MASTERARBEIT | MASTER'S THESIS

Titel | Title

On the Conflict between Nonlocality and Relativity

verfasst von | submitted by

Felix Rutzinger BSc

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien | Vienna, 2026

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 066 876
Studienrichtung lt. Studienblatt | Degree programme as it appears on the student record sheet:	Masterstudium Physics
Betreut von | Supervisor:	Mag. Dr. Beatrix Hiesmayr Privatdoz.

φύσις κρύπτεσθαι φιλεῖ

Dedication

Für meine Eltern, ohne deren beständige Unterstützung diese Arbeit nicht möglich gewesen wäre.

Acknowledgments

I would like to thank my supervisor, Beatrix C. Hiesmayr, for her willingness to supervise this project and her valuable guidance throughout the writing of this thesis. I am grateful to Flavio Del Santo, Sebastian Horvath, and Tobias C. Sutter for their insightful comments, and to Benjamin Kesetovic for his careful feedback on various drafts of this thesis. I am indebted to my friends and family, especially my parents Sabine and Bernd Rutzinger, my sister Lina Rutzinger, and especially my partner Nina Niedermair, for their unwavering support. Finally, I would like to extend my compliments to the great chef.

Abstract

Bell's theorem demonstrates that realist approaches to quantum theory, under reasonable assumptions, must violate locality. But causal influences at spacelike separation appear to conflict with special relativity. While relativistic objective collapse models have demonstrated that nonlocality can be made compatible with a serious notion of relativity, the exact nature and resolution of this tension remains unclear. This thesis clarifies this issue by generalizing an argument due to Nicolas Gisin using causal models, thereby establishing a no-go theorem that elucidates the conflict between nonlocality and relativity. Building on this result, the possibility of nonlocal and relativistic causal explanations of the Bell scenario is investigated. In particular, a uniqueness theorem is proven that, under natural assumptions, singles out a unique causal explanation of the Bell scenario. This causal explanation is shown to be compatible with the account provided by relativistic collapse models, such as rGRWf. These results suggest a broader view of quantum phenomena in which nonlocality and relativity can be reconciled. This reconciliation is achieved by accepting genuine indeterminism and preserving the Lorentz invariance of causal connections while allowing the causal direction of non-local influences to be frame dependent, consistent with the assumption of no retrocausality.

Zusammenfassung

Das Bellsche Theorem zeigt, dass realistische Deutungen der Quantentheorie unter plausiblen Annahmen nichtlokal sein müssen. Kausale Einflüsse zwischen raumartig getrennten Ereignissen scheinen jedoch im Widerspruch zur speziellen Relativitätstheorie zu stehen. Während relativistische objektive Kollapsmodelle gezeigt haben, dass Nichtlokalität mit einer physikalisch sinnvollen Auffassung von Relativität prinzipiell vereinbar ist, bleibt die genaue Natur dieses Konflikts unklar. Die vorliegende Arbeit klärt dieses Problem, indem ein Argument von Nicolas Gisin mithilfe kausaler Modelle verallgemeinert wird. Das so gewonnene No-Go-Theorem präzisiert den Konflikt zwischen Nichtlokalität und relativistischen Prinzipien. Aufbauend auf diesem Resultat wird die Möglichkeit nichtlokaler und relativistischer kausaler Erklärungen des Bell-Szenarios untersucht. Insbesondere wird ein Eindeutigkeitssatz bewiesen, der unter natürlichen Annahmen nur eine einzige Erklärung des Bell-Szenarios als möglich identifiziert. Es zeigt sich, dass diese Erklärung mit der von relativistischen Kollapsmodellen, wie etwa rGRWf, gelieferten Beschreibung übereinstimmt. Diese Ergebnisse legen eine Sicht auf Quantenphänomene nahe, in der Nichtlokalität und Relativität miteinander vereinbar sind. Sie setzt objektiven Indeterminismus voraus, bewahrt die Lorentz-Invarianz kausaler Verbindungen, aber macht die Richtung nichtlokaler Einflüsse vom Bezugssystem abhängig, wodurch retrokausale Verursachung vermieden wird.

Contents

List of Figures

Chapter 1

Introduction

While the apparent incompatibility between quantum theory and general relativity is widely recognized as one of the central problems of contemporary theoretical physics [1, 2], quantum theory and special relativity are usually regarded as compatible, as evidenced by the existence of relativistic quantum field theories.

However, this compatibility largely rests upon a dismissal of realist approaches to quantum theory. If one accepts realism and some reasonable auxiliary assumptions, then Bell's theorem implies nonlocality, i.e., the existence of causal connections between spacelike separated events [3]. Nonlocality conflicts with any understanding of relativity which restricts the causal influence of an event to its future light cone. This problem was recognized early on by Bell himself:

> So one of my missions in life is to get people to see that if they want to talk about the problems of quantum mechanics — the real problems of quantum mechanics — they must be talking about Lorentz invariance. [4]

The tension is further reflected in the development of realist interpretations of quantum theory, which have in many cases proven difficult to convincingly extend to the relativistic case [5]. A prominent example is Bohmian mechanics, whose known relativistic formulations require a preferred foliation of spacetime, a feature that is often regarded as problematic since it appears to conflict with the spirit of relativity [6, 7]. So much the worse for realist approaches to quantum theory, one might argue.

In this thesis, we take a different approach. This is motivated by the existence of realist, nonlocal, and relativistic objective collapse models. Arguably the first such model is rGRWf, a relativistic extension of the GRW model with flash ontology, developed by Roderich Tumulka in 2004 [8]. The existence of these and similar models shows that nonlocality and a serious notion of relativity are not contradictory concepts. The question of the exact nature of the conflict remains. The aim of this thesis is to answer this question by formulating and proving a rigorous no-go theorem that articulates this tension. Building on this result, we analyze the possibility of nonlocal and relativistic explanations of the Bell scenario.

The structure of this investigation is informed by an analogy to the discovery of Bell's theorem. After the advent of quantum theory, it was argued that a realist explanation of quantum phenomena is impossible. This position was held by several of the founders of quantum theory, for example Bohr [9] and Heisenberg [10]. In his famous 1932 book, von Neumann presented an argument that was interpreted by some as offering a rigorous proof of the impossibility of hidden-variable extensions of quantum theory [11][1].

Surprisingly, in two papers published in 1952 [12, 13], Bohm managed to formulate a theory, now known as Bohmian mechanics, that is both realist and empirically equivalent to standard quantum theory. Bohm thereby showed that realism can be made compatible with an empirically adequate description of quantum phenomena. A remarkable aspect of Bohmian mechanics, however, is its grossly nonlocal character. This motivated Bell to search for a similar theory that reproduced the predictions of quantum theory without violating locality [14]:

> And so the question immediately posed itself: Is that inevitable? Can you find another way of refuting von Neumann that does not have this feature of nonlocality? [4]

In 1964, Bell realized that such a theory is impossible under certain reasonable assumptions [3]. Consequently, any reasonable realist theory that reproduces the predictions of quantum theory must be nonlocal. This result is now widely known as Bell's theorem.

Analogously, it has been argued that nonlocality is incompatible with any serious notion of relativity [15]. However, as mentioned above, in 2004 Tumulka was able to formulate rGRWf, which is both nonlocal and genuinely relativistic [8], thereby establishing that nonlocality can be made compatible with a serious notion of relativity. Following the proposed analogy, it is natural to ask whether there is some property that rGRWf and, more generally, any theory that is both nonlocal and relativistic, must exhibit in virtue of this fact. We will argue that there is indeed such a property and hence a no-go result analogous to Bell's theorem. The property in question is shown

[1]For an in-depth discussion of von Neumann's impossibility proof, see Appendix B.

to be the frame dependence of causal structure, and the corresponding no-go theorem generalizes a 2011 result due to Nicolas Gisin [16]. A theory exhibits frame dependence of causal structure if the causal relations between events can differ depending on the choice of Lorentz frame. It should be noted that the frame dependence of causal structure, and in particular the frame dependence of causal direction in rGRWf, have been discussed several times in the recent literature surrounding rGRWf [17, 18, 19].

This analogy between the history of Bell's theorem and the conflict between nonlocality and relativity is represented by the diagram below:

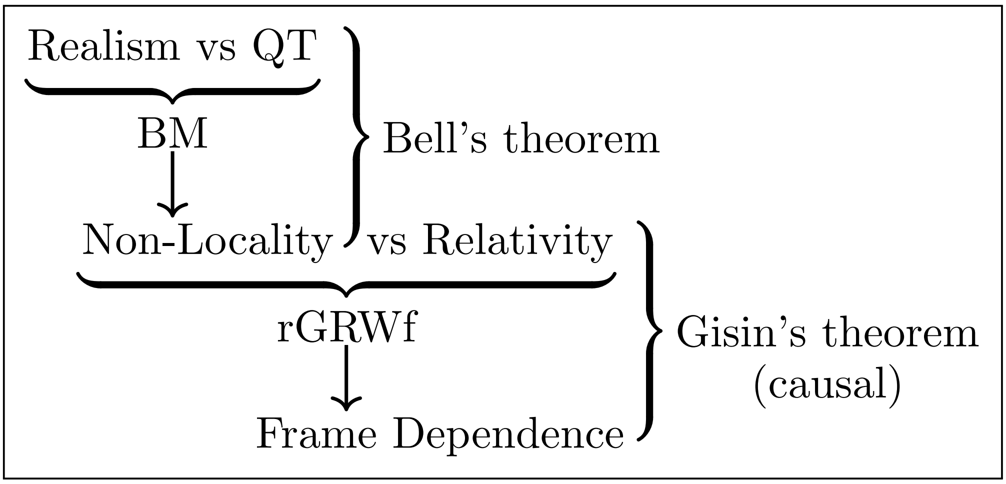


Figure 1.1: An analogy between Bell's and Gisin's theorems.

This diagram visualizes the parallel between Bell's and Gisin's theorems and the fact that Bell's result naturally feeds into Gisin's through the role which nonlocality plays in both.

In this thesis, we establish the following main results: a rigorous formulation of the EPRB argument for n parties using ontological models; a derivation of Bell's theorem from purely causal principles; a generalization of Gisin's argument beyond determinism using causal models; and a uniqueness theorem for nonlocal and relativistic causal explanations of the Bell scenario.

Guided by the analogy proposed above, the thesis is structured as follows: Chapter 2 surveys the necessary technical and conceptual tools. These include elements of probability theory, ontological models, and causal models. The Bell scenario is introduced, and several relevant heuristic principles are discussed.

