be accounted for in terms of common causes, for instance. Sober points out that there are three possible ways of explaining away the correlation:[64]

(i) Postulate the existence of an unobserved common cause.

(ii) The data sample is unrepresentative.

(iii) The data arises from a mixing of populations with different causal structures and correspondingly different probability distributions.

Considering these three options in turn shows, according to Sober, that the Principle of the Common Cause fails. The argument is as follows. First, Sober dismisses option (ii) by pointing out that the correlations in his example do not come out of an unrepresentative sample since data could be spread over a larger time period and the correlations would still be there —I completely agree with this and I will also dismiss option (ii) altogether—.

[64]These three possible explanations had been already suggested by Meek and Glymour after Yule (1926). See also (Sober, 2001, p. 332) and references therein.

Second, option (i) is false in the example *ex hypothesis*. Consequently, Sober takes option (iii) to provide the right (causal) explanation of the correlation. This therefore constitutes a failure of PCC.

Let us now see whether this is a sound argument.

3.5.2 Are Venetian Sea Levels and British Bread Prices Genuinely Correlated?

The first question we must ask is whether the correlation displayed among the 'higher than average values' of the sea levels in Venice and those of the bread prices in Britain are *genuinely* correlated. If the correlation is not *genuine* then there is no reason why a common cause should be required to explain the correlation. Alternatively, if the correlation proves to be *genuine*, a reassessment of the two possible explanations of the correlation (i) and (iii) in the previous section will be in order.[65]

We do not need to provide again the full definition of *genuine* correlation for the purposes of this discussion.[66] It will suffice to recall that we required in the first place that the correlated events belong in the same probability space. Moreover, genuinely correlated events display a probabilistic dependence in that their joint probability differs from their corresponding individual probabilities. (Indeed, it is not possible to make sense of their 'joint probability' if the events do not belong to the same probability space.)

We have seen in the preceding section that the probabilities of 'higher than average values' of Venetian sea levels and British bread prices display what seems to be a probabilistic dependence —by means of expression (3.11) on page 62—. In order to be sure that this is indeed so —that is, that expression (3.11) reflects a probabilistic dependence—, and thus be sure that 'higher than average values' of sea levels and bread prices are correlated, we need to check that the probabilities involved refer to the same probability space. In other words we need to check that 'higher than average values' of sea levels and bread prices belong in the very same probability space.

Nothing in the data sets says whether the probability measure should be the same. In fact, the probabilities for each quantity are derived quite independently (from the relative frequencies of the corresponding measured sea levels and bread prices over a time span). Strictly speaking we should perhaps have initially written them as $p^1(\mathrm{VSL}_i > \langle \mathrm{VSL} \rangle)$ and $p^2(\mathrm{BBP}_i > \langle \mathrm{BBP} \rangle)$,

[65] Recall that we have already dismissed, as Sober does, option (ii) as a plausible explanation of the correlation on the grounds that the correlation would still arise for larger samples.

[66] See Chapter 2, Section 2.3.1 for the details of the definition of correlation.

i.e. as referring to different probability measures p^1 and p^2. And similarly for the joint probability —recall that we calculated the joint probability in the same way as the individual ones, as a relative frequency over the time span (the same time span in fact)—, we should perhaps have written it relative to yet another probability measure $p^3\left[(\mathrm{VSL}_i > \langle\mathrm{VSL}\rangle) \wedge (\mathrm{BBP}_i > \langle\mathrm{BBP}\rangle)\right]$. With these assumptions, relation (3.11) becomes then:

$$p^3\left[(\mathrm{VSL}_i > \langle\mathrm{VSL}\rangle) \wedge (\mathrm{BBP}_i > \langle\mathrm{BBP}\rangle)\right] \\ -p^1(\mathrm{VSL}_i > \langle\mathrm{VSL}\rangle) \cdot p^2(\mathrm{BBP}_i > \langle\mathrm{BBP}\rangle) > 0. \quad (3.12)$$

Thus, the question whether the expression above reflects a correlation between sea level and bread prices may be restated in terms of these three different probability measures. Are p^1, p^2 and p^3 in fact one and the same probability measure?

This is somehow related to Hoover's remarks about the difference between between associations and correlations.[67] Hoover points out that, while associations are a property of the sample, correlations are a property of the probabilistic space used to model it. Hoover rightly assumes that it is only correlations that can reveal 'real' (or physical if we want) properties of the system —although not all of them will do so, as I argued in the previous chapter—. In our case then, if probability measures p^1, p^2 and p^3 are different, expression (3.12) could only be said to reflect some degree of association between sea levels and bread prices, but not a correlation. In order for it to represent a correlation, in Hoover's terms, a consistent probabilistic model —with a single probabilistic measure, that is— must be constructed such that the example's data may be embedded in it.

This is not the case in the 'Venice-Britain' example, in Hoover's view, because the dependence between sea level and bread prices does not represent correlations but mere associations instead. Thus, according to Hoover, since Sober's example does not display a real correlation, 'higher than average' values of sea levels and bread prices cannot be taken as instances of 'improbable coincidences', the kind of dependencies for which Reichenbach demanded common cause explanations. Hoover's conclusion is then that the 'Venice-Britain' scenario does not constitute a counterexample to the Principle of the Common Cause.

But is Hoover right to claim that the data in examples such as the 'Venice-Britain' reflect mere statistical association and hence cannot give rise to correlations? In other words, can the 'Venice-Britain' scenario not be described in a whole single probability space such that the corresponding values or sea levels and bread prices are correlated? I think it can. In particular, while I

[67](Hoover, 2003).

share Hoover's view regarding the difference between mere associations and real correlations, I do not see why the data in Table 3.3 may not be embedded, or modelled if we like, in a single probability space. In fact, one such model is provided by Steel in his critique of Sober,[68] thus rendering Hoover's objection vacuous.

It might be argued that choosing to describe the 'Venice-Britain' data in a single probability space is not sensible. For instance, Sober himself thinks that each data series belongs to different causal structures —this is option (iii) among the possible explanations of the correlation—. But this would not be a very strong objection at all since, as we have seen, causal inference from probabilistic relations is at least controversial. Thus, the fact that two data samples arise from two different causal structures need not be reflected in the data being described in two different probability models. And conversely, the fact that two different data sets belong in the same probability model does not entail that they are the result of the same causal structure.

Summing up, there is no convincing reason why correlations such as those in the 'Venice-Britain' kind of example wouldn't be *genuine*. Sober's examples may then very well be counterexamples to the Principle of the Common Cause. In order to confirm they are, however, we still need to check whether the correlations they involve are the right kind of correlations we would require a (common) causal explanation for, i.e. whether they are *genuinely physical* correlations.

3.5.3 Time Series, Purely Formal and Genuinely Physical Correlations

So we need to find out whether the correlations in the 'Venice-Britain' example and the like are what Reichenbach referred to as 'improbable coincidences' and thus call for a (common) causal explanation. I argued in Section 3.1.1

[68] Steel's model takes advantage of a well known mathematical result so-called the 'mixing theorem'. In brief, the 'mixing theorem' provides us with information about the behaviour of the probability distribution resulting from the mixing of the distributions from two populations, in each of which the variables are probabilistically independent. It tells us, in particular, under what conditions the variables of such 'mixed' probability distribution are probabilistically independent. The theorem then shows that a probability distribution may display dependencies just because it is the result of the mixing of two other probability distributions. Steel claims this is the case in Sober's 'Venice-Britain' example, and constructs a model from two initial sets of data (of both VSL and BBP), each corresponding to a different (distant) time span. If the probability distributions from these two populations are mixed, the resulting distribution displays probabilistic dependencies in just the manner suggested by Sober. See (Steel, 2003) and the references therein for further details.

that Reichenbach's 'improbable coincidences' had to be, not only *genuine* correlations as defined in Section 2.3.1 in the previous chapter, but also what I called *genuinely physical* correlations. That is, correlations that are able to reflect inherent properties of the system at hand —such correlations, more importantly, do not arise as a product of the process of modelling the statistical data[69]—. Hence, this is the issue we must address, I believe, in order to see whether the correlations that Sober's example involves are instances of Reichenbach's 'improbable coincidences'.

So what does the correlation between sea levels and bread prices really say, if anything at all, about the sea in Venice and the bread in Britain? In order to answer this question it seems useful to inspect closely what kind of events we are dealing with. Recall in this respect that the correlated events are not just changes in values of sea levels and bread prices but somehow more sophisticated kind of events, i.e. 'higher than average values' of sea levels '$\mathrm{VSL}_i > \langle\mathrm{VSL}\rangle$' and bread prices '$\mathrm{BBP}_i > \langle\mathrm{BBP}\rangle$'. Note that 'higher than average values' are defined relative to the average of the corresponding quantity over a certain period of time. It is thus clear that the correlated events in the 'Venice-Britain' example have some sort of time dependence, for they include two different time scales, so to speak. More specifically, they involve yearly values on the one hand, and the value of the average over the years within a given time range on the other. The remarkable thing is that the average over the time span is constant (in time) and thus the time dependence of the yearly values is inherited by the 'higher than average' value events.[70]

Time dependent data are also commonly known as non-stationary data. It is also well known that non-stationary data display dependencies that do not always reflect the system's inner structure. For instance, Steel[71] points out that it is a consequence of the so-called 'mixing theorem' of probability theory that two sets of non-stationary data, such as the 'higher than average' values of sea levels and bread prices of Sober's example, display probabilistic dependencies even if each of them refers to a completely different historical period.[72] Also as a consequence of the 'mixing theorem', if the probabilistic

[69]Recall I introduced the distinction between *genuinely physical* correlations in contrast to *purely formal* correlations precisely in order to account for such features of probabilistic modelling of statistical data. See Chapter 2, Section 2.3.2 for the details of such distinction.

[70]Moreover, both sea level and bread prices 'higher than average' values inherit exactly the same time dependence (since the time spans considered are also the same in both series). This point is a very specific feature of the example itself and is not crucial for the main argument here. However it will make the argument look somehow more trivial than it is. See the discussion and example below for details.

[71](Steel, 2003).

[72]See my remarks in Footnote 68 about the 'mixing theorem', and (Steel, 2003) for

dependence of two data series is due to them being non-stationary the correlation will vanish as soon as we describe the data in a probabilistic model in which one of them is no longer non-stationary. And this is in fact what happens with Sober's Venetian sea levels and British bread prices. Let us see how this is so.

Suppose for instance that we include in our model events of the type '$\mathrm{Y}_i > \langle\mathrm{Year}\rangle$', which we may call, following Sober's terminology, 'higher than average time values', or 'higher than average values of years', never mind how strange this may sound. We may then assign probabilities to such events in exactly the same way as we did for 'higher than average' values of sea levels and bread prices, that is by referring to their relative frequencies. Only, we need to make sure that the probability measure is the same for all three values. Thus, from Sober's own data on Table 3.3, we may write

$$p\,(\mathrm{Y}_i > \langle\mathrm{Year}\rangle) = 1/2.$$

If we now take conditional probabilities we obtain, also looking at the data in Table 3.3,

$$p\,(\mathrm{VSL}_i > \langle\mathrm{VSL}\rangle \mid \mathrm{Y}_i > \langle\mathrm{Year}\rangle) = 1, \tag{3.13}$$
$$p\,(\mathrm{BBP}_i > \langle\mathrm{BBP}\rangle \mid \mathrm{Y}_i > \langle\mathrm{Year}\rangle) = 1. \tag{3.14}$$

It is also easy to check that

$$p\,[(\mathrm{VSL}_i > \langle\mathrm{VSL}\rangle) \wedge (\mathrm{BBP}_i > \langle\mathrm{BBP}\rangle) \mid \mathrm{Y}_i > \langle\mathrm{Year}\rangle] = 1. \tag{3.15}$$

It becomes now clear that as soon as we consider the event '$\mathrm{Y}_i > \langle\mathrm{Year}\rangle$' the correlation will vanish. This is because the dependence of the original series washes out conditional on '$\mathrm{Y}_i > \langle\mathrm{Year}\rangle$'. In particular if we define a new probability measure as $p^Y = p(\cdot \mid \mathrm{Y}_i > \langle\mathrm{Year}\rangle)$, the above equations yield

$$\begin{aligned} p^Y\,[(\mathrm{VSL}_i > \langle\mathrm{VSL}\rangle) \wedge (\mathrm{BBP}_i > \langle\mathrm{BBP}\rangle)]& \\ - p^Y(\mathrm{VSL}_i > \langle\mathrm{VSL}\rangle) \cdot p^Y(\mathrm{BBP}_i > \langle\mathrm{BBP}\rangle) &= 0. \end{aligned} \tag{3.16}$$

This example is of course specific for the case at hand, and the 'trick' has been quite the obvious one, since I transformed both the original non-stationary data series into stationary ones by finding the probability space in which all probabilities are one. In this case such probability space is particularly easy (and obvious) to find since the time dependence of both sea

further details and references.

levels and bread prices' higher than average values is exactly the same (see Footnote 70). However obvious and specific this example might be, I hope it illustrates sufficiently the fact that probabilistic dependencies due to the non-stationary character of the data vanish if such data are transformed into stationary. In fact there might be other (less obvious) possible probability spaces for which the same result would obtain. For recall this is a consequence of the 'mixing theorem'.

What the above example and its implications point to is that correlations such as those in the 'Venice-Britain' example arise due to the fact that the series are time dependent, i.e. non-stationary. In other words, sea levels and bread prices are only correlated in virtue of telling us something about time. This time dependence, however, is also model dependent meaning that we may easily construct another model in which the data becomes stationary. In such a model the correlation disappears, as we have seen. This in turn suggests that correlations that arise due to the non-stationary properties of the data do not provide any information whatsoever about the underlying (physical) structure of the system, if there is any system we can speak of. To the contrary, they seem to be a case of what I identified as *purely formal* correlations in the previous chapter (in contrast to *genuinely physical* correlations). Under this view therefore Sober's example does not constitute a genuine counterexample to the Principle of the Common Cause (PCC).

The argument I just presented is very much along the lines of Steel's.[73] Steel however draws a different conclusion. Although he points out that the PCC cannot be applied to non-stationary data series —mainly for the reasons discussed above— he instead takes it that the counterexample is genuine: there is a genuine correlation which is not to be explained in terms of a common cause.

I hope it becomes clear from my discussion that, while I completely agree with the first part of Steel's claim I must disagree with his conclusion. For I take it that the non-applicability of the PCC to non-stationary data is a consequence of the fact that the correlations among such kind of data are not *genuinely physical* correlation, in the sense explained above. Such correlations thus do not call in my view for a common cause explanation in the sense of Reichenbach.

However, if we were to accept Steel's view and insist that the correlation is of the kind that should be explained in terms of a common cause —thus providing the grounds for Sober's counterexample to go through— we may still have a possible defence of the Principle of the Common Cause. It would be a weaker solution, though. We need first to reject the separate cause

[73] (Steel, 2003).

presupposition, just on the same grounds for it to be accepted perhaps, i.e. on metaphysical grounds. That is, we need to reject that *ex hypothesis* there is no common cause of the correlation. We may then use the extensibility and (Reichenbachian) common cause completability results I discussed in Section 3.4 in order to argue for the possibility that a common cause may exist in a greater extended probability space.

It is remarkable that if this the case, the common cause will also be a screening-off common cause. This does not add extra difficulties since the 'Venice-Britain' example applies all the same however common causes are to be characterised.[74] Moreover, the 'Venice-Britain' scenario may be considered to be deterministic (although it need not be). And if such is the case then, as we saw in Section 3.3.4 (see also Appendix B), any common cause model needs to be a screening-off common cause model. Hence, the problems we face have nothing to do with how the postulated common causes are to be characterised.

It will be much more difficult to interpret the screening-off event as a common cause. We can see what the problem is if we go back to my example above, devised to eradicate the non-stationary character of the 'Venice-Britain' data. For, what I provided in such example was nothing more than an extension of the original probability space (with probability measure p^Y) such that it contained screening-off events ('$\mathrm{Y}_i > \langle\mathrm{Year}\rangle$') of the original correlations (between 'higher than average' values of sea levels and bread prices). This can be checked by rewriting equation (3.16) in terms of conditional probabilities and rearranging it as:

$$\begin{aligned} & p\left[(\mathrm{VSL}_i > \langle\mathrm{VSL}\rangle) \wedge (\mathrm{BBP}_i > \langle\mathrm{BBP}\rangle) \mid \mathrm{Y}_i > \langle\mathrm{Year}\rangle\right] \\ &\quad = p(\mathrm{VSL}_i > \langle\mathrm{VSL}\rangle \mid \mathrm{Y}_i > \langle\mathrm{Year}\rangle) \cdot p(\mathrm{BBP}_i > \langle\mathrm{BBP}\rangle \mid \mathrm{Y}_i > \langle\mathrm{Year}\rangle) = 0. \end{aligned}$$

where I have adopted the notation in expressions (3.13), (3.14) and (3.15).

The question is then whether we can make sense of events such as '$\mathrm{Y}_i > \langle\mathrm{Year}\rangle$' as (common) causes. But is time a causal factor? I must admit that I do not have an answer to this question. I find it hard to understand '$\mathrm{Y}_i > \langle\mathrm{Year}\rangle$' as a cause event. Still, perhaps further conceptual innovations may at some stage provide an adequate framework so as to be able to interpret such time dependent events as (common) causes. This interpretational problem will also be present whenever a time dependent event is to be interpreted as a cause. This also poses further conceptual problems to examples involving

[74] In fact, it is not important for Sober whether common causes —when they exist— are screening-off common causes or not. The important fact in his argument is that there are correlations for which no common cause explanation can be provided *ex hypothesis* (the correlations between sea levels in Venice and bread prices in Britain is one such case).

non-stationary data, such as Sober's 'Venetian sea levels and British bread prices'. To sum up, there are some good reasons in my opinion to reject that cases such as 'Venetian sea levels and British bread prices' are definite counterexamples to the Principle of the Common Cause. At least I will assume so in the remainder of this thesis.

3.6 Van Fraassen's View and some Concluding Remarks

I shall close this chapter by commenting very briefly on the position taken by van Fraassen, whose arguments will be discussed at length in the following chapter. This will complete the picture regarding the different views on the status of Reichenbach's Common Cause Principle that I have discussed throughout this chapter. All these views, including van Fraassen's, are summed-up in Table 3.4.

Van Fraassen, like Sober, objects to the metaphysical content of Reichenbach's Common Cause Principle (RCCP), i.e. what I called the Principle of the Common Cause (PCC). His concerns about the status of the principle however differ from Sober's and indeed constitute a completely different position regarding RCCP. While Sober criticises directly the PCC, van Fraassen's critique focuses on Reichenbach's Criterion for Common Causes (the methodological part of RCCP) and the conclusions are later extended to the metaphysical part of RCCP, i.e. the PCC.

A further crucial difference between Sober's view and van Fraassen's is that Sober is not committed at all to indeterminism. In other words, examples of the kind 'Venice-Britain' discussed in the preceding section have the very same reading and implications whether the putative causes act deterministically or purely probabilistically. Van Fraassen, on the other hand, builds his critique of Reichenbach's Common Cause Principle relying on examples involving indeterminism such as that of the 'bombarded atom' introduced in Section 3.2.2. Of special interest to us will be his arguments regarding quantum EPR correlations, which I will discuss further in the next chapter.

The key feature of van Fraassen's view is, as I noted, that he takes it that Reichenbach's Common Cause Principle does not hold at its metaphysical level precisely because it fails at the methodological level. In particular Reichenbach's Common Cause Principle fails, in van Fraassen's opinion, to account for correlations that involve purely indeterministic events. This is in striking contrast with the positions by Salmon, Cartwright and the Budapest School reviewed in the preceding sections, however they might differ between

Common Causes and Determinism/Indeterminism

	PCC			¬ PCC
	RCCC	¬ RCCC		
Determ.	**Salmon** **Cartwright** **Budapest**			**Sober**
Indeterm.	**Budapest**	**Salmon** **Cartwright** $\boxed{\textbf{van Fraassen}}$	⇒	**Sober** **van Fraassen**

Table 3.4: The positions of Sober, van Fraassen, Salmon, Cartwright and the Budapest School towards both the Principle of the Common Cause (PCC) and Reichenbach's Criterion for Common Causes (RCCC).

them (see Table 3.4).

I hope I succeeded in making it clear throughout this chapter that the view that I attributed to the Budapest School seems promising enough if we intend to draw causal conclusions based on Reichenbach's Common Cause Principle. Specially in those domains where we lack our standard causal intuitions (as it is the case of quantum mechanics). Of course this position has its own problems as well —specially at the level of interpretation— and must be taken with the necessary precautions. This is in any case the position I shall adopt in what follows.

Appendix A

Conjunctive Forks and Screening-off

The equivalence between Reichenbach's equations and the screening-off condition can be proved straightaway as follows. Let us first consider he left hand side of Reichenbach's first expression $P(A \wedge B|C) = P(A|C) \cdot P(B|C)$. By applying the 4th axiom of the Probability calculus we find that

$$\begin{aligned} P(A \wedge B|C) &= \frac{P(A \wedge B \wedge C)}{P(C)} \\ &= \frac{P(A \wedge B \wedge C) \cdot P(B \wedge C)}{P(B \wedge C) \cdot P(B \wedge C)} \\ &= P(A|B \wedge C) \cdot P(B|C). \end{aligned}$$

Therefore

$$P(A|B \wedge C) \cdot P(B \wedge B) = P(A \wedge C) \cdot P(B \wedge C),$$

which, as long as $P(B) \neq 0$, yields

$$P(A|B \wedge C) = P(A|C).$$

In a similar fashion the screening-off condition $P(B|A \wedge C) = P(B \wedge C)$ also follows from $P(A \wedge B|C) = P(A|C) \cdot P(B|C)$.

Appendix B

Determinism, Screening-off and Perfect Correlations

I shall show in this appendix the following: (i) a Reichenbachian (screening-off) common cause of a perfect correlation is a deterministic common cause; and (ii) a deterministic common cause of two correlated effects A and B screens-off the correlation. In this case, moreover, A and B are perfectly correlated conditional on the common cause.

B.1 Reichenbachian Common Causes of Perfect Correlations are Deterministic

Already van Fraassen (1982*a*) showed in the context of the EPR experiment that the assumption of *perfect correlation* together with the existence of a two-valued screening-off hidden variable —he termed this last requirement *Causality* in clear reference to common causes—, entailed determinism. A more detailed proof can be found in Graßhoff, Portman and Wüthrich (2005). Such results can be originally traced back to Suppes and Zanotti (1976) and they all involve 'hidden variables' as such.

I shall present here a simplified version of such proofs which only make reference to the particular case of a two-valued common cause 'hidden variable', which may be identified with the usual common cause event (type) as in Reichenbach's original work. This might seem in a first instance too a specific case. However, it can be shown —see, for instance Graßhoff, Portman and Wüthrich (2005, p. 669)— that, for the case of perfect correlations and a multi-valued (common cause) screening-off hidden variable, the values that the hidden variable takes can be split without loss of generality into two disjoint subsets, such that it has the same structure of a two-valued (common cause) hidden variable. Thus, although the following version of the above proofs is not entirely general, it will suffice for the main purposes of this work.

Suppose on the one hand that two event types A and B stand in perfect correlation (PCORR):

$$p(A|B) = p(B|A) = 1. \qquad \text{(B-1)}$$

On the other hand, let C be a Reichenbachian common cause of both A

and B. In particular, since the common cause C screens-off A from B (SO):

$$p(A \wedge B|C) = p(A|C) \cdot p(B|C), \tag{B-2}$$
$$p(A \wedge B|\neg C) = p(A|\neg C) \cdot p(B|\neg C). \tag{B-3}$$

Now equation (B-1) may be rewritten as

$$p(A \wedge B) = p(A) = p(B),$$

which also implies

$$p(A \wedge B \wedge C) = p(A \wedge C) = p(B \wedge C). \tag{B-4}$$

Similarly, equation (B-2) may be written as

$$\frac{p(A \wedge B \wedge C)}{p(C)} = \frac{p(A \wedge C)}{p(C)} \cdot \frac{p(B \wedge C)}{p(C)},$$

which together with the two equalities of equation (B-4) respectively yields

$$\frac{p(A \wedge C)}{p(C)} = \frac{p(A \wedge C)}{p(C)} \cdot \frac{p(B \wedge C)}{p(C)},$$
$$\frac{p(B \wedge C)}{p(C)} = \frac{p(A \wedge C)}{p(C)} \cdot \frac{p(B \wedge C)}{p(C)}.$$

Thus,

$$p(B|C) = 1,$$
$$p(A|C) = 1.$$

This is to say, the presence of the common cause C fully determines that the events A and B must occur with probabilities 0 or 1. In other words, C is a deterministic common cause (DC) of both A and B. This shows the logical implication (3.9) in page 46 holds:

$$\text{PCORR} \wedge \text{SO} \rightarrow \text{DC},$$

where PCORR stands for 'perfect correlation', SO for 'screening-off' and DC for 'deterministic common cause'.

B.2 Deterministic Common Causes Screen-off its (perfectly) Correlated Effects

I shall prove now that all deterministic common causes —again, instances of two-valued common cause hidden variables— are Reichenbachian common causes or, in other words, that they screen-off their effects. Moreover, as a consequence, the effects will be shown to be perfectly correlated conditional on the common cause.

Let C be a deterministic common cause of two events A and B, i.e.

$$p(A|C) = 1,$$
$$p(B|C) = 1.$$

The above equations entail that

$$p(A \wedge C) = p(C),$$
$$p(B \wedge C) = p(C).$$

which may be rewritten as

$$p(B \wedge A \wedge C) = p(B \wedge C),$$
$$p(A \wedge B \wedge C) = p(A \wedge C).$$

Thus

$$p(A \wedge B \wedge C) = p(A \wedge C) = p(B \wedge C) = p(C).$$

We then immediately see that

$$p(A \wedge B|C) = p(A|C) = p(B|C) = 1,$$

and thus the screening-off condition is trivially satisfied:

$$p(A \wedge B|C) = p(A|C) \cdot p(B|C) = 1. \tag{B-5}$$

What this shows then is that

$$\text{DC} \rightarrow \text{SO},$$

where again DC stands for 'deterministic common cause' and SO for 'screening-off'.

It is now almost immediate to see from equation (B-5) above that

$$p(A|B \wedge C) = 1,$$
$$p(B|A \wedge C) = 1,$$

which means that A and B are perfectly correlated conditional on the common cause C, i.e.

$$p_C(A|B) = 1,$$
$$p_C(B|A) = 1.$$

A proof of a similar result in the framework of stochastic hidden variables is provided by Fine, who moreover points out (Fine, 1982*b*, p. 293) that while

> [...] every deterministic hidden-variables model is a factorizable stochastic model.

Note that what Fine refers to here as *factorizability* is not a 'simple' screening-off condition, such as the ones I have used in my proofs. Factorizability is the typical condition imposed on hidden variables in the context of the EPR correlations and the derivation of the Bell's inequalities. It is a screening-off-like condition but it conveys a stronger restriction on common causes than a 'simple' screening-off, due to the fact that measurement settings are explicitly included. It serves however to confirm my point in this section, i.e. that deterministic common causes screen-off their corresponding correlations (since they conform to factorizability).

The passage quoted above thus confirms as well what I pointed out in Section 3.3 regarding Cartwright's assessment of examples of the type of van Fraassen's 'bombarded atom' and the 'Cheap-but-Dirty/Green-and-Clean' factories example.

Chapter 4

Quantum Correlations and Common Causes

The discussion in the previous chapter reviewed the status of Reichenbach's Common Cause Principle (RCCP) —its main withdraws and difficulties— both as a metaphysical claim and regarding the rule for causal inference that it promotes, i.e. the screening-off condition. I also mentioned the case of quantum correlations, which I pointed out are usually taken to be a definitive counterexample to the principle.

I shall discuss quantum correlations in detail in this chapter. More specifically, I shall reassess the implications of Bell's theorem as regards RCCP. We will see that arguments employed in deriving the Bell inequalities, almost exclusively refer to the methodological part of RCCP, i.e. the screening-off condition. However, the conclusions of such arguments are extended to RCCP as a whole, including its metaphysical part. (In particular, finding RCCP methodologically untenable —something that is usually concluded as a consequence of the experimental violation of the Bell inequalities— is commonly taken to mean that its metaphysical claim is also false.) But such a conclusion does not seem warranted, specially in the light of the distinction between the two claims in RCCP made in the previous chapter. For recall that RCCP contains two claims that are logically independent, and finding any one of these to fail does not necessarily entail that the other be false as well.[1]

On the other hand, I shall argue that Reichenbach's Common Cause Principle may also be saved at the methodological level. The argument exploits some subtleties of the results by the Budapest School —particularly regarding common cause completability— in order to show that the usual arguments relating Bell's hidden variables and Reichenbachian common causes

[1]See Chapter 3, Section 3.2 for details.

contain unwarranted assumptions. I would like to stress however that it is not my intention to claim that EPR correlations must be explained above all in terms of Reichenbachian common causes.[2] I am more interested instead in the implications one can draw from considering Reichenbachian common cause explanations as plausible explanations of EPR correlations.

4.1 EPR Quantum Correlations

Conceptual and interpretive issues have always been at stake in quantum mechanics, ever since its very early stages. In the early years of the theory, most interpretive issues were obviously related to the fundamental stochastic character of quantum mechanics, which constituted a major conceptual shift in the understanding of the physical reality. The fundamental (irreducible) indeterminism of the new quantum theory was in striking contrast to the classical conception of physical reality, dominated by a deterministic mechanics, which provided a 'standard' causal picture.

More particularly, the fundamental indeterministic character of quantum mechanics supported by Heisenberg's Uncertainty Principle and Bohr's interpretation of complementary properties was taken to carry along a non-realist interpretation of physical reality. Complementarity issues were at the heart of the famous Bohr-Einstein debate on the completeness of the quantum theory, which had in the so-called EPR paradox perhaps its more significant argument.

4.1.1 Is the Quantum Mechanical Description of Physical Reality Complete?

Heisenberg's Uncertainty Principle states that complementary properties of a physical system, such as momentum and position, can never both be simultaneously determined fully, but only within a range of values according to the expression $\Delta x \Delta p \geqslant h$. There is an inverse relation between the precision with which we can measure such quantities. For instance, exactly determining the value of momentum of a particle entails that its position is completely unknown. Similar results are obtained for other complementary properties such as time and energy, or the components of spin angular momentum along different directions.

In Bohr's interpretation of quantum mechanics —the core to what has come to be known as the Copenhagen interpretation— Heisenberg's Uncer-

[2] Other explanations for EPR correlations may be defended, including non-screening-off common causes *a la* Cartwright, or direct cause explanations.

tainty Principle and the complementarity of quantum 'observables' were understood to result from some uncontrollable disturbances at the time of measurement. In particular, according to Bohr such disturbances are due to the very interaction of the quantum system with the (macroscopic) system that constitutes the measurement apparatus. As a result, the outcomes of measurements are indeterministic, and the values of complementary properties of a system cannot be determined precisely (as per Heisenberg's uncertainty relations). According to this view, moreover, the wave function —the 'object' that encapsulates the full quantum mechanical description of a physical system— was essentially taken to inform us about 'measurement probabilities'. More specifically, the wave function was taken to provide probabilities for the possible outcomes, but only if the appropriate measurements were performed.

One of the most striking consequences of Bohr's interpretation of quantum mechanics is that it implied a complete ignorance about the properties of quantum systems in the absence of measurement. In other words, Bohr's interpretation of the quantum mechanical wave function did not seem to contain information about the physical properties of the system themselves, but only about the possible measurements one could perform on it. As an immediate reaction thus, one could ask whether the quantum wave function provided a complete description of all that there was to the physical system it described. Or in other words, one could ask whether quantum mechanics provided a complete description of physical reality.

4.1.2 The EPR Paradox

The most celebrated work addressing the issue as to whether quantum mechanics could be considered a complete description of physical reality was the 1935 Einstein, Podolsky and Rosen (EPR) paper[3]. Indeed, the EPR paper constituted a turning point in the ongoing debate about the interpretation of the theory. Although the implications of the EPR argument were initially taken to reveal a tension between locality and completeness, later interpretations —specially after the work by Bell— suggested that there is not such a tension, and that quantum mechanics is simply non-local, regardless of any completeness issues. But let us start from the beginning.

The EPR paper proposed an argument specially tailored to show that quantum mechanics did not provide a complete description of a physical system. The original argument in the paper has a complex structure. I

[3](Einstein, Podolsky and Rosen, 1935).

provide here is a more illuminating version of the argument due to Fine.[4]

The EPR argument relies on two premises —which we shall justify later— in order to show that quantum mechanics is incomplete. The first premise states that

Premise 1: *It is the case that either quantum mechanics is incomplete, or that complementary quantities cannot have simultaneous 'reality'.*

(We will see in a moment what the meaning of 'reality' is). The second premise states that

Premise 2: *If quantum mechanics is complete then complementary quantities can have simultaneous real values.*

From these two premises it follows that quantum mechanics must be incomplete. Otherwise a contradiction occurs, i.e. that complementary quantities *can* but *cannot* have simultaneous real values.

The authors justify the above two premises in turn. The first premise, i.e. Premise 1, is justified by appealing first to a reality criterion —so-called 'EPR reality criterion'—, which provides for a sufficient condition for what the authors call 'elements of reality':[5]

> 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 exist an element of reality corresponding to that quantity.

Second, the authors require that for a physical theory to be complete, it is necessary that every 'element of physical reality' (defined with the criterion above) has a counterpart in it. It then follows that, if the values of two complementary properties of a quantum system have simultaneous reality —that is to say that they conform to the 'EPR reality criterion'— then the description that the quantum mechanical wave function provides cannot be complete. This is simply because the wave function cannot have simultaneous eigenvalues of both (complementary) quantities, i.e. it cannot contain simultaneously counterparts for the 'elements of reality' of two complementary quantities.

In order to justify Premise 2, the authors propose what it is now known as the EPR thought experiment. The experiment is aimed at showing that complementary quantities *can* indeed have simultaneous real values in certain

[4] Cf. (Fine, 1996, Ch. 3).

[5] (Einstein, Podolsky and Rosen, 1935, p. 177)

circumstances. It is remarkable, however, that the experiment does not make reference to completeness issues at all.

The original EPR experiment consisted in two systems ('system i' and 'system ii') which were allowed to interact for some period of time. Moreover, two conservation laws were assumed to hold, namely the conservation of momentum and the conservation of the relative position (which entails that the two systems is kept fixed).[6] Once the interaction is over, linear momentum and positions are measured in each of the system (both along the same direction). Because of conservation of linear momentum, determining the value of momentum of 'system i' allows us to infer that of 'system ii'. Similarly, since the relative distance between 'system i' and 'system ii' is conserved, measuring the position for 'system i' would allow us to infer that of 'system ii'. In other words, values of both linear momentum and position are perfectly correlated. In both cases, if the value of momentum (position) is completely determined (with probability equal to one) in 'system i' we will immediately know, also with certainty, the value of momentum (position) of 'system ii' .

Thus, the EPR thought experiment concluded that there exist arrangements in which complementary quantities, such as position and momentum, have simultaneous real values (according to the 'EPR reality criterion'), which, in turn, justified Premise 2.

The EPR thought experiment contains two implicit but crucial assumptions, namely that the systems each constitute a separate (independent) element of reality at the time of measurement, and that they do not interact with each other in any way (this also excludes one system's measurement operations disturbing the other). These two assumptions are commonly known as *separability* and *locality*. We do not need to discuss them at this point but, as we shall see later, they turn out to be central for the subsequent analysis of the consequences of the EPR argument.