Chapter 3 contains a detailed analysis of Bell's theorem. This includes an analysis of the principle of local causality in both the ontological and causal models frameworks and a rigorous reformulation of the EPRB argument. A short summary of the well-known proof of Bell's theorem via ontological models is given and contrasted with a derivation of Bell's theorem using the causal models framework. This is followed by a discussion of the implications of, and responses to, Bell's theorem partly in light of a recent no-go result based on the Wigner's friend scenario.

Chapter 4 is a study of Gisin's argument. Gisin's argument is reconstructed and its assumptions clarified. Past attempts to generalize Gisin's argument beyond the deterministic case, using dilation or ontological models, are found unsatisfactory. Using causal models, Gisin's theorem is formulated for general, including genuinely stochastic, theories. Two proofs are presented, one paralleling the causal derivation of Bell's theorem and the other more conceptual in nature.

Chapter 5 concerns nonlocal and relativistic explanations of the Bell scenario. We advocate a response to Gisin's theorem that accepts the frame dependence of causal structure while retaining the Lorentz invariance of causal connections. This leads to a uniqueness theorem for causal explanations of the Bell scenario. After introducing the rGRWf model explicitly, we show that the unique causal explanation of the Bell scenario is in agreement with the account offered by rGRWf.

The thesis concludes with a summary of the main results and a discussion of their implications for realist approaches to quantum theory. Directions for future research are outlined and related to existing programs in the literature.

Chapter 2

Preliminaries

The aim of this chapter is to introduce the technical and conceptual tools required for the remainder of the thesis. The first section provides the necessary formal background. After fixing notation and recalling some basic elements of probability theory, we introduce two general formal frameworks that can be used to explain operational statistics, namely ontological models and causal models. In the case of causal models, we discuss in some detail the way in which device-independent arguments can be formulated using this framework. We then characterize the Bell scenario, a type of experimental set-up, central to the discussion in later chapters. The second part of this chapter is concerned with heuristic principles that will prove useful in the discussion of the results presented later in the thesis. The heuristics in question are logical conservatism and the ontic symmetry principle. The latter will be compared with the Leibnizian methodological principle and the related notion of faithfulness in causal inference. We will further comment on the connection between the ontic symmetry principle and the special theory of relativity.

2.1 Formal Background

2.1.1 Elements of Probability Theory

Random variables are denoted by uppercase letters (e.g. X), their value sets by boldface symbols (e.g. $\mathbf{X}$), and specific values by lowercase letters (e.g. $x \in \mathbf{X}$). Probability measures are denoted by P, while probabilities are written as $p(x) = P(X = x)$.

Conditional statistical independence of two random variables, X and Y, with respect to a third variable Z is written as $(X \perp\!\!\!\perp Y \mid Z)$ and defined by

$$(X \perp\!\!\!\perp Y \mid Z) \qquad \Longleftrightarrow \qquad p(x, y \mid z) = p(x \mid z)\, p(y \mid z) \qquad \forall x, y, z\,, \tag{2.1}$$

provided that $p(z) > 0$. Whenever the relevant conditional probabilities are well-defined, this is equivalent to

$$p(x \mid y, z) = p(x \mid y', z) \qquad \forall x, y, y', z\,, \tag{2.2}$$

which we often abbreviate as

$$p(x \mid y, z) = p(x \mid z)\,. \tag{2.3}$$

The independence relation is symmetric, meaning that

$$(X \perp\!\!\!\perp Y \mid Z) \qquad \Longleftrightarrow \qquad (Y \perp\!\!\!\perp X \mid Z)\,. \tag{2.4}$$

This property will feature prominently in our discussion of Gisin's argument in Chapter 4.

2.1.2 Ontological Models

We briefly recall the framework of ontological models. For a more in-depth study of ontological models see, for example, [20, 21].

Definition 2.1 (Ontological Model)**.** An *ontological model* of the operational statistics $p(a_1, \ldots, a_n \mid x_1, \ldots, x_m)$ with outcomes $A_1, \ldots, A_n$ and settings $X_1, \ldots, X_m$ is specified by:

(i) A measurable set $\mathbf{\Lambda}$, called the ontic state space. Its elements $\lambda \in \mathbf{\Lambda}$ are called ontic states;

(ii) A conditional probability distribution over the ontic state given the settings $p(\lambda \mid x_1, \ldots, x_m)$;

(iii) A response function $p(a_1, \ldots, a_n \mid x_1, \ldots, x_m, \lambda)$.

Using an ontological model the operational statistics can be expressed via the law of total probability as follows

$$p(a_1, \ldots, a_n \mid x_1, \ldots, x_m) = \int_{\boldsymbol{\Lambda}} \mathrm{d}\lambda \, p(a_1, \ldots, a_n \mid x_1, \ldots, x_m, \lambda) \, p(\lambda \mid x_1, \ldots, x_m) \, . \tag{2.5}$$

The ontic state λ serves as a complete description of the state of the underlying physical system that gives rise to the operational statistics $p(a_1, \ldots, a_n \mid x_1, \ldots, x_m)$.

2.1.3 Causal Models

The framework of causal models aims to formalize causal reasoning. Here, we will mainly follow the exposition in [22]. The notion of causal structure is defined as follows:

Definition 2.2 (Causal Structure)**.** A *causal structure* is given by:

(i) A finite set of random variables $\{X_1, \ldots, X_n\}$;

(ii) A directed acyclic graph (DAG) whose nodes are in one-to-one correspondence with the random variables $X_1, \ldots, X_n$.

For a node X in a causal structure, any node with an arrow pointing into X is called a *parent* of X. The set of parents of X is denoted by $\mathrm{Pa}(X)$. Any node that is reachable from X by a directed path is a *descendant* of X. The set of descendants of X is denoted by $\mathrm{De}(X)$, with $X \notin \mathrm{De}(X)$ by definition. The set of non-descendants of X is denoted by $\mathrm{Nd}(X)$, and similarly $X \notin \mathrm{Nd}(X)$. Using these conventions of kinship between nodes, we define a causal model as follows:

Definition 2.3 (Causal Model)**.** A *causal model* consists of a causal structure with random variables $X_1, \ldots, X_n$ such that:

(i) For each random variable X, there exists a conditional probability distribution $P(X \mid \mathrm{Pa}(X))$.

(ii) The joint distribution over all random variables factorizes according to the *Markov condition*

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid \mathrm{Pa}(X_i)) \,. \tag{2.6}$$

The Markov condition implies that every variable X is conditionally independent of its non-descendants given its parents, i.e.,

$$(X \perp\!\!\!\perp \mathrm{Nd}(X) \mid \mathrm{Pa}(X)) \,. \tag{2.7}$$

In what follows, we will use the causal models framework to construct device-independent arguments via the argument schema given below:

(i) Start with some operational statistics $p(a_1, \ldots, a_n \mid x_1, \ldots, x_m)$.

(ii) Consider the set S of all causal structures which contain $A_1, \ldots, A_n, X_1, \ldots, X_m$.

(iii) Assume causal principles $\phi_1, \ldots, \phi_m$ which constrain the set of causal structures to $S' \subsetneq S$.

(iv) Show that no structure in S' can account for the operational statistics $p(a_1, \ldots, a_n \mid x_1, \ldots, x_m)$.

(v) Conclude that the causal principles $\phi_1, \ldots, \phi_m$ cannot jointly hold.

To express families of causal structures diagrammatically, we use a graded notation that indicates different degrees of causal relation between nodes. We will use the following conventions to represent the possible causal relations between nodes:

(i) *No edge* between X and Y indicates that there is no direct causal relation between X and Y.

Figure 2.1: X and Y have no direct causal relation.

(ii) A *thick directed arrow* from X to Y indicates that X is a cause of Y. This rules out that Y is a cause of X and that X and Y are causally independent as in (i).

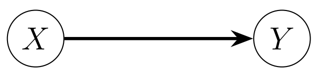


Figure 2.2: X is a cause of Y.

(iii) A *thin directed arrow* from X to Y indicates that X may be a cause of Y, or that there may be no direct causal relation at all between them. What is ruled out is that Y is a cause of X, i.e., a thick arrow from Y to X.

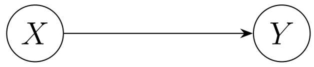


Figure 2.3: X may be the cause of Y.

(iv) An *undirected thin line* indicates that the causal relation between X and Y is unspecified. It is compatible with (i)-(iii) in both directions.

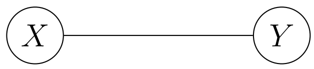


Figure 2.4: X and Y have an unspecified causal relation.

A diagram containing undirected or thin directed edges denotes the set of all DAGs obtained by replacing each undirected or thin directed edge by any of the permitted, definite causal relations. Hence, we often use a single diagram to represent a large set of actual causal structures compatible with it. We will sometimes refer to such diagrams as *generalized causal structures*. But even generalized causal structures always represent families of strictly acyclic causal structures.

2.1.4 The Bell Scenario

In the following, we will often consider the standard *Bell scenario*, defined as follows: Two parties, Alice and Bob, control the settings X and Y, with $\mathbf{X} = \mathbf{Y} = \{0, 1\}$, and observe the outcomes A and B, with $\mathbf{A} = \mathbf{B} = \{-1, 1\}$. They repeat their experiment for many runs, extracting the operational statistics $p(a, b \mid x, y)$.

The spacetime regions in which the values of a random variable are determined are denoted by calligraphic uppercase letters. The past and future light cones of a spacetime region $\mathcal{R}$ are denoted by $J^-(\mathcal{R})$ and $J^+(\mathcal{R})$, respectively.

The experiment is located in spacetime as follows: $\mathcal{A} \subseteq J^+(\mathcal{X})$ and $\mathcal{B} \subseteq J^+(\mathcal{Y})$. Moreover, $\mathcal{A}$ and $\mathcal{X}$ are spacelike separated from $\mathcal{B}$ and $\mathcal{Y}$. A schematic spacetime diagram of the Bell scenario is given in Figure 2.5:

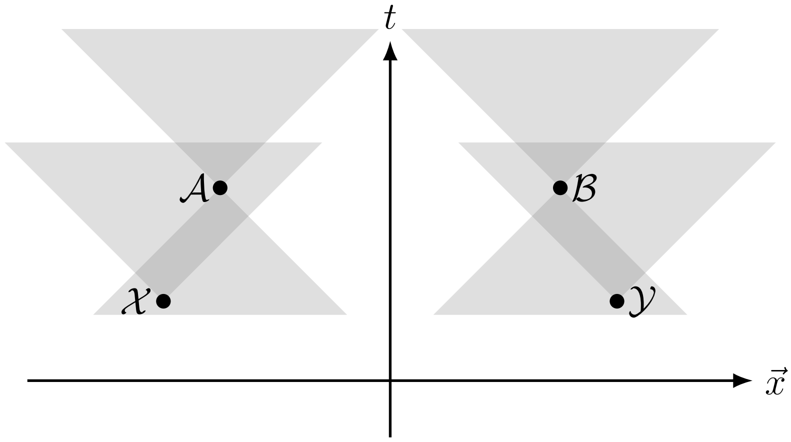


Figure 2.5: Spacetime diagram of the Bell scenario.