It is perhaps worth noting as well that the EPR argument reviewed above did not seem to reflect Einstein's real worries about quantum theory. The EPR argument is usually taken as a critique of Heisenberg's Uncertainty Principle —EPR is usually discussed along these lines—. But, as Fine points out, Einstein's real reservations towards quantum theory seemed to stem more from the 'uncontrollable disturbances' in the interactions of the quantum systems with the measurement apparatus that Bohr defended, and which

[6]The original EPR experimental set-up considerably differs from the one it is usually discussed in the literature —and which I present in the following section— due to Bohm. In particular, it might seem odd that relative position is conserved. However, EPR apparently succeeded in providing a wave function which conformed to both conservation of momentum and relative position. See (Fine, 2004) for details.

provided the grounds for both indeterminism and non-realism.[7]

Also Maudlin[8] suggests, in a similar line of thought, that the EPR argument was originally designed as a 'test' of the non-local character of quantum theory. In Maudlin's view, the argument would go then along the following lines: if it is true that the wave function contains all physical information about the system, quantum mechanics is not complete due to the stochastic character of it. For, given that the evolution of the wave function is governed by Schrödinger's equation, which is entirely deterministic, the stochastic character of the theory comes from the collapse. This would in turn imply that Schrödinger's equation is not a law of nature. Furthermore, Maudlin notes that the above argument is clearly stated in the first part of the EPR paper and would serve on its own but it is later obscured when reference to incompatible, i.e. complementary, observables (linear momentum and position, in particular) are introduced (apparently by Podolsky), invoking the failure of Heisenberg's uncertainty principle.

4.1.3 Bohm's Version of the EPR Experiment

The version of the EPR experiment I will be discussing here is due to Bohm[9]. It is probably the most widely discussed version of EPR in the literature. In an EPR-Bohm experiment (see Figure 4.1) a pair of entangled particles of spin-$\frac{1}{2}$ —I will assume the particles to be electrons— are emitted from a source and move away in opposite directions. The state of such entangled two-particle systems emitted is the so-called *singlet* state:

$$\Psi_s = \frac{1}{\sqrt{2}}(\psi_1^+ \otimes \psi_2^- - \psi_1^- \otimes \psi_2^+).$$

In each wing a measurement of the electron spin is performed by means of a Stern-Gerlach magnet.[10] Moreover, each of the magnets may be oriented in three different directions ϑ_i $(i = 1, 2, 3)$ with respect to the z axis, while kept parallel to the screen. The three possible measurement directions $\vartheta_1, \vartheta_2, \vartheta_3$ are taken to be the same in both wings, however. Let us then denote the event that the measurement apparatus (i.e. the Stern-Gerlach magnet) has been set up to measure in direction ϑ_i $(i = 1, 2, 3)$ in the left wing by L_i.

[7] See (Fine, 1996) for a nice account of Einstein's philosophical worries towards quantum physics.

[8] Cf. (Maudlin, 2002, p.139ff.).

[9] Cf. (Bohm, 1951).

[10] Due to their spin angular momentum, electrons are deflected when flying through an inhomogeneous magnetic field (see Figure 4.1). Spin components are thus determined from such deflected trajectories of the electrons.

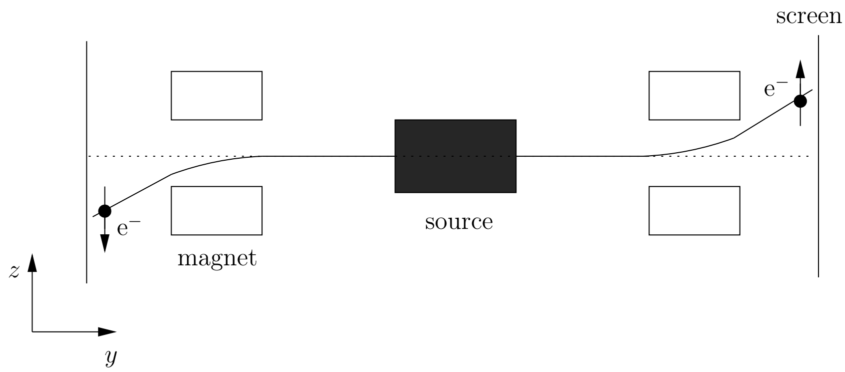


Figure 4.1: Lateral view of a typical EPRB thought experiment set up.

Similarly measurement settings in the right wing of the experiment will be denoted by R_j, where j refers to the corresponding measurement direction ϑ_j $(j = 1, 2, 3)$.[11]

For such a set-up, electrons at each wing of the experiment may be observed (with probability 1/2) to have either *spin-up* (+) or *spin-down* (−) along a given direction. That is to say, two possible outcomes for spin measurements will be observed at each wing for each of the three directions. Let us write L_i^+ for the outcome corresponding of a *spin-up* measurement when the magnet was set in the the direction given by ϑ_i $i = (1, 2, 3)$ in the left wing of the experiment. Similarly R_j^- will denote a *spin-down* measurement in direction j (given by ϑ_j) in the right wing of the experiment. To allow for more general expressions however, I shall write L_i^a for a generic outcome in the left wing of the experiment, where the sub-index $i = 1, 2, 3$ refers to the three possible measurement directions and the super-index $a = +, -$ refers to the *spin-up* (+)/*spin-down* (−) measured component. Thus, L_-^2 denotes the outcome spin-down on a measurement carried out on the left particle along direction 2. Similarly, generic outcomes in the right wing of the experiment will be denoted as R_j^b, where $j = 1, 2, 3$ and $b = +, -$.

It is assumed furthermore that (in each run) events occurring in the left wing of the experiment are space-like separated from those taking place in the right wing (Figure 4.2). In such a situation, causal connections between the different distant events are usually ruled out on relativistic grounds. Space-like separation in the EPR-Bohm set-up is also related to a further (usually implicit) assumption, also present in the original EPR argument: namely that when distant measurements are performed each particle constitutes a

[11]This notation follows that of (van Fraassen, 1982*a*) and (Graßhoff, Portman and Wüthrich, 2005).

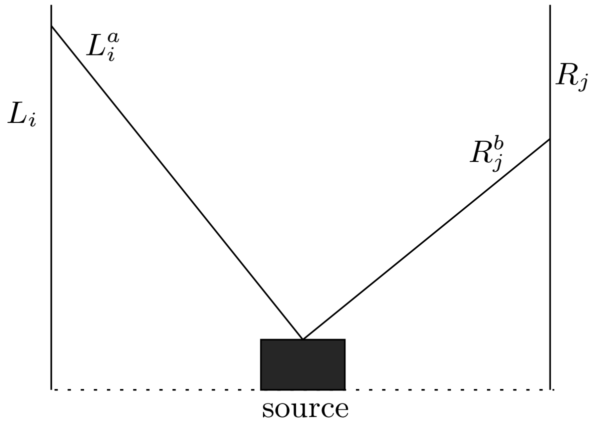


Figure 4.2: Space-time diagram of a typical EPR experiment.

single independent system, i.e. each particle has an independent state. This is also known as *separability*.[12]

After several runs, a statistics is performed of the outcomes of spin measurements in the different directions in each wing. The quantum mechanical formalism assigns probabilities for such outcomes in the form of trace-like quantities. In particular, if we denote by S a generic spin outcome in either wing of the experiment, i.e. $S = \{L_i^a, R_j^b\}$ $(i, j = 1, 2, 3$ and $a, b = +, -)$, its quantum mechanical probability will be given by $\mathrm{Tr}(\hat{W}\hat{S})$, where $\hat{W}$ is the density operator corresponding to the state W of the system (the singlet

[12]Both these assumptions are however controversial. On the one hand, causal claims are always highly sensitive to the theory of causation one is adopting. Counterfactual theories of causation may very well be compatible with distant space-like causal influences. On the other hand, several arguments have been put forward which explore the possibility of superluminal signals in EPR. See for instance (Cartwright and Suárez, 2000) for one such argument.

As for the assumption that the system is *separable*, some may claim that it is not completely justified. Indeed, the fact that the quantum mechanical formalism does not allow us to derive uniquely the complete singlet state wave function Ψ_s from the state of each particle taken individually is very often interpreted in terms of quantum non-separability and holism. Under this view, non-separable models may turn out as a viable alternative for causal explanation of EPR correlations. Healey (1992*b*), for instance, provides one such non-separable model. However, I will not consider this option further here, since my main aim is to study the plausibility of common cause explanations of the EPR correlations, which indeed require separability.

state Ψ_s in our case), and $\hat{S}$ the corresponding projector associated to S.[13,14]

A different matter is the interpretation of quantum probabilities, i.e. trace-like quantities. I will not get into details here about what may their correct interpretation be. I shall just point out that quantum probabilities may be interpreted in terms of relative frequencies, quite similarly to classical probabilities. In particular, quantum probabilities may be viewed as classical conditional probabilities, which allows for an interpretation in terms of relative frequencies.[15] Such an interpretation has been widely adopted in the relevant literature[16] and I shall thus also do it here.

In the case of our spin measurements, recall that spin components were measured (at either wing and in any of the three possible directions) either *spin-up* (+) or *spin-down* (−) with probability 1/2. The quantum mechanical prediction for such spin measurements is given by:

$$\mathrm{Tr}(\hat{W}\hat{L}_i^a) = p(L_i^a|L_i) = \frac{1}{2}, \tag{4.1}$$

$$\mathrm{Tr}(\hat{W}\hat{R}_j^b) = p(R_j^b|R_j) = \frac{1}{2}, \tag{4.2}$$

where $i, j = 1, 2, 3$ and $a, b = +, -$, and all the symbols have their usual meaning (see above for details).

[13]More precisely, the outcome S is identified in this formulation with the projector $\hat{S}$ from the spectral decomposition of the (Hermitian) operator $\hat{s}$ that corresponds to a measurement s, i.e. $s = \{L_i, R_j\}$ $(i, j = 1, 2, 3)$ in the case of spin measurements in typical EPR-Bohm experiments.'

[14]The terminology 'trace-like quantities' for quantum probabilities follows Szabó (2000). This does not suggest in any way that any quantity represented by a trace is to be interpreted as a quantum probability. Only traces such as $\mathrm{Tr}(\hat{W}\hat{S})$ will.

[15]The adopting of such interpretation may be supported by noting that, in fact, quantum mechanics does not provide us with predictions of probabilities for events such as L_i^a, irrespective of the corresponding measurement operation. (Although such probabilities may be deduced, given they are interpreted as relative frequencies, from the corresponding conditional probability expressions, i.e. $p(L_i^a|L_i)$ and $p(L_i)$. Note, however, that $p(L_i)$ depends on the arbitrary choice of the experimenter.) There is a further advantage in adopting such interpretation, namely that measurement becomes explicitly present in our expressions. This will prove of major importance to the foregoing discussion, and in particular in the development of my own common cause model (see Chapter 6).

[16]In fact, all the work on quantum correlations by the Budapest School and related literature mainly follows this view. See for instance (Rédei, 2002), (Hofer-Szabó, Rédei and Szabó, 2000*a*, 2002) or (Szabó, 2000).

Moreover, joint probabilities are also predicted by the theory:

$$p(L_i^+ \wedge R_j^+ | L_i \wedge R_j) = \frac{1}{2}\sin^2\frac{\varphi_{ij}}{2}, \tag{4.3}$$

$$p(L_i^- \wedge R_j^- | L_i \wedge R_j) = \frac{1}{2}\sin^2\frac{\varphi_{ij}}{2}, \tag{4.4}$$

$$p(L_i^+ \wedge R_j^- | L_i \wedge R_j) = \frac{1}{2}\cos^2\frac{\varphi_{ij}}{2}, \tag{4.5}$$

$$p(L_i^- \wedge R_j^+ | L_i \wedge R_j) = \frac{1}{2}\cos^2\frac{\varphi_{ij}}{2}, \tag{4.6}$$

where φ_{ij} is the angle between the two measurement directions i and j.

It follows from expressions (4.1)-(4.6) that quantum mechanics predicts correlations between the different spin outcomes in the distant wings of the experiment:

$$Corr(L_i^a, R_j^b) = p(L_i^a \wedge R_j^b | L_i \wedge R_j) - p(L_i^a|L_i) \cdot p(R_j^b|R_j).$$

The above are the so-called EPR-Bohm correlations. In what follows, I shall be referring to such correlations between spin measurement outcomes also when speaking of EPR-type correlations, or simply EPR correlations.

As special case, *perfect correlations* arise when measurements are taken in both wings with the magnets set in the same direction, i.e. for *parallel settings*. That is, if L_i and R_i are set for measurement then, either L_i^+ and R_i^-, or L_i^- and R_i^+ are obtained as the outcomes in the same run. On the other hand, no correlation is found for outcomes L_i^+ and R_i^+ in the same run, neither are L_i^- and R_i^- correlated. As we saw in the preceding section quantum mechanical perfect correlations (both of values of linear momentum and relative positions) are at the heart of the EPR argument against the completeness of quantum mechanics.[17] And in fact the EPR argument follows as well for perfect correlations of spin components.

4.2 Hidden Variables and the Bell Inequalities

In putting forward the EPR argument, it was not perhaps Einstein's aim to suggest that quantum mechanics be completed in any particular way.[18] The

[17]We will see as well in the foregoing discussion that considerations about perfect correlations turn out to be crucial in most of the arguments aimed at ruling out common cause explanations of EPR correlations (see in particular Section 4.3 and Chapter 5).

[18]See the remarks at the end of Section 4.1.2 about Einstein's real worries regarding quantum mechanics.

truth is that the EPR argument, however cogent, opened the door somehow to the development of a *hidden variable* program. The EPR paper, in fact, ended with an invitation to develop theories which, unlike quantum mechanics, would provide a complete description of reality.[19]

4.2.1 Hidden Variables

Hidden variable theories were proposed as a solution for the incompleteness of quantum mechanics. The aim of such theories was, in particular, to postulate additional 'elements of reality' that jointly with the standard quantum wave function would provide a complete description of physical reality. The name *hidden variables* was due to the fact that these new 'elements of reality' were not present in the standard quantum formalism. Neither was there any empirical evidence for them.[20]

Determinism is also a motivation behind hidden variables. In Section 4.1 we saw how issues about completeness of quantum mechanics and its fundamental indeterministic character were thought to be closely related. Worries regarding the (irreducible) indeterminism of quantum mechanics are also central to the EPR argument, and in particular to the 'EPR reality criterion' (see quotation on page 80). Thus, by adding extra factors (hidden variables), hidden variable theories aimed both to complete quantum mechanics, and to restore the deterministic causal picture of classical theories. Bohm's quantum mechanics is a paradigmatic example.[21]

With the exception of Bohm's quantum mechanics, hidden variable theories, although popular, have not been successful. Perhaps, due to the strong and influential arguments that several so-called *no-go* theorems have provided for the impossibility of such theories.[22] Typically, *no-go* theorems

[19]The very last paragraph of the paper seems very compelling in this regard:

> While we have thus shown that the wave function does not provide a complete description of the physical reality, we left open the question of whether or not such a description exists. We believe, however, that such a theory is possible. (Einstein, Podolsky and Rosen, 1935, p. 780)

[20]Such hidden factors may even be unobservable. However, in the standard treatment of the problem, there is no need to assume hidden variables to be unobservable. The meaning of the term 'hidden' simply refers to the extra factors not being present in the description of the physical system.

[21]See (Bohm, 1952). Quite often Bohmian mechanics is referred to as an 'Bohm's interpretation' of quantum mechanics. However, both Bohm's and de Broglie's approach to quantum mechanics can be considered, specially at the ontological level, as essentially different theories to the standard quantum theory.

[22]Von Neumann (1932) provided one of the first and for quite some time one of the

usually address the issue asking what a hidden variable theory should look like under standard in principle reasonable assumptions, such as locality or separability —which recall were already present, even if implicitly, in the EPR argument—. As a result a set of restrictions on the hidden variables are derived. The typical *no-go* argument would then attempt to show that quantum mechanics does not allow for the hidden variables to satisfy the derived restrictions.

Bell's theorem[23] is one more argument of the kind as well, despite Bell's apparent reservations.[24] Shortly after Kochen and Specker[25] provided a generalisation of Bell's theorem, which indeed constitutes a very compelling argument against the possibility of certain hidden variable theories.[26]

4.2.2 Bell on Hidden Variables

The Bell inequalities set restrictions on EPR-type correlations under a very minimal set of intuitively reasonable assumptions:

(i) The *existence of a hidden variable* λ (or set of variables $\{\lambda_i\}$).

(ii) The so called *Separability* assumption.

(iii) A locality condition (which I will refer to as *Factorizability*).

In Bell's argument, the three assumptions above are sufficient for the derivation of a set of inequalities which impose an upper limit on the EPR correlations. Moreover, it is a virtue of Bell's theorem that it assumes very little about the postulated hidden variables. But let us have a closer look at these assumptions in turn.

It is remarkable in the first place that the hidden variables may be assumed either deterministic or stochastic. I pointed out in the preceding

most influential *no-go* theorem.

[23]Cf. (Bell, 1964).

[24]In fact, one of Bell motivation in addressing the issue was to reassess von Neumann's *no-go* theorem. I shall discuss Bell's theorem and its implications more in detail in the following two sections.

[25]Cf. (Kochen and Specker, 1967).

[26]Strictly speaking we should distinguish here between so-called *contextual* and *non-contextual* hidden variable theories. For Bell-type (*no-go*) theorems, such as the Kochen-Specker theorem, seem only conclusive in the case of non-contextual hidden variables. It is widely accepted that while contextual hidden variable theories may be fully consistent with the standard quantum predictions, and the corresponding empirical data —Bohm's quantum mechanics is one such theory—, there are no non-contextual hidden variable theories that reproduce such predictions and data.

section that hidden variables were partly motivated by a deterministic conception of physical reality. We might then be tempted to think that indeterministic hidden variable models could be provided that avoid the difficulties posed by the *no-go* theorems. However, Bell's theorem turns out to be equally valid whether the hidden variables are assumed to be deterministic or stochastic.[27] In the remainder of this thesis I will assume a stochastic notion of hidden variable. Furthermore, I will assume for the sake of simplicity, that λ is a continuous hidden variable.[28]

Bell's *Separability*, also present in the original EPR argument, states that each particle at each wing of the experiment constitutes an individual system, which is physically distinguishable from the one at the other wing. I briefly pointed out (see Footnote 12 in page 84) that *Separability* is by no means uncontroversial. However, as I also said, it will be necessary for our present purpose to assume it actually holds. I shall thus assume it so in what follows.

Finally, locality is, in Bell's own words, the "vital assumption" in the argument:[29]

> The vital assumption is that the result B for particle 2 does not depend on the setting $\mathbf{a}$, of the magnet for particle 1, nor A on $\mathbf{b}$.

It is remarkable that Bell's locality assumption borrows directly from Einstein, who is quite clear about the fact that the the independence between (both) distant outcomes and settings is due to their space-like separation.[30]

Bell's locality assumption is usually expressed in terms of probabilistic independence between the outcomes at each wing of the experiment and the measurement settings in the distant wing. The resulting expression is what Butterfield, for instance, calls *Factorizability*:[31]

Factorizability:

$$p_\lambda(L_i^a \wedge R_j^b | L_i \wedge R_j) = p_\lambda(L_i^a | L_i) \cdot p_\lambda(R_j^b | R_j),$$

[27]See Appendix C for the details of the distinction between deterministic and stochastic hidden variable.

[28]This will not make a difference to our conclusions. For, as Bell himself stresses (Bell, 1964, p. 196), his results, i.e. the inequalities, holds whether a single hidden variable or a set of them is assumed.

[29](Bell, 1964, p. 15 of the 1987 reprint).

[30]Cf. note 2 in (Bell, 1964, p. 20 of the 1987 reprint).

[31]Cf. (Butterfield, 1989). *Factorizability* is also found in the literature under the name of *Local Causality* (Clauser and Horne, 1974) or *Strong Locality* (Jarrett, 1984). However, it seems to me that *Factorizability* is the most appropriate terminology for our purposes since it refers exclusively to the formal structure of the condition (see below) and it remains completely neutral as to how it should be interpreted.

where $i, j = 1, 2, 3$ and $a, b = +, -$; and $p_\lambda(\cdot)$ denotes the probability with which the hidden variable λ determines the outcomes. That is to say, the probability of the outcomes given that the hidden variable λ is present. The other symbols have their usual meaning (see the notation introduced in section 4.1.3).

Since in the larger probability space $p_\lambda(\cdot)$ is equivalent to a conditional probability (conditional on λ), i.e. $p_\lambda(\cdot) = p(\cdot|\lambda)$, the expression above can also be written as

$$p(L_i^a \wedge R_j^b | L_i \wedge R_j \wedge \lambda) = p(L_i^a | L_i \wedge \lambda) \cdot p(R_j^b | R_j \wedge \lambda), \tag{4.7}$$

which is in fact the expression I will refer to as *Factorizability* in what follows.

The expression above says that the probability (given the hidden variable λ) for the outcome of a measurement in the left wing depends *only* on the measurement settings chosen in the left wing. Similarly, the probability (given the hidden variable λ) for the outcome of a measurement in the right wing depends *only* on the parameter chosen for measurement in the right wing. Moreover, because of such outcome-distant settings independence, the joint probability of any two outcomes in an EPR experiment must factorise —given, that is, that the hidden variable λ is present, and the corresponding measurement settings—. Or in other words, given the whole state of the system, which now includes the hidden variable λ, the probability of any two outcomes factorises.

As I said, *Factorizability* is generally taken to reflect our locality intuitions. However, some authors take this to be a controversial claim, specially in relation to Einstein's ideas about physical locality.[32] I shall not discuss locality issues at this point, neither shall I assess here whether or not *Factorizability* amounts to some notion of locality. In what follows (at least for the remaining part of this chapter) I shall mean 'local' when I write 'factorizable' and *vice versa*, unless explicitly stated otherwise.

All in all, it is important to stress that *Factorizability* is a necessary assumption for the derivation of the Bell inequalities. We will see in the following subsection (as well as in Section 4.3) that *Factorizability* may also be recovered from two further separate independent conditions, which are arguably interpreted as reflecting causal intuitions and physical locality respectively.

4.2.3 The Bell Inequalities

Factorizability is then the crucial assumption that allows Bell to derive his now famous inequalities. It is not needed for my present purposes to pro-

[32] See in particular (Fine, 1996).

vide a derivation of the Bell inequalities. It will suffice, in order to see the implications of Bell's theorem to state the final result, i.e. the inequalities, although in a later version derived by Wigner[33].

What Bell showed was that *factorizable* hidden variables committed to certain relations between the EPR correlations, such as:

$$p(L_1^+ \wedge R_2^+ | L_1 \wedge R_2) + p(L_2^+ \wedge R_3^+ | L_2 \wedge R_3) \geq p(L_1^+ \wedge R_3^+ | L_1 \wedge R_3). \quad (4.8)$$

I still follow here the notation introduced in Section 4.1.3, page 82. (The other inequalities may be obtained by permutations of the numbers 1, 2 and 3.)

As I already pointed out, the inequality above, and the other the Bell inequalities restrict the possible range that the probabilities that characterise the correlations of EPR-Bohm experiment outcomes can have. In particular, the Bell inequalities require spin outcome correlations to be smaller than what the actual quantum mechanical predictions suggest (see equations (4.3)-(4.6) in page 86). In this sense, it is generally acknowledged that the Bell inequalities demand weaker correlations than quantum mechanics provides. What is clear, in any case, is that quantum mechanical predictions of the EPR correlations do not satisfy the Bell inequalities. We can see this for the particular Bell inequality above.

Recall that the quantum mechanical predictions for the outcomes entering equation (4.8) are given by equations (4.3)-(4.6):

$$p(L_1^+ \wedge R_2^+ | L_1 \wedge R_2) = \frac{1}{2}\sin^2\frac{\varphi_{12}}{2}, \quad (4.9)$$

$$p(L_2^+ \wedge R_3^+ | L_2 \wedge R_3) = \frac{1}{2}\sin^2\frac{\varphi_{23}}{2}, \quad (4.10)$$

$$p(L_1^+ \wedge R_3^+ | L_1 \wedge R_3) = \frac{1}{2}\sin^2\frac{\varphi_{13}}{2}. \quad (4.11)$$

From the above expressions, the Bell inequality (4.8) reads

$$\frac{1}{2}\sin^2\frac{\varphi_{12}}{2} + \frac{1}{2}\sin^2\frac{\varphi_{23}}{2} \geq \frac{1}{2}\sin^2\frac{\varphi_{13}}{2}. \quad (4.12)$$

As suggested by Wigner[34], this equation is best understood if, being the three directions coplanar, direction 2 bisects the angle between direction 1 and 3. In this case we have $\varphi_{12} = \varphi_{23} = \frac{1}{2}\varphi_{13}$, and equation (4.12) can be rewritten as

$$\sin^2\frac{\varphi_{12}}{2} \geq \frac{1}{2}\sin^2\frac{\varphi_{13}}{2} = 2\sin^2\frac{\varphi_{12}}{2}\cos^2\frac{\varphi_{12}}{2}.$$

[33] See (Wigner, 1970). This version of the derivation of the Bell inequalities is also presented in (van Fraassen, 1982*a*), which I shall discuss in detail in the Section 4.3.

[34] See (Wigner, 1970).

which entails that $\varphi_{12} \geq \frac{\pi}{2}$, or $\varphi_{13} \leq \pi$.

This is the final result as well even if direction 2 does not bisect the angle formed by directions 1 and 3. Thus, one derives a contradiction (since the only possible values the equation above holds for are $\varphi_{12} = \varphi_{23} = \frac{1}{2}\varphi_{13} = \pi$).[35] The calculation above shows that the quantum mechanical predictions for EPR-Bohm spins violate the Bell inequality (4.8). This theoretical result has been confirmed by experiment.

As a consequence, if we are to avoid the contradiction that the Bell inequalities entail, one of the assumptions that allow for their derivation must be dropped. The common view in this respect puts the blame entirely on *Factorizability*. Although *Separability* is hardly justified in most cases, it is rarely blamed.[36] I shall endorse here the received view, and assume that *Separability* holds. My main interest is to study whether common cause explanations may be appropriate for the EPR correlations. If *Separability* did not hold, it seems to me that remote outcome events occurring in a given run of an EPR experiment should be considered as one and the same event. Such correlations hence would be straightforwardly explained by event identity and would not call (or need) for a common cause explanation whatsoever.

What does it mean then that *Factorizability* fails? If *Factorizability* reflects our intuitions about locality then the violation of the Bell inequalities entails that quantum mechanics is 'non-local', meaning 'non-factorizable'.

In some cases this claim has been further refined by considering the *Factorizability* expression as the conjunction of two independent conditions, so-called *Outcome Independence* and *Parameter Independence*:[37]

Outcome Independence:

$$p(L_i^a \wedge R_j^b | L_i \wedge R_j \wedge \lambda) = p(L_i^a | L_i \wedge R_j \wedge \lambda) \cdot p(R_j^b | L_i \wedge R_j \wedge \lambda). \qquad (4.13)$$

Parameter Independence:

$$p(L_i^a | L_i \wedge R_j \wedge \lambda) = p(L_i^a | L_i \wedge \lambda), \qquad (4.14)$$

$$p(R_j^b | L_i \wedge R_j \wedge \lambda) = p(R_j^b | R_j \wedge \lambda). \qquad (4.15)$$

[35] Wigner (1970) notes that this result is not restricted to the case of coplanar measuring directions. However, recall that in the EPR-Bohm experimental setup (see Section 4.1.3) the three directions are in the same plane —that perpendicular to the direction of the emitted beam—. Hence, I shall only refer here to the coplanar case.

[36] Healey (1992*b*) is an example.

[37] The terminology used here is that of Shimony (1993). Again, these two further conditions can be found in the literature under the names of *Completeness* and *Locality* (Jarrett, 1984) or, as we will see in Section 4.3, *Causality* and *Hidden Locality* (van Fraassen, 1982*a*).

(Where all the symbols have their usual meaning, as in the notation introduced in Section 4.1.3)

Thus, if the conjunction of *Outcome Independence* (OI) and *Parameter Independence* (PI) entails *Factorizability*,[38] the failure of *Factorizability* entails that either OI or PI (or both) fails. The common view is that due to the spherical symmetry of the spin singlet state, standard quantum mechanics satisfies PI, and therefore OI must fail. As part of the this influential view, moreover, PI is usually interpreted as a locality condition, while OI is identified with causation, which has an obvious reading, i.e. that (standard) quantum mechanics does not allow for an underlying causal structure. This conclusion is very much based upon a very influential argument by van Fraassen[39] which I will discus in Section 4.3. But of course, this line of thought is not free from controversy. For, there are also other accounts of quantum phenomena in which it is PI that is violated, while OI holds. This is the case, for instance, in Bohmian mechanics and it will be the case as well in my own common cause model, which I will be presenting in Chapter 6.

To sum up, that quantum mechanics is non-local, i.e. non-factorizable, is probably the most important insight of Bell's theorem. It helped, in particular, to dismiss the belief that the EPR thought experiment entailed a dilemma between locality and completeness. Bell's theorem shows that there is no such dilemma. For it concludes that, regardless of what our description of quantum mechanics is, i.e. regardless of whether we describe it in terms of hidden variables or not, what we surely find is that quantum mechanics is non-local (non-factorizable), as a matter of fact.

4.2.4 Nature (also) Violates the Bell Inequalities

As we saw in the previous section, the Bell inequalities place restrictions on the correlations deriving from EPR-type experiments. We saw as well in section 4.1.3 that equations (4.3)-(4.6) on page 86 expressed the quantum mechanical predictions of the conditional probabilities entering such correlations. Finally, it was found that such quantum mechanical predictions imply higher correlations than those allowed by Bell's theorem. It remained to be seen thus whether Nature agreed with quantum mechanics.

Some time after Bell's theorem was published and its implications already well known and accepted, a number of EPR-type experiments were

[38] Jarrett (1984) claimed that the conjunction of OI and PI not only entails *Factorizability*, but is moreover logically equivalent to it. Logical equivalence does not follow however, as it is shown in Maudlin (2002). Logical equivalence is not needed, however, for the usual argument to go trough.

[39] Cf. (van Fraassen, 1982*a*).

carried out in order to test the predictions of quantum theory. Already in 1974, Clauser and Horne[40], provided the guidelines for so-called Bell experiments. It was not until the work of Aspect, Dalibard and Roger[41], however, that the actual experiment was performed. The Aspect et al. experiments thus became probably the the most significant of all EPR-type experiments. Without going into the details, the actual experiment consisted in correlation measurements of the polarisation angles of two photons, which were simultaneously emitted from a source and then flew apart in opposite directions to be measured thereafter. Although photons are spin-less particles, the polarisation state of such pairs of photons is also the singlet state and therefore EPR-type correlations arise all the same.

The results of the original Aspect et al. experiment were taken to agree with the quantum mechanical predictions, i.e. equations (4.3)-(4.6) on page 86. As a consequence, the Bell inequalities were taken to be experimentally refuted. In spite of the good agreement with the theory, however, some doubts were cast over the reliability of the Aspect et al. experiments, due to high inefficiencies of the polariser detectors. In fact, hidden variable models have been suggested that take advantage of such inefficiencies and which would be completely consistent with the obtained experimental data.[42]

Despite the mentioned drawbacks, however, it is mostly accepted among both physicists and philosophers of science that the violation of the Bell inequalities is confirmed by experiment, in agreement with the predictions of quantum mechanics. Thus, the implications of Bell's theorem outlined in the previous section are to be taken perhaps even more seriously in the face of our experimental observations. Albert[43] puts it very concisely:

> What Bell has given us is a proof that there is as a matter of fact a genuine nonlocality in the actual workings of nature, *however* we attempt to describe it, period. That nonlocality is, to begin with, a feature of quantum mechanics itself, and it turns out (via Bell's theorem) that it is also a feature of every possible manner of calculating (without or with superpositions) which produces the same statistical predictions as quantum mechanics does; and those predictions are now experimentally known to be correct.

Thus, not only quantum mechanics is non-local (in the sense of Bell's *Factorizability*), but Nature also seems to be so.

[40](Clauser and Horne, 1974).

[41](Aspect, Dalibard and Roger, 1982).

[42]This is for instance the case of Fine's 'prism' models (Fine, 1982*c*, 1989). See also (Szabó and Fine, 2002).

[43](Albert, 1993, p. 70)

4.3 Van Fraassen on EPR. Common Causes as Hidden Variables

Due to to its profound conceptual implications Bell's theorem caught the attention of a great number of physicists and philosophers of science. Several variations of Bell's original argument were proposed in an effort to better understand the foundations of quantum mechanics.[44]

A particularly interesting version of Bell's theorem is due to van Fraassen[45], who identified for the first time hidden variables with common causes. Van Fraassen's argument, together with the already known first experimental results of the violation of the Bell inequalities,[46] paved the way for a series of arguments aimed at showing the impossibility of a common cause description of quantum correlations, in particular EPR-type correlations. This was in turn taken to imply that Reichenbach's Common Cause Principle does not hold as a general fundamental principle.

Van Fraassen's argument against common causes is framed within an anti-realist agenda. The argument constituted a very effective critique against scientific realism since it claimed to have shown that in some cases, such as those to which quantum mechanics refers to, Nature does not admit causal explanations —common cause explanations in particular—. As we shall see in a moment, however, van Fraassen's argument refers exclusively to the methodological part of Reichenbach's Common Cause Principle, i.e. Reichenbach's Criterion for Common Causes (RCCC), as I called it, following Suárez, in the preceding chapter (see Section 3.1.1 in page 28). I argued there that RCCC's failure as a sufficient or necessary condition on common causes does not in principle have an impact on the status of the metaphysical content of Reichenbach's Common Cause Principle —which I referred to as the Principle of the Common Cause— since they constitute two completely independent claims.

4.3.1 Perfect Correlations and *Surface Locality*

Van Fraassen's argument involves correlations which display exactly the same structure as those in EPR-Bohm thought experiments. In fact, an explicit reference to quantum mechanical correlations is made. Thus I shall review van Fraassen's argument in the framework of EPR-Bohm correlations. I have

[44]See for instance (Clauser *et al.*, 1969; Clauser and Horne, 1974; Clauser and Shimony, 1978) and (Mermin, 1986), and (van Fraassen, 1982*a*), which I shall discuss in detail now.

[45](van Fraassen, 1982*a*).

[46]See Section 4.2.4 and (Aspect, Dalibard and Roger, 1982) for details.

given a detailed description of the corresponding EPR-Bohm experimental set-up in Section 4.1.3 and we only need to recall now that spin components may be measured in three different directions in each wing, arbitrarily chosen by the experimenter. Following the notation introduced there, L_i (R_i) will denote the type event that the left (right) measurement apparatus is set to measure the spin in direction $i = 1, 2, 3$. Also, the event that the outcome of a measurement in direction $i = 1, 2, 3$ in the left (right) wing is $a = +, -$ —corresponding to spin-up and spin-down outcomes respectively— will be written as L_i^a (R_i^a).[47]

Van Fraassen begins by considering a special case for the 'surface states' —given by the probabilities derived from the corresponding observed frequencies— which are postulated to display, on the one hand, *perfectly anti-correlated* outcomes, i.e. *perfectly correlated* opposite outcomes, when parallel experiments are performed. Such assumption may be written probabilistically as

Perfect Correlation:

$$p(L_i^a \wedge R_i^a | L_i \wedge R_i) = 0. \tag{4.16}$$

Moreover, a so-called *Surface Locality* principle is satisfied by the 'surface states' which requires that "the *outcomes* at either apparatus are statistically independent of the *settings* of the other."[48]

Surface Locality:

$$p(L_i^a | L_i \wedge R_j) = p(L_i^a | L_i), \tag{4.17}$$
$$p(R_j^b | L_i \wedge R_j) = p(R_j^b | R_j). \tag{4.18}$$

Surface Locality as formulated above seems a completely general assumption and may be satisfied in any EPR-Bohm experimental set-up regardless of any special measurement setting arrangements. Note that this is not however the case of van Fraassen's *Perfect Correlation* condition above. In particular, the correlations considered by van Fraassen in the first place are not general instances of EPR-Bohm correlations but a specific subset of those, i.e. perfect correlations. In the next chapter, I shall discuss at length what is the role in my opinion of a *Perfect Correlation* assumption for the derivation of a Bell-type inequality.

[47] Note that this notation is similar to van Fraassen's own. See (van Fraassen, 1982*a*) for details.

[48] (van Fraassen, 1982*a*, p. 103 of the 1989 reprint).

By demanding a causal explanation for the above (perfect) correlations the ideas of 'hidden variables' and 'common causes' are brought together in the argument. Van Fraassen has of course already discarded direct cause explanations in virtue of their apparent conflict with special relativity (due to the *space-like* separation of both wings of the experiment). In particular, *Surface Locality* is taken to capture the intuitions behind those special relativity restrictions, since it rules out superluminal signals between the experimenters.

That the *space-like* separation of the correlated events (and their respective measurement operations) rules out direct causal influences between the two wings of the experiment was already assumed in Bell's original argument. It is a standard assumption in the usual derivations of the Bell inequalities. However, it is a debatable assumption, as I pointed out before (see Footnote 12 in page 84). On the one hand, there are theories of causation that do not require energy or matter transference for the interaction to be causal. This is the case of counterfactual theories of causation. Van Fraassen, however, advises to resist counterfactual inferences of this kind. On the other hand, superluminal processes —capable of producing causal influence— cannot be completely ruled out, even on empirical grounds. In fact, direct cause models have been provided for EPR-type correlations, which explore this possibility.[49]

4.3.2 Common Causes as Hidden Variables

Thus van Fraassen's argument brings together the ideas of 'common cause' and 'hidden variable'. The identification is particularly clear in the following quotation,[50] where van Fraassen suggests that a causal theory aimed to explain such correlations would

> (...) either postulate or exhibit a factor, associated with the particle source, that acts as *common cause* of the two separate outcomes (...) I shall refer to this as *the hidden factor*; not because I assume that we cannot have experimental or observational access to it, but because it does not appear in the surface description (i.e. in the statement of the problem). [The italics are mine.]

Van Fraassen's argument rests on three more assumptions, so-called *Causality*, *Hidden Locality* and *Hidden Autonomy*, which he claims a proper

[49]See (Cartwright and Suárez, 2000) for a direct (superluminal) cause model. A good discussion of such issues may be also found in (Suárez, 2007).

[50](van Fraassen, 1982*a*, p. 104 of the 1989 reprint).

causal theory must satisfy. The first assumption, *Causality* is clearly an application of Reichenbach's Criterion for Common Causes, i.e. the methodological recipe in Reichenbach's Common Cause Principle, to the case of EPR-type correlations:

Causality:

$$p(L_i^a \wedge R_j^b | L_i \wedge R_j \wedge C) = p(L_i^a | L_i \wedge R_j \wedge C) \cdot p(R_j^b | L_i \wedge R_j \wedge C). \quad (4.19)$$

In the expression above all the symbols have their usual meaning (see the notation introduced in Section 4.1.3) and C represents a common cause event type.[51] Two remarks are in order regarding *Causality*. In the first place, I shall point out once more that *Causality* as expressed above is a condition on generic EPR correlation outcomes and not only for the particular case addressed by van Fraassen, i.e. perfect correlations. Secondly, note that although the above expression can certainly be seen as a screening-off condition, the screening-off event is not what van Fraassen assumes to be the cause —which I have written above as the event type C— but a more complex event formed by the conjunction of the postulated common cause and *both* the measurement setting events, i.e. $L_i \wedge R_j \wedge C$. In spite of this we are told that it is C by itself, and only C, which is responsible for both the distant outcomes of the experiment:[52]

> (...) *Hidden Locality* and *Hidden Autonomy*, are meant to spin out implications of the idea that it is the common cause alone, and not special arrangements or relationships between the two separate experimental setups, that accounts for the correlation.

As we can see in the quotation above, the claim that C is the only causal factor of (both) the outcomes is supported by appealing to two further conditions, *Hidden Locality* and *Hidden Autonomy*. The first of these two assumptions, *Hidden Locality* establishes the probabilistic independence of each of the outcomes and their correspondingly opposite measurement settings:

Hidden Locality:

$$p(L_i^a | L_i \wedge R_j \wedge C) = p(L_i^a | L_i \wedge C), \quad (4.20)$$
$$p(R_j^b | L_i \wedge R_j \wedge C) = p(R_j^b | R_j \wedge C). \quad (4.21)$$

[51] Van Fraassen, however, does consider in the first instance a common cause variable A, which can take values q. In his formalism Aq represents the common cause event type that for simplicity I call C here, i.e. $Aq \equiv C$.

[52] (van Fraassen, 1982*a*, p. 105 of the 1989 reprint).

Hidden Autonomy, on the other hand, expresses the probabilistic independence of the (hidden) common cause C and the chosen measurement settings. That is,

Hidden Autonomy:

$$p(C|L_i \wedge R_j) = p(C). \tag{4.22}$$

One can clearly see the similarity between van Fraassen's *Causality* and *Hidden Locality* and the conditions introduced in Section 4.2.3 as *Outcome Independence* (OI) and *Parameter Independence* (PI), respectively. Indeed, they turn out to be equivalent. I pointed out that the logical conjunction of OI and PI entails Bell's Factorizability, which is sufficient for the derivation of the Bell inequalities. Thus, it might seem striking that van Fraassen introduces a further assumption, i.e. *Hidden Autonomy*, in order to derive the Bell inequalities. The reason for that, I think, is simply that *Hidden Autonomy* is introduced in the argument aiming to show explicitly that the common cause C alone —and not in conjunction with anything else (i.e. the measurement setting events)— is responsible for the outcomes of the experiment. Perhaps van Fraassen was not fully aware that the logical conjunction of his *Completeness* and *Hidden Locality* was sufficient for the derivation of the Bell inequalities. For it was not until some time after the first publication of van Fraassen's argument in 1982 that Jarrett[53] showed that Bell's *Factorizability* could be recovered from these two independent conditions.

4.3.3 Derivation of the Bell Inequalities

Van Fraassen derives the Bell inequalities in three steps. In the first step we are shown that for parallel setting experiments, the *Causality* condition entails what it is called *partial determinism*. In particular, experiments for which both *Perfect Correlation* and *Causality* hold have all their possible outcomes determined with certainty by the common cause, i.e. have either probability *one* or *zero*. We have thus in this first step the partial result that

$$\begin{aligned} p(L_i^a|L_i \wedge R_i \wedge C) &= 0 \text{ or } 1, \\ p(R_j^b|L_j \wedge R_j \wedge C) &= 0 \text{ or } 1. \end{aligned}$$

In other words, *partial determinism* says that, conditional on identical measurement settings and the common cause, the probabilities of all possible outcomes are either one or zero.

[53]Cf. (Jarrett, 1984).

In a second step, *partial determinism* is generalised to experiments with different measurement settings as a consequence of *Hidden Locality*. In particular, it is pointed out that *Hidden Locality* expresses the fact that the choice of measurement settings in one of the wings of the experiment is statistically irrelevant for the outcome in the opposite wing. That means then that *partial deterministic* outcomes are deterministic regardless of the measurement choice in the distant wing. Thus so-called *complete determinism* is derived, meaning that, given the common cause is present, any possible experimental outcome (at any of the wings of the experiment) is determined with certainty. This can be written probabilistically as

$$p(L_i^a|L_i \wedge C) = p(L_i^a|L_i \wedge R_i \wedge C) = 0 \text{ or } 1, \tag{4.23}$$

$$p(R_j^b|R_j \wedge C) = p(R_j^b|L_j \wedge R_j \wedge C) = 0 \text{ or } 1. \tag{4.24}$$

In the third and final step joint probabilities for experiments involving different measurement settings in the two wings, i.e. $i \neq j$, are derived. In order to do this, the result of *complete determinism* is first used to classify all possible common causes of a generic EPR experiment into different types, each characterised by a specific set of outcomes for the three possible measurement directions in a given wing. For instance, a common cause of type C^{+++} is said to be such that it brings about *spin-up* outcomes in all three measurement directions in the left wing, say. That is:

$$C^{+++} = C \text{ such that } p(L_1^+|L_1 \wedge C) = p(L_2^+|L_2 \wedge C) = p(L_3^+|L_3 \wedge C) = 1.$$

It can be easily shown that there are eight possible such types of common cause that exhaust all possible outcomes of an EPR experiment (see Appendix D for their derivation).

Looking at these eight types of common cause we see now that, for instance, the experimental outcome $L_1^+ \wedge R_2^+$ may be instantiated either by C^{+-+} or C^{+--}. Thus, the joint probability $p(L_1^+ \wedge R_2^+|L_1 \wedge R_2)$ will be given by:

$$\begin{aligned} p(L_1^+ \wedge R_2^+|L_1 \wedge R_2) = p(L_1^+ \wedge R_2^+|L_1 \wedge R_2 \wedge C^{+-+}) \cdot p(C^{+-+}) \\ + p(L_1^+ \wedge R_2^+|L_1 \wedge R_2 \wedge C^{+--}) \cdot p(C^{+--}). \end{aligned} \tag{4.25}$$

And since the conditional probabilities in the right hand of the expression above are all 1 —as a consequence of the *complete determinism* result in the second step— one may write:

$$p(L_1^+ \wedge R_2^+|L_1 \wedge R_2) = p(C^{+-+}) + p(C^{+--}). \tag{4.26}$$

In a similar way $p(L_2^+ \wedge R_3^+|L_2 \wedge R_3)$ and $p(L_1^+ \wedge R_3^+|L_1 \wedge R_3)$ may be derived:

$$p(L_2^+ \wedge R_3^+|L_2 \wedge R_3) = p(C^{++-}) + p(C^{-+-}), \tag{4.27}$$
$$p(L_1^+ \wedge R_3^+|L_1 \wedge R_3) = p(C^{++-}) + p(C^{+--}). \tag{4.28}$$

Combining equations (4.26), (4.27) and (4.28) we have

$$p(L_1^+ \wedge R_2^+|L_1 \wedge R_2) + p(L_2^+ \wedge R_3^+|L_2 \wedge R_3) \geq p(L_1^+ \wedge R_3^+|L_1 \wedge R_3),$$

which is a Bell inequality.

It is worth pointing out that this third step of the derivation has two implicit assumptions, of which, only one is acknowledged by van Fraassen, namely that the argument implicitly assumes that *Hidden Autonomy* holds. Were this not be so, equation (4.25) could not have been written since the common causes C^{+-+} and C^{+--} could not have been taken as independent of the corresponding measurement setting operation events. Thus, if *Hidden Autonomy* did not hold the Bell inequality could not have been derived. What it is not acknowledged by van Fraassen is a further issue in connection with the postulated common causes, i.e. C^{+-+}, C^{+--} and the like. Such common causes belong, as we will see, to a particular class of Reichenbachian (screening-off) common causes, so-called Reichenbachian *common*-common causes. The identifying feature of a Reichenbachian *common*-common cause C is that it screens-off not just a single correlation but a set of them. The distinction between *common*-common causes and *individual*-common causes, will make a difference when it comes to results about the existence of *screening-off* common causes. A particularly important case is the applicability of the *extensibility* theorem reviewed in the previous chapter. Such a distinction will be seen to be crucial for a defence of Reichenbach's Common Cause Principle against van Fraassen's argument. I will discuss this issue in detail in following section.

To conclude, van Fraassen stresses that the Bell inequalities involve 'surface' probabilities, meaning that they do not depend at all on any hidden variable whatsoever. Since these probabilities are empirically testable, van Fraassen arrives at the 'inevitable' conclusion that the experimental violation of the Bell inequalities rules out a common cause explanation of the EPR-Bohm correlations.

4.4 Causal Realism in Quantum Mechanics

As I already pointed out, van Fraassen's argument against Reichenbach's Common Cause Principle presented in the preceding section has been hugely influential among philosophers of science. More particularly, van Fraassen's argument has contributed a great deal to the background assumption among philosophers of physics that quantum mechanics is not causally interpretable. Van Fraassen's argument, however, is not as compelling as is usually thought. In particular there are at least three possible responses in defence of Reichenbach's Common Cause Principle.

The first option would be to show the plausibility of direct causal models for EPR correlations. I already stressed that my main interest focuses on common cause explanations rather than direct cause ones. Thus, I will not discuss in detail this first possible response to van Fraassen. However, some remarks about this line of thought are in order.

In the first place, recall that van Fraassen's argument —as well as the usual derivations of the Bell inequalities found in the literature— takes it that the *space-like* separation of both wings of the experiment rules out direct causal influences between them.

Were such a successful direct cause model be provided, the consequences for van Fraassen's kind of argument would, of course, be devastating. In particular, providing a direct cause explanation of the EPR correlations would achieve two things. On the one hand, a successful direct cause model for the EPR correlations would show that the correlations involved are not a case of what Reichenbach called 'improbable coincidences' (since both correlated events are already causally related). Thus van Fraassen's critique of Reichenbach's Common Cause Principle would not go through. But on the other hand, such a direct cause model for EPR correlations would automatically provide quantum mechanics with a causal (although non-local in the sense of superluminal) interpretation. This would in turn undermine, not only the specific argument against Reichenbach's Common Cause Principle, but also the more crucial anti-realist argument, taken as a cornerstone of the corresponding larger philosophical program.

The second and third possible arguments in defence of Reichenbach's Common Cause principle have as a starting point the same question, namely whether or not the common causes considered in van Fraassen's derivation of the Bell inequality are the 'right' common cause events for an appropriate causal explanation of the EPR correlations. Both arguments similarly arrive at the same overall answer but on different grounds and therefore with different implications. In fact, the two arguments differ in what such 'right characterisation' is of the common cause events for an appropriate common cause explanation. Let me explain.

Recall that in the previous chapter the metaphysical content in Reichenbach's Common Cause Principle (RCCP) was distinguished and shown independent from the methodological recipe which allowed for the characterisation of the postulated common causes —I referred to these two (independent) claims as the Principle of the Common Cause (PCC) and Reichenbach's Criterion for Common Causes (RCCC)—. In the light of such distinction we may then ask whether van Fraassen's arguments aimed at the PCC, RCCC or both. The answer, I think, is clear: van Fraassen's critique addresses exclusively with RCCC, the methodological part of Reichenbach's Common Cause Principle.

Let us assume for the sake of the argument that van Fraassen argument is right as far as the existence of *screening-off* (Reichenbachian) common cause explanations for EPR correlations concerns. In other words, let us assume, following van Fraassen, that there is no plausible *screening-off* (Reichenbachian) common cause explanation of the EPR correlations. Does that mean in addition that there does not exist a common cause explanation of such correlations? This will depend on whether all common cause explanations are screening-off common cause explanations. We saw in the previous chapter that, as soon as we distinguish carefully between the metaphysical and methodological parts of Reichenbach's Common Cause Principle, it becomes possible to assert that not all common cause explanations involve screening-off common causes. Recall that Cartwright, for instance, exploits such a distinction to argue in favour of *non-screening-off* common causes.[54] The important point here is that a generalised framework for the characterisation of common causes —such as Cartwright's— allows for non-screening-off common cause explanations of the EPR phenomena. And the mere possibility they exist renders van Fraassen's argument inconclusive, at least.

In the previous chapter I also discussed and defended the Budapest School proposal that a *screening-off* (Reichenbachian) common cause may *always* be found for any (genuinely physical) correlation.[55] It is precisely this result by the Budapest School that takes us to the third possible argument in defence of Reichenbach's Common Cause Principle. This third alternative is also aimed to show that van Fraassen's derivation is not considering the 'right' common causes —thus failing to be conclusive as to whether a common cause explanation exist for the EPR correlations. The reasons, however, are different than those above and have to do with so-called *common*-common causes. I shall discuss this option in more detail in the following section.

Before proceeding, however, it is perhaps worth mentioning that there is a fourth possible route if we are to insist on the validity of Reichenbach's Common Cause Principle. This fourth alternative would exploit the vi-

[54] We do not need to review the details and motivations of Cartwright's proposal here again. See Section 3.3 for the detailed discussion.

[55] See Section 3.4.

olation of *Hidden Autonomy* and thus imply, in the usual interpretation, backwards in time causal influences, also identified in some cases with some universal conspiracies. Backwards in time causation is of course controversial. However, several arguments have been advanced that suggest that the (time-symmetrical) theoretical structure of microphysics does indeed allow for such backwards in time phenomena. It has even been claimed that such arguments would even provide, if taken seriously, a causal realist interpretation of quantum mechanics.[56] As for the supposedly conspiratorial features of such models, I shall provide in Chapter 6 a reinterpretation of the violation of conditions such as *Hidden Autonomy*, which will not endorse backwards in time causation, nor any kind of strange universal conspiracy. The violation of *Hidden Autonomy* will be seen there to provide what in my view is a natural Reichenbachian common cause model of the EPR correlations.

4.5 *Common*-common Causes Instead of *Individual*-common Causes

Of the three (four) arguments in defence of Reichenbach Common Cause Principle I mentioned in the previous section, the one involving what I called *common*-common causes is the most interesting for my purposes. The '*common*-common cause' defence is particularly interesting in that, as we will see, it attempts to rescue Reichenbach's Common Cause Principle as a whole (both its metaphysical and methodological claims). As a result, I hope, the possibility will remain open that a Reichenbachian (screening-off) common cause model can fit the EPR correlations.

The '*common*-common cause' argument was originally proposed by Hofer-Szabó, Rédei and Szabó[57] and in this section I shall discuss its main features. Before doing that, however, a small amount of technical work is needed so that the idea of *common*-common cause is properly (and formally) defined and its implications stated as regards common cause completability.

4.5.1 *Common*-common Causes of Quantum Correlations

First of all it should be recalled that Reichenbach's notion of common cause was originally formulated in terms of classical, commutative probability the-

[56]See for instance (Price, 1994, 1996) for such arguments.

[57]See (Hofer-Szabó, Rédei and Szabó, 1999, 2000*a*,*b*) and (Hofer-Szabó, Rédei and Szabó, 2002) for the details of the main ideas and arguments discussed in this section.

ory.[58] But note that we are now dealing with correlations between quantum mechanical events (corresponding to measurement outcomes of certain observables $\mathcal{A}_i$), to which quantum mechanics assigns as probabilities the traces $Tr(\hat{W}\hat{A}_i)$ (for $\hat{W}$ the density operator representing the quantum mechanical state of the system at hand). These probabilities require non-commutative probability theory for their proper description.[59]

Thus, in order to work properly with quantum correlations and quantum (common) causes, the definition of Reichenbachian common cause, as well as other related concepts needs to be 'translated' into non-commutative probability theory. As a fist step, classical Boolean algebras $\mathcal{S}$ need now to be replaced by non-distributive orthomodular lattices $\mathcal{L}$ which contain the quantum events A_i. Secondly, the classical probability measures p will now be replaced by an additive state function ϕ on $\mathcal{L}$ in order to assign the corresponding probabilities $Tr(\hat{W}\hat{A}_i)$.

In this framework the quantum probability of a generic event $A \in \mathcal{L}$ is to be expressed as

$$\phi(A) = Tr(\hat{W}\hat{A}),$$

where $\hat{A}$ is the projector operator associated to A and $\hat{W}$ is the density operator corresponding to the state W of the system.

The definition of (positive) correlation between 'quantum events' that belong in a quantum probability space $(\mathcal{L}, \phi)$ may be now expressed as:[60]

$$\phi(A \wedge B) > \phi(A) \cdot \phi(B).$$

Equivalent definitions and results may also be obtained for the other concepts relevant to our discussion, such as the notion of Reichenbachian common 'quantum' cause or the theorems that lead to common cause completability, which I discussed at length in the previous chapter. Thus, the definition of 'quantum' Reichenbachian common cause may be given as the

[58]See Chapter 3, Section 3.1 for a discussion of Reichenbach's original concept of common cause.

[59]Very succinctly, non-commutative probability theory constitutes a generalisation of classical probabilities to non-abelian (i.e. non-commutative) algebras. This allows to accommodate quantum concepts and features that have no counterpart in classical physics. Cf. (Rédei and Summers, 2007) for further reading.

[60]Note that the idea of correlation here is essentially the same as in classical probability spaces, where a (positive) correlation was characterised by the inequality $p(A \wedge B) > p(A) \cdot p(B)$. Note as well, as it was briefly pointed in the previous chapter, that the arguments here do not lose any generality due to the fact that only positive correlations are considered. In fact, as it was mentioned, similar results are achieved by considering negative correlations as well.

following four (independent) conditions:

$$\phi(A \wedge B|C) = \phi(A|C) \cdot \phi(B|C),$$
$$\phi(A \wedge B|\neg C) = \phi(A|\neg C) \cdot \phi(B|\neg C),$$
$$\phi(A|C) > \phi(A|\neg C),$$
$$\phi(B|C) > \phi(B|\neg C).$$

A quantum mechanical version of this notion has also been worked out.[61] In particular, the following proposition can be proved:[62]

Proposition 1 *Every quantum probability space* $(\mathcal{L}, \phi)$ *is common cause completable with respect to the set of pairs of events that are correlated in the state* ϕ.

The proposition above, similarly to the case of classical probability spaces, says that given two correlated events A and B in quantum mechanical probability space $(\mathcal{L}, \phi)$, an extension $(\mathcal{L}', \phi')$ can be found such that it contains a common cause of the correlation $Corr(A, B)$. Thus, it makes perfect sense to speak, also in the quantum case, of *extensible quantum probability spaces* and, most crucially, *common cause completability*.[63] We shall now see how this notion of common cause completability may be used to argue in defence of Reichenbach's Common Cause Principle in the face of van Fraassen's critique.

4.5.2 Not All Common Causes are *Common*-common Causes

We saw in the preceding chapter that Reichenbachian, i.e. *screening-off*, common causes may be found under the right conditions for any given classical

[61]See e.g. (Hofer-Szabó, Rédei and Szabó, 1999; Rédei, 2002).

[62]For the details of the quantum mechanical versions of the concepts of correlation, common cause, etc., as well as for the detailed proof of the following proposition, see (Hofer-Szabó, Rédei and Szabó, 1999).

[63]Note that above I expressed quantum mechanical probabilities as $\phi(\cdot)$. However, and for the sake of consistency with the notation previously introduced, in what follows I shall write quantum mechanical probabilities (as well as classical ones) as $p(\cdot)$. Whether $p(\cdot)$ refers to a classical or a quantum probability, shall become clear from the context. Which is which will also become evident when measurement operations are made explicit. For, recall that I pointed out in Section 4.1.3 that quantum probabilities may be viewed as classical conditional probabilities —the conditioning events are the (classical) events of actually measuring a given observable $\mathcal{A}$—. That is to say, $Tr(\hat{W}\hat{A}) = p(A|a)$, where a represents the event that a certain measurement has been performed.

correlation. The idea of extension of a (classical) probability space, and that of (Reichenbachian) common cause completability, were crucial in this regard. Similar results may be provided, I pointed out in the previous section, for quantum correlations as well. The Budapest School results thus seem be in contradiction with van Fraassen's conclusions about the existence of Reichenbachian common cause explanation of quantum EPR correlations. So what is really going on? May EPR correlations be explained in terms of Reichenbachian common causes?

The above is, perhaps, a too loose formulation of the problem we are pursuing here. For, before any attempt is made to answer questions such as the above, we need to have a precise knowledge of the properties of the Reichenbachian common causes we would demand for a satisfactory (and reasonable) explanation of the EPR correlations. We shall require, for instance, that the common causes be local in some sense, or that their presence does not influence the (free will) decisions of the experimenters when setting the measurement apparatus. We shall also be interested in some other aspects that regard the structure of the common cause. For instance, we would like to know whether the common causes are deterministic or stochastic, or whether their presence would screen-off one or more correlations.

In the usual approach these intuitions about the postulated common causes are incorporated as extra requirements in the form of probabilistic relations. In other words, besides Reichenbach's Criterion for Common Causes, further probabilistic conditions are tacitly imposed on Reichenbachian common causes for a satisfactory explanation of EPR. This was indeed the case in van Fraassen's argument where, as we saw, *Causality* —identified with the *screening-off* condition characteristic of Reichenbachian common causes— was required alongside *Hidden Locality* and *Hidden Autonomy*. And recall that these two extra assumptions were invoked as reflecting some locality and measurement decisions independence intuitions respectively.

This is why van Fraassen's derivation of the Bell inequality —and his 'inevitable' conclusions about the non existence of Reichenbachian common cause explanations for EPR— can be made compatible with the results on common cause completability by the Budapest School. For, van Fraassen's Reichenbachian common causes are not merely screening-off common causes (since they also satisfy other extra requirements). But, as we have seen, common cause completability only applies in principle to 'unrestricted' Reichenbachian common causes, that is common causes that satisfy Reichenbach's Criterion for Common Causes —i.e. *screening-off* and statistical relevance— and nothing more. Thus there is no such contradiction here.

The immediate question is whether the extra assumptions about the postulated common causes —reflecting locality, for instance— are at all reasonable. Putting it the other way around, do Reichenbachian common cause explanations exist which satisfy further reasonable extra conditions that reflect our standard intuitions? Or is van Fraassen right to claim that his derivation rules out Reichenbachian common cause explanation of the EPR correlations instead?

The questions above might be partially answered by appealing to the notion of *common*-common cause introduced by Hofer-Szabó, Rédei and Szabó[64]. The general idea is fairly simple and straightforward since it merely concerns the number of correlations an event is a common cause explanation of. More particularly, in Reichenbach's definition (characterisation) of common cause we found that *a* common cause C was postulated when an 'improbable coincidence' was observed.[65] Nothing was said, however, as to whether the postulated common cause C can also be used to explain other correlations. It is feasible that a particular correlation may be screened-off by more than one Reichenbachian common cause. Conversely, a particular common cause may screen-off more than one correlation.[66]

I shall mark the distinction by referring to common cause events which screen-off one and only one correlation as *individual*-common causes, or simply (Reichenbachian) common causes.[67] By contrast, I shall refer to common cause events which screen-off a set of two or more correlations as *common*-common causes. Of course the distinction between *individual*-common causes and *common*-common causes is equally meaningful in both classical probability spaces and von Neumann lattices, i.e. quantum probability spaces.

Why is then the notion of *common*-common cause important and how does it affect van Fraassen's own and other and similar arguments against Reichenbach's Principle of the Common Cause? Note first that the results of extensibility and common cause completability —both for classical and quantum probability spaces— refer to, and thus are essentially valid for, *in-*

[64](Hofer-Szabó, Rédei and Szabó, 2002).

[65]Recall that 'improbable coincidence' is the idea behind our correlations $Corr(A, B)$. See Chapter 3, Section 3.1 for details.

[66]Note that the idea of *common*-common cause is in principle quite a general idea about how many correlations an event is a common cause of. However, as I said, the definition given here corresponds to the original notion, in the context of Reichenbachian (screening-off) common causes. But of course, other definitions/characterisations of common cause would have their corresponding definition of *common*-common cause. I shall not consider however other such alternatives here.

[67]This characterisation is also found in the literature under the name of *separate*-common cause. See for instance (Szabó, 2000; Graßhoff, Portman and Wüthrich, 2005) or (Hofer-Szabó, 2007*b*). In my opinion, however, the term *individual*-common cause gives a better sense of the idea that it purports, i.e. that the common cause screens-off a unique correlation.

dividual-common causes. We may then ask whether introducing the notion of *common*-common cause should make a difference and, in particular, whether *common*-common cause completability may hold. But it turns out that, while any given probability space —regardless of whether is a classical or a quantum probability space— may be *individual*-common cause completable with respect to a set of correlations, this is not true in general for *common*-common causes. In other words, it is not true that *common*-common causes can be found for a any given set of correlations by extending the original probability space.[68]

The upshot is that requiring that a set of correlations be explained in terms of *common*-common causes instead of *individual*-common causes is likely to result in failure to provide such explanation. And this is in fact what Hofer-Szabó, Rédei and Szabó[69] claim van Fraassen requires, although implicitly, in his derivation of the Bell inequalities. More specifically, van Fraassen's argument implicitly assumes that the common causes that should explain the EPR correlations are not just *individual*-common causes but *common*-common causes. Although I already pointed this out when reviewing van Fraassen's argument, I did not give a detailed account of why this is so. I shall then make those remarks more precise now.

4.5.3 Van Fraassen's *Common*-common Causes

In Section 4.3.3 I noted that van Fraassen proceeds in three consecutive steps in order to derive the Bell inequalities. Recall in particular that in the third step of the derivation (see page 101) all possible common causes that could be postulated to explain the EPR correlations were classified into eight types. Each such type was characterised by the particular outcomes for the three possible measurement directions in a given wing. (For instance, C^{+-+} was defined to be the common cause type such that, if it were present, $L_1^+(R_1^-)$, $L_2^-(R_2^+)$ and $L_3^+(R_3^-)$ would be brought about with certainty under the corresponding measurement operations $L_1(R_1)$, $L_2(R_2)$ and $L_3(R_3)$. Similarly, under the influence of a common cause of type C^{+--}, the outcomes $L_1^+(R_1^-)$, $L_2^-(R_1^+)$ and $L_3^-(R_1^+)$ would be brought about, also with certainty, correspondingly under measurement operations $L_1(R_1)$, $L_2(R_2)$ and $L_3(R_3)$, etc.)[70]

Recall finally that this particular way of classifying all possible common causes in an EPR experiment is what, together with other partial results,

[68] A formal proof supporting this claim is provided in (Hofer-Szabó, Rédei and Szabó, 2002).

[69] (Hofer-Szabó, Rédei and Szabó, 2002).

[70] See Appendix D for details of how the eight types of common cause are derived.

allows van Fraassen to derive of the joint probabilities $p(L_i^a \wedge R_j^b | L_i \wedge R_j)$ entering the Bell inequalities. For instance

$$p(L_1^+ \wedge R_2^+ | L_1 \wedge R_2) = p(C^{+-+}) + p(C^{+--}), \tag{4.29}$$

since an EPR experiment with outcomes L_1^+ and R_2^+ may be instantiated by either C^{+-+} or C^{++-}. Similarly van Fraassen derives all other joint probabilities. However, we might now look at van Fraassen's definition of common causes from another perspective, and ask how many correlations C^{+-+}, for instance, will be a screening-off common cause of. Of course, $Corr(L_1^+, R_2^+)$ is one of them, as we just have seen, but C^{+-+} screens-off *three* more correlations, i.e. $Corr(L_1^+, R_3^-)$, $Corr(R_1^-, L_2^-)$ and $Corr(R_1^-, L_3^+)$. This follows quite straightforwardly from the definition of the eight types of common cause. It is thus clear that the common causes postulated by van Fraassen are *common*-common causes in that they screen-off more than a single correlation. A similar argument may be followed for the remaining seven common causes considered by van Fraassen.

I would like to stress that the assumption that the common causes belong to one of these types is a requirement besides that they satisfy *Causality*, *Hidden Locality* and *Hidden Autonomy*. In other words, what amounts to the fact that the common causes be *common*-common causes instead of *simple*-common causes is, at least in principle, independent of the fact that the common causes fulfil the required assumptions about locality or independence of the measurement choices. Hofer-Szabó, Rédei and Szabó thus point out that the assumption that the postulated common causes are also *common*-common causes is an unwarranted assumption. Moreover, since *common*-common cause explanations of a set of correlations are not to be expected in general (for just the reasons given above) van Fraassen's argument turns out to be inconclusive as for the existence of *individual*-common cause explanations of EPR correlations.

Of course, I have not shown that the Bell inequalities are not be derived assuming just *individual*-common causes rather than *common*-common causes. Nor am I aware of any such proof. However, the fact that the most common derivations of the Bell inequalities do assume, even if implicitly —just as van Fraassen does—, *common*-common causes instead of *individual*-common causes seems to suggest that a *common*-common cause assumption might be necessary for the derivation of the Bell inequalities. So it might be thought that the key to Bell's theorem lies here, rather than in any matter concerning locality, or causality *per se*. Some recent work by Graßhoff, Portman and Wüthrich questions this line of reasoning. They claim to offer a derivation of the Bell inequalities which explicitly assumes the common causes of the EPR correlations to be *individual*-common causes (and not *common*-common causes). I will discuss Graßhoff, Portman and Wüthrich's derivation in detail in the next chapter and argue that the combination of the different assumptions employed amount somehow to assuming *common*-common causes all the same. So my thesis will indeed be investigating whether the Bell inequalities only refute the use of a specific kind of common cause, namely, *common*-common causes.

Appendix C

Deterministic and Stochastic Hidden Variables

A *deterministic* hidden variable λ is defined such that outcomes of the experiment would be determined with total certainty by the fact that λ is present or not. That is, the overall probabilities will be given by

$$p(L_i^a \wedge R_j^b | L_i \wedge R_j) = \sum_k \lambda_k L(i, \lambda_k) R(j, \lambda_k) \tag{C-1}$$

for a discrete λ distribution λ_k $(k = 1, \ldots, n)$, or

$$p(L_i^a \wedge R_j^b | L_i \wedge R_j) = \int d\lambda \rho(\lambda) L(i, \lambda) R(j, \lambda) \tag{C-2}$$

if λ assumes a continuous value distribution $\rho(\lambda)$. In both cases $L(i, \lambda)$ and $R(j, \lambda)$ are characteristic functions, which represent the event "measuring *spin-up*" (or *spin-down*) in each of the wings of the experiment, i.e. $L(i, \lambda) = 1$ denotes that the *spin-up* $(a = +)$ was measured in direction i in the left wing and $L(i, \lambda) = 0$ that *spin-down* $(a = -)$ was measured in direction i, also in the left wing.[71] Similarly, $R(j, \lambda)$ may take values 1 and 0.

On the other hand, in a stochastic hidden variable model, λ would assign a certain probability $p_\lambda(\cdot)$ for the different possible outcomes to take place. Hence the overall probability would be written as

$$p(L_i^a \wedge R_j^b | L_i \wedge R_j) = \sum_k \lambda_k p_{\lambda_k}(L_i^a \wedge R_j^b | L_i \wedge R_j) \tag{C-3}$$

if λ takes discrete values, or

$$p(L_i^a \wedge R_j^b | L_i \wedge R_j) = \int d\lambda \rho(\lambda) p_\lambda(L_i^a \wedge R_j^b | L_i \wedge R_j) \tag{C-4}$$

in the case λ is a continuous variable.

[71] This would be written in the notation introduced before as L_i^+ and L_i^- respectively.

Appendix D

Eight Types of Common Causes for an EPR Experiment

In this Appendix I derive the eight types of common cause used by van Fraassen in his derivation of the Bell inequalities.

Van Fraassen starts by noting that the outcomes of an EPR experiment respond to the following two questions:[72]

(i) Suppose that L_i. Is it the case that L_i^+?

(ii) Suppose that R_j. Is it the case that R_j^+?