Considering only the operationally accessible random variables, the set of possible causal structures underlying the Bell scenario can be represented by the following generalized causal structure:

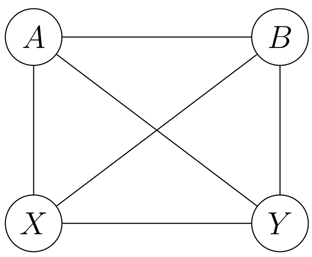


Figure 2.6: The most permissive generalized causal structure compatible with the Bell scenario with no latent variables.

To arrive at the most general causal explanation of the Bell scenario, we consider a latent variable Λ. Note that Λ is entirely unrestricted with regard to its possible values $\lambda \in \mathbf{\Lambda}$ or to its localization in spacetime. Since multiple latent variables can always be combined into a single composite latent variable, it is sufficient to assume only one latent cause. This leads to the following generalized causal structure:

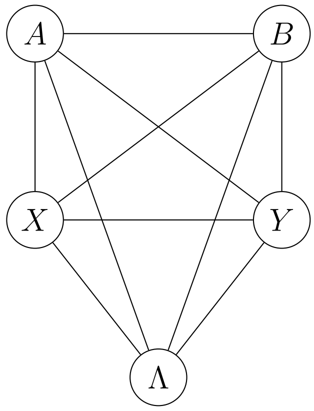


Figure 2.7: The most permissive generalized causal structure compatible with the Bell scenario.

2.2 Heuristic Principles

2.2.1 Logical Conservatism

We will often make use of heuristic principles to assess the plausibility of different responses to no-go theorems. Results such as Bell's theorem usually rule out certain classes of explanations of the operational statistics but do not give a positive account of the actual explanation, which creates a need for heuristics.

Consider Bell's theorem formulated within the ontological models framework. It shows that any ontological model satisfying both statistical independence and locality is empirically inadequate. However, the theorem does not tell us whether statistical independence, locality, or both are violated in nature. A heuristic, called *logical conservatism* (LC), offers the following guidance in situations like these:

Definition 2.4 (Logical Conservatism)**.** If a set of plausible physical principles $\phi_1, \ldots, \phi_n$ is jointly empirically inadequate, then, all else being equal, one should abandon as few of these principles as possible.

The motivation is straightforward. If all of $\phi_1, \ldots, \phi_n$ are plausible, giving up any one of them is associated with a dialectical cost. To minimize this cost we will, if there are no additional reasons not to do so, give up as few principles as possible in order to avoid inconsistency. Hence, in the case of Bell's theorem mentioned above, LC disfavors giving up both statistical independence and locality.

2.2.2 The Leibnizian Methodological Principle

The *Leibnizian methodological principle* (LMP) is a reformulation, due to Spekkens [23], of the metaphysical principle of the identity of indiscernibles, prominently espoused by Gottfried Wilhelm Leibniz [24], as a heuristic for theory choice. Spekkens offers the following formulation:

> If an ontological theory implies the existence of two scenarios that are empirically indistinguishable in principle but ontologically distinct (where both the indistinguishability and distinctness are evaluated by the lights of the theory in question), then the ontological theory should be rejected and replaced with one relative to which the two scenarios are ontologically identical. [23]

Spekkens argues that the LMP played an important role in the history of physics, especially through the work of Einstein who 'applied it repeatedly to great effect [...] in his development of the special and general theories of relativity' [23]. Especially relevant for our present purposes is that the LMP rules out the existence of an unobservable preferred foliation of spacetime since it would render operationally indistinguishable scenarios ontologically distinct in virtue of their relation to the preferred foliation.

In the causal modeling framework, the LMP implies a condition known as *faithfulness* or *no fine-tuning* [25, 26].

Definition 2.5 (Faithfulness)**.** Every conditional independence relation in the joint probability distribution of the causal model is implied by the causal structure via the Markov condition.

The connection to the LMP becomes clear when we consider violations of faithfulness. If faithfulness fails, conditional independence relations among the random variables do not only arise from the causal structure but also from a fine-tuned choice of model parameters. Hence, it is possible for empirically indistinguishable scenarios to correspond to different underlying causal structures, making them ontologically distinct and therefore contradicting the LMP.

Besides its historical pedigree, the LMP can be justified as an instance of inference to the best explanation. Spekkens argues that '[t]he best explanation of the fact that two scenarios are in-principle observationally indiscernible is that they are associated to the same physical state of affairs'.

It has been shown that every classical causal explanation of the Bell scenario that accounts for the observed statistics, such as nonlocal or superdeterministic explanations, must violate faithfulness [26]. This has partly motivated the development of quantum causal models as they are able to provide a causal explanation of the Bell scenario that is not fine-tuned [25, 27]. By generalizing the notion of causal models one can hold on to faithfulness and thereby to the LMP.

Despite its appeal, we will in the following reject the LMP. The reason for this is twofold. Firstly, the LMP rules out nonlocal explanations of the Bell scenario, which would make our project a non-starter. Secondly, recently published arguments suggest that in the extended Wigner's friend scenario (EWFS), a generalization of the Bell scenario, there is, under reasonable assumptions, no way to save faithfulness even by adopting non-classical causal models [28]. If one accepts this conclusion, then the LMP must be rejected, since the LMP implies faithfulness. We will return to these issues in our discussion of the possible responses to Bell's theorem in Chapter 3.

2.2.3 The Ontic Symmetry Principle

We propose a related heuristic, called the *ontic symmetry principle* (OSP), that is in many respects similar to the LMP without implying faithfulness. The OSP is also a principle of theory choice. It prefers theories in which observationally symmetric situations are explained in an ontologically symmetric way.

Definition 2.6 (Ontic Symmetry Principle)**.** Physical theories in which operational symmetries are, all else being equal, preserved on the ontic level are preferable to physical theories in which this is not the case.

Although not as restrictive as the LMP, the OSP can still help with evaluating different causal explanations. To illustrate this, consider the following causal explanation of the Bell scenario:

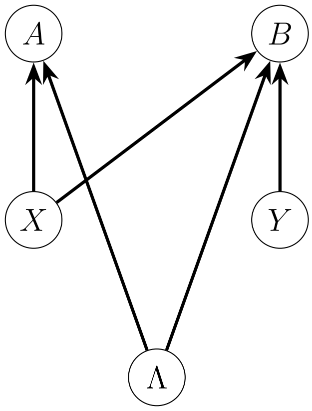


Figure 2.8: An asymmetric, nonlocal explanation of the Bell scenario.

Any theory which employs this causal structure to explain the operational statistics in the Bell scenario is disfavored by the OSP. This is because the operational statistics $p(a, b \mid x, y)$, as predicted by quantum theory, are symmetric under a swap of parties, while this causal structure is not. The causal structure given below is not open to the same objection:

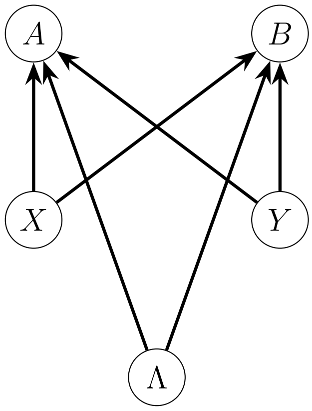


Figure 2.9: A symmetric, nonlocal explanation of the Bell scenario.

Here, the causal explanation is symmetric under exchange of parties and hence compatible with the OSP. The OSP does not imply faithfulness as it does not prohibit nonlocal explanations. It merely enforces that the nonlocal causal structure exhibits the same symmetries as the operational statistics it gives rise to.

The OSP can be motivated in ways analogous to the LMP. It has been argued that the OSP plays a role in the history of physics similar to that ascribed to the LMP by Spekkens. For example, Zahar attributes the following heuristic principle to Einstein:

> All observationally revealed symmetries signify fundamental symmetries at the ontological level. Hence the heuristic rule: replace any theory which does not explain symmetrical observational situations as the manifestation of deeper symmetries [. . .]. [29]

Most relevant for our purposes is the connection between the OSP and the principle of relativity, which implies that no Lorentz frame is physically preferred. Since no experiment has revealed evidence of a preferred frame, it is plausible that all operational statistics are symmetric under a change of Lorentz frame. The OSP then tells us that theories which preserve this operational symmetry on an ontic level should be generally preferred over those which do not preserve this symmetry. If the underlying ontological theory were to pick out a preferred Lorentz frame, this would violate the operational symmetry and would thus be disfavored by the OSP. Hence the OSP in conjunction with the empirical symmetry of Lorentz frames favors theories which respect the principle of relativity.

The OSP can also be viewed as an inference to the simplest explanation. The simplest explanation of a symmetry on the operational level is, everything else being equal, that the underlying physical reality exhibits the same symmetry. If this were not the case, one would have to devise some explanation for why the ontic asymmetry is unobservable on the operational level. Beyond its relation to special relativity, the OSP will play a role in the context of our analysis of possible causal explanations of the Bell scenario in Chapter 5.

Chapter 3

Bell's Theorem

In this chapter, we offer a detailed analysis of Bell's theorem. We begin by introducing the principle of local causality through historical examples and formulate it within the frameworks of ontological and causal models. This leads to a brief discussion of the EPRB argument. We show that, within the ontological models framework, probabilistic local causality entails locality. Locality, together with statistical independence, implies Bell locality. We then derive Bell locality from purely causal principles: weak measurement independence, local causality, and no-backward causality are shown to jointly imply Bell locality. We establish that Bell-local probability distributions must satisfy the CHSH inequality, and that quantum mechanics and experiments violate the CHSH inequality up to the Tsirelson bound. These results yield two formulations of Bell's theorem, one within the ontological models framework and one via the causal models framework. The latter will be most relevant for our purposes. Finally, we examine how one should respond to Bell's theorem. We briefly survey the main options. In this context, the Leibnizian methodological principle, as well as the related notion of faithfulness in causal models, play important roles. These considerations set the stage for treating nonlocality in the remainder of the thesis as a fact of nature.

3.1 The Principle of Local Causality

In the history of physics, it has been commonplace to argue that all physical phenomena must respect some sort of locality condition, which we will refer to here as the *principle of local causality.* A famous example of this view is provided by Newton in the context of his own theory of universal gravitation:

> It is inconceivable that inanimate brute matter should, without the mediation of something else which is not material, operate upon and affect other matter, without mutual contact [. . .]. [S]o that one body may act upon another at a distance, through a vacuum, without the mediation of anything else, by and through which their action and force may be conveyed from one to another, is to me so great an absurdity, that I believe no man who has in philosophical matters a competent faculty of thinking can ever fall into it. [30]

In his article 'Quanten-Mechanik und Wirklichkeit', Einstein argues that one of the main flaws of quantum mechanics is that it seems to violate the 'Prinzip der Nahewirkung':

> Für die relative Unabhängigkeit räumlich distanter Dinge (A und B) ist die Idee characteristisch: äussere Beeinflussung von A hat keinen unmittelbaren Einfluss auf B; dies ist als 'Prinzip der Nahewirkung' bekannt [. . .]. Völlige Aufhebung dieses Grundsatzes würde die Idee von der Existenz (quasi-)abgeschlossener Systeme und damit die Aufstellung empirisch prüfbarer Gesetze in dem uns geläufigen Sinne unmöglich machen. [31]

For Newton, local causality is simply obvious, while Einstein argues that it is vital for the project of natural science[1]. Furthermore, it has been argued that local causality is a direct consequence of special relativity [15]; a point which we dispute and revisit in the following chapters.

The historical examples of Newton and Einstein show that local causality has long served as a guiding notion in the development of physical theory. We now seek to give

[1]What is telling about this quote is that Einstein speaks of a '[v]öllige Aufhebung dieses Grundsatzes' as in conflict with the possibility of experimental science. This leaves open the possibility that a partial suspension of local causality may be compatible with the reliability of experiment. If the nonlocal influences were to be negligible in most experimental situations, then their existence would not necessarily undermine the project of experimental science as a whole. Indeed, if there are nonlocal influences, as for example in Bohmian mechanics, they are very subtle and practically irrelevant for most situations.

a precise formulation of this principle. Here, we will largely follow Bell's exposition given in [32]. Bell initially states the physical idea of local causality as follows:

> The direct causes (and effects) of events are near by, and even the indirect causes (and effects) are no further away than permitted by the velocity of light. [32]

But this formulation is not yet sharp enough to impose meaningful restrictions on the possible statistics in operational situations. Bell therefore provides the following more helpful version of the principle, which we refer to as *probabilistic local causality*[2] [32]:

Definition 3.1 (Probabilistic Local Causality)**.** For all spacetime regions $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$, where $\mathcal{Z}$ is a slice of $J^-(\mathcal{X})$ which does not intersect $J^-(\mathcal{Y})$, if $\mathcal{X}$ and $\mathcal{Y}$ are spacelike separated, then the probability of any local physical quantity X at $\mathcal{X}$ attaining some value x is unchanged by the value y of any local physical quantity Y at $\mathcal{Y}$ given a full physical description of $\mathcal{Z}$

$$p(x \mid y, \zeta) = p(x \mid \zeta)\,, \tag{3.1}$$

where ζ is the full physical description of $\mathcal{Z}$.

Note that a local physical quantity is any physical quantity that can be ascribed to a bounded spacetime region. What local physical quantities there are and what constitutes a full physical description of a spacetime region, have to be specified by the relevant theory itself.

There is another way to formalize the intuition behind the principle of local causality in a more straightforward way. This is possible by articulating the principle not in terms of probabilistic, but causal notions.

Definition 3.2 (Local Causality)**.** For all spacetime regions $\mathcal{X}$, $\mathcal{Y}$ and any local physical quantities X and Y in $\mathcal{X}$ and $\mathcal{Y}$ respectively, if $\mathcal{X}$ and $\mathcal{Y}$ are spacelike separated, then X is not a cause of Y.

Because this formulation will play an important role in the following investigation we will simply refer to it as *local causality* (LC). This version of the principle applies only to theories or explanations that posit causal structure in addition to probabilities. Hence, its scope is narrower than that of the statistical formulation.

In the sections to come, we will explore how one can leverage these statistical and causal principles of local causality to constrain the statistics observed in the Bell scenario. This will be achieved via the ontological and causal models frameworks, respectively.

[2] This version of the principle is not identical in form to the one found in [32], but in content.

3.2 The EPRB Argument

The Einstein-Podolsky-Rosen-Bohm (EPRB) argument has been at the center of much controversy since its initial publication [33, 34]. What should be uncontroversial is that it played a major role in the discovery of Bell's theorem as Bell, by his own admission, took the EPRB scenario and modified it for his purposes [3]. Furthermore, Bell himself claims in places that the EPRB argument is the first part of a two-step argument [3, 32]. Initially, the EPRB argument is used to infer the existence of something akin to deterministic hidden variables from locality. Then, it can be inferred that local deterministic hidden variables are unable to explain certain quantum-mechanical predictions, which have since been experimentally verified. Hence, nature herself must be nonlocal.

This presentation of Bell's line of argument has been echoed in the literature [35, 36] but also heavily criticized [37]. What is at stake here is whether Bell's theorem needs to assume some form of realism, such as (deterministic) hidden variables, in order to get off the ground. According to opposing accounts of Bell's theorem, it merely shows the impossibility of local realism, a conjunctive condition consisting of locality and realism[3]. This seems to open the door to responses to Bell's theorem which deny realism in order to uphold locality. In the following, we will analyze the EPRB argument in order to ascertain if the assumption of realism involved in Bell's proof has to be stipulated or can be inferred via the EPRB argument. Our account is a simplified retelling of the original argument as presented by EPR [33] and by Bohm [34].

The argument makes use of the following experimental setup: Two parties, Alice and Bob, are spacelike separated. They share a singlet state

$$\left|\psi^-\right\rangle_{AB} = \frac{1}{\sqrt{2}}\left(\left|0\right\rangle_A \left|1\right\rangle_B - \left|1\right\rangle_A \left|0\right\rangle_B\right), \tag{3.2}$$

here written in the computational basis. If Alice and Bob both perform a spin measurement in the same basis, the outcomes are perfectly anti-correlated. Because the singlet state is invariant under any change of orthonormal basis, as long as Alice and Bob measure in the same direction, they will always obtain perfectly anti-correlated outcomes.

The EPRB argument then assumes a *locality condition*, which states that a measurement on a system does not disturb a spacelike separated system. Furthermore, the

[3]The precise meaning of realism has been the subject of considerable debate; see, for instance, [37, 38, 39, 40]. A discussion of the role of realism in Bell's theorem is given at the end of Chapter 3.

reality criterion is required, according to which the value of any quantity that can be predicted with probability one, without disturbing the system, must be predetermined. Assume that both Alice and Bob measure their qubits in the computational basis and that Alice obtains -1 as her outcome. Then she can predict with probability one that Bob will obtain 1. By the locality condition, Alice's measurement cannot disturb Bob's system, as they are spacelike separated. Hence, by the reality criterion, Bob's outcome must be predetermined. The same reasoning applies to Alice's outcome, implying that both outcomes are predetermined when the parties measure along the same direction.

In order to extend this inference to measurements along arbitrary directions, the argument further relies on a form of *counterfactual definiteness*, according to which such predetermined values are independent of the particular experimental context in which they are revealed. Hence, the experimental setup in conjunction with these assumptions implies that the values of the spin along all directions for both Alice and Bob must be predetermined. It is easy to see that the non-realist would not be swayed by this version of the EPRB argument. The problem is that counterfactual definiteness [9, 41] and even the reality criterion [42] are in general not acceptable for non-realists.

Our discussion has shown that it is indeed possible to infer the predetermination of the outcomes of spin measurements in the EPRB scenario from a suitable locality condition. But the required additional assumptions are so strong that they are unacceptable for the non-realist. Hence, it seems that the claim of inferred realism does not hold water. Another rather different version of the EPRB argument is suggested by Bell in the following quote:

> Of course, mere correlation between distant events does not by itself imply action at a distance, but only correlation between the signals reaching the two places. These signals, in the idealized example of Bohm, must be sufficient to determine whether the particles go up or down. For any residual undeterminism could only spoil the perfect correlation. [3]

It turns out that this insight is indeed correct and can be given a formal proof via the ontological models framework. For these purposes we will redefine the experimental setup slightly. We will refer to the *EPRB scenario* simply as the Bell scenario with trivial settings. Furthermore, we only need a certain kind of locality condition which amounts to outcome independence with trivial settings:

$$p(a, b \mid \lambda) = p(a \mid \lambda)\, p(b \mid \lambda)\,. \tag{3.3}$$

This condition can, for the EPRB scenario, be straightforwardly derived from probabilistic local causality which will be done in the following section. This enables us to recast Bell's remarks as the following theorem:

Theorem 3.3. *Any outcome-independent ontological model of the EPRB scenario that reproduces perfect (anti-)correlations is locally deterministic, meaning that for almost all* $\lambda \in \mathbf{\Lambda}$*, the conditional probabilities satisfy*

$$p(a \mid \lambda) \in \{0, 1\}\,, \qquad p(b \mid \lambda) \in \{0, 1\}\,. \tag{3.4}$$

Proof. We begin by defining a linear functional, that we will call the EPRB functional,

$$S_{\mathrm{EPRB}} = \sum_{a,b} ab\, p(a, b) = p(-1, -1) - p(-1, 1) - p(1, -1) + p(1, 1)\,. \tag{3.5}$$

Consider then a perfectly correlated behavior

$$p(a, b) = \mu\, \delta_{a,-1}\delta_{b,-1} + (1 - \mu)\, \delta_{a,1}\delta_{b,1}\,, \tag{3.6}$$

with $\mu \in [0, 1]$. For any such behavior, it is easy to check that

$$S_{\mathrm{EPRB}} = 1\,. \tag{3.7}$$

Assume an outcome-independent ontological model of the EPRB scenario,

$$p(a, b) = \int_{\mathbf{\Lambda}} \mathrm{d}\lambda\, p(\lambda)\, p(a \mid \lambda)\, p(b \mid \lambda)\,, \qquad a, b \in \{-1, 1\}\,. \tag{3.8}$$

Now evaluate S_{EPRB} in the outcome-independent ontological model:

$$\begin{aligned} S_{\mathrm{EPRB}} &= \sum_{a,b} ab \int_{\mathbf{\Lambda}} \mathrm{d}\lambda\, p(\lambda)\, p(a \mid \lambda)\, p(b \mid \lambda) && (3.9)\\ &= \int_{\mathbf{\Lambda}} \mathrm{d}\lambda\, p(\lambda)\, f(\lambda)\,, && (3.10) \end{aligned}$$

where

$$f(\lambda) = \sum_{a,b} ab\, p(a \mid \lambda)\, p(b \mid \lambda)\,. \tag{3.11}$$