Consider first answers to the first kind of questions, i.e. the (i) question, and assign a common cause C according to the possible answers. Such common cause C will then be defined to be of type C^{abc}, where $a, b, c = +, -$, depending what these answers are:

$$\begin{aligned}
C^{+++} = C: &\ p(L_1^+|L_1 \wedge C) = p(L_2^+|L_2 \wedge C) = p(L_3^+|L_3 \wedge C) = 1,\\
C^{++-} = C: &\ p(L_1^+|L_1 \wedge C) = p(L_2^+|L_2 \wedge C) = p(L_3^-|L_3 \wedge C) = 1,\\
C^{+-+} = C: &\ p(L_1^+|L_1 \wedge C) = p(L_2^-|L_2 \wedge C) = p(L_3^+|L_3 \wedge C) = 1,\\
C^{+--} = C: &\ p(L_1^+|L_1 \wedge C) = p(L_2^-|L_2 \wedge C) = p(L_3^-|L_3 \wedge C) = 1,\\
C^{-++} = C: &\ p(L_1^-|L_1 \wedge C) = p(L_2^+|L_2 \wedge C) = p(L_3^+|L_3 \wedge C) = 1,\\
C^{-+-} = C: &\ p(L_1^-|L_1 \wedge C) = p(L_2^+|L_2 \wedge C) = p(L_3^-|L_3 \wedge C) = 1,\\
C^{--+} = C: &\ p(L_1^-|L_1 \wedge C) = p(L_2^-|L_2 \wedge C) = p(L_3^+|L_3 \wedge C) = 1,\\
C^{---} = C: &\ p(L_1^-|L_1 \wedge C) = p(L_2^-|L_2 \wedge C) = p(L_3^-|L_3 \wedge C) = 1.
\end{aligned}$$

Thus there are eight types of common cause, which describe all possible answers to the first of our questions. As for the second question, i.e. question (ii), we are in a similar situation. Note however that when *complete determinism* holds —which is a partial result derived in the second step of the derivation (see Section 4.3.3)—, answers to the (i)-type questions completely determine answers to the (ii)-type questions.[73] This is easily seen if we consider one of the results expression (4.23) in page 100 refers to. Setting $i = 1$ and $a = +$, for instance, we have:

$$p(L_1^+|L_1 \wedge C) = p(L_1^+|L_1 \wedge R_1 \wedge C) = 1.$$

[72](van Fraassen, 1982*a*, p. 107 of the 1989 reprint).

[73]This is explicitly shown in van Fraassen (1982*a*, p. 107 of the 1989 reprint).

Now since *Perfect Correlation* holds, we will have that

$$p(L_1^+ \wedge R_1^- | L_1 \wedge R_1 \wedge C) = p(L_1^+ | L_1 \wedge R_1 \wedge C) = 1,$$

and also that

$$p(R_1^- | L_1 \wedge R_1 \wedge C) = p(L_1^+ \wedge R_1^- | L_1 \wedge R_1 \wedge C) = 1.$$

The above equation is a particular case of expression (4.24) in page 100, i.e.

$$p(R_1^- | R_1 \wedge C) = p(R_1^- | L_1 \wedge R_1 \wedge C) = 1.$$

Thus

$$p(L_1^+ | L_1 \wedge C) = p(R_1^- | R_1 \wedge C) = 1 - p(R_1^+ | R_1 \wedge C),$$

which means that if the common cause C instantiates L_1^+ it also does instantiate R_1^-.

Therefore, if C^{+++} is the type of common cause for which the left wing outcomes L_1^+, L_2^+ and L_2^+ are predicted, it must also be the one for which the right wing outcomes R_1^-, R_2^- and R_3^- are predicted. This means that the the eight possible types of common causes for the right wing are exactly the same types we have found for the left wing of the experiment.

Chapter 5

Conspiracy, Perfect Correlations and the Bell Inequalities

In the previous chapter I have given an account of the more common arguments relating Reichenbach's Principle of the Common Cause (RPCC) and the EPR quantum correlations. I pointed out that distinguishing *individual*-common causes and *common*-common causes[1] is particularly important in this respect. Moreover, the fact that *individual*-common causes are not correctly identified in the standard derivations of the Bell inequalities seems to suggest that a causal explanation of EPR correlations still remains possible. This approach was, in fact, exploited by Szabó, who proposed an *individual*-common cause model for EPR. Szabó's model, however, had a serious drawback since it featured some 'conspiratorial' character as he himself pointed out.[2]

So-called 'no-conspiracy' conditions have become, in one way or another, a standard in the derivations of the Bell inequalities. However, my view is that violations of such conditions might be understood in two alternative ways, which do not entail any sort of conspiratorial behaviour, as it is usually taken. This is partly the reason for renaming such conditions.[3] I will be discussing

[1]See Chapter 4 Section 4.5 for details. See also (Hofer-Szabó, Rédei and Szabó, 2000*b*) and (Hofer-Szabó, Rédei and Szabó, 2002) for a full detailed account of such distinction.

[2]See (Szabó, 2000).

[3]Of course, if such conditions are reinterpreted such that they do not entail conspiratorial behaviour, the name 'no-conspiracy' is simply misleading. This is why I think it is more adequate to rename such conditions for clarity. I will introduce the term 'measurement independence' to refer to the usual 'no-conspiracy' conditions. More on this in a moment.

the first of such interpretations in this chapter and leave the second for the following chapter, where it will be central to my own *individual*-common cause model of EPR.

It is remarkable that the 'conspiratorial' consequences of Szabó's model served as a motivation for a new (minimal assumption) derivation of a Bell-type inequality, which does not assume *common*-common causes.[4] The bulk of this chapter is in fact a critical discussion of this derivation. We will see that a closer analysis of the argument by Graßhoff, Portman and Wüthrich casts reasonable doubts on the applicability of their derivation. In particular, the central assumption of perfect correlations renders the whole argument problematic. I will argue that precisely because of the assumption of perfect correlations, several underlying assumptions hold which either point to *common*-common causes or constrain severely the applicability of the derivation. Finally I shall also discuss the necessity of assuming perfect correlations for the derivation of the Bell inequalities. I will close the chapter relating the 'mixing' of different perfectly correlated runs of an EPR-Bohm experiment to the existence of *purely formal* correlations —in contrast to *genuinely physical* correlations[5]— that I suggest might be behind conspiracy issues.

5.1 Szabó's Common Cause Model

A very significant effort to give a satisfactory account of EPR/EPRB correlations in terms of common causes can be found in the work of Szabó[6]. In fact, the common cause model proposed by Szabó constituted a shift in the way the EPR paradox had been treated so far, as regards to common causes.

Szabó's approach to EPR relied on a reinterpretation of the concept of 'quantum probability' on the one hand, and on a clear cut distinction between *individual* and *common*-common causes on the other. In particular, the model was grounded upon and made use of the implications of the work by the Budapest School.[7] The goal of the model was to resolve the paradox by explicitly constructing an *individual*-common cause model for the EPR/EPRB correlations with the help of Hofer-Szabó et al. extensibility theorems.

I am not interested in evaluating Szabó's model at this point, i.e. whether or not it succeeded in providing a proper *individual*-common cause model

[4](Graßhoff, Portman and Wüthrich, 2005).

[5]Recall the distinction between *purely formal* and *genuinely physical* correlations I introduced in Chapter 2.

[6](Szabó, 2000).

[7]See Chapter 3, Section 3.4, and (Hofer-Szabó, Rédei and Szabó, 2000*b*) for details.

of the EPR correlations. This section is less critical and hopefully more constructive. For the aim in this section is to point out and discuss the novelties and new insights that the model introduced with respect to prior discussions of the problem. This will help in developing and applying the idea of Reichenbachian common cause in the particular case of the EPR correlations. Thus it will not be necessary to reproduce here the technical details of the model.[8]

There are two aspects in Szabó's model that I think are worth a careful analysis. In the first place, as I already pointed out, reinterpreting the idea of 'quantum probability' was central to the model. The second aspect which makes Szabó's model specially interesting is the explicit requirement of the so-called *no-conspiracy* assumption. *No-conspiracy* is required, along with some locality restrictions on the common causes, aiming to capture the intuition that there be a statistical independence between the common causes and the measurement settings. It is remarkable that the main drawback of Szabó's model is that, in spite of such restriction, it features a deeper kind of 'conspiracy'.

An additional issue, perhaps related to Szabó's model 'conspiratorial' character, seems worth a close look. It is remarkable that in setting the appropriate probability space for the proper description of the EPR-Bohm experiment, two qualitatively different types of correlations (or statistical dependencies) can be distinguished. The distinction —which is usually tacitly assumed and hence goes unnoticed (Szabó does not point it out either)— identifies and differentiates purely formal features from physical features of the description. I will suggest that such a distinction is crucial when discussing EPR/EPRB correlations. For it might help us explain violations of the referred *no-conspiracy* assumptions.

Let us start by looking at Szabó's proposal for a correct interpretation of 'quantum probabilities'.

5.1.1 Quantum Probabilities are Conditional Probabilities

Szabó's reinterpretation of the idea of 'quantum probability' gives a central role in the specification of measurement settings. It is rightly pointed out by Szabó that the usual treatment of the problem does not take this issue properly into account.

[8]That is, the way the *individual*-common causes are found in it. It will enough to point out to this end that this is actually achieved by systematically applying the above mentioned extensibility theorems. See, again, (Szabó, 2000) for the details.

In the usual approach to the problem, the 'quantum probability' $p(A)$ of a particular property A —say A represents for instance the spin of the particle in the right wing— is identified with the quantum mechanical probability $Tr(\hat{W}\hat{A})$ provided by the theory. In this approach clearly measurement operations, performed with a certain probability within the whole experiment, are not taken into account. In this sense Szabó points out that under such an interpretation —an precisely because measurement operations are not taken into account— 'quantum probabilities' are non existent (in reality) and cannot be understood as relative frequencies. More particularly, if measurement specifications are not taken into account in our description, the probabilities entering Bell's theorem cannot be interpreted as relative frequencies.

This renders the question as to whether a common cause explanation for EPR correlations exist nonsense. In Szabó's words:[9]

> [...] this question [whether a common cause exist for the EPR correlations], in the above formulation, is meaningless. I am convinced that we must not ignore the fact that [...] the numbers $q(A)$, $q(A')$, $q(B)$, ... are not interpretable as relative frequencies and do not form a Kolmogorovian probability system that the Reichenbach axioms could apply to. There are no events —and in principle, there cannot exist events in reality the probabilities (relative frequencies) of which would be equal to "quantum probabilities" $q(A)$, $q(A')$, $q(B)$, ...

The right approach would hence require that 'quantum mechanical' probabilities, i.e trace-like quantities $Tr(\hat{W}\hat{A})$, be interpreted as *conditional* probabilities —conditional, that is, on the measurement operations.[10]— For instance, the 'quantum probability' that the particle in the left wing of the experiment is measured with spin-up will be given by $p(L_i^+|L_i)$. And the probability for a spin outcome measured along one of the possible measurement directions i in the left wing will also take into account that measurement has been performed in that particular direction i. That is

$$p(L_i^+) = p(L_i^+|L_i) \cdot p(L_i).$$

Under this interpretation, the question as to whether a common cause exist for EPR correlations, can be reformulated properly (and meaningfully) in terms of conditional probabilities.

[9](Szabó, 2000, pp. 4 and 5).

[10]The difference between 'measurement operations' and 'measurement setting operations' (or 'measurement settings' for short) will not play any important role for the purposes of the discussion here. I shall thus use those two expressions equivalently.

In this sense Szabó's model is a step forward towards a clarification of some terms often very loosely used in deriving and discussing Bell's theorem. In the next chapter I will try to go a bit further with the idea that quantum mechanical probabilities, i.e. trace-like quantities, shall only be interpreted as conditional probabilities —again, conditional on the specific measurement choices—. I will be suggesting then that measurement choices are important not only descriptively, in clarifying the properties of the probability space where EPR correlations are reproduced. Also physically, I will claim, measurement choices are to be taken seriously into account. I will suggest, in particular, that measurement choices must be an important part of the causal structure, if any, underlying the EPR correlations. But let us leave it here for the moment and turn to another crucial assumption in Szabó's common cause model, i.e. *no-conspiracy.*

5.1.2 No-Conspiracy and Measurement Independence

The starting point for Szabó's *No-conspiracy* assumption is the underlying intuition that the postulated common causes must not have an influence, or be otherwise influenced, by the choices of measurement settings. The presence of any of these kinds of influence in the model at hand would amount to a strange sort of conspiratorial behaviour. It would be conspiratorial in the sense that the presence of the common cause would 'force' or determine to some extent the presumably free independent decisions of the experimenter on measurement.

Hence, the fact that we regard the experimenter's decisions leading to specific measurement settings in an EPR experiment as free will decisions suffices to rule out such influences. Non-conspiratorial behaviour of this sort can be explicitly ruled out by imposing a statistical independence between the postulated common causes C and the measurement settings in each wing of the experiment.[11] That is

$$\begin{aligned} p(L_i \wedge C) &= p(L_i) \cdot p(C), \\ p(R_j \wedge C) &= p(R_j) \cdot p(C). \end{aligned} \tag{5.1}$$

Conditions of this sort are common in the standard derivations of the Bell inequalities but they are only assumed implicitly in most cases. One of the

[11] Once more I am assuming here that, at least for the purposes of the argument, stochastic relations of this sort have a causal reading. We have seen that causal inference from probabilities is highly problematic and that this assumption involves considerable judgement. However, as I have discussed in Chapters 2 and 3, one can under certain conditions, reliably establish a more or less faithful correspondence between causal connections and stochastic relations. The argument would be compromised fatally if this were not assumed.

few cases in which an assumption similar to Szabó's *no-conspiracy* is made explicit is van Fraassen's *Hidden Autonomy*.[12]

Since I want to claim that violations of conditions such as the above can be meaningfully interpreted and do not necessarily entail a conspiracy, I shall rename Szabó's *no-conspiracy* condition. From now on, unless stated otherwise, I will refer to the condition expressed by equations (5.1) as *measurement independence*.[13]

I have already pointed out that Szabó's model fulfils the above *measurement independence* condition. It turns out however that it features other dependencies involving combinations (both conjunctions and disjunctions) of the different (explicitly) calculated *individual*-common causes. These dependencies, of course, are taken to be conspiratorial: such is the usual (and Szabó's as well) interpretation.

Thus, while Szabó's model fulfils what we could call 'simple' *measurement independence* conditions —equations (5.1)— it is found to violate other more general *measurement independence* conditions, which I will refer to as 'generalised' *measurement independence* conditions in that they involve all possible combinations of the postulated common causes.[14] That is:

$$p(C_{ii}^{+-} \wedge \neg C_{jj}^{+-} \wedge L_i \wedge R_j) = p(C_{ii}^{+-} \wedge \neg C_{jj}^{+-}) \cdot p(L_i \wedge R_j). \tag{5.2}$$

As I said, the fact that violations of such 'generalised' *measurement independence* conditions are taken to be 'conspiratorial' has a crucial impact on Szabó's solution of the EPR problem. As Szabó himself confesses, numerical calculations suggest that removing those more general conspiratorial features is a hopeless task.

However, we need not pursue this 'hopeless' task, after all. For there are, as I see it, two possible ways out this. On the one hand, it is worth noting that the significance of the *measurement independence* assumptions such as the above heavily rely on another underlying implicit assumption. This is the assumption that the postulated common causes take place (or happen) not only *before* measurement operations are performed, but also

[12]See Chapter 4 Section 4.3 and (van Fraassen, 1982*a*) for details.

[13]In the spirit of Shimony's *outcome* and *parameter independence*, the expression *measurement independence* tries to capture the idea that the postulated common cause is independent of both the measurement settings.

[14]The idea of 'generalised' *measurement independence* condition, which Szabó's model would not fulfil, is discussed in detail in (Hofer-Szabó, 2007*a*). It is also present as an explicit assumption ('NO-CONS') in the minimal assumption derivation of the Bell inequalities by Graßhoff et al., which I will discuss in the next section. Note that both (Hofer-Szabó, 2007*a*) and (Graßhoff, Portman and Wüthrich, 2005) follow the usual interpretation of equations (5.1) and thus refer to such conditions as *no-conspiracy*.

before the apparatus are set for measurement (always in the rest-frame of the laboratory).

It is indeed remarkable that such an assumption is present, although usually implicitly, in all derivations of the Bell inequalities known to date.[15] And it is in this context that the *measurement independence* assumption makes sense. However, such an assumption is not necessarily forced upon us.

In the following chapter I will discuss in detail the possibility that the common cause be conceived as an event which takes place, or acts, at the time of measurement. I will argue that this option is plausible if we are to 'integrate' the influence of measurement operations as causally relevant to the outcomes (both of them) of an EPR experiment. Under such interpretation the situation is then completely different. Since measurement events are taken to be an integral part of the common cause itself the violation of equations (5.1) does not constitute any kind of strange universal conspiracy. It just expresses the fact that the common causes are indeed dependent on measurement, which is a necessary consequence of the structure of the common cause so defined.

A second way out of the violation of a 'generalised' *measurement independence* condition makes use of the distinction between *purely formal* and *genuinely physical* correlations that in the model.

5.1.3 *Purely Formal* Correlations and *Genuinely Physical* Correlations Revisited

It is pointed out in Szabó[16] that the probability structure used to reproduce the EPR experiment contains some correlations between the different measurement choices in each wing. In particular, it is noted that correlations such as $Corr(L_i, L_j)$ for $i \neq j$ exist in the model's underlying probability structure.

But of course such correlations are purely artificial. They are the result of the framework employed to formally model the problem, namely: a probability distribution defined over *all* variables, including conditional probabilities between measurement set-up variables.[17]

[15] The assumption that common causes take place before measurement setting operations —and therefore also before measurement operations as such— is so rooted in our minds that models have even been proposed that allow causal influences to propagate backwards in time in an attempt to by-pass the consequences of Bell's theorem. I shall briefly discuss such kind of models in the following chapter. See also (Price, 1996) and (Cramer, 1986).

[16] (Szabó, 2000, p. 8).

[17] In this sense, these correlations can be viewed fundamentally as a (somehow 'irreducible') formal structural feature reflecting the fact that we are to describe the EPR

Such considerations motivated the distinction between *purely formal* and *genuinely physical* correlations introduced in Chapter 2. Recall that we considered, on the one hand, *purely formal* correlations —or simply *formal* correlations— arising in a the context of a particular description or interpretation of a set of statistical data. Such correlations appeared in the very process of formalising the description of the data at hand —in the 'embedding' of the data in a model, that is—. However, we saw that they do not reflect the possible physical underlying features that might be responsible for the data. By contrast, correlations that could be attributed to the physical properties underlying the data were referred to as *genuinely physical* correlations.

Recall as well that it was a remarkable fact about *purely formal* correlations that they arose whenever 'incompatible' measurement operations were performed in the same experiment. And, as we have seen, this is indeed the case as well for some of the correlations that EPR experiment involve, such as those above pointed out by Szabó. In Szabó's model correlations between different measurement settings in the same wing, although explicitly noted, are nevertheless not taken to be important. Nor are their possible implications investigated. The point is simply made that they are not to be dealt with since they do not conflict with locality intuitions in any way.[18]

However, whether or not the correlations between different measurement setting events in the same wing have an impact on the model seems in itself a worthy and interesting question to pursue. Moreover there is a sense in which such correlations can be seen to be, at least, partly related to measurement dependence issues, i.e. to the violation of Szabó's *No-conspiracy*. For note that *purely formal* correlations could be responsible for certain (statistical) dependencies among the measurement settings in the *distant* wings. Therefore it will be crucial to clearly identify *purely formal* correlations and set them aside to avoid mixing them with other correlations that may very well reflect genuine causal features of the system. Failing to do this may result in the misinterpretation of *purely formal* correlations (or correlations derived from these) in terms of causal influences. An example can be found in EPR-Bohm experiments in which measurement settings are always chosen to be parallel, i.e. perfect correlated experiments.[19] Let us see how this is so.

---

experiment and correlations in a probability space in which *all* three possible measurements and their respective outcomes are represented. These correlations are not however exclusive of EPR experiments. On the contrary, they are a quite general feature of many similar descriptions and thus it seems they cannot be avoided at any rate in this particular case.

[18]See (Szabó, 2000, p. 8).

[19]The argument to follow can be found to be related as well to the derivation of a

5.1.4 *Purely Formal* Correlations and Measurement Dependence

Say for instance that we consider an EPR-Bohm experiment consisting in several runs of two parallel setting measurements, i and j.[20] Say that measurement is performed in direction i (in both wings) in n runs of the experiment and call these *type-i* runs. Also, suppose that a number m of runs are performed in the experiment such that the $j \neq i$ direction is chosen for measurement. Call these *type-j* runs.

In a probabilistic description of the whole experiment (that is of the $m + n$ runs) we will find different correlations, for different combinations of measurement settings and outcomes. The correlations I am interested in at this stage are correlations between the measurement choices, for they are what we have characterised as *purely formal* correlations. Let me explain.

Due to the very structure of our EPR-Bohm experiment correlations between the two possible measurement settings in each wing arise. For instance, in the left wing we get correlations such as

$$Corr(L_i, L_j) \neq 0 \quad i \neq j.$$

These are clearly *purely formal* correlations since they arise as a result of embedding both the descriptions of *type-i* and the *type-j* runs in the same whole probabilistic model. Moreover, such correlations do not reflect any physical property of the EPR-Bohm system at all. Such correlations do not seem to be problematic in any sense, either. This is why Szabó does not pay much attention to them. In particular, in Szabó's opinion, the above correlations can be completely dismissed since they do not threaten in any way our locality intuitions.

The existence of correlations such as the above, however, already gives us a hint that correlations will arise as well between different measurement settings in both wings of the experiment. In particular, since our EPR-Bohm experiment is a collection of two types of parallel measurement setting runs only, *type-i* and *type-j* —in which the corresponding perfect correlations arise— we will observe correlations between distant wing's settings. That is:

$$Corr(L_i, R_j) \neq 0 \quad i \neq j.$$

Again, it is clear that such correlations are *purely formal* since they arise exactly in the same way as $Corr(L_i, L_j)$ above. In fact, we will have that

$$Corr(L_i, R_j) = Corr(L_i, L_j) \neq 0 \quad i \neq j.$$

Bell-type inequality I will be discussing in the following section.

[20] We may assume for instance that the measurement devices are such that the two possible measurement directions cannot be set in each wing independently.

It is at this point that the distinction between *purely formal* and *genuinely physical* correlations becomes crucial. For if we took the expression $Corr(L_i, R_j) \neq 0 \quad i \neq j$ as a *genuinely physical* correlation we would be suggesting that there are physical dependencies between remote measurement settings. In other words, if the above correlations are not properly understood as a *purely formal* feature of our description they may be easily interpreted as a violation of locality! Moreover, such a violation of locality, i.e. such remote measurement setting dependencies, may be interpreted, under the right conditions, as causal.

Thus, the above example suggests that, a statistical dependence relation in our model can be easily overlooked as *genuinely physical* —and perhaps causal, under the right conditions— while having a *purely formal* origin. And this might well be, for instance, the origin of a violation of *measurement independence* conditions in Szabó's model —which as we saw are taken in the standard view to reflect some conspiratorial character—. The reason is simply that once we accept that there are dependencies between distant measurement settings we seem bound to accept as well that there are dependencies between the measurement settings and the corresponding distant outcomes. And since outcomes clearly depend on the postulated common causes, it seems plausible then that common causes also display probabilistic dependencies with the measurement settings.[21] This is however just a conjecture at this stage —although a quite plausible one, in my view— and more work is needed in order to investigate whether statistical measurement dependencies in Szabó's can be traced back to *purely formal* correlations.

Of course, the above only follows if our EPR-Bohm experiment *contains*[22] runs in which the same measurement settings are chosen in both wings, i.e. runs with parallel settings which moreover display perfect correlations among outcomes. In other words, if our EPR-Bohm experiment consists in runs in which different measurement directions are chosen in each wing each time, *purely formal* correlations such as $Corr(L_i, R_j) \neq 0$ will not exist and the argument above will not obtain. Under this view, therefore, if our EPR-Bohm experiment *contains* parallel settings runs, there is nothing conspiratorial in the fact that *measurement independence* is violated. Indeed, measurement dependencies should be in such cases to be as expected —just as 'normal', that is— as the *purely formal* dependencies among different measurement settings in each individual wing, which we cannot get by

[21] As we will see later, violations of *measurement independence* constitute a case for the violation of *Parameter Independence*. See Chapter 6 for details.

[22] We do not need that the experiment be exclusively formed by such kind of runs, as in the example above. It is only needed that some of the runs in the experiment are parallel settings runs so that *purely formal* correlations such as $Corr(L_i, R_j) \neq 0$ arise.

without.

A reaction to my argument above would be to point out that the correlations involved in Bell's theorem are not, so to speak, mixtures of perfectly correlated experiments. On the contrary, one could say, the Bell inequalities tell us about correlations that arise in EPR-type experiments for which different settings are chosen in both wings.[23] And in such cases, as I just pointed out, *purely formal* correlations such as $Corr(L_i, R_j) \neq 0$ do not arise. I agree but I think such a reaction only reinforces my point. We need to make sure that our EPR-type experiments do not involve *purely formal* correlations among distant measurement settings as the above. Or, alternatively, if it is the case that such correlations are present in the model, then we need to make sure that they are properly identified and their impact traced back, whether it is in the form of a violation of *measurement independence* or other statistical dependencies.

5.2 Yet Another Derivation of the Bell Inequalities

From the analysis carried out so far, it is still not very clear whether a proper and sensible *individual*-common cause model for EPR/EPRB can be provided. On the one hand, we have seen, the distinction between *individual*-common cause and *common*-common cause has proved very effective in avoiding the most compelling arguments against common causes, such as van Fraassen's. On the other hand, the main model proposed which explicitly builds on this distinction (Szabó's model) features unavoidable dependencies between certain combinations of common causes and measurement. And this may be interpreted as some kind of 'conspiracy'.[24]

In the previous section I gave an argument against the 'conspiracy' interpretation of the violation of *measurement independence* conditions. I also pointed out that there was another argument by which a natural *individual*-common cause model can be provided that does not fulfil *measurement independence*. As I said, I shall discuss such a possibility in the next chapter. It might seem then of little interest if we go back now to the 'conspiracy' (standard) interpretation of the violation of *measurement independence* and

[23] I will sharpen this point in the following section, where I discuss a recent derivation of a Bell-type inequality. There, I will focus precisely on the fact that perfect correlations are necessarily assumed.

[24] Although, as pointed out by Szabó at the end of the revised version of his (Szabó, 2000), even if numerical calculations suggest that such conspiracies cannot be removed, those numerical calculations are not decisive or conclusive.

keep asking whether an *individual*-common cause model may exist that does not contain violations of 'generic' *measurement independence* conditions —or conspiracies, in its own terms—.[25] But it is interesting to do so in so far new insights can be gained by pushing the concept of Reichenbachian common cause all the way. This will test the limits of the concept.

Let us hence discuss a further attempt to answer the question above and see what we can learn from it. The attempt I am referring to is a paper by Graßhoff, Portman and Wüthrich[26] in which a negative answer to the question —as to whether an only *individual*-common cause model can be given that does not contain the above mentioned dependencies— is given. Graßhoff, Portman and Wüthrich claim, in particular, to have provided a derivation of a Bell-type inequality that does not assume *common*-common causes. This, of course, would constitute a strong claim in the usual interpretation. Indeed, if proved valid, such a derivation may very well be taken to confirm the non validity of Reichenbach's Common Cause Principle as regards quantum correlations.[27]

In this section I will present in detail the Graßhoff et al. 'minimal assumption' derivation of a Bell-type inequality. A critical discussion will be carried out in next two sections.

5.2.1 Perfect Correlations and Common Causes

In what follows, I will be referring only to quantum correlations that arise in a typical EPR-Bohm experiment.The notation used in the description of EPR-Bohm correlations will also be used here.[28] Once more, it will be further assumed that the term *event* stands for *event type* and that any such event A 'takes place' in a quantum probability space $(\mathcal{L}, p)$ described by a von-Neumann lattice $\mathcal{L}$ and an additive function state p.

The derivation of the Bell inequalities provided by Graßhoff, Portman and Wüthrich[29] starts by considering the case of *perfect correlations*. I will discuss

[25] This question left open at the end of Szabó's paper (Szabó, 2000).

[26] See (Graßhoff, Portman and Wüthrich, 2005).

[27] However, that RCCP is found not to hold for quantum phenomena would constitute a strong argument against its non-fundamental character in general —in a similar vein than van Fraassen's argument in the previous chapter—. This would also raise some questions as to the real significance of the results by Hofer-Szabó, Rédei and Szabó's presented in Chapter 3 —see also (Hofer-Szabó, Rédei and Szabó, 1999, 2000*b*; Rédei, 2002)—, i.e. what conditions on common causes are really implied by those and whether these constraints can be thought to be reasonable for quantum phenomena.

[28] See Chapter 4, Section 4.1.3 for a detailed account of the EPRB experimental set-up and its main features.

[29] (Graßhoff, Portman and Wüthrich, 2005).

later on what is, in my opinion, the significance of taking perfect correlations as an assumption for the derivation of a Bell-type inequality. For the moment it is enough to point out that the whole argument by Graßhoff et al. takes this assumption for granted. This fact is even acknowledged by the authors in the discussion at the end of the referred paper.[30] At any rate, the *Perfect Correlation* assumption is presented by the authors as:[31]

Perfect Correlation (PCORR):

$$p(R_i^-|L_i^+ \wedge L_i \wedge R_i) = 1,$$
$$p(L_i^+|R_i^- \wedge L_i \wedge R_i) = 1. \tag{5.3}$$

Besides the above PCORR, the authors assume a *Separability* (SEP) condition:

Separability (SEP): *The coinciding instances of L_i^a and R_j^b are distinct events.*

The authors' *Separability* condition above is motivated by the observation that events such as L_i^a and R_j^b can take place simultaneously while located at large spatial separated points. This is taken then to suggests that L_i^a and R_j^b are indeed distinct events.[32]

Furthermore, given that the large spatial separation of L_i^a and R_j^b is even space-like, i.e. both events lie outside each other's light-cone, they are both taken to be *causally irrelevant* to each other. Conflict with special relativity is avoided in this way.[33] This turns out to be the authors' first *locality*

[30](Graßhoff, Portman and Wüthrich, 2005, pp. 676-677).

[31](Graßhoff, Portman and Wüthrich, 2005, p. 668).

[32]As we have seen (see Chapter 4) conditions similar to the above *Separability* are also assumed in the standard arguments for the derivation of the Bell inequalities. Recall however that most of those derivations introduce this kind of assumption implicitly, or at least without proper justification. I also pointed out that it is remarkable that Bohr's first response to the EPR argument was that the spin-singlet system be treated as a whole. This may be interpreted as suggesting some non-separable character of the spin-singlet state. Discussions on whether quantum mechanics may imply that certain systems are non-separable give rise to claims of quantum holism. I would like to point out once more that, even if I will be assuming in what follows the EPR remote outcomes to be different 'separate' events, this assumption may certainly be challenged. But the purpose of this work is not to explore the consequences of this assumption.

[33]Again, we should bear in mind that this is the received interpretation of the situation. There are several causal models which explore the possibility that there being causal influence, after all, between the remote outcomes of an EPR experiment. This option may even be conceived such as it manages to avoid contradictions with relativistic restrictions. See, for instance, (Cartwright and Suárez, 2000; Suárez, 2007). However, since my aim here is to explore the possibility of a common cause explanation of the EPR correlated outcomes, I shall explicitly require that both the outcomes are different events (that is assuming *separability*) and that they do not causally influence each other in any way.

assumption.[34]

Locality 1 (LOC1): *No L_i^a or R_j^b is causally relevant for the other.*

The assumptions of *Separability* and LOC1 exclude *event identity* and *direct causation* explanations respectively. Everything points hence towards a *common cause* explanation of the EPRB correlations. This justifies introducing the *screening-off* condition[35] of Reichenbach's Common Cause Principle. Screening-off is referred to as the *Principle of Common Cause*:[36]

Principle of the Common Cause (PCC): *If two event types A and B are correlated and the correlation cannot be explained by direct causation nor by event identity, then there exists a common cause variable V, with values $q \in I = q_1, q_2, q_3, \ldots, q_k$ such that $\sum_q p(Vq) = 1$ and $p(A \wedge B|Vq) = p(A|Vq) \cdot p(B|Vq)$, $\forall q$.*

The postulated common cause variable is shown, following Suppes and Zanotti[37], to be a two-valued variable in the case of perfect correlations. This is easily seen by observing that the set I of all values V of the hidden variable can be split into two disjoint subsets I^+ and I^- such that if $q \in I^+$ then $p(A \wedge q) \neq 0$; and if $q \in I^-$ then $p(A \wedge q) = 0$.

This allows the authors to define a two valued common cause variable for a generic perfect correlation $p(A|B) = p(B|A) = 1$.[38] A *common cause event type C* can also be defined in this context as the disjunction of all events for which the common cause variable takes the value $q \in I^+$:

$$C = \vee_{q \in I^+} Vq,$$
$$\neg C = \vee_{q \in I^-} Vq.$$

The common cause so defined is indeed an *individual*-common cause since it screens-off only the corresponding perfect correlation.

[34](Graßhoff, Portman and Wüthrich, 2005, p. 669).

[35]See equation (3.1) in page 33.

[36](Graßhoff, Portman and Wüthrich, 2005, p. 669). Note that the authors' *Principle of the Common Cause* is what I called *Reichenbach's Common Cause Principle*, i.e. both the metaphysical and the methodological parts.

[37]Cf. (Suppes and Zanotti, 1976).

[38]The fact that the common cause variable can be conceived as a two valued variable is not exclusive of the perfect correlations. On the contrary, it seems quite natural to think of a common cause variable as a two valued variable. This is, in fact, Reichenbach's original definition (see Chapter 3, Section 3.1).

5.2.2 Determinism

It is important to note that the hidden variable so defined (under the SEP, PCORR and PCC assumptions) is also 'deterministic' when applied to the perfect correlated outcomes.[39] This is a crucial feature of the argument since it will allow the authors to derive their so-called minimal theories (MTH), as will be shown later on.

The above results are valid for a generic perfect correlation $p(A|B) = p(B|A) = 1$. These results are thus completely general. In the case of the EPRB correlations we can write for a given perfect correlation $p_{ii}(L_i^+|R_i^-) = p_{ii}(R_i^-|L_i^+) = 1$[40],

$$p_{ii}(L_i^+|C_{ii}^{+-}) = p_{ii}(R_i^-|C_{ii}^{+-}) = 1, \tag{5.4}$$
$$p_{ii}(L_i^+|\neg C_{ii}^{+-}) = p_{ii}(R_i^-|\neg C_{ii}^{+-}) = 0. \tag{5.5}$$

Summing up, in a first stage Graßhoff et al. propose an event algebra in which each of the EPRB perfect correlations has an *individual*-common cause. Those *individual*-common causes are deterministic with respect to the experiment outcomes. More particularly, the PCORR and PCC assumptions 'add up' so to speak to yield an event algebra that fulfils the following relations:

$$p_{ii}(L_i^+|R_i^-) = p_{ii}(R_i^-|L_i^+) = 1, \tag{5.6}$$
$$p_{ii}(\neg L_i^+|\neg R_i^-) = p_{ii}(\neg R_i^-|\neg L_i^+) = 0, \tag{5.7}$$
$$p_{ii}(L_i^+|C_{ii}^{+-}) = p_{ii}(R_i^-|C_{ii}^{+-}) = 1, \tag{5.8}$$
$$p_{ii}(L_i^+|\neg C_{ii}^{+-}) = p_{ii}(R_i^-|\neg C_{ii}^{+-}) = 0. \tag{5.9}$$

This is, of course, as long as L_i^+ and R_i^- are distinct events (which is justified by means of the *separability* assumption).

At this point and in order to exclude common cause models that take advantage of the inefficiencies of detectors in real experiments[41], a new assumption is introduced. Namely, it is assumed that every spin measurement has an outcome and that it is either + or −. This is called the *Exactly one of exactly two possible outcomes* assumption (EX):[42]

[39] I would like to stress that it is the perfect correlation assumption that is mainly responsible for the derived 'determinism' of both the outcomes and the corresponding common cause. See Appendix B in Chapter 3 for details.

[40] For consistency with the notation introduced in the previous sections, sub and superscripts refer here to possible directions and possible outcomes respectively ($i = 1, 2, 3$ and $a, b = +, -$). Thus C_{ii}^{ab} stands for the common cause that screens-off a *spin-a* and a *spin-b* in the left and right wings respectively, in a measurement of spin along the same direction i in both wings.

[41] See (Szabó and Fine, 2002) for instance.

[42] (Graßhoff, Portman and Wüthrich, 2005, p. 671).

Exactly one of exactly two possible outcomes (EX):

$$p_{ii}(L_i^+) + p_{ii}(L_i^-) = 1, \quad p_{ii}(L_i^+ \wedge L_i^-) = 0,$$
$$p_{ii}(R_i^+) + p_{ii}(R_i^-) = 1, \quad p_{ii}(R_i^+ \wedge R_i^-) = 0.$$

The EX assumption may not be taken as legitimate by many authors. Fine's *Prism Models*[43], for instance, explicitly repudiate it. In fact, since real EPR-type experiments are performed with real measurement apparatus, detector inefficiencies can never be avoided. In real experiments therefore, since not all fired electrons can be detected by the Stern-Gerlach apparatus, assuming EX seems not to be justified. Thus we may only justify EX, and set these issues aside, in idealised EPR-type experiments (with 100% efficiency). In what follows I shall assume, for the sake of the argument, that this is indeed the case.

By introducing the EX assumption Graßhoff et al. ensure that, while C_{ii}^{+-} is necessary and sufficient for L_i^+ and R_i^-, $\neg C_{ii}^{+-}$ is also necessary and sufficient for the opposite outcomes, i.e. $\neg L_i^+ \equiv L_i^-$ and $\neg R_i^- \equiv R_i^+$.

The conjunction of EX, SEP, PCORR and PCC generates an event algebra which is *deterministic*, not only regarding the outcomes of the experiment, L_i^a and R_i^b, but also the common cause C_{ii}^{ab}. Moreover, the negation of the common cause $\neg C_{ii}^{ab}$ is also deterministic with respect to the opposite outcomes $\neg L_i^a$ and $\neg R_i^b$.[44] In other words, the common cause event C_{ii}^{ab} is both *necessary* and *sufficient* for the events L_i^a and R_i^b which are screened-off by it. But also, $\neg C_{ii}^{ab}$ is *necessary* and *sufficient* for $\neg L_i^a$ and $\neg R_i^b$.

Such an event structure is depicted in Figure 5.1. The diagram represents, in particular, the event structure of any given perfect correlated run of the experiment (given that a parallel setting measurement $L_i \wedge R_i$ has been performed). Apart from the events representing measurement settings L_i and R_i, all other possible events, i.e. spin outcomes and common causes, take place within the intersection $L_i \wedge R_i$. In fact this *measurement* event intersection provides the complete probability space of a given such run.[45] We can

[43]See for instance (Fine, 1982*c*).

[44]Recall that $a, b = +, -$. Hence writing $\neg L_i^{a=+}$ amounts to writing $L_i^{a=-}$. Similarly for the right wing events.

[45]Note that in the diagram of Figure 5.1 I am already assuming that the probability measure p is faithful. This is not so straightforward but it will serve for now. I will discuss on the faithfulness of the probability measure and the correspondence between probabilistic and algebraic relations in a moment (in Section 5.3.1). It is remarkable that if this is the case, i.e. if we grant faithfulness, the above assumptions imply that some of the events in the algebra, such as $L_i^+ \wedge L_i \wedge R_i$ and $R_i^- \wedge L_i \wedge R_i$ are identical. This is specially interesting in the light of the SEP assumption that presupposes those events in particular to be different events (see SEP in page 127).

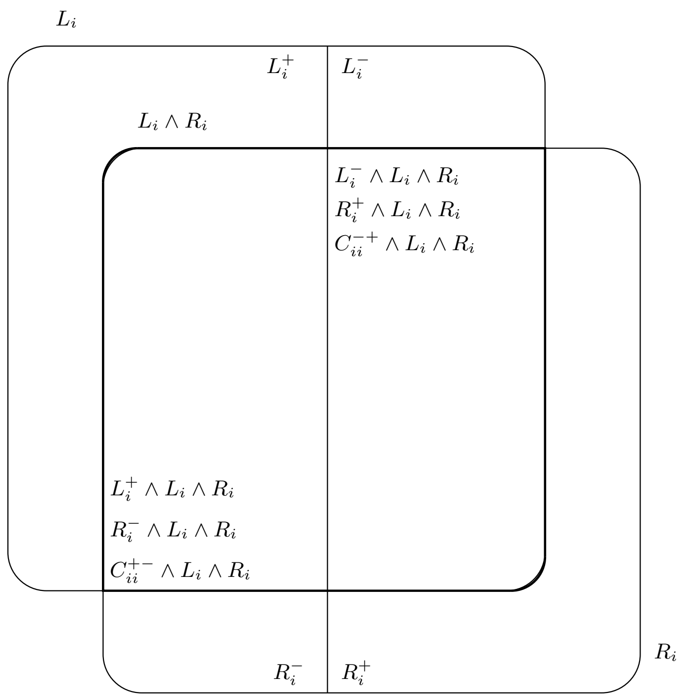


Figure 5.1: The event structure defined by SEP, PCORR, PCC and EX.

write, for each of those perfect correlated events, the following probabilistic relations:

$$p(L_i^+|R_i^- \wedge L_i \wedge R_i) = p(R_i^-|L_i^+ \wedge L_i \wedge R_i) = 1 \tag{5.10}$$
$$p(\neg L_i^+|R_i^- \wedge L_i \wedge R_i) = p(R_i^-|\neg L_i^+ \wedge L_i \wedge R_i) = 0 \tag{5.11}$$
$$p(L_i^+|\neg R_i^- \wedge L_i \wedge R_i) = p(\neg R_i^-|L_i^+ \wedge L_i \wedge R_i) = 0 \tag{5.12}$$
$$p(\neg L_i^+|\neg R_i^- \wedge L_i \wedge R_i) = p(\neg R_i^-|\neg L_i^+ \wedge L_i \wedge R_i) = 1. \tag{5.13}$$

As we can see, not only L_i^+ is perfectly correlated with R_i^- but also $\neg L_i^+$ is with $\neg R_i^-$. Observe as well that, on the other hand, L_i^+ is perfectly anti-correlated with $\neg R_i^-$, as well as $\neg L_i^+$ with R_i^-.

Moreover, the postulated common causes (and their negations) are also perfectly correlated with both the corresponding outcomes (and their opposite outcomes). In particular:

$$p(L_i^+|C_{ii}^{+-} \wedge L_i \wedge R_i) = p(C_{ii}^{+-}|L_i^+ \wedge L_i \wedge R_i) = 1, \tag{5.14}$$
$$p(R_i^-|C_{ii}^{+-} \wedge L_i \wedge R_i) = p(C_{ii}^{+-}|R_i^- \wedge L_i \wedge R_i) = 1, \tag{5.15}$$
$$p(L_i^+|\neg C_{ii}^{+-} \wedge L_i \wedge R_i) = p(\neg C_{ii}^{+-}|L_i^+ \wedge L_i \wedge R_i) = 0, \tag{5.16}$$
$$p(R_i^-|\neg C_{ii}^{+-} \wedge L_i \wedge R_i) = p(\neg C_{ii}^{+-}|R_i^- \wedge L_i \wedge R_i) = 0, \tag{5.17}$$
$$p(\neg L_i^+|\neg C_{ii}^{+-} \wedge L_i \wedge R_i) = p(\neg C_{ii}^{+-}|\neg L_i^+ \wedge L_i \wedge R_i) = 1, \tag{5.18}$$
$$p(\neg R_i^-|\neg C_{ii}^{+-} \wedge L_i \wedge R_i) = p(\neg C_{ii}^{+-}|\neg R_i^- \wedge L_i \wedge R_i) = 1, \tag{5.19}$$
$$p(\neg L_i^+|C_{ii}^{+-} \wedge L_i \wedge R_I) = p(C_{ii}^{+-}|\neg L_i^+ \wedge L_i \wedge R_I) = 0, \tag{5.20}$$
$$p(\neg R_i^-|C_{ii}^{+-} \wedge L_i \wedge R_i) = p(C_{ii}^{+-}|\neg R_i^- \wedge L_i \wedge R_i) = 0. \tag{5.21}$$

Recall once more that here $\neg L_i^+ \equiv L_i^-$, $\neg R_i^- \equiv R_i^+$ and $\neg C_{ii}^{+-} \equiv C_{ii}^{-+}$, which is a particular feature of perfect correlations. This will not be so in the case of non-perfect correlations.

5.2.3 Locality and Minimal Theories

Once the event structure described above is obtained two further conditions aimed at representing *locality* are required by the authors. These two locality assumptions will prove crucial in the argument, since they are the conditions that allow for the derivation of the so-called minimal theories (MTH) —we come to this in a moment—. I will discuss the MTH's in more detail in the next section so it is only needed to present them at this stage.

The first of the locality conditions I am referring to is the so-called LOC2:[46]

[46](Graßhoff, Portman and Wüthrich, 2005, p. 672).

Locality 2 (LOC2): *If $L_i \wedge R_i \wedge X$ is sufficient for L_i^+, then $L_i \wedge X$ alone is sufficient for L_i^+; and similarly for R_j^+, i.e. if $L_j \wedge R_j \wedge Y$ is sufficient for R_j^+, then $R_j \wedge Y$ alone is sufficient for R_j^+.*

If X is identified with C_{ii}^{+-}, i.e. $X \equiv C_{ii}^{+-}$, LOC2 tells us that $L_i \wedge C_{ii}^{+-}$ is sufficient for L_i^+. On the other hand, if $Y \equiv \neg C_{jj}^{+-}$ we have then by LOC2 that $R_j \wedge \neg C_{jj}^{+-}$ is sufficient for R_j^+. It is pointed out that $L_i \wedge C_{ii}^{+-}$ is furthermore *minimally* sufficient for L_i^+ (as well as $R_j \wedge \neg C_{jj}^{+-}$ is *minimally* sufficient for R_j^+). Such *minimality* is readily justified by pointing out that "none of its parts is sufficient on its own". That is, neither L_i nor C_{ii}^{+-} on their own is sufficient for L_i^+ (and neither R_j nor $\neg C_{jj}^{+-}$ on their own is sufficient for R_j^+).

In the next section I will discuss in detail the meaning of LOC2. Let me stress now, however, the following: while the antecedents in LOC2 take advantage of the properties of perfect correlated outcomes —they certainly hold in such cases— the consequents seem to be made extensive to *any* EPR/EPRB correlation, regardless of whether it is perfect or not..

With a similar motivation, another locality condition, LOC3, is introduced in the argument. The aim now is to establish a necessity criterion for L_i^+ (R_j^+). This is in turn achieved by looking for a sufficiency condition for the negation of L_i^+ (R_j^+), i.e. $\neg L_i^+$ ($\neg R_j^+$):[47]

Locality 3 (LOC3): *If $L_i \wedge R_i \wedge X$ is sufficient for $\neg L_i^+$, then $L_i \wedge X$ alone is sufficient for $\neg L_i^+$; and similarly for $\neg R_j^+$, i.e. if $L_i \wedge R_i \wedge Y$ is sufficient for $\neg R_j^+$, then $R_j \wedge Y$ alone is sufficient for $\neg R_j^+$.*

The conclusion that $L_i \wedge C_{ii}^{+-}$ is indeed necessary for L_i^+ (and similarly $R_j \wedge \neg C_{jj}^{+-}$ is necessary for R_j^+) is reached by complementing LOC3 with the assumption that outcomes will only take place if a measurement is performed, or *No outcome without measurement* (NOWM), as the authors call it:[48]

No Outcome Without Measurement (NOWM):

$$p(L_i^+ \wedge \neg L_i) = 0 \quad p(L_i^- \wedge \neg L_i) = 0,$$
$$p(R_j^+ \wedge \neg R_j) = 0 \quad p(R_j^- \wedge \neg R_j) = 0.$$

It is by means of the above assumptions, LOC3 and NOWM, and some simple algebraic manipulation that the necessity conditions obtain. Again the necessity is claimed to be *minimal* in the same sense as pointed out above.

[47](Graßhoff, Portman and Wüthrich, 2005, p. 672).

[48]Cf. (Graßhoff, Portman and Wüthrich, 2005, p. 672).

The sufficiency and necessity conditions just calculated give rise to what Graßhoff et al. call *Minimal Theories* (MTH):

$$L_1^+ \leftrightarrow (L_1 \wedge C_{11}^{+-}), \tag{5.22}$$
$$R_1^+ \leftrightarrow (R_1 \wedge \neg C_{11}^{+-}), \tag{5.23}$$
$$L_2^+ \leftrightarrow (L_2 \wedge C_{22}^{+-}), \tag{5.24}$$
$$R_2^+ \leftrightarrow (R_2 \wedge \neg C_{22}^{+-}), \tag{5.25}$$
$$L_3^+ \leftrightarrow (L_3 \wedge C_{33}^{+-}), \tag{5.26}$$
$$R_3^+ \leftrightarrow (R_3 \wedge \neg C_{33}^{+-}). \tag{5.27}$$

Four of these *Minimal Theories* may be further used to derive probabilistic expressions that define correlations between *mixed perfect correlated* outcomes.[49] For instance, combining equations (5.22) and (5.25), (5.24) and (5.27), and ((5.22) and (5.27) respectively one gets the probabilistic relations:

$$p(L_1^+ \wedge R_2^+) = p(L_1 \wedge C_{11}^{+-} \wedge R_2 \wedge \neg C_{22}^{+-}), \tag{5.28}$$
$$p(L_2^+ \wedge R_3^+) = p(L_2 \wedge C_{22}^{+-} \wedge R_3 \wedge \neg C_{33}^{+-}), \tag{5.29}$$
$$p(L_1^+ \wedge R_3^+) = p(L_1 \wedge C_{11}^{+-} \wedge R_3 \wedge \neg C_{33}^{+-}), \tag{5.30}$$

which, in turn, will allow us to derive the Bell inequality in a more or less straightforward way. The final part of the derivation makes the further assumption that the postulated common causes and their conjunctions are not statistically dependent on the measurement settings.[50] Such an assumption

[49]It is remarkable that only three such *Minimal Theories* are independent, meaning that the remaining expressions can be derived from them. For example, the logical identity (5.24), $L_2^+ \leftrightarrow (L_2 \wedge C_{22}^{+-})$, expresses the same as $R_2^+ \leftrightarrow (R_2 \wedge \neg C_{22}^{+-})$ (equation (5.25)), given that the experiment outcomes are *deterministic* (which follows mainly from the PCORR assumption). In this sense equations (5.24) and (5.25) are not independent. The same applies to equations (5.22)-(5.23), and equations (5.26)-((5.27).

However, as it has been mentioned above, four such relations are needed in order to derive the probabilistic expressions (5.28)-(5.30). This suggests that there is some redundancy in such expressions. This fact has been acknowledged by the authors in personal communication but taken not to be of crucial importance. There is, nonetheless, one respect in which this observation is important. The point is simply that the fact that only three among the six *Minimal Theories* are independent (in the sense above) suggests that each of those three minimal theories (each of the three pairs of equations really, i.e equations (5.22)-(5.23), ((5.24)-(5.25) and (5.26)-(5.27)) correspond to a *separate* perfect correlation event structure. As we will see, this will be crucial when assessing the real significance of the derivation in Section 5.4.

[50]See (Graßhoff, Portman and Wüthrich, 2005, pp. 675-676) for details of the whole derivation.

thus explicitly rules out violations of the *measurement independence* condition, such as those found in Szabó's model (see Section 5.1.2). In fact, the authors' *No-conspiracy* assumption (NO-CONS) is nothing more than a 'generalised' *measurement independence* assumption (see equation (5.2) in page 120). For it requires that both the *individual*-common causes on their own and their combinations be statistically independent from measurement choices.[51] That is:

No-conspiracy (NO-CONS):

$$p(C_{ii}^{+-} \wedge \neg C_{jj}^{+-} | L_i \wedge R_j) = p(C_{ii}^{+-} \wedge \neg C_{jj}^{+-}) \tag{5.31}$$

I already commented that the charge of conspiracy against Szabó's model can be avoided by means of a reinterpretation of the 'no-conspiracy' condition.[52] A similar strategy might be used here. However, since the role of the NO-CONS assumption in Graßhoff et al. is not as interesting as some other aspects, I shall not discuss that option. It will be enough to recall once more our conclusions in Section 5.1.3: One should be very careful when assuming certain statistical independence relations to hold that aim at providing physical independence requirements. For these statistical independence relations may very well have a *purely formal* structural origin, i.e. they may be features of the representation and not of the underlying physics of the system.

5.3 Preliminary Comments on the Graßhoff's et al. Derivation

It will be useful before we proceed to briefly summarise the assumptions made by the authors and the structure of their argument.

In a first step, recall, determinism for perfectly correlated outcomes and their corresponding common causes was derived from *Separability* (SEP), *Perfect Correlation* (PCORR), *Locality* (LOC1), and the *Principle of the Common Cause* (PCC). This determinism is extended to the negations of outcomes and common causes by assuming *Exactly one of exactly two possible outcomes* (EX). In a second step, Graßhoff et al. impose further locality conditions, LOC2 and LOC3 so as to derive their *Minimal Theories* (MTH), as a partial result of the derivation. Finally, *No-conspiracy* (NO-CONS) is

[51]This is also pointed out in (Hofer-Szabó, 2007*a*).

[52]See Section 5.1.

assumed so that the *Minimal Theories* (MTH) may be combined in such a way that a Bell-type inequality is obtained.[53]

Several of the above assumptions, however, have been contested. I already commented (briefly) on some of the controversies introduced by conditions such as SEP, LOC1 or EX.[54] Of course, failure of either SEP, LOC1 or EX would make common cause explanations entirely compatible with the EPR framework.[55]. But I shall not consider these issues further, since my aim in this chapter is to assess whether the argument by Graßhoff, Portman and Wüthrich itself is valid, independently of the truth of the conclusion as for the existence of an *individual*-common cause model for EPR. In other words, I want to show that, even if all the assumptions in the Graßhoff et al. argument were true, their conclusions about the non-existence of *individual*-common cause models of EPR are strictly speaking not correct.

I shall start by discussing some qualitative and formal features of the derivation. Some preliminary comments are in order, regarding the logical structure and the core strategy of the paper.[56] These comments, although important, cannot be said to invalidate the derivation. In the following section, however, I will turn to the real significance of the derivation (by means of an analysis of the applicability of its central locality assumptions).

5.3.1 On Probabilistic and Logical Equivalence

It seems remarkable that the initial —and to my mind crucial— assumption of the argument, *Perfect Correlation* (PCORR), is given as a probabilistic condition (see PCORR on page 127). In other words PCORR is an assumption regarding the statistical dependencies between the events, not the events themselves. At some point, however, the argument qualitatively jumps down to the algebra level (i.e. to the level of events) where the MTH result is obtained (see page 134), as well as the other main assumptions, in particular LOC2 and LOC3. And later on the argument jumps once more from the

[53]See the previous section and (Graßhoff, Portman and Wüthrich, 2005) for details of the several assumptions and the argument itself.

[54]See footnotes 32 and 33 in page 127, as well as Section 5.2.2, in page 130.

[55]For instance, if the EX does not hold, *individual*-common cause models may be provided that partly rely on detector inefficiencies. This is the case, I pointed out, of Fine's *Prism Models* (Fine, 1982*c*). Other possibilities include direct (superluminal) causal models if LOC1 did not hold. (Cartwright and Suárez, 2000) provide an example of such models.

[56]Some of the ideas in this and the following sections are not just my own but the result of a collaboration with Miklós Rédei and Gábor Hofer-Szabó. In particular, Sections 5.3.1 and 5.3.2 reflect also Rédei's main worries regarding the formal structure of some of the claims by Graßhoff et al.

level of events back to the probabilistic level, in which the Bell inequalities are finally derived.

It is not my intention to judge here whether the above is an adequate strategy or not, i.e. whether probabilistic expressions shall be uniquely identified with logical expressions involving the events the former probabilities refer to, and conversely. In case this strategy needs to be pursued, the identification of probabilistic and logical expressions needs to be done with extreme care, in order to guarantee that both descriptions actually account for the same event structure. Special attention should be paid in this regard to objects that cannot be dismissed at an algebraic level but may be irrelevant at the level of probabilities, such as zero probability measure events. For, such events legitimately belong in the algebra but they could be overlooked in a probabilistic description of the event structure.[57]

A clear example of such an identification can be found in the Graßhoff et al. argument, when *sufficiency* and *necessity* relations are established. For instance, the (physical) significance of both the locality conditions LOC2 and LOC3 (see page 133) relies on the precise meaning of "event A being sufficient for event B". Therefore, if we expect to extract some physical (locality) moral from LOC2 and LOC3 we need to understand exactly what it is meant by "event A being sufficient for event B". We need, in other words, to be given a precise unambiguous definition for this.

This is not what Graßhoff et al. do. On the contrary, the paper seems to handle two different notions of sufficiency. One of them is laid out in a probability language and the other in terms of logical relations. We find for instance the following reference to sufficiency relations:[58]

> Together with the previous equation this implies that Vq is sufficient for B for all $q \in I^+$:
>
> $$p(B|Vq) = 1, \quad \forall q \in I^+.$$

And right after we can read

> [...] in turn implies that
>
> $$p(B|Vq) = 0, \quad \forall q \in I^+,$$
>
> which means that Vq with $q \in I^+$ is also necessary for B.

[57] This point was specially stressed by Miklós Rédei.

[58] (Graßhoff, Portman and Wüthrich, 2005, p. 670).

In the above two quotations, Graßhoff et al. clearly take sufficiency and necessity to be defined in terms of probabilities, perhaps along the following lines:[59]

Definition 8 *Given a probability space* $(\mathcal{S}, p)$ *event* $A \in \mathcal{S}$ *is* Sufficient^p *for event* $B \in \mathcal{S}$ *if and only if* $p(B|A) = 1$. *Event* A *is* Necessary^p *for event* B *if and only if* $p(B|\neg A) = 0$.

The notation —with super-script p— used above makes explicit the entirely probabilistic character of the notions of sufficiency and necessity so defined, i.e. *Sufficient*^{p} and *Necessary*^{p}.

However, we can also find at some point later in the paper a completely different reference to sufficiency:[60]

> [...] $\neg C_{ii}^{+-} \wedge L_i \wedge R_i$ implies $\neg L_i^+$:
>
> $$\neg C_{ii}^{+-} \wedge L_i \wedge R_i \rightarrow \neg L_i^+.$$
>
> Again we propose a locality condition based on the idea that the measurement choice in one wing should be causally irrelevant for the outcomes (and the choices) in the other wing: if $\neg C_{ii}^{+-} \wedge L_i \wedge R_i$ is sufficient for $\neg L_i^+$, then $\neg C_{ii}^{+-} \wedge L_i$ *alone* should be sufficient for $\neg L_i^+$.

And further down we read:[61]

> [...] L_i is necessary for L_i^+. That means $L_i^+ \rightarrow L_i$ [...].

The above seem to fit into a standard definition of sufficiency (necessity) in terms of the algebra events and their logical relations (that can be written as set operations):[62]

Definition 9 *Event* A *is* Sufficient^l *for event* B *if and only if* $A \subseteq B$. *Event* A *is* Necessary^l *for event* B *if and only if* $B \subseteq A$.

---

[59] The following definitions of sufficiency (necessity) have been suggested by Miklós Rédei (personal communication).

[60] (Graßhoff, Portman and Wüthrich, 2005, p. 672).

[61] (Graßhoff, Portman and Wüthrich, 2005, p. 673).

[62] Note that set algebraic relations are taken here to be equivalent to logical relations. In particular $A \rightarrow B$ is understood as equivalent to $A \subseteq B$, and $A \leftarrow B$ to $A \supseteq B$. This is not strictly correct since the equivalence is between set relations and logical relations between set elements, i.e. $A \subseteq B$ if and only if $\forall a(a \in A \rightarrow b \in B)$. However, we do not need make that explicit for our own purposes here. Thus, in what follows I will use set algebraic relations and logical relations interchangeably.

Once more, the super-script l in the definition above expresses the exclusively logical character of the notions of sufficiency and necessity so defined.

Of course the two different definitions of sufficiency and necessity are related to each other. Their relation is not, so to speak symmetrical. In particular, if A is *Sufficient*^{l} for B then A is *Sufficient*^{p} for B for any p. However, the converse does not hold in general.

More specifically, we can write, in general:

$$A \subseteq B \text{ entails } p(B|A) = \frac{p(B \wedge A)}{p(A)} = \frac{p(A)}{p(A)} = 1.$$

However,

$$p(B|A) = 1 \text{ entails } A \subseteq B$$

only if

$$p\,[A \setminus (A \wedge B)] = 0 \text{ entails } [A \setminus (A \wedge B)] = \emptyset.$$ [63]

This will happen for instance in case p is faithful, but in general having $p\,[A \setminus (A \wedge B)] = 0$ will *not* necessarily entail $[A \setminus (A \wedge B)] = \emptyset$.

Similarly, if A is *Necessary*^{l} for B then A is *Necessary*^{p} for B for any p, but not conversely.

Since sufficiency and necessity in the sense of Definitions 8 and 9 are not equivalent notions, one must be very careful when using them indistinguishably. In particular, as I pointed out earlier in this section special attention must be paid to structures that may contain empty sets with associated non-zero probability.

5.3.2 A Problematic Inference

One cannot find in Graßhoff et al. any sort of justification for the fact that Definitions 8 and 9 of sufficiency and necessity are used indistinguishably. This is of special significance since some steps in the argument rely precisely on such identification. The inferences involved in such steps can be thought for this reason to be highly questionable.

An example of one such inference can be found when we read at some stage:[64]

[63] Recall that, if $p(A) \neq 0$, the conditional probability $p(B|A) = 1$ is equivalent to $p(B \wedge A) = p(A)$.

[64] (Graßhoff, Portman and Wüthrich, 2005, p. 672).

> [...] we saw in section 4.2 that if parallel settings are chosen and $\neg C_{ii}^{+-}$ is instantiated an event of type L_i^+ never occurs. In other words, $\neg C_{ii}^{+-} \wedge L_i \wedge R_i$ implies $\neg L_i^+$:
>
> $$\neg C_{ii}^{+-} \wedge L_i \wedge R_i \rightarrow \neg L_i^+.$$

In this passage the authors start from a probabilistic notion of sufficiency to arrive at an implication, which we just saw is an expression of sufficiency in terms of algebraic elements. Let us see in more detail how this is done.

First, the authors point out that in their section 4.2 (pp. 669-70 of the paper) it was found that "if parallel settings are chosen and $\neg C_{ii}^{+-}$ is instantiated an event of type L_i^+ never occurs". This is expressed by means of equation (25), also of the paper,[65] which reads:

$$p_{ii}(L_i^+|\neg C_{ii}^{+-}) = p_{ii}(R_i^-|\neg C_{ii}^{+-}) = 0.$$

Using the previous section definitions, the above equation says that

$$\neg C_{ii}^{+-} \wedge L_i \wedge R_i \;\text{ is }\; \textit{Necessary}^p \;\text{ for }\; L_i^+$$

(and for R_i^- as well). This is the same as saying that

$$\neg C_{ii}^{+-} \wedge L_i \wedge R_i \;\text{ is }\; \textit{Sufficient}^p \;\text{ for }\; \neg L_i^+.$$[66]

The authors, however, take this to mean that

$$\neg C_{ii}^{+-} \wedge L_i \wedge R_i \rightarrow \neg L_i^+.$$

But, this means that $\neg C_{ii}^{+-} \wedge L_i \wedge R_i \subseteq \neg L_i^+$ or, in other words, that

$$\neg C_{ii}^{+-} \wedge L_i \wedge R_i \;\text{ is }\; \textit{Sufficient}^l \;\text{ for }\; \neg L_i^+.$$

We see therefore that in the above quoted passage "A is *Sufficient*^{l} for B" is inferred from "A is *Sufficient*^{p} for B". The paper contains at least two more steps in which inferences of this sort are made.

I pointed out that whether such inferences are justified depends sensitively on the properties of the probability measure p. But since the probability measure p is not completely known —we only know some of the conditional probabilities it determines but ignore for instance how it behaves on events

[65](Graßhoff, Portman and Wüthrich, 2005, p.671).

[66]This is true, of course, under the assumptions made in the argument. Specially, under the PCORR, PCC and EX assumptions (see Section 5.2.1 for details).

that are related to the hypothetical common cause events C_{ii}^{+-}— this inference seems not to be warranted in any way. The whole argument is therefore compromised by such inferences.

I would like to stress as well that switching between the two levels (algebraic and probabilistic) of description is confusing, specially when it comes to accounting for some particular crucial assumptions in the argument. An example of such an ambiguity is found in the case of the notion of *perfect correlation* (PCORR), which, I have already pointed out, is introduced as a probabilistic condition. Its implications are nevertheless presented in the argument at the level of events (by logical relations) in virtue of its relation with an implicit extra condition, i.e. the *parallel settings* condition, an algebraic condition that just refers to events and not to probabilities at all.[67] The question we may ask would then be: to what extent the logical implications of a probabilistic assumption such as PCORR reflect the intuitions that the original assumption contains? A adequate answer to this question would be desirable in the argument for clarity, and in order to legitimate further claims based on our assumptions.

Let us nonetheless assume for the sake of the argument that the identifications between logical and probabilistic notions of sufficiency and necessity are correct. Even so, we shall see that the argument by Graßhoff et al. may be questioned.

5.4 Perfect Correlations and the Bell Inequalities

As I pointed out, even if the problems discussed in the previous section are set to one side, there are still, in my opinion, some remarkable features of the derivation by Graßhoff et al. that might be challenged, casting serious doubts on its applicability and significance. There are in particular two issues

[67] Although *parallel settings* is not explicitly required anywhere in the argument, Graßhoff et al. introduce it as the 'measurement scenario' for PCORR. In this sense, *parallel settings* seems to play the role of a pre-condition for PCORR. The real underlying assumption here is that a certain (minimal perhaps) event structure, such as that provided by measurement settings, needs to be characterised even before physical intuitions can be incorporated in the form of probabilistic expressions. This would be in fact another reading of the distinction between *purely formal* and *genuinely physical* correlations introduced in Section 5.1.3. And if this is so, the resulting event structure will clearly feature both physical intuitions and formal structural constrains. However, Graßhoff et al. seem not to acknowledge this and take PCORR later in the argument as some sort of stand-alone condition that can be assumed to hold regardless of its 'measurement scenario'. I will discuss this in more detail in the next section.

I want to address. On the one hand, I shall investigate in what circumstances the minimal theories (MTH), i.e. relations (5.22)-(5.27), do really hold. In case they hold, I shall ask furthermore, whether they can be *mixed* to yield relations (5.28)-(5.30), which make it possible to derive the Bell inequalities.

On the other hand, I shall investigate whether the argument takes formal aspects of the EPR-Bohm experimental description as physical. In particular, I shall point out that Graßhoff et al. rely, at least partially, on *purely formal* correlations in order to obtain the *Minimal Theories*.[68]

5.4.1 Locality and the Minimal Theories

The key for the derivation of the *Minimal Theories* (MTH) is in the notions of locality that conditions LOC2 and LOC3 express. Recall moreover that LOC2 and LOC3 are in turn proposed taking as a reference the results of the *perfect correlation* assumption (PCORR).

Let us set aside faithfulness issues such as those discussed in the previous section.[69] We need, in order to assess the significance of the *Minimal Theories* (MTH), to carefully analyse whether the two referred locality assumptions are reasonable. My main claim will be that LOC2 and LOC3 are reasonable constraints on *individual*-common causes only for the case of perfect correlations. Alternatively, if one insists in that LOC2 and LOC3 hold in general —that is, for both perfect and non-perfect correlations—, a *common*-common cause structure must be implicitly assumed.

As we will see, in the first case both LOC2 and LOC3 are found to be vacuous —since they can be inferred from the set structure forced by the assumption PCORR and the existence of a common cause (PCC)— and the Bell inequalities are trivially satisfied. In the second case, if a *common*-common cause structure is assumed —even if implicitly—, the original claim

[68]I pointed out in Section 5.1.3 that the logical structure used in the description of EPR phenomena forces us to admit some kind of (formal) correlation between the different (arbitrarily chosen) measurement settings of two runs of the experiment in one of the wings. That is, correlations such as

$$Corr(L_i, L_j), \quad i \neq j \quad \text{and} \quad i, j = 1, 2, 3.$$

I want to recall once more that such correlations are merely *purely formal*. That is, the correlations are an artifact of the formal representation.

In view of such a feature, which is clearly and explicitly pointed out by Szabó, I have asked whether we might expect some apparent 'conspiratorial' behaviour in our model.

[69]In fact, in what follows I will take, in virtue of p being *faithful*, Definitions 8 and 9 of sufficiency and necessity introduced in Section 5.3.1 to be equivalent. That is, A is *sufficient* for B, $p(B|A) = 1$, $A \subseteq B$ and $A \rightarrow B$ (and correspondingly A is *necessary* for B, $p(B|\neg A) = 0$, $A \supseteq B$ and $A \leftarrow B$) will be used indistinguishably.

by Graßhoff et al., of course, does not follow.

I will start by noting that both LOC2 and LOC3 (see page 133) contain two separate claims each.[70] In particular, LOC2 consists of

LOC2 (i): *If $L_i \wedge R_i \wedge X$ is sufficient for L_i^+, then $L_i \wedge X$ alone is sufficient for L_i^+.*

LOC2 (ii): *If $L_j \wedge R_j \wedge Y$ is sufficient for R_j^+, then $R_j \wedge Y$ alone is sufficient for R_j^+.*

Similarly, LOC3 may be split into

LOC3 (i): *If $L_i \wedge R_i \wedge X'$ is sufficient for $\neg L_i^+$, then $L_i \wedge X'$ alone is sufficient for $\neg L_i^+$.*

LOC3 (ii): *If $L_j \wedge R_j \wedge Y'$ is sufficient for $\neg R_j^+$, then $R_j \wedge Y'$ alone is sufficient for $\neg R_j^+$.*

First of all, regarding the logical structure of the above, it is remarkable that LOC2(i), LOC2(ii), LOC3(i) and LOC3(ii) are conditional statements with a sufficiency premise in the antecedent. For instance, the antecedent of LOC2(i) depends on the fact that $L_i \wedge R_i \wedge X$ is sufficient for L_i^+. Therefore, LOC2(i) is vacuously true if $L_i \wedge R_i \wedge X$ is *not* sufficient for L_i^+. We will see in a moment under what circumstances such premises hold, i.e. under what circumstances it is meaningful to appeal to the above locality conditions.