Set $x = p(a = 1 \mid \lambda)$ and $y = p(b = 1 \mid \lambda)$. Then

$$\begin{aligned} f(\lambda) &= (1 - x)(1 - y) - (1 - x)y - x(1 - y) + xy && (3.12)\\ &= (1 - 2x)(1 - 2y)\,. && (3.13) \end{aligned}$$

Since $x, y \in [0,1]$, it follows that $1-2x, 1-2y \in [-1,1]$ and hence

$$-1 \leq f(\lambda) \leq 1\,. \tag{3.14}$$

Moreover, the extremal values $f(\lambda) = \pm 1$ occur if and only if $x, y \in \{0,1\}$, that is, if and only if $p(a \mid \lambda)$ and $p(b \mid \lambda)$ are deterministic. From perfect correlation we have

$$S_{\text{EPRB}} = 1\,, \tag{3.15}$$

while from the ontological model we have

$$S_{\text{EPRB}} = \int_{\mathbf{\Lambda}} \mathrm{d}\lambda\, p(\lambda)\, f(\lambda)\,. \tag{3.16}$$

Using $f(\lambda) \leq 1$ for all λ and $\int_{\mathbf{\Lambda}} \mathrm{d}\lambda\, p(\lambda) = 1$ with $p(\lambda) \geq 0$, we obtain the inequality

$$S_{\text{EPRB}} = \int_{\mathbf{\Lambda}} \mathrm{d}\lambda\, p(\lambda)\, f(\lambda) \leq \int_{\mathbf{\Lambda}} \mathrm{d}\lambda\, p(\lambda) = 1\,. \tag{3.17}$$

Since S_{EPRB} is also equal to 1 operationally, the inequality must be saturated. This is possible only if

$$f(\lambda) = 1 \tag{3.18}$$

for almost all $\lambda \in \mathbf{\Lambda}$. As noted above, $f(\lambda) = 1$ implies $x, y \in \{0,1\}$, meaning that $p(a \mid \lambda), p(b \mid \lambda) \in \{0,1\}$ for almost all λ. Hence the model is locally deterministic in the case of perfect correlation. The argument for perfect anti-correlation is analogous, yielding $f(\lambda) = -1$ almost everywhere on $\mathbf{\Lambda}$ and again implying local determinism. This concludes the proof. □

This argument can be generalized to an arbitrary number of parties, which is done in Appendix A. The trouble with this version of the EPRB argument is that it makes use of the ontological models framework. This is not a problem in itself, but it makes the argument irrelevant when it comes to inferring realism via locality and perfect (anti-)correlations on pain of circularity. This is because the assumption of an ontological model is itself a prominent form of realism. Hence, even though it is possible to derive determinism on the ontic level from perfect correlation or anti-correlation, some sort of realism has to be presupposed.

3.3 Bell Locality in Ontological Models

In the following, we consider an ontological model of the Bell scenario. We begin by arguing that probabilistic local causality implies outcome and parameter independence which jointly entail locality. Locality, under the assumption of statistical independence, is then shown to imply Bell locality for the operational statistics. *Outcome independence* and *parameter independence*, in the context of an ontological model of the Bell scenario, can be stated as follows:

Definition 3.4 (Outcome Independence)**.**

$$p(a \mid b, x, y, \lambda) = p(a \mid x, y, \lambda)\,. \tag{3.19}$$

Definition 3.5 (Parameter Independence)**.**

$$p(a \mid x, y, \lambda) = p(a \mid x, \lambda)\,, \tag{3.20}$$

and

$$p(b \mid x, y, \lambda) = p(b \mid y, \lambda)\,. \tag{3.21}$$

To see how outcome independence follows from probabilistic local causality, note that the ontic state λ acts for both outcomes a and b as a full physical description as required by probabilistic local causality. Furthermore, a and x are spacelike separated from b and y. Hence, according to probabilistic local causality, $p(a \mid b, x, y, \lambda)$ cannot depend on b and analogously $p(b \mid a, x, y, \lambda)$ cannot depend on a. This amounts to outcome independence. For the same reasons, $p(a \mid x, y, \lambda)$ cannot depend on y and $p(b \mid x, y, \lambda)$ on x, which is equivalent to parameter independence.

Locality states that in any ontological model of the Bell scenario, the probability $p(a, b \mid x, y, \lambda)$ factorizes conditional on the ontic state λ:

Definition 3.6 (Locality)**.**

$$p(a, b \mid x, y, \lambda) = p(a \mid x, \lambda)\, p(b \mid y, \lambda)\,. \tag{3.22}$$

It can be shown that outcome and parameter independence jointly imply locality:

Lemma 3.7. *Any ontological model of the Bell scenario that satisfies*

(i) Outcome Independence,

(ii) Parameter Independence,

is local.

Proof. Assume an ontological model of the Bell scenario that respects (i) and (ii). The ontological model ensures the existence of the conditional probability distribution $p(a, b \mid x, y, \lambda)$. By the definition of conditional probability we find that

$$p(a, b \mid x, y, \lambda) = p(a \mid b, x, y, \lambda)\, p(b \mid x, y, \lambda)\,. \tag{3.23}$$

Using outcome independence, we replace the first factor

$$p(a \mid b, x, y, \lambda)\, p(b \mid x, y, \lambda) = p(a \mid x, y, \lambda)\, p(b \mid x, y, \lambda)\,. \tag{3.24}$$

Applying parameter independence to each factor yields

$$p(a \mid x, y, \lambda)\, p(b \mid x, y, \lambda) = p(a \mid x, \lambda)\, p(b \mid y, \lambda)\,. \tag{3.25}$$

Hence, we conclude

$$p(a, b \mid x, y, \lambda) = p(a \mid x, \lambda)\, p(b \mid y, \lambda)\,. \tag{3.26}$$

□

In the following, we will make use of an additional assumption unrelated to probabilistic local causality:

Definition 3.8 (Statistical Independence)**.**

$$p(\lambda \mid x, y) = p(\lambda)\,. \tag{3.27}$$

Statistical independence asserts that the ontic state λ is independent of the settings, or equivalently, that the settings are independent of the ontic state. This is often motivated by our common-sense belief that experimenters can freely choose their measurement settings independently of the experiment they perform[4]. The final notion that we will introduce in this section is that of *Bell locality*:

[4]In our causal formulation of Bell's theorem, we will find that this way of motivating statistical independence is not quite sufficient. To guarantee that λ is statistically independent of the settings, it is not enough to argue that the settings should not be influenced by λ; one must also exclude the possibility that the settings influence λ in such a way as to reproduce the observed statistics locally. It turns out that, in order to rule out such causal pathways, one has to assume that there are no causal influences from a local physical quantity into its past light cone.

Definition 3.9 (Bell Locality)**.**

$$p(a, b \mid x, y) = \int_{\boldsymbol{\Lambda}} \mathrm{d}\lambda \, p(a \mid x, \lambda) \, p(b \mid y, \lambda) \, p(\lambda) \, . \tag{3.28}$$

Bell locality is now finally a restriction on the observable statistics $p(a, b \mid x, y)$. It can be shown that any local model that satisfies statistical independence must also be Bell-local:

Lemma 3.10. *An ontological model of the Bell scenario that satisfies*

(i) Statistical Independence,

(ii) Locality,

is Bell-local.

Proof. We begin by assuming that the Bell scenario is described by an ontological model and further that (i) and (ii) hold. The ontological model ensures that

$$p(a, b \mid x, y) = \int_{\boldsymbol{\Lambda}} \mathrm{d}\lambda \, p(a, b \mid x, y, \lambda) \, p(\lambda \mid x, y) \, . \tag{3.29}$$

By statistical independence we have that

$$p(\lambda \mid x, y) = p(\lambda) \, . \tag{3.30}$$

Locality states that

$$p(a, b \mid x, y, \lambda) = p(a \mid x, \lambda) \, p(b \mid y, \lambda) \, . \tag{3.31}$$

Substituting equations (3.30) and (3.31) into equation (3.29) we obtain

$$p(a, b \mid x, y) = \int_{\boldsymbol{\Lambda}} \mathrm{d}\lambda \, p(a \mid x, \lambda) \, p(b \mid y, \lambda) \, p(\lambda) \, , \tag{3.32}$$

which is the defining condition of Bell locality. □

3.4 Bell Locality in Causal Models

3.4.1 The Required Causal Principles

In the following, we discuss the causal principles required to derive Bell-local statistics within the causal models framework. We introduce and motivate a causal principle called measurement independence, of which there exists both a weak and a strong formulation. Measurement independence guarantees the experimenters' freedom to choose their measurement settings by limiting the causes of the settings X and Y within the Bell scenario. The assumption of no-backward causality forbids causal arrows from a variable X at a spacetime region $\mathcal{X}$ to any variable Y localized in the backward light cone of $\mathcal{X}$. Later, we will also discuss how LC and no-backward causality relate to the broader notion of relativistic causality.

The general idea, that we take to motivate both versions of measurement independence, is called the *principle of free experimentation* and can be stated as follows:

> An experimenter can choose the settings of their experiment independently of the physical system under investigation.

Although somewhat vague, principles of this kind have long been argued to be essential for the very possibility of empirical science. Consider, for example, the following remark of Anton Zeilinger as a concise way of articulating this point of view:

> We always implicitly assume the freedom of the experimentalist [...]. This fundamental assumption is essential to doing science. If this were not true, then, I suggest, it would make no sense at all to ask nature questions in an experiment, since then nature could determine what our questions are, and that could guide our questions such that we arrive at a false picture of nature. [43]

In the context of the Bell scenario, as mentioned above, free experimentation suggests a condition we will call *measurement independence*, which comes in a *weak* (WMI) and a *strong* (SMI) form:

Definition 3.11 (Weak Measurement Independence)**.** In the causal structure underlying the Bell scenario, Λ is not a cause of either X or Y.

Definition 3.12 (Strong Measurement Independence)**.** In the causal structure underlying the Bell scenario, neither X nor Y has any causes.

For our derivation of Bell locality from causal principles, the weak version suffices, while the strong formulation will become relevant only later.

For measurement independence to be plausible, the ontic state Λ must be sufficiently localized in spacetime[5]. If Λ were identified with the state of the entire universe, then it would almost certainly act as a cause of both X and Y, leading to a violation of measurement independence and, consequently, of free experimentation[6]. Hence, for these principles to make sense, the underlying physical state relevant to the outcomes should be taken to describe only the subsystem involved in the Bell experiment, in a way that excludes causal influences on the measurement settings.

This issue is also related to failures of LC. If LC fails and spacelike separated regions can influence the outcomes of some experiment, then it becomes less clear that Λ can be localized in the required manner. But this does not mean it is impossible. Bohmian mechanics, for example, is manifestly nonlocal, yet it still provides a notion of subsystem. Hence, it least in the case of the Bell scenario, one does not [44].