We will now see with the help of some set diagrams that LOC2 and LOC3 are not in general reasonable locality assumptions. Take LOC2(i), for instance, which lies behind the (minimal) sufficiency of $L_i \wedge C_{ii}^{+-}$ for L_i^+.[71] That is, if LOC2(i) holds we can already write, for $X \equiv C_{11}^{+-}$ and $i = 1$,

$$L_1 \wedge C_{11}^{+-} \rightarrow L_1^+.$$

In the Graßhoff et al. derivation, the above expression is completed with a necessity condition (of $L_i \wedge C_{ii}^{+-}$ for L_i^+) to yield a *Minimal Theory* (equation (5.22) in particular). Such necessity condition is derived however from another sufficiency condition, implied by LOC3(i). In particular, from LOC3(i) the following expression can be derived ($X' \equiv \neg C_{11}^{+-}$ and $i = 1$):

$$L_1 \wedge \neg C_{11}^{+-} \rightarrow \neg L_1^+,$$

[70] The LOC2 and LOC3 statements are slightly different to those in (Graßhoff, Portman and Wüthrich, 2005). For the purpose of the argument here I want to distinguish the event X from X'; and Y from Y' (in LOC2 and LOC3 respectively). This distinction does not change the content of the original claims, however. See also LOC2 and LOC3 in page 133 for the exact formulation of the locality conditions.

[71] This is even already pointed out by the authors. See (Graßhoff, Portman and Wüthrich, 2005, p.672) for details.

from which it follows that

$$L_1 \wedge C_{11}^{+-} \leftarrow L_1^+.$$

The final result, as implied from LOC3(i), is thus a necessary condition: $L_i \wedge C_{ii}^{+-}$ is necessary for L_i^+.[72]

Altogether one obtains in this way, by combining the implications of LOC2(i) and LOC3(i) (for measurement setting $i = 1$), the *Minimal Theory* (5.22) in page 134:

$$L_1 \wedge C_{11}^{+-} \leftrightarrow L_1^+.$$

In a completely equivalent way, equation (5.23), as it stands, is derived by successively applying LOC2(ii) and LOC3(ii) for measurement setting $i = 1$, and identifying $Y \equiv \neg C_{11}^{+-}$ and $Y' \equiv C_{11}^{+-}$. In this case, however, LOC2(ii) is what backs up the necessity of $R_1 \wedge \neg C_{11}^{+-}$ for R_1^+. And LOC3(ii) behind the corresponding sufficiency relation. That is, in virtue of LOC2(ii) we may write

$$R_1 \wedge \neg C_{11}^{+-} \leftarrow R_1^+,$$

while, applying LOC3(ii) gives us

$$R_1 \wedge \neg C_{11}^{+-} \rightarrow R_1^+.$$

Both expressions can be combined to yield the corresponding *Minimal Theory*, i.e. equation (5.23):

$$R_1 \wedge \neg C_{11}^{+-} \leftrightarrow R_1^+.$$

It also becomes clear that equations (5.24)-(5.25) and (5.26)-(5.27) are exactly the same expressions for $i = 2$ and 3 respectively.

5.4.2 Are LOC2 and LOC3 Reasonable Locality Conditions?

Now, with the aid of Figure 5.2 we can see that the premise in LOC2(i) holds trivially, since the triple intersection $L_i \wedge R_i \wedge C_{ii}^{+-}$ is always contained in L_i^+.[73] However, its implication ($L_i \wedge C_{ii}^{+-} \rightarrow L_i^+$) means that the shadowed region in the left hand side of the diagram must be the empty set. This seems intuitively acceptable. In particular, it seems natural to think that

[72]See (Graßhoff, Portman and Wüthrich, 2005, pp. 672-3) for the details of how this expression is obtained.

[73]It can be shown with a few calculations that such a set arrangement is the result of the assumptions in (Graßhoff, Portman and Wüthrich, 2005).

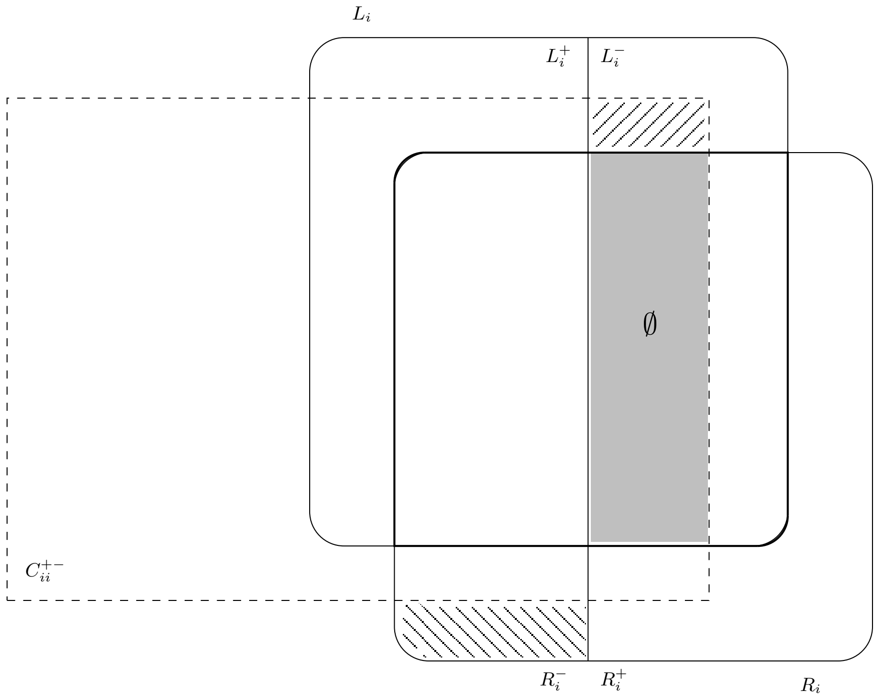


Figure 5.2: Set diagram for LOC2 (see page 133). If $L_i \wedge C_{ii}^{+-} \rightarrow L_i^+$ and $R_i \wedge \neg C_{ii}^{+-} \rightarrow R_i^+$ the shadowed regions must have probability zero, i.e. must be the zero set.

if it is the case that a measurement is performed, say in *direction 1* on the left wing, given the common cause C_{11}^{+-} is present, the outcome will be L_1^+, *regardless* of what happens on the other wing of the experiment. Thus, I conclude that LOC2(i) is a reasonable general claim, both for perfect and non-perfect correlations. In general thus, we may write

$$L_i \wedge C_{ii}^{ab} \rightarrow L_i^a,$$

where $i = 1, 2, 3$ and $a, b = +, -$.

Turning now to LOC2(ii) (which grounds the necessity of $L_i \wedge C_{ii}^{ab}$ for L_i^a), although it is apparently similar to LOC2(i), it turns out to have quite different implications. As I pointed out before, the event Y in LOC2(ii) seems to be identified with $\neg C_{ii}^{+-}$.[74]

Again, with the aid of Figure 5.2 we see that the premise holds trivially. As for the implication $R_i \wedge \neg C_{ii}^{+-} \rightarrow R_i^+$, it means that the shadowed region in the right hand side of the picture must be the empty set. It is not clear,

[74]See also (Graßhoff, Portman and Wüthrich, 2005, p. 672).

however, why this should be the case in general. In particular, for non-perfect correlations, R_i^- may be instantiated by the presence of C_{ji}^{a-} ($j \neq i$ and $a = +,-$)[75], which are contained in $\neg C_{ii}^{+-}$. And it is, in fact, those events that the shadowed region under consideration represents. Hence $R_i \wedge \neg C_{ii}^{+-} \rightarrow R_i^+$ seems to hold only in the $L_i \wedge R_i$ overlapping region of the diagram, i.e. for the perfect correlation case only!

We may still want to require that LOC2(ii) be valid in general, i.e. valid for correlations other than perfect as well. But if we do, then we will have *ipso facto* assumed a *common*-common cause structure. This can be see as follows:

Consider first a non-perfect correlation of the type $Corr(L_j^a, R_i^-)$, with $i \neq j$ and $a = +,-$. For such a correlation a (Reichenbachian) *individual*-common cause C_{ji}^{a-} can then be postulated. Now insisting that $R_i \wedge \neg C_{ii}^{+-} \rightarrow R_i^+$ holds in general, i.e. also for non-perfect correlations, means that if measurement R_i is performed and C_{ii}^{+-} is not present then the outcome R_i^+ is always instantiated. In particular, if an event such as C_{ji}^{a-} is present and measurement R_i is preformed we must expect R_i^+ as the outcome.[76] But this is very counter-intuitive unless, of course, we assume C_{ji}^{a-} to be contained in C_{ii}^{-+}, which in virtue of PCORR guarantees that R_i^+ is obtained deterministically, i.e. in all cases.

If this is so we then have that C_{ji}^{a-} is not only a common cause of $Corr(L_j^a, R_i^-)$ but also of $Corr(L_i^+, R_i^-)$. That is C_{ji}^{a-} is a *common*-common cause, as it stands.

Somehow symmetrical results are obtained when analysing LOC3. In this case, while LOC3(ii) seems sensible, whether perfect or non-perfect correlations are considered, LOC3(i) turns out to be non reasonable as a general locality claim. (The corresponding diagram can be seen in Figure 5.3.) In particular, I do not think that the following claim is justified in general:[77]

[75]Specifically, the probability of finding say R_1^- as an outcome will be given by

$$p(R_1^-) = p(L_1 \wedge R_1 \wedge C_{11}^{+-}) + p(L_1 \wedge R_1 \wedge C_{11}^{--}) + p(L_2 \wedge R_1 \wedge C_{21}^{+-}) + \\ p(L_2 \wedge R_1 \wedge C_{21}^{--}) + p(L_3 \wedge R_1 \wedge C_{31}^{+-}) + p(L_3 \wedge R_1 \wedge C_{31}^{--}).$$

Note that $p(L_1 \wedge R_1 \wedge C_{11}^{--})$ must be zero since outcomes L_1^- and R_1^- cannot be simultaneously instantiated in the same run.

[76]This is so because C_{ji}^{a-} is part of $\neg C_{ii}^{+-}$, i.e. $C_{ji}^{a-} \subseteq \neg C_{ii}^{+-}$.

[77]I would like to point out that the claim in the following quote (Graßhoff, Portman and Wüthrich, 2005, p. 672) is a new version of a former claim that appeared in an earlier version of the Graßhoff et al. paper. One could see from that former claim what is the intended underlying intuition of LOC3(i). The claim I am referring to went along the following lines: "[...] by a locality argument [LOC3(i)] C_{ii}^{+-} should also be necessary if in the remote wing a setting other than parallel is chosen ($R_j,\ j \neq i$). That is $L_i^+ \wedge L_i$

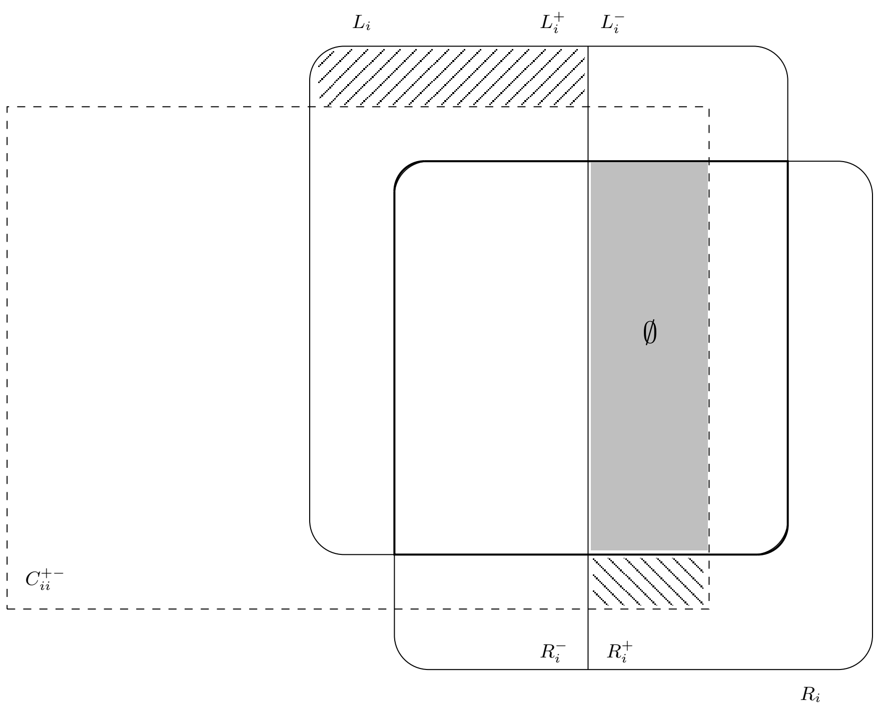


Figure 5.3: Set diagram for LOC3 (see page 133). If $L_i \wedge \neg C_{ii}^{+-} \to \neg L_i^+$ and $R_i \wedge C_{ii}^{+-} \to \neg R_i^+$ the shadowed regions must have probability zero, i.e. must be the zero set.

> [...] if $\neg C_{ii}^{+-} \wedge L_i \wedge R_i$ is sufficient for $\neg L_i^+$, then $\neg C_{ii}^{+-} \wedge L_i$ *alone* should be sufficient for $\neg L_i^+$.

Once more, if we insist in taking claims as the above as valid general claims we have to acknowledge that some of our postulated common causes are not *individual*-common causes any more, but *common*-common causes instead.

The argument I just provided takes it that the C_{ji}^{a-} ($j \neq i$ and $a = +,-$) events are contained in $\neg C_{ii}^{+-}$ (recall, in particular Footnote 76 above). This may be further challenged on the grounds that there is nothing in the Graßhoff et al. argument that blocks the possibility that the above postulated C_{ji}^{a-} ($j \neq i$ and $a = +,-$) *be* in fact in the $R_i^- \wedge C_{ii}^{+-}$ intersection (see Figure 5.2).[78] I shall show, however, that this observation does not help in avoiding *common*-common causes.

The fact that C_{ji}^{a-} ($j \neq i$ and $a = +,-$) are not contained in $\neg C_{ii}^{+-}$ would mean that LOC2(ii) holds, i.e. $R_i \wedge \neg C_{ii}^{+-} \subseteq R_i^-$, but also that (for a given fixed i)

$$C_{ji}^{a-} \subseteq C_{ii}^{+-} \qquad (j \neq i \text{ and } a = +,-). \tag{5.32}$$

In particular for $i = 1$

$$\{C_{21}^{+-}, C_{21}^{--}, C_{31}^{+-}, C_{31}^{--}\} \subseteq C_{11}^{+-}. \tag{5.33}$$

Similarly, for $i = 2$

$$\{C_{12}^{+-}, C_{12}^{--}, C_{32}^{+-}, C_{32}^{--}\} \subseteq C_{22}^{+-}, \tag{5.34}$$

and for $i = 3$

$$\{C_{13}^{+-}, C_{13}^{--}, C_{23}^{+-}, C_{23}^{--}\} \subseteq C_{33}^{+-}. \tag{5.35}$$

This is also what it means, by the way, that the shadowed region in the right hand side of Figure 5.2 is the empty set, i.e. $\neg C_{ii}^{+-} \wedge R_i^- \equiv \emptyset$.

In a similar way, if LOC3(i) holds, then

$$\{C_{12}^{++}, C_{12}^{+-}, C_{13}^{++}, C_{13}^{+-}\} \subseteq C_{11}^{+-}; \tag{5.36}$$
$$\{C_{21}^{++}, C_{21}^{+-}, C_{23}^{++}, C_{23}^{+-}\} \subseteq C_{22}^{+-}; \tag{5.37}$$
$$\{C_{31}^{++}, C_{31}^{+-}, C_{32}^{++}, C_{32}^{+-}\} \subseteq C_{33}^{+-}. \tag{5.38}$$

Expressions (5.33)-(5.35) and (5.36)-(5.38) entail then that

$$\{C_{21}^{+-}, C_{21}^{--}, C_{31}^{+-}, C_{31}^{--}, C_{12}^{++}, C_{12}^{+-}, C_{13}^{++}, C_{13}^{+-}\} \subseteq C_{11}^{+-}; \tag{5.39}$$
$$\{C_{12}^{+-}, C_{12}^{--}, C_{32}^{+-}, C_{32}^{--}, C_{21}^{++}, C_{21}^{+-}, C_{23}^{++}, C_{23}^{+-}\} \subseteq C_{22}^{+-}; \tag{5.40}$$
$$\{C_{13}^{+-}, C_{13}^{--}, C_{23}^{+-}, C_{23}^{--}, C_{31}^{++}, C_{31}^{+-}, C_{32}^{++}, C_{32}^{+-}\} \subseteq C_{33}^{+-}. \tag{5.41}$$

---

alone imply C_{ii}^{+-}. " (The square brackets are mine).

[78] This challenge was suggested by Hofer-Szabó and Rédei (personal communication).

It is now straightforward to see that C_{11}^{+-}, C_{22}^{+-} and C_{33}^{+-} have non-empty intersections. In particular,

$$\{C_{21}^{+-}, C_{12}^{+-}\} \subseteq C_{11}^{+-} \wedge C_{22}^{+-}; \tag{5.42}$$
$$\{C_{31}^{+-}, C_{13}^{+-}\} \subseteq C_{11}^{+-} \wedge C_{33}^{+-}; \tag{5.43}$$
$$\{C_{32}^{+-}, C_{23}^{+-}\} \subseteq C_{22}^{+-} \wedge C_{33}^{+-}; \tag{5.44}$$

which means that $C_{21}^{+-}, C_{12}^{+-}, C_{31}^{+-}, C_{13}^{+-}, C_{32}^{+-}$ and C_{23}^{+-} are nothing more than *common*-common causes!

In particular, C_{21}^{+-} screens-off $Corr(L_2^+, R_1^-)$, $Corr(L_1^+, R_1^-)$, but it also screens-off $Corr(L_2^+, R_2^-)$. The same holds for the remaining five common causes as well.

Therefore, as I said, the fact that C_{ji}^{a-} ($j \neq i$ and $a = +, -$) are not contained in $\neg C_{ii}^{+-}$ does not help in avoiding my *common*-common cause critique. If anything, it helps in making my claims even more general. For in the light of these observations the statement above (page 146) that "[...] for non-perfect correlations, R_i^- may be instantiated by the presence of C_{ji}^{a-} ($j \neq i$ and $a = +, -$), which *are contained* in $\neg C_{ii}^{+-}$." is not even strictly speaking correct.

In the *most general* case, it may be reasonable to think that the C_{ji}^{a-} ($j \neq i$ and $a = +, -$) have a non-zero intersection with C_{ii}^{+-}. But it does not follow that the C_{ji}^{a-} should be, strictly, subsets of C_{ii}^{+-} either. In other words, and referring once more to the set diagram (Figure 5.2), I don't see why the shadowed region in the right hand side of the picture should be the empty set.[79] In my view, the C_{ji}^{a-} ($j \neq i$ and $a = +, -$) should be assumed to have *at most* a non-zero intersection with $\neg C_{ii}^{+-}$.

Therefore, we may conclude that:

> In particular, for non-perfect correlations, R_i^- may be instantiated by the presence of C_{ji}^{a-} ($j \neq i$ and $a = +, -$), which are ***in the most general case at least partly contained*** in $\neg C_{ii}^{+-}$.

As I said, taking the above into account however does not affect my main claims regarding LOC2 and LOC3. Summing up: on the one hand we found that LOC2 and LOC3 can be claimed to hold in general only if *common*-common causes are —either explicitly or implicitly— assumed. On the other hand, if we stick to *individual*-common causes, LOC2 and LOC3 only hold for the particular case of perfect correlations.

We will see in the next section that the Bell inequalities can be derived in both cases by 'mixing' the resulting *Minimal Theories* (MTH). However,

[79] Even though this would not constitute a serious drawback to *common*-common cause critique.

the analysis above on the status of LOC2 and LOC3 severely restricts their significance and applicability.

5.4.3 Minimal Theories and the Bell Inequalities

In Section 5.4.1 I explained how the *Minimal Theories* were to be derived from assumptions LOC2 and LOC3. It is important to recall that both of the locality conditions are necessary for the *Minimal Theories* to obtain. In particular, recall that each of the expressions of the *Minimal Theories*, i.e. equations (5.22)-(5.27), was derived by combining either LOC2(i) and LOC3(i) or LOC2(ii) and LOC3(ii). Recall finally that from the discussion in the previous section we concluded that LOC2(ii) and LOC3(i) could only be taken as valid locality claims in general if *common*-common causes were tacitly assumed. Therefore, since all the *Minimal Theories* need either LOC2(ii) or LOC3(i) to be derived all of them inherit in one way or another the limitations we found for LOC2 and LOC3.

We then seem to face a dichotomy: on the one hand the *Minimal Theories* (MTH) may be taken to be valid in general (for all possible events in an EPR experiment), in which case the common causes that they contain are really *common*-common causes. On the other hand, if we restrict the *Minimal Theories* to exclusively *individual*-common causes, then expressions (5.22)-(5.27) are limited to the corresponding perfectly correlated runs. Each run corresponding to one of the three possible measurement directions of the EPR experiment.

The first horn of the dichotomy leads to the usual well known arguments for the derivation of the Bell inequalities, in which *common*-common causes are assumed. The second horn of the dichotomy is however much more interesting since it has implications regarding the information about the existence of *individual*-common causes. Let me explain better.

In Section 5.2.2 we saw that the assumptions of PCORR, SEP, LOC, PCC and EX gave rise to a very particular probability space which was defined in the overlap $L_i \wedge R_i$, and therefore referred to a given type of perfectly correlated runs of an EPR/EPRB experiment (see Figure 5.1 in page 131). The specific characteristics of such a probability space are not of crucial importance now. It is worth recalling however that the events in it had very particular deterministic relations and that the common causes were defined as *individual*-common causes.[80]

A generic such 'overlapping' region $L_i \wedge R_i$ is depicted in Figure 5.4 (a). We can conclude from the above that it is in such a region, i.e. for that

[80] See Sections 5.2.1 and 5.2.2 for details.

specific perfect correlation, and only in it, that the *Minimal Theories* hold.

Now, in a first step, in order to obtain the Bell inequality Graßhoff et al. need to 'mix' their *Minimal Theories* in pairs —combining, for instance, equations (5.22) and (5.23)[81]—. This amounts to superimposing two of the referred diagrams. The result can be seen in Figure 5.4(b). It follows that, because $L_i \wedge L_j = \emptyset$ (and $R_i \wedge R_j = \emptyset$), the resulting joint probabilities entering the Bell inequality are all zero, i.e. $p(L_1^+ \wedge R_2^+) = 0$, $p(L_1^+ \wedge R_3^+) = 0$ and $p(L_2^+ \wedge R_3^+) = 0$. Thus the Bell inequalities are trivially satisfied, and no contradiction is found.

There is nothing strange in the above result. Recall that the result is conditional on PCORR. Why then should we expect to observe correlations among outcomes such as L_i^a and R_j^b, i.e. $Corr(L_i^a, R_j^b)$, with $i \neq j$ when we are assuming an EPR-Bohm experiment with perfect correlations?

There is only one way for such correlations to occur. It has to do with the underlying formal description of the experiment. I discussed in Section 5.1.3 how *purely formal* correlations may very well be misinterpreted and be taken to reflect some physical features of the system at hand. I pointed out that this was specially likely if dealing with 'mixtures' of 'perfectly correlated' statistics as it is the case here.

In sum, we have found that if no *common*-common causes are assumed the Graßhoff et al *Minimal Theories* are only valid when they refer separately to *different perfectly correlated runs of an EPR-Bohm experiment*. In such cases no correlations of the type $Corr(L_i^a, R_j^b)$ can be observed in the same run, with $i, j = 1, 2, 3$ $(i \neq j)$ and $a, b = +, -$. When derived in this way, the Bell inequalities are vacuous since they involve probability-zero elements (such as $p(L_1^+ \wedge R_2^+) = 0$) and hence are trivially satisfied. On the other hand, if we are to derive the Bell inequalities meaningfully by means of the *Minimal Theories* (5.22)-(5.27), assuming *common*-common causes still seems unavoidable.

The result by Graßhoff et al. hence fails to say anything definite on the non-existence of a *local individual*-common cause that would explain EPR-Bohm correlations. And Graßhoff et al. would seem to accept as much when they write:[82]

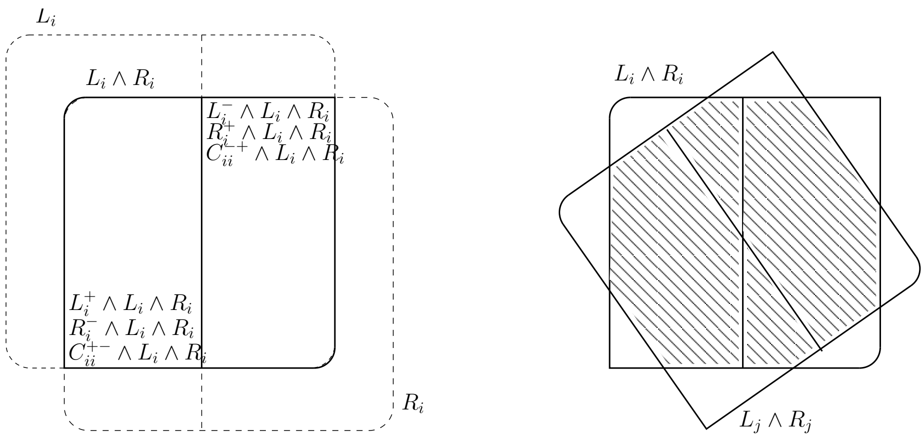
> We have not been able to derive a Bell-type inequality ruling out perfect correlations and allowing different [individual] common

[81] More specifically, Graßhoff et al. derive the Bell inequalities using expressions such as

$$p(L_1^+ \wedge R_2^+) = p(L_1 \wedge C_{11}^{+-} \wedge R_2 \wedge \neg C_{22}^{+-}). \tag{5.45}$$

See (Graßhoff, Portman and Wüthrich, 2005) for details.

[82] (Graßhoff, Portman and Wüthrich, 2005, p. 672). The square brackets are mine.



(a) (b)

Figure 5.4: The overlap of two minimal theories has zero probability since $L_i \wedge L_j = \emptyset$ for $i \neq j$.

cause variables.

The quotation above points as well at the assumption of *Perfect Correlation* (PCORR) —one of the central assumptions of the argument— as responsible for the relative significance of their derivation. I shall close this chapter with some brief comments about such an assumption.

5.4.4 Why Perfect Correlation?

Most derivations of the Bell inequalities rely heavily on assumptions similar to the Graßhoff et al.'s PCORR.[83] Taking *Prefect Correlation* as a reasonable assumption is usually justified on the grounds that such kinds of correlations exist (theoretically) in certain EPR experimental set-ups. I would like to revise here this approach. The main reason is that I do not see why conditions such as *Perfect Correlation* should be assumed to hold in a derivation of the Bell inequalities when the final result —the Bell inequalities themselves— is supposed to hold generally, i.e. for both perfect and non-perfect EPR correlations.

A preliminary answer to my concerns can be given by pointing out that PCORR is a legitimate assumption in that it reflects the existence —at least

[83]See for instance (van Fraassen, 1982*a*), which I discussed in detail in the previous chapter.

at a theoretical level— of such kind of correlations, i.e. perfect correlations, in an EPR experiment. But, is this really so? To require that conditions such as PCORR hold throughout an argument, does it merely tell us about the existence of a particular kind of EPR correlations or does it convey something else? I believe I have already answered this question with my discussion in the previous section. We saw there, in particular, that the implications of PCORR reach deep into the argument.

Thus, I think the question is not any more whether EPR perfect correlations exist but whether the partial results found for such particular kinds of correlations apply as well to more general correlations. That is to say, can we simply generalise the structure that PCORR forces on the EPR events to the whole EPR experiment? Or in other words, do the common causes postulated under the constraint of PCORR have the same properties that the common causes of more general, non-perfect EPR correlations? I do not think the answer to these questions is so straightforwardly positive. On the one hand, we already saw in Chapter 3 how special is the case of perfect correlations as regards the existence of Reichenbachian common causes. We saw there, in particular, that Reichenbachian common causes of perfect correlations are very specific, since they are bear much stronger constraints than those of any generic non-perfect correlation. Indeed, we found that such common causes must be deterministic. And on the other hand, my arguments in the previous section, once more, suggest that generalising the (partial) results obtained under the PCORR assumption, i.e. specific to EPR perfect correlations, is problematic and clearly compromises the significance of the whole derivation.[84] Hence, in my opinion, the correct strategy for the derivation of the Bell inequalities should not need to refer to the notion of perfect correlation, perhaps in the vein of Clauser and Horne.[85]

Finally even Graßhoff et al. acknowledge that assuming *Perfect Correlation* does not give a satisfactory derivation of the Bell inequality, although for other reasons than those explained above. In this regard, it is pointed out in the final discussion of the paper[86] that their derivation might be challenged on the grounds that PCORR is a necessary condition for it to obtain. If this is so, they go on, and given that perfect correlations are not experimentally

[84]Recall that I showed that at least one of the two claims contained in each LOC2 and LOC3, i.e. LOC2(ii) and LOC3(i), only held if PCORR held as well. That means that PCORR is indeed necessary for each LOC2(ii) and LOC3(i). And if this is so then PCORR is also necessary for the *Minimal Theories* to hold. What this means is that PCORR is assumed throughout the derivation, carried along, so to speak by the locality assumptions.

[85]Recall that the Clauser-Horne inequality is a Bell-type inequality which is derived without considering perfect correlations at all. Cf. (Clauser and Horne, 1974) for details.

[86](Graßhoff, Portman and Wüthrich, 2005, p. 677).

observed, a common cause model might be provided:

> If PCORR is indeed a necessary assumption for our derivation of the Bell inequality, it should be possible to construct a model in which PCORR does not hold (being violated by an arbitrary small deviation, say). Since the actually measured correlations are never perfect —a fact that is usually attributed to experimental imperfections— it is not obvious how such a model could be refuted.

Indeed, this observation was the motivation for further work in which a minimal assumption derivation of a Clauser-Horne inequality is provided.[87] In a similar vein, Hofer-Szabó[88] also claims to have obtained the Bell inequalities without the assumption of *Perfect Correlations*. Hence the issues are complex and the debate continues. But at least I hope to have shown that the Graßhoff et al. putative proof of the impossibility of common causes is at the very least questionable.

[87](Portman and Wüthrich, 2006).
[88](Hofer-Szabó, 2007*b*).

Chapter 6

A Common Cause Model for EPR Correlations

> "Quantum Physics has presented us with very complex processes and to meet them we must further enlarge and refine our concept of causality."
>
> — Albert Einstein

In this chapter I shall provide a Reichenbachian common cause for the EPR correlations. The model is motivated by the results of both Szabó and Graßhoff et al. discussed in the previous chapter, and builds on the intuition that measurement operations are crucial to the measured outcomes. The importance of measurement in the model though is not merely 'operational' or observational, so to speak, in that it constitutes the 'vehicle' for the particular outcomes to be observed. More importantly, measurement operations have a causal reading in the model, and as such will be considered (explicitly) to be part of the common causes. As a result the model turns out to be non-factorizable and may be rejected on the grounds that it does not conform our standard (locality) intuitions. Yet I shall suggest, and I hope I can sufficiently motivate, that the non-factorizable common causes of the model are indeed a legitimate (and intuitively appealing) alternative for a causal explanation of the EPR correlations.

6.1 Summary of Results and Motivation

Up to this point I have given sufficient reasons (I hope) to cast doubts on the usual arguments claiming that Reichenbachian common cause explanations of the EPR correlations are not possible. I did so in various stages:

I explained first why arguments such as van Fraassen's are not strictly valid inasmuch they contain the assumption that the (Reichenbachian) common causes be indeed *common*-common causes. Such an assumption was seen to be too restrictive and ultimately unwarranted. This suggested that (Reichenbachian) *individual*-common cause models of the EPR correlations that conformed to the standard intuitions about locality, etc. could still be possible. That was indeed the motivation behind Szabó's attempt to construct one such model explicitly —the results by the Budapest School on Reichenbachian common cause completability were central to this end—. Szabó's model however turned out to contain certain dependencies between the postulated common causes and the measurement setting events, which were interpreted as 'conspiratorial'. It was then left as an open question whether such *individual*-common cause explanations of the EPR correlations were still possible.

As we saw, the issue was later re-taken in a paper by Graßhoff, Portman and Wüthrich, who claimed to have responded Szabó's open question in the negative. In particular, Graßhoff, Portman and Wüthrich claimed to have derived the Bell inequalities without the assumption of *common*-common causes, i.e. the postulated common causes were explicitly assumed to be *individual*-common causes. However, a closer analysis of the Graßhoff, Portman and Wüthrich argument suggested, I argued, that the derivation contains assumptions which combined together implicitly amount to that of *common*-common causes —I expressed that as a sort of dichotomy by pointing out that the derivation could only be taken to be of general significance if the assumption of *common*-common cause was implicit in it. On the contrary, insisting on the common causes being *individual*-common causes throughout, the derivation could only be considered to hold for a particular subset of the the EPR correlations, i.e. perfect EPR correlations—. Thus, I concluded that the Graßhoff, Portman and Wüthrich derivation is not conclusive either. It does not determine whether *individual*-common cause explanations of EPR correlations subject to reasonable assumptions (of locality, etc.) may be given.

Finally, I pointed out that quite recently Portman and Wüthrich and independently Hofer-Szabó have claimed that, after all, no *individual*-common cause explanations of the EPR correlations are possible under standard locality assumptions —once more they rely on the derivation of the Bell inequalities—. These results are too recent for me to be able to provide an assessment here.

If such were the case, we would be faced, once more, with the task of reassessing the assumptions under which such a conclusion is valid, and ultimately dismissing one of them so as to avoid the charge of Bell's theorem. Recall that the most commonly accepted move is to drop the assumption that Reichenbachian common causes that explain the correlations exist. Thus, if we are to provide a sensible *individual*-common cause model we will need to argue that one (or more) of the remaining assumptions fails to hold for some reason. This may be achieved by suggesting a new reading or interpretation —different to the standard— of the assumption(s) in question. Offering a sensible Reichenbachian common cause model at this point then responds to the fact that the reviewed arguments all contain assumptions which may have other alternative readings. Of course, reassessing such assumptions may have an impact on the concepts that the new model involves. Perhaps, even a revision of the underlying ontology of the model might be in need.

All in all, if we succeed in providing such a common cause model we will certainly have made a step forward in building a common-causal explanation of EPR correlations —even if the common causes turn out to be non-factorizable in the sense of Bell—. Furthermore, we will have provided also an argument in defence of Reichenbach's Common Cause Principle (RCCP) as regards quantum mechanical correlations.

6.2 Measurement Operations, Quantum Probabilities and Causation

The common cause model that I am just about to propose has three key ingredients besides the fact that the common causes are explicitly *individual*-common causes. In the first place, it will be important to note again that Reichenbach's Common Cause Principle (RCCP) is originally conceived as a classical principle of causal inference. In particular, RCCP is spelt out in terms of classical probabilities.[1] Recall that this is what was behind Szabó's remarks regarding the interpretation of quantum probabilities and the correlations involved in EPR experiments.[2] In my view, Szabó's work also shows how important measurement is. Thus, making the role of measurement explicit will be crucial in developing the model. This will be our second ingredient. Finally, I shall propose a reinterpretation of the *Measure-*

[1] Recall the formal account of RCCP in Chapter 3.

[2] See Section 5.1 and (Szabó, 2000) for details.

ment Independence conditions, which would allow for their violation without entailing the usual 'conspiracy' implications. Let us consider these three issues in turn.

6.2.1 EPR Correlations from a Classical Perspective

I just pointed out that Reichenbach's Common Cause Principle (RCCP) was originally proposed to account for classical correlations (and common causes). In dealing with quantum correlations the question immediately arises as to whether RCCP needs to be modified in some particular way (formally perhaps) to conform to the quantum mechanical probabilistic formalism. As we saw in Chapter 5, Section 5.1, Szabó points out that this question needs to be assessed before addressing the applicability of RCCP for quantum correlations. Recall that Szabó's response is to suggest that quantum probabilities, i.e. trace-like quantities, must be interpreted as classical conditional probabilities —conditional on the corresponding measurement operations, that is—. In doing that, Szabó adapted, so to speak, the quantum formalism to the classical talk of RCCP. In other words, Szabó's approach provided a classical description of the quantum correlations such that a (classical) common cause explanation could be attempted by means of RCCP, a classical principle.

Thus, what Szabó offered really was a reformulation of the problem in which measurement operations were explicitly present in the corresponding probability relations. As a consequence, the question as to whether a Reichenbachian common cause existed for any two EPR correlated outcomes, say L_i^a and R_j^b (where $i, j = 1, 2, 3$ and $a, b = +, -$), was accordingly reformulated. We were then urged to ask instead whether a Reichenbachian common cause existed of such outcomes *given that the corresponding measurement operations had taken place*. The model proposed by Szabó postulated, as a result, a set of *individual*-common causes which explicitly referred to both outcomes and measurement settings. These are as well the C_{ij}^{ab} in the previous chapter discussion of the Graßhoff, Portman and Wüthrich argument.

I agreed that that was the right approach if we are to provide an *individual*-common cause model of the EPR correlations. However, I would like to tackle the issue from a somehow different perspective. More specifically, I shall not look at the EPR correlations as 'quantum correlations' —i.e. as correlations defined between conditional probabilities, as Szabó suggested—, and postulate then the corresponding *individual*-common causes. Instead, I shall look at the EPR correlations from a completely classical perspective first, look for the corresponding common causes, and then provide them with a quantum mechanical interpretation. We may see how this is to be done with the aid of a simple example:

Suppose that we do not know anything about quantum mechanics but that a good friend of us does. Our friend has designed a device which collects the information of an EPR experiment and 'translates' it into light flashes. The device has two identical panels ('PANEL A' and 'PANEL B') with green and red lights in three rows. Each row, in particular, has a pair of lights, one green and the other red. The device is somehow connected to the EPR detectors. Spin measurement outcomes in the left wing of the EPR experiment then make the lights in 'PANEL A' flash. Say, for instance, that if the outcome L_1^+ has been measured, then the green light in the first row of 'PANEL A' flashes; or that if L_2^- is obtained in the EPR experiment, the second row's red light in 'PANEL A' flashes. Similarly for the right wing outcomes. Say, for instance, that the outcome R_3^- has occurred, then the red light of the third row in 'PANEL B' flashes. Moreover, the device is such that the causal connections between the EPR events and their corresponding flashing light events are known to be deterministic and work completely independently for each panel.[3] It is clear then that the correlations between the two panels' flashing light events exactly mirror those of the EPR experiment.

As I said, we do not know anything about quantum EPR, nor do we know what is the origin of the correlations observed between the lights. We just observe the correlation. Say however that we heard about Reichenbach's Common Cause Principle and are then ready to postulate a common cause for each of the observed correlations. I shall stress that to us the whole issue is entirely classical, i.e. classical correlations (between lights) are observed and classical Reichenbachian common causes are to be postulated. Thus, it is not difficult to come up with such *individual*-common cause in the light of results such as 'common cause completability', of which we are aware.[4]

Our good friend intends to use the information that we may provide about our postulated common causes and attempt at an explanation of the EPR correlations. Yet, she is aware of an extra issue that might be problematic, namely that unlike the correlations between the flashing lights observed by us, which are not space-like separated, those in the EPR experiment are. On reflection, however, she realises that it should not make a difference. For she knows that each EPR outcome is the one and only cause of its corresponding flashing light event —she knows, in particular, the detailed causal structure that relates each of the outcome events in the EPR experiment and the

[3]Say for instance that the mechanisms connecting each wing with its corresponding panel are completely reliable and operate independently.

[4]Of course, we may face problems when physically interpreting the postulated common cause events. I already discussed such an issue in Chapter 3 and suggested possible solutions. I shall assume thus for the sake of the argument that we are able to deal with such problems.

corresponding flashing light event—. She has then come to the conclusion that the postulated common causes for the flashing light correlations must be the same that the common causes for the corresponding EPR correlations.

Our friend thus sets for a common cause explanation of her EPR correlations in terms of our postulated common causes. Only, she will later need to give an appropriate interpretation of such common causes in order to accommodate the specific features of the EPR experiment, such as the space-like separation of the to wings, etc.

The first thing that our friend notices is that our postulated common causes are Reichenbachian *individual*-common causes. This is to say, they are common causes which screen-off one, and only one, of the correlations. Consequently, the postulated common causes have a label which identifies the specific correlation (between flashing lights in our case, or the corresponding EPR outcomes in her case). Let us denote such a common cause, for instance, $\mathsf{C}^{\mathsf{ab}}_{\mathsf{ij}}$. We will then have:

$$p(L^a_i \wedge R^b_j|\mathsf{C}^{\mathsf{ab}}_{\mathsf{ij}}) = p(L^a_i|\mathsf{C}^{\mathsf{ab}}_{\mathsf{ij}}) \cdot p(R^b_j|\mathsf{C}^{\mathsf{ab}}_{\mathsf{ij}}), \tag{6.1}$$

$$p(L^a_i \wedge R^b_j|\neg\mathsf{C}^{\mathsf{ab}}_{\mathsf{ij}}) = p(L^a_i|\neg\mathsf{C}^{\mathsf{ab}}_{\mathsf{ij}}) \cdot p(R^b_j|\neg\mathsf{C}^{\mathsf{ab}}_{\mathsf{ij}}), \tag{6.2}$$

$$p(L^a_i|\mathsf{C}^{\mathsf{ab}}_{\mathsf{ij}}) > p(L^a_i|\neg\mathsf{C}^{\mathsf{ab}}_{\mathsf{ij}}), \tag{6.3}$$

$$p(R^b_j|\mathsf{C}^{\mathsf{ab}}_{\mathsf{ij}}) > p(R^b_j|\neg\mathsf{C}^{\mathsf{ab}}_{\mathsf{ij}}). \tag{6.4}$$

Note that I am using here basically the same notation to that in the two preceding chapters. I am just denoting the common cause of the classical correlations with a serif font $\mathsf{C}^{\mathsf{ab}}_{\mathsf{ij}}$ instead of the usual *italic* C^{ab}_{ij} I used for the common causes of the EPR correlations up to now. As we will see in a moment this slight change of notation is justified since the $\mathsf{C}^{\mathsf{ab}}_{\mathsf{ij}}$ are really different events from the C^{ab}_{ij} in the preceding two chapters. Still the sub-indexed and super-indexes keep the same meaning, i.e $i, j = 1, 2, 3$ and $a, b = +, -$ represent the three possible measurement directions and the two possible spin outcomes respectively.

Finally, note as well that the postulated common causes are not necessarily deterministic, i.e. their occurrence does not necessarily entail that the corresponding outcomes will occur with certainty. This remark as we will see in not crucial, however.

6.2.2 Measurement and Causation

Our good friend may now ask what is the structure of the $\mathsf{C}^{\mathsf{ab}}_{\mathsf{ij}}$, or in what sense these common causes are associated to the quantum mechanical system, i.e. to the singlet state. She notes two remarkable features of the postulated common causes when addressing such a question. On the one hand, the postulated common cause includes information —expressed by the labels— not only about the outcomes, but also about the specific measurement settings under which the outcomes are detected. On the other hand, our friend recalls that the common causes were postulated for classical correlations (between the different light flashes) and that she was led to think that the same common cause would explain both the classical correlations and the corresponding quantum ones. These observations —perhaps the latter in particular— seem to suggest that the $\mathsf{C}_{\mathsf{ij}}^{\mathsf{ab}}$ is an event that somehow *contains* the measurement operations. This is indeed a remarkable feature of the postulated common causes. In particular, this seems to be a feature exclusive of quantum mechanics since classically, by contrast, measurement operations do not usually causally influence experimental outcomes. Such a dependence, nevertheless, may be motivated classically as well.[5]

Classically, we might imagine the following fictitious situation. Suppose we have some identically looking spherical objects the nature of whose we would like to understand better. In order to do so we chose to measure something we call ‘elasticity’. Suppose for the sake of the argument that ‘elasticity’ is a property that all objects have and which takes only one of two possible values when measured, either ‘0’ or ‘1’. Thus, when it comes to ‘elasticity’ we can only find ‘0-elasticity’ objects and ‘1-elasticity’ objects. We want to find out then whether our spheres have ‘0-elasticity’ or ‘1-elasticity’ so we set up a simple experiment to this end. As it happens, though, ‘elasticity’

[5]Measurement dependence may also motivated quantum mechanically somehow (following the Copenhagen interpretation) by appealing to the collapse of the wave function. The argument would go along the following lines. Since, as I pointed out, the above postulated common cause $\mathsf{C}_{\mathsf{ij}}^{\mathsf{ab}}$ is a completely classical event but with a certain quantum mechanical origin, it may very well be thought of as a collapsed ‘quantum common cause’ event. In particular, we might take it that there is a certain (hidden) dynamical variable in the singlet state wave function that, on measurement, collapses to give the postulated common cause as a (hidden) outcome, i.e. the $\mathsf{C}_{\mathsf{ij}}^{\mathsf{ab}}$. It is thus clear that if such were the case the $\mathsf{C}_{\mathsf{ij}}^{\mathsf{ab}}$ would refer to a specific measurement operation.

Such an interpretation, however, has in my opinion serious problems. More in particular, the idea of collapse is clearly problematic and is not properly understood —it is not know what is the mechanism that might be behind it—. Thus, relying on the collapse of wave function in order to make causal factors arise does not seem to provide anything fundamental about the possible underlying causal structure of the system, and seems, moreover, obscure somehow. We might recall to this end Bell’s own critique in “Against Measurement” (Bell, 1987, pp. 213-31 of the 2004 revised edition) along this lines pointing out that a theory that contains elements the interpretation of which relies on measurement cannot be a fundamental theory. I shall then not pursue this option here. The option I do explore accounts for the measurement dependence in a much more natural way, as we will see.

is a property which is not 'directly' observable and needs to be inferred from other experimental facts. Our experiment then consists in hitting the spheres, one at a time, with a hammer —the same hammer and with the same strength each time—. The recorded events are of two kinds: we measure the hammer's recoil after hitting the sphere as well as the sound produced —we can suppose for simplicity that typically we will find the hammer recoiling in two modes, so to speak, a 'large recoil' and a 'short recoil'. Similarly, the sounds emitted on hitting the spheres with the hammer turn out to be of two kinds only, one 'higher pitched' and the other 'lower pitched'—. These two bits of information therefore are enough to tell whether our spheres have '0-elasticity' or '1-elasticity'. That is, if a 'short' hammer recoil and a 'high pitch' emitted sound are measured we can safely infer that the sphere has '0-elasticity'. Otherwise, the sphere has '1-elasticity'.

This example I think illustrates sufficiently how measurement may have a causal role in a causal explanation. In particular, our measurement operations —the hammer hits— are responsible, causally responsible that is, for the observed outcomes that allow us in turn to infer a property of the system at hand. This is not to say, of course, that the hammering is the cause of the sphere having a particular value of the property 'elasticity', i.e. having '0-elasticity' or '1-elasticity'. But once more, what the hammering is causally relevant for is the particular sound and degree of recoil which allow us to make an inference regarding the 'elasticity' of the spheres. We could perhaps say the our measurement operations *reveal* what the value of the 'elasticity' of the spheres is by *causing* further events (related to that property).

My view is that spin measurement experiments may be taken to some extent —I will explain to what extent in a moment— to be similar to 'elasticity' measurements in the example above, at least empirically, i.e. as far as empirical observations are concerned. Recall, for instance that in an EPR experiment spin components (values) of the electrons are *detected* by letting the particles through an inhomogeneous magnetic field (a Stern-Gerlach magnet) and observing them flashing on the screen, either on its upper or lower side. The electron spin component is thus *inferred* from the effect that the Stern-Gerlach magnet has on the particle. Thus, the magnetic field may be taken not to cause the electron to have a particular spin value but to *reveal* such a spin value by causing a change of the electron trajectory (in either way). Of course, such a view would be soon rejected by proponents of the Copenhagen interpretation of quantum mechanics on the grounds that no such property exists, really, before measurement.[6] The orthodox would

[6]Recall that the interpretation of Heisenberg's Uncertainty Principle conveys that incompatible properties (and by extension all properties in general) of a physical system

then claim that no property such as spin really exist prior to measurement and regard the above as irrelevant. My view, by contrast, is that the orthodox's critique, while it would surely undermine the identification I made above between the case of 'elasticity' and spin to the extent that both are real physical properties of the system, would leave untouched the claim that measurement settings may be taken to be causally relevant to the outcomes (whatever the underlying reality of the system is). And recall that this was in fact the claim I aimed to motivate through the example above. Thus, we may still stick to such an idea in what follows and need not worry for the moment about the degree of realism we are committed to.

A final remark concerning measurement is in order before proceeding. Note that measurement operations as considered above are not to be taken as the only relevant causal factors of the observed outcomes. In the classical case, for instance, outcomes do also arise due to the property we were aiming to study. In particular, whether the spheres have '0-elasticity' or '1-elasticity' clearly contributes as well to the fact that a particular sound and degree of recoil is observed. And such a contribution, I would say, seems to be causal as well. The case of spin measurements is different just for the reasons given above concerning the orthodox interpretation of Heisenberg's uncertainty relations. As I pointed out, the uncertainty relations are usually interpreted as a bound on the physical reality of the properties of a system. Such a view makes it difficult thus to maintain that a property is responsible for an outcome. However, there seem to be good intuitive grounds to say that if there is some causal connection between the unmeasured quantum system and the measured outcomes this has to include some feature exclusive to the system itself. In the EPR context, it is usually claimed that the spin-singlet state itself —or perhaps some factor closely related to it, and characteristic of it— is causally relevant to the EPR outcomes.

6.3 A non-Factorizable Common Cause Model for EPR

The conclusions of the discussion in the previous section about measurement and the structure of the postulated common causes can be expressed formally, in terms of algebra events and probabilistic relations.

have no simultaneous reality. That is, Heisenberg's Uncertainty Principle is viewed, under the orthodox interpretation, as a bound on the physical reality of the properties of a system.

6.3.1 *Individual*-Common Causes and Outcome Independence

Recall first that the common causes were each postulated to screen-off a single correlation —these are what I called *individual*-common causes—. I expressed that by means of the screening-off conditions (6.1)-(6.2) in page 160:

$$p(L_i^a \wedge R_j^b | \mathsf{C}_{\mathsf{ij}}^{\mathsf{ab}}) = p(L_i^a | \mathsf{C}_{\mathsf{ij}}^{\mathsf{ab}}) \cdot p(R_j^b | \mathsf{C}_{\mathsf{ij}}^{\mathsf{ab}}), \tag{6.1}$$

$$p(L_i^a \wedge R_j^b | \neg\mathsf{C}_{\mathsf{ij}}^{\mathsf{ab}}) = p(L_i^a | \neg\mathsf{C}_{\mathsf{ij}}^{\mathsf{ab}}) \cdot p(R_j^b | \neg\mathsf{C}_{\mathsf{ij}}^{\mathsf{ab}}). \tag{6.2}$$

Now the role of measurement in the model can be explicitly accounted for if we think of the postulated *individual*-common causes $\mathsf{C}_{\mathsf{ij}}^{\mathsf{ab}}$ as the result of a conjunction of the specific measurement operations L_i and R_j and some other causally relevant factor Λ. That is:

$$\mathsf{C}_{\mathsf{ij}}^{\mathsf{ab}} \subset L_i \wedge R_j \wedge \Lambda. \tag{6.5}$$

Typically, Λ will most probably be associated to the singlet state Ψ_s itself, and perhaps to some other relevant causal factors prior to the preparation of the entangled system.[7] In some sense Λ seems much deeply related to the quantum mechanical structure of the system (of the experiment) than the postulated common causes $\mathsf{C}_{\mathsf{ij}}^{\mathsf{ab}}$. This observation is also supported by the fact that the $\mathsf{C}_{\mathsf{ij}}^{\mathsf{ab}}$ are explicitly classical in the same sense that in the case of the light panels example in the previous section. The factor Λ however has been proposed in the context of the quantum mechanical system, i.e. the spin singlet state Ψ_s.

We may assume furthermore that such causal factor Λ is common to several (or even to all) correlations in the EPR experiment. It is important to note however that Λ, in contrast to the $\mathsf{C}_{\mathsf{ij}}^{\mathsf{ab}}$, will not in general screen-off the correlations. I shall explicitly require this in order to avoid problems concerning *common*-common causes. That such problems may arise is clear, since if Λ is, as I said, a causal factor common to all the possible outcomes of the experiment, and if it were a screening-off event, it would be a *common*-common cause of all the outcome correlations. Such *common*-common causes we saw in Chapter 4 may be problematic in the sense that a model which

[7]This characterisation of Λ somehow reminds to the kind of events Cartwright considers that must be the right common cause events for EPR. That the similarity is quite so it will become clear in a moment, since I will be requiring (or at least allowing) that the Λ be non-screening-off events, just like in Cartwright's common cause account of EPR. See for instance (Cartwright, 1987; Cartwright and Jones, 1991) and (Chang and Cartwright, 1993).

relies on them cannot be guaranteed to exist.[8] Moreover, in the EPR context, assuming *common*-common causes leads quite straightforwardly to the Bell inequalities. Thus, requiring that Λ is not a screener-off event is crucial in order to avoid such problems, as well as the charge of Bell's theorem (of course, quite trivially, Λ will not conform to the Bell inequalities from the very beginning, since it is defined as a non-screening-off hidden variable).

Similar reasons lead us to require as well that the conjunction of Λ with the measurement operation events, i.e. $L_i \wedge R_j \wedge \Lambda$, be non-screening-off events. In particular, $L_i \wedge R_j \wedge \Lambda$ is assumed to be a causal factor common to all the correlations that involve these particular measurement settings (which in general will not screen them off). This suggests that $L_i \wedge R_j \wedge \Lambda$ contains all the postulated common causes $\mathsf{C}_{\mathsf{ij}}^{\mathsf{ab}}$ of the correlations displayed for these specific measurement settings. That is:

$$L_i \wedge R_j \wedge \Lambda = \mathsf{C}_{\mathsf{ij}}^{++} \vee \mathsf{C}_{\mathsf{ij}}^{+-} \vee \mathsf{C}_{\mathsf{ij}}^{-+} \vee \mathsf{C}_{\mathsf{ij}}^{--}. \tag{6.6}$$

One may however think that the above may entail that the $\mathsf{C}_{\mathsf{ij}}^{\mathsf{ab}}$ be *common*-common causes. In order to avoid that and stick to the idea that $\mathsf{C}_{\mathsf{ij}}^{\mathsf{ab}}$ be *individual*-common causes, we shall require furthermore that they be mutually exclusive, i.e.

$$\mathsf{C}_{\mathsf{ij}}^{\mathsf{ab}} \wedge \mathsf{C}_{\mathsf{ij}}^{\mathsf{a'b'}} = \emptyset, \tag{6.7}$$

where $a, b, a', b' = +, -$ and $ab \neq a'b'$.

The *individual*-common cause structure resulting from the above considerations can be seen in Figure 6.1.

A remarkable feature of the resulting event structure is that our postulated common causes $\mathsf{C}_{\mathsf{ij}}^{\mathsf{ab}}$ satisfy a yet familiar condition very closely related to the derivation of the Bell inequalities, namely *Outcome Independence*[9]. This is indeed what expressions (6.1) and (6.2) encapsulate. In fact, as we saw, it was one of van Fraassen's insights to identify *Outcome Independence* —which he called *Causality*— with Reichenbach's Common Cause Principle screening-off condition.[10] But note as well that, on the other hand, the event Λ (and also $L_i \wedge R_j \wedge \Lambda$) will not in general satisfy such constraint. In particular, $L_i \wedge R_j \wedge \Lambda$ will in general violate *Outcome Independence* due to the requirement that it shall not screen-off the correlations. (Recall that

[8] Recall that this result was in contrast to the idea of common cause completability which guaranteed that *individual*-common causes could always be found for any correlation. See Chapters 3 and 4 for details.

[9] See Section 4.2.3, page 92.

[10] See (van Fraassen, 1982*a*) and my own discussion in Chapter 4 for further details.

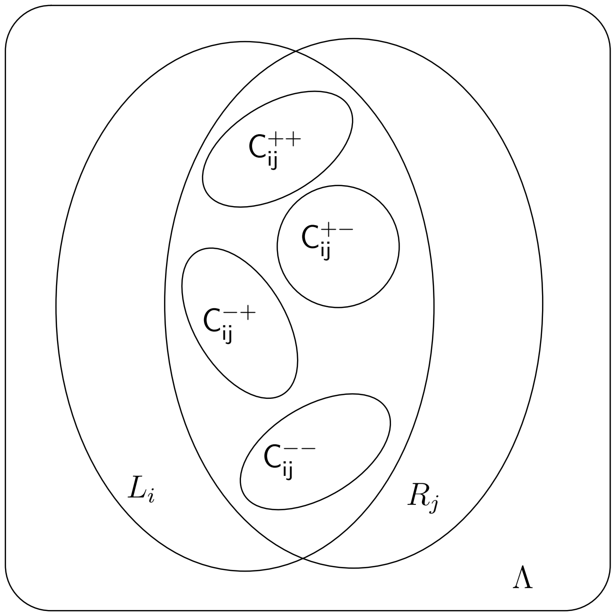


Figure 6.1: The *individual*-common cause model event structure for EPR correlations with measurement settings L_i and R_j.

I explicitly required that in order to avoid the usual problems faced by the assumption *common*-common causes.)

What it is interesting about these remarks is that the features of the model so far are able to accommodate the usual interpretation of the violations of the Bell inequalities, i.e. the violations of Bell's *Factorizability*, in terms of the violation on either *Outcome Independence* or *Parameter Independence*. In this sense, it is usually acknowledged that the reason why quantum mechanics does not conform to the Bell inequalities is that the formalism entails a violation of *Outcome Independence*, while it is compatible with *Parameter Independence*.[11] Thus, the fact that the *Outcome Independence* condition is violated when Λ is taken as the hidden variable is on complete agreement with the view that quantum mechanics violates that condition as well. For recall that we considered Λ to be deeply related to the quantum mechanical structure of the system. The postulated common causes C_{ij}^{ab}, on the other hand, were taken as classical events which in principle do not belong to the 'inner' quantum structure of the system and may thus be thought to satisfy *Outcome Independence* without any problem.

[11] Recall that such claims are normally supported by the fact that the spin-singlet state has a spherical symmetry.

6.3.2 Measurement Dependence is not a Conspiracy

A second remarkable feature of the model just outlined is that, because the dependence of the common causes on measurement, the $\mathsf{C}_{\mathsf{ij}}^{\mathsf{ab}}$ violate what I called *Measurement Independence* in the previous chapter. In particular, the $\mathsf{C}_{\mathsf{ij}}^{\mathsf{ab}}$ violate both what I referred to as 'simple' and 'generalised' versions of *Measurement Independence* (see equations (5.1) and (5.2) in page 119). However, it will suffice for my main purpose here to refer to the violations of the 'simple' version of the condition.[12] That is:

$$
\begin{aligned}
p(\mathsf{C}_{\mathsf{ij}}^{\mathsf{ab}}|L_i) &\neq p(\mathsf{C}_{\mathsf{ij}}^{\mathsf{ab}}) \cdot p(L_i), \\
p(\mathsf{C}_{\mathsf{ij}}^{\mathsf{ab}}|R_j) &\neq p(\mathsf{C}_{\mathsf{ij}}^{\mathsf{ab}}) \cdot p(R_j).
\end{aligned}
\tag{6.8}
$$

Recall that *Measurement Independence* was a condition explicitly required in Szabó's model —under the name of *No-conspiracy*— on the grounds that its violation would amount to some sort of 'universal conspiracy'.[13] The question then seems quite straightforward: In the light of expression (6.8), is our model a 'conspiratorial' model (in the sense above)? Or in other words, is *Measurement Independence* a necessary assumption for common causes of EPR correlations? The answer, I think, is negative since, as I also pointed out in Chapter 5, violations of *Measurement Independence* conditions may be reinterpreted such as to avoid the 'charge of conspiracy', as Szabó puts it.[14] Let us see more in detail how this may be done.

One of the things I noted when discussing the meaning and role of Szabó *No-conspiracy* assumption, i.e. my *Measurement Independence* condition, was that such a condition seemed appropriate only if we assumed that the common causes take place prior to the measurement apparatus is set up. For, it is only in those cases that probabilistic dependencies between the

[12] In fact, it is quite straightforward to see that a violation of the 'simple' version of *Measurement Independence* entails a violation of the 'generalised' version as well. See Chapter 5, Section 5.1 for further details.

[13] *Measurement Independence* was assumed as well in the derivation of the Bell inequalities by Graßhoff, Portman and Wüthrich, as *No-conspiracy*; and earlier by van Fraassen under the name of *Hidden Autonomy*. Van Fraassen however did not associate his *Measurement Independence* to non-conspiratorial features of the postulated EPR causal structure. See (Szabó, 2000; Graßhoff, Portman and Wüthrich, 2005), and Chapter 5, as well as (van Fraassen, 1982*a*) and Chapter 4, for details.

[14] Recall as well that Szabó's 'charge of conspiracy' may be avoided if we were to admit the possibility of backwards in time causation. I shall not purse this option further here, however. Just to remind the reader that common cause models have been proposed that explore such possibility. See, for instance (Price, 1994) for one such model. See also Chapter 5, Section 5.1 for more detailed remarks about the issue and Section 6.3.3 for the similarities between Price's model and mine.

common causes and measurement settings, i.e. violations of the corresponding *Measurement Independence* conditions, may be sensibly interpreted as some sort of ‘universal conspiracy’ —meaning that measurement operations are somehow pre-settled by the particular values that the common cause hidden variable takes—. In principle, however, there is nothing in the notion of common cause or in the very structure of the EPR experiment that forces us to stick to the idea that the common cause must take place prior to measurement (and therefore prior to the setting-up of the measurement devices). This has become clear, I hope, from the flashing lights example in Section 6.2.1. So, we may perfectly allow instead that the common causes take place after the measurement devices have been set-up. In such a case, requiring *Measurement Independence* conditions to hold does not seem particularly appealing —and in fact, expressions such as (6.8) *do not* suggest strange conspiracies of any sort—. What we can be sure is that for such an interpretation *Measurement Independence* is not a necessary condition for the common causes. On the contrary, violations of *Measurement Independence* will necessarily happen as long as measurement has some sort of (causal) influence on the common causes. And this is what our model features.

In sum, we have seen that our model explicitly violates *Measurement Independence*. This is however not due to any sort of conspiratorial behaviour by which the common causes pre-fix the measurement settings. On the contrary, the violation of *Measurement Independence* is quite naturally seen as a consequence of the fact that measurement operations are causally relevant to the outcomes and thus ‘contribute’, so to speak, to the postulated common causes.

6.3.3 Parameter Dependence and non-Factorizability

As a direct consequence of the violation of *Measurement Independence*, the model turns out to be non-factorizable, i.e. Bell’s *Factorizability* is also violated. That this is so can clearly be seen from the fact that *Measurement Independence* is necessary for a more general independence condition, namely *Parameter Independence*, which is in turn necessary for *Factorizability*. In particular, a violation of *Measurement Independence* entails, as we have seen in the previous section, that the common cause is not independent of *both* the measurement settings. That is:

$$p(\mathsf{C}_{\mathsf{ij}}^{\mathsf{ab}}|L_i) \neq p(\mathsf{C}_{\mathsf{ij}}^{\mathsf{ab}}) \cdot p(L_i),$$
$$p(\mathsf{C}_{\mathsf{ij}}^{\mathsf{ab}}|R_j) \neq p(\mathsf{C}_{\mathsf{ij}}^{\mathsf{ab}}) \cdot p(R_j).$$

This in turn means that the outcomes will also statistically depend on *both* measurement settings, since they obviously depend on the common cause, i.e.

$$p(L_i^a|L_i \wedge R_j \wedge \mathsf{C}_{\mathsf{ij}}^{\mathsf{ab}}) \neq p(L_i^a|L_i \wedge \mathsf{C}_{\mathsf{ij}}^{\mathsf{ab}}),$$
$$p(R_j^b|L_i \wedge R_j \wedge \mathsf{C}_{\mathsf{ij}}^{\mathsf{ab}}) \neq p(R_j^b|R_j \wedge \mathsf{C}_{\mathsf{ij}}^{\mathsf{ab}}).$$

And this expression is nothing more than the violation of *Parameter Independence* (see Section 4.2.3, page 92). Thus, violations of *Measurement Independence* entail that *Parameter Independence* is also violated. Furthermore, since *Parameter Independence* is necessary for Bell's *Factorizability*, the above just means that a failure of *Measurement Independence* entails a violation of *Factorizability* as well.

The rejection of *Factorizability* has two obvious consequences, one which may be seen to support the model, the other which might be problematic for it. Firstly, it is obvious that the Bell inequalities can not be derived. This is certainly a good sign since it means that the model, i.e. the common cause structure, avoids the implications of the Bell inequalities, and more importantly that it is perfectly compatible with the (empirically confirmed) quantum mechanical predictions for the EPR correlations. Thus, the model presented here may very well be empirically correct.

On the other hand, as I pointed out in several places already, the violation of Bell's *Factorizability* is usually interpreted as a sign of non-local behaviour. In this sense it is often claimed that what the Bell inequalities show is that quantum mechanics is non-local. The standard view takes it in particular that the non-local character of quantum mechanics is due to a failure of *Outcome Independence*, while *Parameter Independence* is believed to hold.[15] Such interpretation is moreover in good agreement with the idea that there do not exist superluminal causal influences between the wings of an EPR experiment. (In fact, as we have seen, some authors[16], go as far as to claim that due to failure of *Outcome Independence* there are no superluminal causal influences at all in the EPR experiment.) Of course such a claim relies in the fact that *Outcome Independence* reflects some fundamental aspect of causation, while *Parameter Independence* guarantees locality. The received view thus denies superluminal causal influences at the expense of retaining a local underlying event structure.

Now, since our model, contrary to such received view, violates *Parameter Independence* and conforms to *Outcome Independence*, we seem bounded to explain in some way or another the existence of non-local superluminal causal

[15]I already pointed out that this is claimed on the grounds that the spin singlet state is spherically symmetric. See Section 4.2.3, page 93.

[16]Van Fraassen (1982*a*), in particular.

influences. This, of course, unless we are prepared to endorse other more radical and perhaps less intuitively appealing views concerning the implications of the failure of *Measurement Independence* (or *Parameter Independence*). One such option would be to interpret violations of *Measurement Independence* as implying backwards in time causation. The possibility of backwards in time influences is not, by any means, new. For instance, Cramer's so-called Transactional Interpretation of quantum mechanics[17] accounts quite naturally for such influences. The difficulty would be rather to provide these with a causal interpretation. This possibility was explored in a common cause model suggested by Price.[18] Indeed Price's model needs to assume backwards in time influences in order for the postulated common causes operate locally. In particular, Price's model assumes that the common cause events take place in the overlap of the backwards light-cones of the correlated EPR outcomes —this guarantees that the causal influences are local—. But this suggests as well that violations of *Measurement Independence* can only obtain if the influence of the settings on the common causes operates backwards in time.

What it is remarkable, I think, is that the probabilistic event structure of Price's model and my own presented here is exactly the same. Only, the interpretation of the events is different. In this sense, I believe, my model has an (intuitive) advantage over that of Price, namely that appeal to backwards in time causation is not needed. This is due, of course, to the mere fact that my model's common causes are postulated to take place after the measurement operations. The pay-off is, as I already pointed out, that we have explicit non-local behaviour. This is not new either. A similar non-locality is found in Bohm's quantum mechanics which, just as in the model outlined here explicitly violates *Parameter Independence*, while it conforms to *Outcome Independence*. I shall not however, pursue further the possible similarities between Bohmian mechanics and the model I am sketching here. It will suffice, I believe, to try to explain the non-local character of the model by attempting at a reinterpretation of our common causes at an ontological level.

---

[17] Cf. (Cramer, 1986).

[18] Cf. (Price, 1994). Although the model was initially presented as a common cause model (containing backwards in time causal influences), Price seemed later to retract from interpreting it as causal (which perhaps allowed it also to fit better with Cramer's Transactional Interpretation). In particular, Price (1996) seems to suggest that the backwards in time influences of the model be of no causal origin. An explicit causal interpretation of Price's model is given in (Suárez, 2007). For the sake of my argument here, I shall assume that such model may indeed be causally interpreted.

6.4 The Model's Ontological Implications

The model I just presented provides in my view a quite plausible explanation of the EPR correlations in terms of *individual*-common causes. This is not to say that such an explanation is free of controversial aspects. One of the model's most controversial features is probably the fact that the common causes in the model are explicitly *non-factorizable*. Non-factorizability, however, has arisen in the context of the intuition that measurement is causally relevant to the EPR correlated out comes. In fact, I really think that it is a virtue of the model that it makes explicit the role of measurement as (partially) causally responsible for the correlation. But there is also a pay-off. In particular, the fact that our common causes are postulated as events that include information about measurement at both wings of the EPR experiment —which, recall are space-like separated— suggests that they arise non-locally. Moreover the common causes operate non-locally as well to produce their corresponding correlated (space-like) outcome events. Thus, although, as I said, the model's common causes may be seen as classical events —the instances of which would in principle thought to be well located in space-time and propagate locally— they inherit the quantum 'weirdness' of the spin singlet state.

I can think, in principle, of two different ways to interpret such 'strange' features that our common causes display. These ways require us to revise our classical concepts regarding events or causal powers. Of course, ontological concerns (and difficulties) are no new to quantum mechanics. In fact, as we shall see in a moment, the ontological readings that the model suggests have already been discussed in the literature, more in particular in relation to the conceptual implications of the idea of fundamental particle in quantum field theory. I shall not discuss such issues in detail but just point out to what extent the results of such a debate are relevant to the model presented here.

Finally, I would like to stress that it is not my intention here to endorse any of the two possible interpretations of the model's common causes in particular. I believe that, in order to evaluate them further work needs to be carried out, which I leave for a future occasion.