The last remaining assumption is that of *no-backward causality* (NBC), which is stated below:

Definition 3.13 (No-Backward Causality)**.** For all spacetime regions $\mathcal{X}$, $\mathcal{Y}$ and any local physical quantities X and Y in $\mathcal{X}$ and $\mathcal{Y}$ respectively, if Y is located in the backward light cone of $\mathcal{X}$, meaning $\mathcal{Y} \subseteq J^-(\mathcal{X})$, then X is not a cause of Y.

One reason why NBC is needed in addition to LC and WMI is to ensure that the settings do not retroactively cause the hidden variable. Without this assumption, one can construct causal structures that satisfy both LC and WMI, while violating Bell locality via a failure of statistical independence.

Consider for example the following causal structure, which respects both LC and WMI if one considers Λ to be localized in the past light cone of both $\mathcal{X}$ and $\mathcal{Y}$, meaning $\Lambda \subseteq J^-(\mathcal{X}) \cap J^-(\mathcal{Y})$:

[5]Strictly speaking, the following arguments implicitly assume that the latent variable Λ, or at least parts of it, can be localized in spacetime. This assumption can be relaxed without loosing the thrust of the arguments presented below, as shown in Appendix B.

[6]This is not necessarily so. It is also possible that the values of X and Y are truly random and hence entirely uncaused.

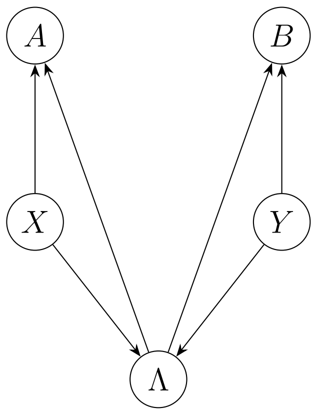


Figure 3.1: A causal structure compatible with measurement independence and LC which violates Bell locality via backward causality.

It can be checked that this causal structure can account for violations of Bell locality. To rule out such cases can make use of NBC which prohibits the influence from X and Y to Λ.

NBC, as well as LC, can be seen to follow from the same underlying, namely from the principle of *relativistic causality*:

Definition 3.14 (Relativistic Causality)**.** For all spacetime regions $\mathcal{X}$, $\mathcal{Y}$ and any local physical quantities X and Y in $\mathcal{X}$ and $\mathcal{Y}$ respectively, X can be a cause of Y only if $\mathcal{Y}$ is located in the forward light cone of $\mathcal{X}$, meaning $\mathcal{Y} \subseteq J^+(\mathcal{X})$.

Further, LC and NBC are jointly equivalent to relativistic causality. To see this, consider two random variables X and Y in $\mathcal{X}$ and $\mathcal{Y}$ respectively. Relativistic causality forbids a causal arrow from X to Y except if $\mathcal{Y} \subseteq J^+(\mathcal{X})$. NBC forbids the arrow if $\mathcal{Y} \subseteq J^-(\mathcal{X})$, and LC forbids the arrow if $\mathcal{X}$ and $\mathcal{Y}$ are spacelike separated. Thus, arrows are permitted only in the case $\mathcal{Y} \subseteq J^+(\mathcal{X})$, which is exactly the content of relativistic causality, as visualized below:

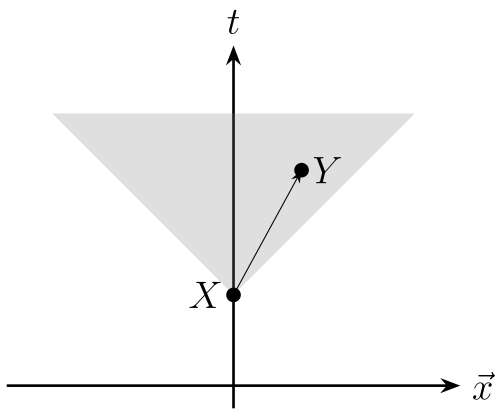


Figure 3.2: Spacetime diagram illustrating relativistic causality.

Hence, with regard to the derivation of Bell locality from WMI, LC, and NBC, the reasoning can be phrased more succinctly: WMI together with relativistic causality implies Bell locality. Showing that this, and some closely related results, hold is the goal of the next subsection.

3.4.2 Deriving Bell Locality from Causal Principles

We begin our derivation from Bell locality from causal principles by establishing the following lemma: any causal structure where there is no directed path from Alice's side of the experiment to Bob's outcome and vice versa, and the settings have no parents, implies Bell locality via the Markov condition. Then we will argue that any causal explanation of the Bell scenario is jointly restricted by WMI, LC and NBC to only such structures that satisfy the conditions of the previously mentioned lemma. Hence, WMI, LC and NBC jointly imply Bell locality in causal models.

Lemma 3.15. *Any causal explanation of the Bell scenario where there is*

(i) no directed path from A or X to B,

(ii) no directed path from B or Y to A,

(iii) and the settings have no parents, meaning $\mathrm{Pa}(X) = \mathrm{Pa}(Y) = \emptyset$*,*

must give rise to Bell-local statistics.

Proof. Assumptions (i) and (ii) forbid any directed path from A or X to B, and from B or Y to A. In particular, the direct arrows $A \to B$, $X \to B$, $B \to A$, and $Y \to A$ are excluded. Due to (iii) there can be no arrow from A, B, or Λ to X or Y. This leaves us with the following list of arrows which are not directly prohibited by (i)–(iii):

(1) $A \to \Lambda$ (3) $X \to \Lambda$ (5) $\Lambda \to A$ (7) $X \to A$

(2) $B \to \Lambda$ (4) $Y \to \Lambda$ (6) $\Lambda \to B$ (8) $Y \to B$

Now, due to (i) and (ii) and the fact that cycles are disallowed, not all combinations of (1)–(8) are possible. Consider the arrows (5) and (6). They potentially enable a connection form one side of the experiment to the other through the latent cause Λ. Logically, there are four possibilities:

(I) $\neg(5) \wedge \neg(6)$. (III) $\neg(5) \wedge (6)$.

(II) $(5) \wedge \neg(6)$. (IV) $(5) \wedge (6)$.

We will now go through each of the possibilities (I)–(IV) and show that in each case there exists exactly one maximally inclusive compatible set of causal relations. We begin by noting that

(5) is mutually incompatible with (1), (2), and (4).

Similarly,

(6) is mutually incompatible with (1), (2), and (3).

(5) is incompatible with (1) because cycles are prohibited, (2) due to (ii), and (4) also due to (ii). Analogously, (6) is incompatible with (2) because cycles are prohibited, (1) due to (i), and (4) also due to (i).

We now consider the case of (I). Because (5) and (6) are both ruled out, and all other allowed arrows are mutually compatible we find the following maximally inclusive compatible set of causal relations: {(1), (2), (3), (4), (7), (8)}. In the case of (II), only (6) is ruled out while (5) is assumed. Hence, (1), (2), and (4) are ruled out. The remaining arrows are compatible and form the set {(3), (5), (7), (8)}. The case of (III) is completely analogous to the case of (II) but with (5) being ruled out instead of (6). Therefore, we are left with {(4), (6), (7), (8)}. In the case of (IV), both (5) and (6) are assumed. This rules out (1), (2), (3) and (4) which leaves the set {(5), (6), (7), (8)}, all of which are compatible.

Hence, we have established the existence of four maximally inclusive sets of causal relations, each corresponding to one of the four possibilities described above:

(I) {(1), (2), (3), (4), (7), (8)}, (III) {(4), (6), (7), (8)},

(II) {(3), (5), (7), (8)}, (IV) {(5), (6), (7), (8)}.

They are maximally inclusive as every set of causal relations compatible with (i)–(iii) must be a subset of (I)–(IV). These sets of causal relations give rise to the following four causal structures:

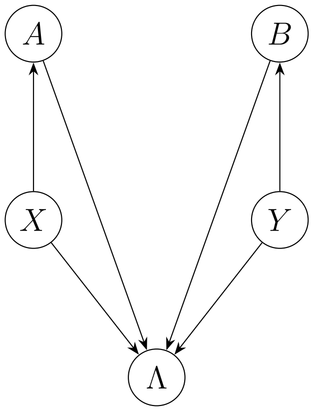


Figure 3.3: Causal structure (I).

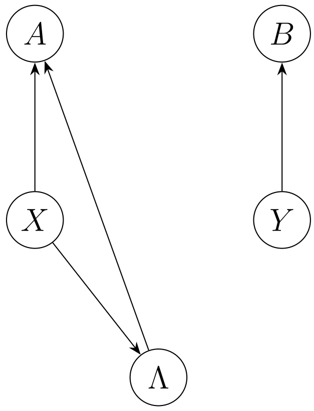


Figure 3.4: Causal structure (II).

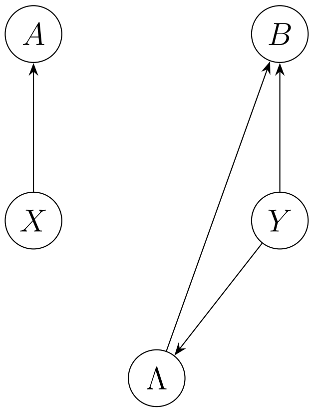


Figure 3.5: Causal structure (III).

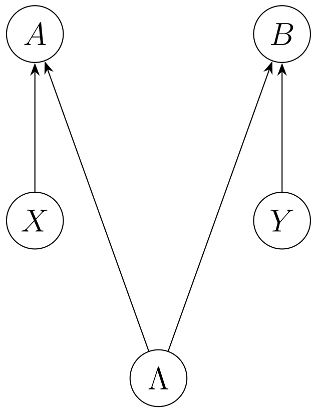


Figure 3.6: Causal structure (IV).