6.4.1 A non-Local Events Ontology

One possible interpretation of the model would take it that the common causes C_{ij}^{ab} are non-localised events which somehow spread over space and time. Such non-localised events allow for all causal influences to propagate locally. Consider the causal 'chain' of a given EPR experiment as constituted by two steps. In a first stage we may interpret the common cause C_{ij}^{ab} as

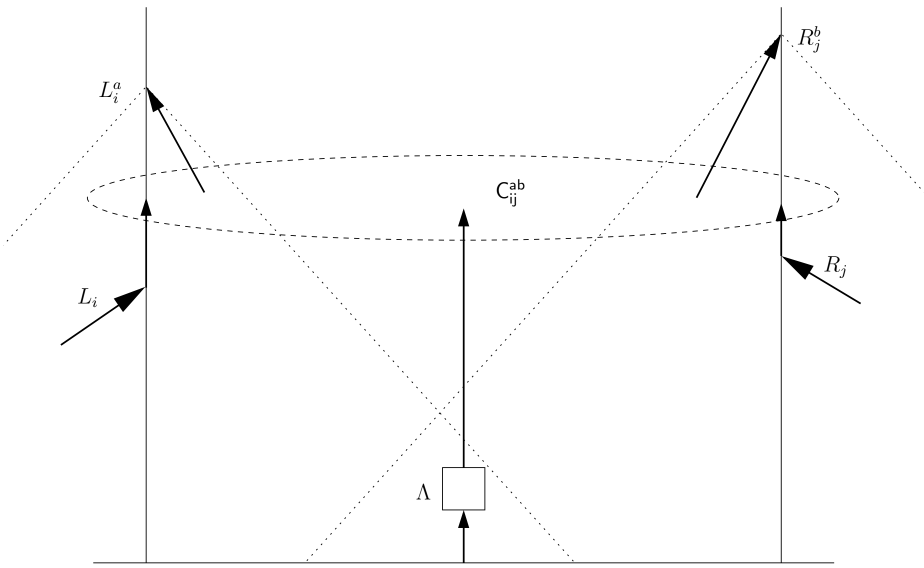


Figure 6.2: Space-time diagram of the model under the view that the common cause $\mathsf{C}_{\mathsf{ij}}^{\mathsf{ab}}$ are non-localised events which may however locally cause the corresponding EPR outcomes L_i^a and R_j^b. (Causal influences are represented by continuous lines).

an effect of the events representing measurement operations on both wings, i.e. L_i and R_j, and the event Λ associated to the spin singlet state —plus some other causally relevant factors, perhaps—. Each of these operate (causally) *locally* but their conjunction has as a result the non-localised common cause $\mathsf{C}_{\mathsf{ij}}^{\mathsf{ab}}$, which is spread out in space-time. In a second step the non-localised common cause acts *locally* to produce the corresponding outcome events. The resulting causal structure is represented in the space-time diagram in Figure 6.2 (where solid lines with arrows represent causal influences). As we can see there, it is also remarkable that the non-localised $\mathsf{C}_{\mathsf{ij}}^{\mathsf{ab}}$ do not necessarily spread out to cover whole slices of the outcome events backward light cones (represented with dotted lines in the diagram). If such were the case, the $\mathsf{C}_{\mathsf{ij}}^{\mathsf{ab}}$ would deterministically cause the corresponding outcomes. As we saw, the model keeps entirely neutral in this respect (which I think is also one of its virtues).

This idea of non-localised event is very close to what Teller[19] calls *quanta* in his interpretation of the notion of fundamental particle that arises in quantum field theory.[20] This is not to say, of course, that our common causes are

[19](Teller, 1995).

[20]Without going into details, a somehow simplistic definition of *quanta* may be given as

(or represent) fundamental particles, i.e. excitation states of quantum fields —recall moreover that I have considered the common cause to be fundamentally classical events, which would make such a view even more difficult to maintain—. It seems more appropriate to say that the common cause events inherit, so to speak, the typical properties of Teller's *quanta* —electrons, photons and the like— that we may suppose are involved in EPR experiments. So our common causes will inherit, as a matter of fact, the conceptual advantages of *quanta*, but also their interpretative difficulties.

As advantage, our common causes will not be in conflict with relativity in that, as I said, causes propagate locally throughout. The main difficulty of such an interpretation will stem perhaps from the fact that *quanta* are typically claimed to suffer from being indistinguishable precisely due to the fact that they cannot be located in space-time. This will be thus a problem for our common causes as well. We may think that such a problem is not as crucial in our case however, since, after all, our common cause events are event types, and as such need not be located in space-time —in fact, the lack of a space-time location is what allows us to call them event types—. However, even if this is so, it is not less true that the causal relations that we attribute to such event types are based on the particular causal relations among the token events that may be taken to constitute them.[21] Thus, the problem still stands as to how space-time non-localised events may be identified.

Bartels[22] offers a solution for the problem of the indistinguishability of *quanta* by pointing out that the identification of *quanta* does not necessarily require that specific space-time locations are provided. Bartels' suggestion is that *quanta* be individually identified by their *causal* location, i.e. through their *causal* past and future.[23] Bartels' idea of *quanta* seems then even more appropriate for the interpretation of the common causes that the model postulates.

---

the occurrences of well defined excitation states of the quantum field. See (Teller, 1995) for details.

[21] Recall that in Chapter 2, when discussing the relation between the idea of type event and token event, I argued that it was quite intuitively appealing to take event types as collections of tokens. I also pointed out that that was in fact the interpretation, if any, I was to endorse throughout.

[22] (Bartels, 1999).

[23] Bartels' idea of *causal* identification goes back originally to Davidson's *causal* identity criterion of events. See (Bartels, 1999, p. S181) for further details and references therein.

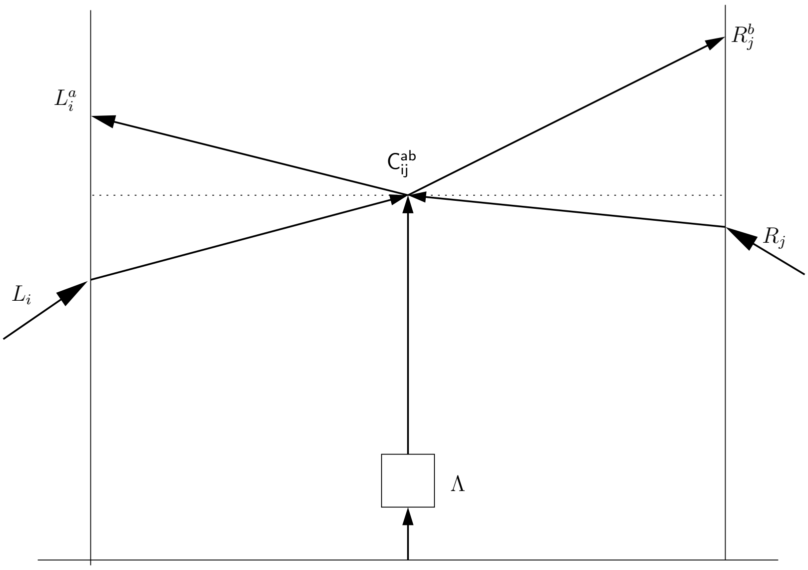


Figure 6.3: Space-time diagram of the alternative ontological interpretation of the model. The common causes $\mathsf{C}_{\mathsf{ij}}^{\mathsf{ab}}$ are now space-time localised events which act non-locally to cause the corresponding EPR outcomes L_i^a and R_j^b. (Causal influences are represented by continuous lines)

6.4.2 Non-Local Causal Influence

The second alternative interpretation of our model's common causes retains the more widespread view of events as spatiotemporally well localised. Again, I would like to stress that although the idea of space-time localisation is usually only applicable to token events, I am taking it that event types, such as our common causes or the EPR outcomes (or the corresponding measurement operations), are simply sets of tokens. It is in this sense that the expression 'spatiotemporally localised common causes' must be taken.

In any case, if the model's common causes are taken to be localised events we seem forced to accepts that their actions are not so localised. It is remarkable as well that such common cause events may also be seen as effects arising from non-local causal influences. In particular, although the event Λ associated to the singlet state may very well be taken to operate locally as a (partial) cause of $\mathsf{C}_{\mathsf{ij}}^{\mathsf{ab}}$, the measurement operations L_i and R_j (or one of them at least) need to operate non-locally for the common cause to be the conjunction $L_i \wedge R_j \wedge \Lambda$, as the model assumes. Such a situation is represented in the diagram of Figure 6.3 (where once more solid lines represent direct causal influences).

Unlike the previous interpretation of the common causes as non-localised events, in this case, the space-time location of the C_{ij}^{ab} guarantees their identity. The main problem we are to face with such 'non-local causal powers' interpretation is that superluminal influences may be in conflict with relativity theory. The key issue in this regard is whether the so-interpreted model's causal structure would allow, not only for superluminal 'bare' influences, but also for superluminal signaling. For it is the possibility of superluminal signaling that is really in conflict with relativity theory.

There are at least two possible moves in order to avoid such difficulties. On the one hand we may note that the notion of signalling may be sensitive to the concept of causation we are endorsing. The notion of signalling is usually taken to involve transfer of some physical quantity —energy, matter, etc.— and if this is so, counterfactual causation, for instance, would not seem to involve signalling even in the case of superluminal causal influences. Another possible defence of the 'non-local causal powers' interpretation would simply consist in noting that, even if the model allowed for superluminal signalling, the existence of entities such as tachyons —which are able to propagate superluminally— has not been empirically refuted to date, nor it seems to be in clear conflict with special relativity.[24] To sum up, localised common cause events with non-local causal powers provide a possible interpretation of the causal structure in the model.

As I said, at the beginning of the section, it is not my intention to endorse any of the two interpretations of the model discussed here in particular. The aim of these two last sections was rather to suggest that the model's implications may be interpreted in some way or another. A detailed discussion of the respective merits of both interpretations must await further work.

[24]Maudlin (2002, Ch. 3) provides a nice account of such views.

Conclusions

The conclusions of this thesis come in two parts. The first refers exclusively to the adequacy of Reichenbach's Common Cause Principle (RCCP) for causal inference, both in deterministic and indeterministic contexts. The second is specific to the applicability of RCCP to EPR quantum correlations, and the implications of Bell's theorem.

If we are to provide correct causal inferences from statistical facts we need first to clearly identify the relevant probabilistic relations that they involve. Thus an accurate formal definition of *genuine* correlation was provided in the first place (see Definition 1 on page 16). Two types of correlations were further distinguished. On the one hand we found that some *genuine* correlations were seen to arise as purely formal constructs in the process of modelling empirical data. We saw that this kind of correlations do not reflect any specific (physical) property of the system associated to the data sample. Thus, I called them *purely formally descriptive*, or just *purely formal*, correlations. By contrast, we saw that there are other kind of *genuine* correlations which do indeed reflect in some way or another the underlying (physical) properties of the system associated to the data sample. I hence referred to those as *genuinely physical* correlations. The distinction is important because only this later kind of correlations, i.e. *genuinely physical* correlations, should be expected to reflect causal relations. Thus, causal inference —by means of RCCP or any other methods— should only be attempted for *genuinely physical* correlations.

As for the status of Reichenbach's Common Cause Principle (RCCP), a clear cut distinction between its metaphysical and methodological parts —I called them *Principle of the Common Cause* (PCC) and *Reichenbach's Criterion for Common Causes* (RCCC) respectively— was crucial in assessing it. I followed Suárez in this and concluded that, since those two parts indeed constitute two logically independent claims, arguments against the status of one of them need not necessarily affect the status of the other. This is applicable to the most common arguments against the status of Reichenbach's Common Cause Principle (RCCP), which usually relate exclusively to RCCC, its methodological part (even if their conclusions are usually extended to RCCP as a whole). An important exception is Sober's 'Venetian Sea Levels and British Bread Prices' example, which aims at the *Principle of the Common Cause* (PCC), the metaphysical content of RCCP. Sober's example, however, involves time-dependent data series, whose correlations I suggested are *purely formal*. Thus counterexamples to RCCP of that kind, I argued, may be effectively avoided by appealing to the distinction between *purely formal* and *genuinely physical* correlations.

Regarding, more specifically, Reichenbach's Criterion for Common Causes (RCCC), we saw that it can be considered in general to be neither sufficient nor necessary for common causes. We also saw that restricting further the conditions on common causes —so as to provide a sufficient condition on them— did not seem a particularly appealing idea, since it would seem *ad hoc* in most cases. We were then left to either weaken RCCC —in order to achieve a necessary condition on common causes—, or keep the criterion and see how far we can go with it. The move to weaken Reichenbach's criterion answers to the need to account for indeterminism: while it is widely accepted that RCCC constitutes a necessary condition on deterministic common causes (and can thus be considered a valid reliable method for causal inference), the standard view takes it that it fails for genuinely indeterministic ones. Contrary to this standard view, however, I defended that RCCC is compatible with indeterministic common causes —as well as deterministic ones, of course— and hence that a generalisation of RCCC to indeterministic contexts was not fully justified. Moreover, no generalisation of RCCC seems to constitute a necessary condition on common causes either. I thus proposed, following Williamson, to take RCCP —and RCCC in particular— as a default methodological rule for causal inference.

These results were complemented by the notions of *extensibility* and common cause *completability* introduced by the Budapest School in order to motivate the idea that it is always possible to find a screening-off event for any given correlation. Despite of the highly formal character of the notions of *extensibility* and common cause *completability* I argued that it is conceivable that screening-off events found by these means may be physically interpreted in our cases of interest —and under the right conditions, and with the necessary precautions— as common causes. This, I stressed, was equally valid for both deterministic and indeterministic common causes and, as such, we could take advantage of such results in order to address our main topic, i.e. quantum correlations.

The second part of the thesis aimed to reassess the standard claim that quantum EPR correlations constitute a definitive counterexample to Reichenbach's Common Cause Principle (RCCP). The usual argument identifies Reichenbachian common causes as hidden variables that, under standard assumptions —of locality, separation, etc.—, conform to the premises of Bell's theorem and thus lead to the derivation of the Bell inequalities. It is thus argued that such (hidden variable) common cause models of EPR correlations cannot exist (in the light of the violation of the inequalities).

Van Fraassen was one of the first philosophers to identify Reichenbachian common causes with Bell's hidden variables, and his critique of RCCP has been highly influential. We saw however that several assumptions in van Fraassen's derivation of the Bell inequalities are questionable. The resulting critique of RCCP is thus not as severe as it might have been initially thought. I focused, in particular, on the fact that the common causes postulated by van Fraassen are a very special case of Reichenbachian common causes, so-called *common*-common causes. I followed the arguments of the Budapest School in claiming that this is an unwarranted assumption. Moreover, since the results of *extensibility* and common cause *completability* do not in general hold for such specific common causes, van Fraassen's formal results were seen to follow quite naturally. But their interpretation does not do so. More particularly, I argued that van Fraassen's argument shows that the EPR correlations cannot be explained in terms of *common*-common causes. However, it is inconclusive as regards the existence of Reichenbachian common cause explanations for EPR correlations.

I then discussed a further recent attempt by Graßhoff, Portman and Wüthrich to derive the Bell inequalities assuming simply *individual* Reichenbachian common causes instead of *common*-common causes. I argued however that the derivation hinges on a set of assumptions which turn out to be as strong as the assumption that the postulated Reichenbachian common causes are *common*-common causes. The alternative entails a severe restriction on the proof's applicability. I put this result as a dichotomy: if the Graßhoff, Portman and Wüthrich derivation is to be taken as a general result on EPR correlations, its assumptions are tantamount (even if implicitly) to the assumption of *common*-common causes. The other horn of the dichotomy stated that if Graßhoff, Portman and Wüthrich insist that their common causes are *individual*-common causes then the derivation is only applicable to EPR *perfect* correlations, which renders the corresponding Bell inequalities vacuous. Thus, I concluded, the argument of Graßhoff, Portman and Wüthrich is not conclusive either.

An essential ingredient in all derivations of the Bell's inequalities reviewed in this thesis is the assumption of *Measurement Independence*. This assumption carries different names in different arguments, i.e. *Hidden Autonomy* (in van Fraassen's argument) or *No-conspiracy* (in the derivation of Graßhoff, Portman and Wüthrich) but it always requires that the postulated common causes are independent of the measurement operations. Violations of *Measurement Independence* are usually interpreted in terms of some 'universal conspiracy' or backwards causation. I proposed however to look at the putative causal structure of an EPR experiment as incorporating information about measurement. In order words, I suggested that the postulated common causes of EPR correlations include measurement operations (as a part of their causal identity, that is). I justified that this may be so partly by noting that Reichenbachian common causes, as defined by Reichenbach himself, are classical events, which in the quantum mechanical context only arise after measurement is performed. Under this view then, violations of *Measurement Independence* are seen as natural and should be expected since the common causes are defined as containing information about measurement, i.e. are measurement dependent.

A common cause model for EPR correlations was sketched under such premises. An immediate consequence of the model was the violation of *Parameter Independence* —by means of the violation of *Measurement Independence*— which amounted to the model's common causes having some sort of non-local character (in the sense of Bell's factorizability). This demanded for a revision of the ontology of the postulated common causes, either in terms of spatiotemporally well localised events with non-local causal powers, or as non-localised events causally interacting as is standard, by contiguity.

The proposed model however was very tentative and sketchy. Its consequences, especially as regards its ontological interpretation, need further research. Moreover it would be interesting to see the model's implications extended beyond standard non-relativistic quantum mechanics. In this regard, we saw that the two possible interpretations of the model's common cause events were close to some proposed ontologies for the interpretation of the notion of 'fundamental particle' in quantum field theory (Teller's and Bartel's ideas of *quanta* are two such proposals). This suggests that further research efforts might be invested, not only on understanding better the notion(s) of locality that the model suggests, but also on investigating whether the model can be extended to quantum field theory. More particularly, it seems interesting to investigate what the consequences are of the violation of *Measurement Independence* in quantum field theory. This seems worth a research program on its own and is thus matter of future investigation.

Bibliography

Albert, D. Z. (1993) *Quantum Mechanics and Experience*. Harvard University Press.

Arntzenius, F. (2005) “Reichenbach’s Common Cause Principle”. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2005 Edition). URL <http://plato.stanford.edu/archives/spr2005/entries/physics-Rpcc/>.

Aspect, A., Dalibard, J. and Roger, G. (1982) “Experimental Test of Bell’s Inequalities Using Time-Varying Analyzers”. *Physical Review Letters*, **49**: 1804–7.

Ballentine, L. E. (1986) “Probability Theory in Quantum Mechanics”. *American Journal of Physics*, **54**: 883–9.

Ballentine, L. E. and Jarret, J. P. (1987) “Bell’s Theorem: Does Quantum Mechanics Contradict Relativity”. *American Journal of Physics*, **55**: 696–701.

Bartels, A. (1999) “Objects or Events?: Towards and Ontology for Quantum Field Theory”. *Philosophy of Science*, **66**, Supplement. Proceedings of the 1998 Biennal Meetings of the Philosophy of Science Asoiciation (Part I: Contributed Papers): S170–84.

Bell, J. S. (1964) “On the Einstein-Podolsky-Rosen Paradox”. *Physics*, **1**: 195–200. Reprinted in Bell (1987), pp. 14–21.

Bell, J. S. (1971) “Introduction to the Hidden-Variable Question”. In B. d’ Espagnat (Ed.), *Foundations of Quantum Mechanics. Proceedings of the International School of Physics ‘Enrico Fermi’*. Academic Press, pp. 171–81. Reprinted in Bell (1987), pp. 29–39.

Bell, J. S. (1982) “On the Impossible Pilot Wave”. *Foundations of Physics*, **12**: 989–99. Reprinted in Bell (1987), pp. 159–68.

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

Bennett, J. (1988) *Events and Their Names.* Hackett.

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

Bohm, D. (1952) "A Suggested Interpretation of Quantum Theory in Terms of Hidden Variables, I and II". *Physical Review*, **85**: 166–93 and 369–96.

Bohm, D. and Hiley, B. (1993) *The Undivided Universe.* Routledge.

Butterfield, J. (1989) "A Space-Time Approach to the Bell Inequality". In J. Cushing and E. McMullin (Eds.), *Philosophical Consequences of Quantum Theory.* University of Notre Dame Press, pp. 114–44.

Butterfield, J. (1990) "Causal Independence in EPR Arguments". In *Proceedings of the 1990 Bienal Meeting of the Philosophy of Science Association,* vol. I. pp. 213–25.

Butterfield, J. (1992) "David Lewis meets John Bell". *Philosophy of Science*, **59**: 26–43.

Butterfield, J. (2007) "Stochastic Einstein's Locality Revisited". Discussion Paper 01/07, LSE Centre for the Philosophy of the Natural and Social Sciences.

Cartwright, N. (1987) "How To Tell a Common Cause: Generalizations of the Conjunctive Fork Criterion". In J. H. Fetzer (Ed.), *Probability and Causality.* Reidel Pub. Co., pp. 181–8.

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

Cartwright, N. (1990) "Quantum Causes: The Lesson of the Bell Inequalities". In *Philosophy of the Natural Sciences: Proceedings of the 13th International Wittgenstein Symposium.* Hölderlin-Pichler-Tempsky.

Cartwright, N. (1993) "Marks and Probabilities: Two Ways to Find Causal Structure". In F. Stadler (Ed.), *Scientific Philosophy: Origins and Developments,* Yearbook 1/93, Institute Vienna Circle. Kluwer Academic Publishers, pp. 113–9.

Cartwright, N. (1999*a*) "Causal Diversity and the Markov Condition". *Synthese*, **121**: 3–27.

Cartwright, N. (1999*b*) *The Dappled World: A Study of the Boudaries of Science*. Cambridge University Press.

Cartwright, N. (2002) "Against Modularity, the Causal Markov Condition, and Any Link Between the Two: Comments on Hausman and Woodward". *The British Journal for the Philosophy of Science*, **53**: 411–53.

Cartwright, N. and Jones, M. (1991) "How To Hunt Quantum Causes". *Erkenntnis*, **35**: 205–31.

Cartwright, N. and Suárez, M. (2000) "A Causal Model for EPR". Discussion Paper 50/00, LSE Centre for the Philosophy of the Natural and Social Sciences.

Chang, H. and Cartwright, N. (1993) "Causality and Realism in the EPR Experiment". *Erkenntnis*, **38**: 169–90.

Clauser, J. F. and Horne, M. A. (1974) "Experimental Consequences of Objective Local Theories". *Physical Review D*, **10**: 526–35.

Clauser, J. F. and Shimony, A. (1978) "Bell's Theorem: Experimental Tests and Implications". *Reports on Progress in Physics*, **41**: 1881–927.

Clauser, J. F. *et al.* (1969) "Proposed Experiment to Test Local Hidden-variable Theories". *Physical Review Letters*, **23**: 880–4.

Cramer, J. G. (1986) "The Transactional Interpretation of Quantum Mechanics". *Reviews of Modern Physics*, **58**: 647–87.

Cushing, J. and McMullin, E. (Eds.) (1989) *Philosophical Consequences of Quantum Theory*. University of Notre Dame Press.

Earman, J. (1986) *A Primer on Determinsm*. Reidel.

Eells, E. (1991) *Probabilistic Causality*. Cambridge University Press.

Einstein, A., Podolsky, B. and Rosen, N. (1935) "Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?" *Physical Review*, **47**: 777–80.

Elby, A. (1992) "Should We Explain the EPR Correlations Causally?" *Philosophy of Science*, **56**: 16–25.

Elby, A. (1993) "Why Local Realistic Theories Violate, Nontrivially, the Quantum Mechanical EPR Perfect Correlations". *The British Journal for the Philosophy of Science*, **44**: 213–30.

Esfeld, M. (2001) “Lewis’ Causation and Quantum Correlations”. In W. Spohn, M. Ledwig and M. Esfeld (Eds.), *Current Issues in Causation*. Paderborn, pp. 175–89.

Fine, A. (1981) “Correlations and Physical Locality”. In P. Asquith and R. Giere (Eds.), *Proceedings of the 1980 Biennal Meeting of the Philosophy of Science Association,* vol. II. Lansing, pp. 535–56.

Fine, A. (1982*a*) “Hidden Variables, Joint Probability, and the Bell Inequalities”. *Physical Review Letters*, **48**: 291–5.

Fine, A. (1982*b*) “Joint Distributions, Quantum Correlations, and Commuting Observables”. *Journal of Mathematical Physics*, **23**: 1306–10.

Fine, A. (1982*c*) “Some Local Models for Correlation Experiments”. *Synthese*, **50**: 279–94.

Fine, A. (1989) “Correlations and Efficiency: Testing the Bell Inequalities”. *Foundations of Physics*, **19**: 453–78.

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

Fine, A. (2004) “The Einstein-Podolsky-Rosen Argument in Quantum Theory”. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy.* URL `<http://plato.stanford.edu/entries/qt-epr/>`.

Glymour, C., Spirtes, P. and Scheines, R. (1991) “Causal Inference”. *Erkenntnis*, **35**: 151–89.

Graßhoff, G., Portman, S. and Wüthrich, A. (2005) “Minimal Assumption Derivation of a Bell-type Inequality”. *The British Journal for the Philsophy of Science*, **56**: 663–80.

Gyenis, B. and Rédei, M. (2004) “When can Statistical Theories be Causally Closed?” *Foundations of Physics*, **34**: 1285–303.

Hausman, D. (1998) *Causal Asymmetries.* Cambridge University Press.

Hausman, D. (1999) “Lessons from Quantum Mechanics”. *Synthese*, **121**: 79–92.

Hausman, D. M. and Woodward, J. (1999) “Independence, Invariance and the Causal Markov Condition”. *The British Journal for the Philosophy of Science*, **50**: 521–83.

Healey, R. (1992*a*) "Causation, Robustness, and EPR". *Philosophy of Science*, **59**: 282–92.

Healey, R. (1992*b*) "Chasing Quantum Causes: How Wild is the Goose?" *Philosophical Topics*, **20**: 181–204.

Heisenberg, W. (1958) *Physics and Philosophy*. Harper & Row. Reprinted in Penguin Classics, 2000.

Hitchcock, C. (1993) "A Generalized Probabilistic Theory of Causal Relevance". *Synthese*, **97**: 335–64.

Hitchcock, C. (2007) "Probabilistic Causation". In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2007 Edition), forthcoming. URL `<http://plato.stanford.edu/archives/fall2007/entries/causation-probabilistic/>`.

Hoefer, C. (2004) "Causality and Determinism: Tension, or Outright Conflict?" *Revista de Filosofía*, **29**: 99–115.

Hofer-Szabó, G. (2007*a*) "Bell Inequalities and Common Cause Systems". In S. Aerts and C. de Ronde (Eds.), *Perspectives on Understanding Quantum Mechanics*, (forthcomming).

Hofer-Szabó, G. (2007*b*) "Separate- versus *Common*-common-cause-type derivations of the Bell inequalities". *Synthese* (forthcomming).

Hofer-Szabó, G., Rédei, M. and Szabó, L. E. (1999) "On Reichenbach's Common Cause Principle and Reichenbach's Notion of Common Cause". *The British Journal for the Philsophy of Science*, **50**: 377–99.

Hofer-Szabó, G., Rédei, M. and Szabó, L. E. (2000*a*) "Common Cause Completability of Classical and Quantum Probability Spaces". *International Journal of Theoretical Physics*, **39**: 913–9.

Hofer-Szabó, G., Rédei, M. and Szabó, L. E. (2000*b*) "Reichenbach's Common Cause Principle: Recent Results and Open Questions". *Reports on Philosophy*, **20**: 85–107.

Hofer-Szabó, G., Rédei, M. and Szabó, L. E. (2002) "Common Causes are not Common-common Causes". *Philosophy of Science*, **69**: 623–36.

Hoover, K. D. (2003) "Nonstataionary Time Series, Cointegration, and the Principle of the Common Cause". *The British Journal for the Philosophy of Science*, **54**: 527–51.

Howard, D. (1989) "Holism, Separability, and the Methaphisical Implications of the Bell Experiments". In J. Cushing and E. McMullin (Eds.), *Philosophical Consequences of Quantum Theory*. University of Notre Dame Press, pp. 224–53.

Jarrett, J. P. (1984) "On the Physical Significance of the Locality Conditions in the Bell Arguments". *Noûs*, **18**: 569–89.

Jarrett, J. P. (1989) "Bell's Theorem: A Guide to the Implications". In J. Cushing and E. McMullin (Eds.), *Philosophical Consequences of Quantum Theory*. University of Notre Dame Press, pp. 60–79.

Kochen, S. and Specker, E. P. (1967) "The Problem of Hidden Variables in Quantum Mechanics". *Journal of Mathematics and Mechanics*, **17**: 59–87.

Maudlin, T. (2002) *Quantum Non-Locality and Relativity* (2^{nd} ed). Blackwell Publishing.

Mermin, N. D. (1986) "Generalizations of Bell's Theorem to Higher Spins and Higher Correlations". In L. Roth and A. Inomato (Eds.), *Fundamental Questions in Quantum Mechanics*. Gordon and Breach, pp. 7–20.

Papineau, D. (1990) "Causes and Mixed Probabilities". *International Studies in the Philosophy of Science*, **4**: 79–88.

Portman, S. and Wüthrich, A. (2006) "Minimal Assumption Derivation of a Weak Clauser-Horne Inequality". Preprint. URL `http://www.arxiv.org/quant-ph/0604216`.

Price, H. (1994) "A Neglected Route to Realism About Quantum Mechanics". *Mind*, **103**: 303–36.

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

Rédei, M. (2002) "Reichenbach's Common Cause Principle and Quantum Correlations". In T. Placek and J. Butterfield (Eds.), *Non-locality, Modality and Bell's Theorem*. Kluwer, pp. 259–270.

Rédei, M. and Summers, S. J. (2007) "Quantum Probability Theory". *Studies in the History and Philosophy of Modern Physics*, **38**: 390–417.

Redhead, M. (1987) *Incompleteness, Nonlocality and Realism*. Oxford Calrendon Press.

Redhead, M. (1989*a*) "The Nature of Reality". *British Journal for the Philosophy of Science*, **40**: 429–41.

Redhead, M. L. G. (1989*b*) "Nonfactorizability, Stochastic Causality and Passion-at-a-distance". In J. Cushing and E. McMullin (Eds.), *Philosophical Consequences of Quantum Theory.* University of Notre Dame Press, pp. 145–53.

Reichenbach, H. (1956) *The Direction of Time*; Edited by Maria Reichenbach. Unabridged Dover, 1999 (republication).

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

San Pedro, I. (2004) "Reichenbach's Common Cause Principle and Quantum Correlations". Research Report, University of the Basque Country, UPV-EHU. Unpublished.

Shimony, A. (1984*a*) "Contextual Hidden Variables Theories and Bell's Inequalities". *British Journal for the Philosophy of Science*, **35**: 25–45.

Shimony, A. (1984*b*) "Controllable and Uncontrollable Non-Locality". In Kamefuchi *et al.* (Ed.), *Proceedings of the International Symposium: Foundations of Quantum Mechanics in the Light of New Technology.* Physical Society of Japan, pp. 225–30.

Shimony, A. (1986) "Events and Processes in the Quantum World". In R. Penrose and C. Isham (Eds.), *Quantum Concepts in Space and Time.* Oxford University Press, pp. 182–203. Reprinted in Shimony (1993), pp. 140–62.

Shimony, A. (1993) *Search for a Naturalistic World View, Volume II: Natural Science and Metaphysics.* Cambridge University Press.

Skyrms, B. (1984) "EPR: Lessons for Metaphysics". In P. A. French and T. E. Uehling, Jr (Eds.), *Causation and Causal Theories. Midwest Studies in Philosophy,* **9**. University of Minnesota Press, pp. 245–55.

Sober, E. (1985) "Two Conepts of Cause". In *Proceedings of the 1984 Bienal Meeting of the Philosophy of Science Association,* vol. II. pp. 405–24.

Sober, E. (1987) "The Principle of the Common Cause". In J. H. Fetzer (Ed.), *Probability and Causality: Essays in Honor of Wesley Salmon.* Reidel, pp. 211–28.

Sober, E. (2001) “Venetian Sea Levels, British Bread Prices, and the Principle of the Common Cause”. *The British Journal for the Philosophy of Science*, **52**: 331–46.

Spohn, W. (1990) “Direct and Indirect Causes”. *Topoi*, **9**: 125–45.

Spohn, W. (1994) “On Reichenbach’s Principle of the Common Cause”. In W. Salmon and G. Wolters (Eds.), *Logic, Language, and the Structure of Scientific Theories*. Pittsburgh University Press, pp. 215–39.

Steel, D. (2003) “Making Time Stand Still: A Response to Sober’s Counter-Example to the Principle of the Common Cause”. *The British Journal for the Philosophy of Science*, **54**: 309–17.

Steel, D. (2005) “Indeterminism and the Causal Markov Condition”. *British Journal for the Philosophy of Science*, **56**: 3–26.

Suárez, M. (1997) *Models of the World, Data Models and the Practice of Science: the Semantics of Quantum Theory*. Ph.D. thesis, University of London, LSE.

Suárez, M. (2000) “The Many Faces of Non-Locality: Dickson on the Quantum Correlations”. *British Journal for the Philosophy of Science*, **51**: 882–92.

Suárez, M. (2007) “Causal Inference in Quantum Mechanics: A Reassessment”. In F. Russo and J. Williamson (Eds.), *Causality and Probability in the Sciences*. London College, pp. 65–106.

Suárez, M. and San Pedro, I. (2007) “EPR, Robustness and the Causal Markov Condition”. Discussion Paper 04/07, LSE Centre for the Philosophy of the Natural and Social Sciences.

Suppes, P. and Zanotti, M. (1976) “On the Determinism of Hidden Variable Theories with Strict Correlation and Conditional Statistical Independence of Observables”. In P. Suppes (Ed.), *Logic and Probability in Quantum Mechanics*. Reidel, pp. 445–55.

Suppes, P. and Zanotti, M. (1981) “When are probabilistic explanations possible?” *Synthese*, **48**: 191–9.

Szabó, L. E. (1995) “Is Quantum Mechanics Compatible with a Deterministic Universe? Two Interpretations of Quantum Probability”. *Foundations of Physics Letters*, **8**: 421–40.

Szabó, L. E. (2000) "On an Attempt to Resolve the EPR-Bell Paradox via Reichenbachian Concept of Common Cause". *International Journal of Theoretical Physics*, **39**: 911–26.

Szabó, L. E. and Fine, A. (2002) "A Local Hidden Variable Theory for the GHZ Experiment". *Physics Letters A*, **295**: 229–40.

Teller, P. (1986) "Relational Holism and Quantum Mechanics". *The British Journal for the Philosophy of Science*, **37**: 71–81.

Teller, P. (1989) "Relativity, Relational Holism, and the Bell Inequalities". In J. Cushing and E. McMullin (Eds.), *Philosophical Consequences of Quantum Theory*. University of Notre Dame Press, pp. 208–23.

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

van Fraassen, B. C. (1982*a*) "The Charybdis of Realism: Epistemological Implications of Bell's Inequality". *Synthese*, **52**: 25–38. Reprinted with corrections in Cushing and McMullin (1989), pp. 97–113.

van Fraassen, B. C. (1982*b*) "Rational Belief and the Common Cause Principle". In R. McLaughlin (Ed.), *What? Where? When? Why? Essays in Honour of Wesley Salmon*. Reidel, pp. 193–209.

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

von Neumann, J. (1932) *Mathematische Grundlagen der Quantenmechanik*. Springer. English translation, *Mathematical Foundations of Quantum Mechanics*, Princeton University Press, 1955.

Wessels, L. (1985) "Locality, Factorability and the Bell Inequalities". *Noûs*, **19**: 481–519.

Wetzel, L. (2007) "Types and Tokens". In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2005 Edition). URL `<http://plato.stanford.edu/archives/win2007/entries/types-tokens/>`.

Wigner, E. P. (1970) "On Hidden Variables and Quantum Mechanical Probabilities". *American Journal of Physics*, **38**: 1005–9.

Williamson, J. (2005) *Bayesian Nets and Causality: Philosophical and Computational Foundations*. Oxford University Press.

Yule, G. U. (1926) "Why do We Sometimes get Nonsensical Relations Between Time Series? A Study of Sampling and the Nature of Time Series". *Journal of the Royal Statistical Society*, **89**: 1–64.

end