We now apply the Markov condition to all of them to show that they imply Bell locality. For (I), we obtain the following joint distribution

$$p(a,b,x,y,\lambda) = p(a \mid x)\, p(b \mid y)\, p(x)\, p(y)\, p(\lambda \mid x,y,a,b)\,. \tag{3.33}$$

Summing over a, b, and integrating over λ gives

$$p(x,y) = \sum_{a,b\in\{-1,1\}} \int_{\mathbf{\Lambda}} \mathrm{d}\lambda\, p(a,b,x,y,\lambda) = p(x)\, p(y)\,, \tag{3.34}$$

and therefore

$$p(a,b \mid x,y) = p(a \mid x)\, p(b \mid y)\,, \tag{3.35}$$

which is trivially Bell-local. In case (II) the Markov condition gives

$$p(a, b, x, y, \lambda) = p(a \mid x, \lambda)\, p(b \mid y)\, p(x)\, p(y)\, p(\lambda \mid x)\,. \tag{3.36}$$

Again, we find that $p(x, y) = p(x)\, p(y)$, and thus

$$p(a, b \mid x, y) = \int_{\mathbf{\Lambda}} \mathrm{d}\lambda\, p(a \mid x, \lambda)\, p(\lambda \mid x)\, p(b \mid y) \tag{3.37}$$

$$= p(a \mid x)\, p(b \mid y)\,, \tag{3.38}$$

which is also trivially Bell-local. Case (III) is simply (II) under a swap of parties and therefore also trivially Bell-local. For the common cause structure (IV) we have

$$p(a, b, x, y, \lambda) = p(a \mid x, \lambda)\, p(b \mid y, \lambda)\, p(x)\, p(y)\, p(\lambda)\,. \tag{3.39}$$

Again $p(x, y) = p(x)\, p(y)$, so

$$p(a, b \mid x, y) = \int_{\mathbf{\Lambda}} \mathrm{d}\lambda\, p(\lambda)\, p(a \mid x, \lambda)\, p(b \mid y, \lambda)\,, \tag{3.40}$$

which is the defining condition of Bell locality. Thus all maximally compatible structures (I)–(IV) respect Bell locality.

Any causal structure satisfying (i)–(iii) is obtained from one of (I)–(IV) by removing arrows. Removing arrows introduces additional conditional independence relations via the Markov condition and therefore cannot generate correlations that violate Bell locality. Consequently, every such structure must also yield Bell-local statistics. This completes the proof. □

Now it remains to be shown that the relevant causal principles actually constrain the causal explanations of the Bell scenario to causal structures that satisfy the conditions of the lemma:

Lemma 3.16. *Any causal model of the Bell scenario that satisfies:*

(i) Weak Measurement Independence,

(ii) Local Causality,

(iii) No-Backward Causality,

is Bell-local.

Proof. We begin with the permissive generalized causal structure compatible with the Bell scenario:

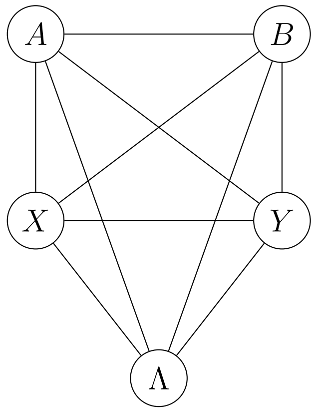


Figure 3.7: The most permissive generalized causal structure compatible with the operational Bell scenario.

Imposing WMI removes the arrows from Λ into X and Y. This restricts the general structure to:

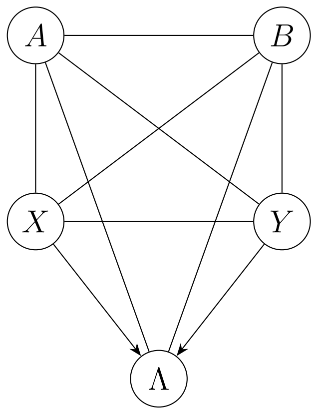


Figure 3.8: The generalized causal structure constrained by WMI.

Next, we impose LC and NBC on the operational variables. Since Λ may be arbitrarily localized, we, at first, only apply the spacetime constraints to A, B, X, and Y:

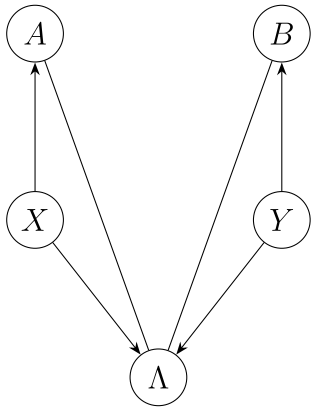


Figure 3.9: The generalized causal structure further constrained by LC and NBC for the operational variables.

In this structure, the settings have no parents. Thus, by Lemma 3.15, Bell locality follows once it is shown that there is no directed path from X or A to B, and/or from Y or B to A. Any such path must pass through Λ. We begin by examining whether the following directed path is possible given our causal assumptions:

$$X \to \Lambda \to B\,. \tag{3.41}$$

Assume, for an indirect proof, that $X \to \Lambda$ and $\Lambda \to B$ are both present. This yields the following causal structure:

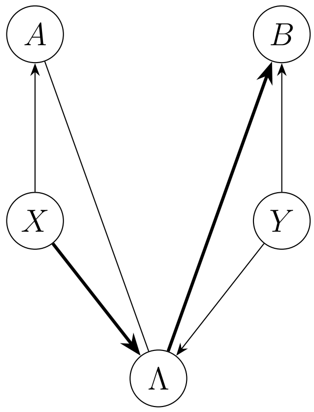


Figure 3.10: Causal structure assumed for an indirect proof.

Then LC and NBC jointly imply:

$$\mathcal{R}_\Lambda \subseteq J^+(\mathcal{X})\,, \qquad \mathcal{B} \subseteq J^+(\mathcal{R}_\Lambda)\,, \tag{3.42}$$

and hence

$$\mathcal{B} \subseteq J^+(\mathcal{X})\,. \tag{3.43}$$

Here we have used $\mathcal{R}_\Lambda$ for the spacetime region in which the relevant part of Λ is located. But the Bell scenario requires X and B to be located in spacelike separated regions, which implies that

$$\mathcal{B} \nsubseteq J^+(\mathcal{X})\,. \tag{3.44}$$

We have reached a contradiction and conclude by indirect proof that the path

$$X \to \Lambda \to B\,, \tag{3.45}$$

is impossible under (i)–(iii). By analogous arguments, the following remaining options are also ruled out:

(1) $A \to \Lambda \to B$,

(2) $Y \to \Lambda \to A$,

(3) $B \to \Lambda \to A$.

Thus, no directed path from X or A to B, and/or from Y or B to A is compatible with the assumptions (i)–(iii). Together with the already noted fact that the settings X and Y have no parents, Lemma 3.15 applies. Since we began with the most general causal structure and showed that any structure satisfying (i)–(iii) must also satisfy the conditions of Lemma 3.15, we conclude that every causal explanation of the Bell scenario compatible with (i)–(iii) must be Bell-local. □

3.5 The CHSH Inequality

Having seen that different sets of assumptions in both ontological and causal models restrict the statistics of the Bell scenario to be Bell-local, we now show that Bell locality implies the Clauser-Horne-Shimony-Holt (CHSH) inequality. We then show that quantum theory, as well as experiment, violates the CHSH inequality up to the Tsirelson bound. It follows that neither an ontological nor a causal model can reproduce the operational statistics without violating at least one of the aforementioned assumptions.[7]. For our present purposes, it is useful to recall the definition of the

[7]All the results and proofs in this section are standard and can, for example, be found in [45].

bipartite correlation function E:

$$E(x,y) = \sum_{a,b} ab\, p(a,b \mid x,y)\,. \tag{3.46}$$

Using E, we are now in the position to state that any $p(a,b \mid x,y) \in \mathcal{L}$ must satisfy the CHSH inequality:

Lemma 3.17 (CHSH)**.** *Any behavior $p(a,b \mid x,y)$ that is Bell-local satisfies the CHSH inequality*

$$|S_{\text{CHSH}}| = |E(0,0) + E(0,1) + E(1,0) - E(1,1)| \leq 2\,. \tag{3.47}$$

Proof. For a detailed proof see Appendix A. □

The CHSH functional S_{CHSH} can maximally attain an absolute value of 2 over the local set, i.e., the set of all probabilities $p(a,b \mid x,y)$ which are Bell-local. The algebraic maximum of $|S_{\text{CHSH}}|$ is 4. It is achieved when $E(0,0) = E(0,1) = E(1,0) = -E(1,1) \in \{-1,1\}$. This value is also the maximal value compatible with the *no-signaling* constraints which are given below:

Definition 3.18 (No-Signaling)**.**

$$p(a \mid x,y) = p(a \mid x)\,, \qquad p(b \mid x,y) = p(b \mid y)\,. \tag{3.48}$$

A no-signaling behavior that can saturate the algebraic bound of the CHSH expression is the Popescu–Rohrlich (PR) box [46][8]

$$p(a,b \mid x,y) = \frac{1}{4}(1 + ab\,(-1)^{xy})\,. \tag{3.50}$$

Quantum mechanics allows violations up to $2\sqrt{2}$, the *Tsirelson bound* [47]:

Theorem 3.19 (Tsirelson)**.** *The maximal violation of the CHSH inequality achievable by quantum theory is $2\sqrt{2}$.*

Proof. For a detailed proof see Appendix A. □

[8]It is sometimes more convenient to make use of a representation of the PR box for $a,b,x,y \in \{0,1\}$. In this case the PR box is given as follows:

$$p(a,b \mid x,y) = \frac{1}{2}\delta_{a\oplus b,xy}\,. \tag{3.49}$$

A great number of experimental tests have been performed which yielded violations of the CHSH inequality. But only in the last decade have loophole-free Bell tests been successfully performed [48, 49, 50]. These loophole-free Bell tests strongly support the conclusion that the observed CHSH violations are due to a failure of the relevant assumptions and not some experimental imperfection.

3.6 Two Formulations of Bell's Theorem

The results discussed in the sections above can be summarized in the following two no-go theorems:

Theorem 3.20 (Probabilistic Bell)**.** *There is no ontological model of the Bell scenario that accounts for the operational statistics and satisfies:*

(i) Statistical Independence,

(ii) Locality.

Proof. Assume, for an indirect proof, that there exists an ontological model of the Bell scenario that satisfies (i) and (ii) and reproduces the operational statistics. By Lemma 3.10, any ontological model satisfying (i) and (ii) must be Bell-local. But every Bell-local behavior satisfies the CHSH inequality by Lemma 3.17. Since experiments violate the CHSH inequality, the operational statistics cannot be Bell-local. This contradicts our initial assumption. By indirect proof, we conclude that no ontological model satisfying (i) and (ii) can account for the operational statistics of the Bell scenario. □

Theorem 3.21 (Causal Bell)**.** *There is no causal model of the Bell scenario that accounts for the operational statistics and satisfies:*

(i) Weak Measurement Independence,

(ii) Local Causality,

(iii) No-Backward Causality.

Proof. The proof is completely analogous to the proof of Theorem 3.20. □

In the following we will mostly orient our discussion around Theorem 3.21, to which we will simply refer as *Bell's theorem*. The main reason is that causal models can provide a more detailed analysis than ontological models, due to the fact that causal language can express the direction of influences. This fact will become especially relevant in the following chapters.

These results pose a great challenge for our common-sense understanding of the physical world. They force us to choose between assumptions which all seem highly reasonable. Nonetheless, something has to give. In the following section, we sketch the major responses to Bell's theorem in the light of methodological principles presented in Chapter 2. We will tentatively argue that the most reasonable way to proceed, given the current dialectical situation, is to accept nonlocality as a fact of nature.

3.7 Responses to Bell's Theorem

3.7.1 Realism, Exotic Realism and Anti-Realism

Before we consider the possible responses to Bell's theorem, we want to clear up a common source of confusion surrounding Bell's theorem. This concerns the role of realism in Bell's theorem. The assumptions of Bell's theorem can be listed as follows:

(i) The existence of an applicable causal model of the Bell scenario,

(ii) Weak Measurement Independence,

(iii) Local Causality,

(iv) No-Backward Causality.

Hence, it is natural to identify *realism* with assumption (i): the existence of a causal, or sometimes ontological, model that correctly describes the Bell scenario. This is how we will use the notion of realism in the remainder of the thesis. As shown earlier, Bell's theorem forces us to reject at least one of these assumptions.[9] We will briefly go through all four possibilities and comment on most obvious issues associated with them.

The most involved case is that of the rejection of realism. This is because, in contrast to the other assumptions, which are well-defined conditions in general frameworks,

[9] In the spirit of logical conservatism, we will only consider the rejection of a single assumption at a time.

the rejection of realism involves the rejection of a general framework as a whole. We distinguish between *anti-realism* and *exotic realism*[10]. This distinction has its roots in the structure of ontological and causal models. As presented in Chapter 2, both frameworks involve two assumptions of distinct types: The first assumption guarantees the existence of some underlying objective reality. In the case of ontological models this is given by the ontic state and state space, while in causal models this role is played by the relevant causal structure. The second assumption relates the underlying objective reality to the observed probabilities. In ontological models this is done via equation (2.5); in causal models this is accomplished via the Markov condition. It is important to notice that the second assumption presupposes the first. Therefore, one can either reject both assumptions or only the second one. The first option is identified with anti-realism, as it rejects realism wholesale; the second with exotic realism, since it preserves the existence of some underlying objective reality, while that reality is related to the operational statistics in a non-classical way.

For the purposes of this thesis, it should be noted that by adopting anti-realism the classification of theories into local or nonlocal becomes meaningless. This is because these notions presuppose the existence of some underlying objective reality. For example, the definition of locality in the ontological model framework explicitly refers to the ontic state. Similarly, in causal models, LC prohibits causal influences at spacelike separation, thereby invoking the notion of causal structure. Hence, anti-realism cannot properly be classified as local or nonlocal, since it lacks the theoretical resources to make sense of this distinction. For this reason, we will not engage closely with anti-realism in what follows, as it removes the structure needed for the tension between nonlocality and relativity to arise.

Exotic realism comes in different forms. Here we will briefly mention two of the most prominent: the *Everettian interpretation* of quantum theory and *quantum causal models*. The way in which the Everettian interpretation modifies standard realism is by denying that experiments have singular outcomes. A number of challenges have been raised against this view, such as the status of probabilities and the emergence of the classical world from the universal quantum state. These issues have been discussed extensively in the literature and will not be revisited here; for an overview, see [51]. The question of whether Everettian quantum theory should be regarded as local remains controversial. The relation between (non-)locality and relativity in the Everettian interpretation is outside the scope of this thesis.

Quantum causal models drastically alter the rules that connect causal structures to observable probabilities in the presence of a so-called quantum cause. They thereby

[10] The term exotic realism is courtesy of Matthew Leifer.

allow for local, common cause explanations of the Bell scenario while accounting for CHSH violations. Again, for details on how quantum causal models achieve this feat, we refer to the literature; see for example [22][52]. An argument against preferring quantum over classical causal models, based on some recent no-go results, is given in the following subsection.

3.7.2 Faithfulness and Wigner's Friend

A critique of all realist responses to Bell's theorem is achieved via an appeal to faithfulness. As discussed in Chapter 2, any empirically adequate causal explanation of the Bell scenario violates faithfulness. If one wants to hold on to faithfulness, possibly due to a commitment to the LMP, one needs to abandon classical causal models. If anti-realism is further found to be unacceptable, exotic realism in the form of non-classical causal models, especially quantum causal models, provides a way out.

But there is a major problem with this approach due to an important generalization of Bell's theorem, which is known as the local friendliness theorem. It is based on the *extended Wigner's friend scenario* (EWFS)[11]. Here hermetically sealed laboratories house two agents, Charlie and Debbie, who perform a fixed measurement on a shared system. Two further agents, Alice and Bob, are positioned outside Charlie's and Debbie's laboratories respectively. Alice and Charlie are spacelike separated from Bob and Debbie.

Instead of positing an ontic state λ, one assumes that every observed event is absolute and not relative to any observer. Furthermore, it is assumed that there exists a joint probability distribution over all observed events which implies the existence of $p(a, b \mid x, y, c, d)$. This compound assumption is called the *absoluteness of observed events* (AOE). Put simply, the measurement outcomes c, d of Charlie and Debbie in the EWFS play an analogous role to the ontic state λ in the Bell scenario. But there is an important difference between the two scenarios: In the case where Alice sets $x = 1$, she opens Charlie's laboratory and sets her outcome a equal to Charlie's outcome c, meaning that $p(a \mid b, x = 1, y, c, d) = \delta_{a,c}$. Bob proceeds analogously. As in the case of Bell's theorem, one further assumes locality and statistical independence. The conjunction of these assumptions is called *local-friendliness*. It turns out that the set of behaviors satisfying local-friendliness forms a polytope which is in general a strict superset of the local polytope [53].

The EWFS has also been subject to causal analysis. As in the case of Bell's theorem, every causal explanation of the EWFS must violate faithfulness. But unlike the case

[11] In the following we will follow the presentation of [53].

of the Bell scenario, quantum causal models cannot provide a faithful explanation of the EWFS under the assumption of AOE [28]. The same holds true for compositional causal models, a very permissive and general class of causal models that encompass not only classical and quantum causal models but arbitrary GPT-systems and even causal loops [28]. These new developments make it plausible that, if the AOE holds, there are scenarios in which the underlying causal structure must violate faithfulness. This constitutes a serious argument against the LMP, especially from a broadly realist standpoint. It further undermines one of the main motivations for quantum causal models as they cannot be used to save faithfulness while preserving realist intuitions by holding on to the existence of an objective causal structure.

From a realist standpoint, this situation makes it tempting to reject the LMP and faithfulness in order to preserve the AOE. As argued in Chapter 2, one can replace the LMP with the OSP, which can be motivated in a similar way and, as in the case of relativity theory, plays a similar dialectical role. However, the OSP, unlike the LMP, does not imply faithfulness and is hence unfazed by the no-go theorems arising from the EWFS.

3.7.3 Superdeterminism, Backward Causality and Nonlocality

In what follows, we will give an overview of the realist options in response to Bell's theorem. They are:

(i) *Superdeterminism*: Rejection of Weak Measurement Independence.

(ii) *Backward Causality*: Rejection of No-Backward Causality.

(iii) *Nonlocality*: Rejection of Locality.

We will briefly outline some of the most common problems associated with all of these three possibilities, starting with (i). Superdeterminism amounts to the claim that Λ is a cause of the measurement settings X and/or Y. Two popular objections to superdeterminism are that it undermines the reliability of experimental science and that there are no fully developed superdeterministic theories available. Whether superdeterminism really undermines the reliability of experimental science depends heavily on the exact form under consideration. If, by some mechanism, superdeterminism only becomes relevant in those situations in which a nonlocal theory such as Bohmian mechanics must make use of its nonlocal resources, then it can hardly be thought to undermine the reliability of experimental science any more than Bohmian mechanics does. General superdeterminism though, which is not restricted to these kinds of instances, could potentially lead to a systematically distorted picture of the world.

The lack of a clear mechanism specifying when superdeterminism takes effect, such that these skeptical conclusions can be avoided, brings us to the second main objection against superdeterminism. There is currently no fully developed superdeterministic interpretation of quantum mechanics.[12] This fact makes superdeterminism appear more like a merely theoretical possibility than a live one. We conclude with a quote from Bell, which seems to us to be a fair and open-minded summary of our cursory encounter with superdeterminism:

> A theory may appear in which such conspiracies inevitably occur, and these conspiracies may then seem more digestible than the nonlocalities of other theories. When that theory is announced I will not refuse to listen, either on methodological or other grounds. But I will not myself try to make such a theory. [57]

Another possible realist response to Bell's theorem is to accept backward causality and allow causal influences to propagate into their past light cones. We address two objections to backward causality: the possibility of causal paradoxes and the conflict with common sense and relativistic notions of causation.

The existence of causal cycles does not necessarily lead to logical contradictions. For example, there exist extensions of the causal model program presented above that allow cyclic causal structures. However, allowing causal cycles restricts the allowed statistics in a non-trivial way [58]. Alternatively, causal structures that allow backward causality could be prohibited from containing causal cycles altogether. However, this raises the question by what mechanism the occurrence of (paradoxical) causal cycles is ruled out.

The second kind of criticism has to do with the fact that accounts that involve backward causality must reject the principle of relativistic causality. This objection is also prominently leveled against nonlocal theories. It should be noted that backward causality and nonlocality violate relativistic causality in different ways, as backward causality preserves the light-cone structure while nonlocality allows causal influences that are not confined to it. However, the notion of cause is heavily associated with temporal order, in the sense that causes are usually taken to precede their effects. This connection, which can be seen to be integral to the notion of causation, is violated

[12] There are some efforts in this direction, most notably by Gerard 't Hooft [54, 55, 56].

by backward causality, but not necessarily by nonlocality[13]. Hence, it is not entirely obvious which violation of relativistic causality should be regarded as the more severe.

The last remaining option is that of nonlocality, which amounts to admitting the existence of causal influence at spacelike separation. Similar to superdeterminism, nonlocality is sometimes claimed to undermine the reliability of experimental science. Whether this is correct depends on how pathological the version of nonlocality under consideration is. In contrast to superdeterminism, there exist nonlocal realist models that recover the predictions of standard quantum theory exactly, as in Bohmian mechanics, or approximately, as in objective collapse theories within appropriate domains of applicability. Hence, the argument that nonlocality undermines experimental science does not seem to go through.

The strongest and most prominent argument against nonlocality is its perceived incompatibility with the special theory of relativity. This tension will occupy us throughout the remainder of the thesis. However, as already mentioned in Chapter 2, the mere existence of rGRWf and other relativistic nonlocal theories significantly weakens this argument. If one could identify general ways of reconciling nonlocality with relativity, this would considerably strengthen the case for taking nonlocality to be the lesson of Bell's theorem. In what follows we are primarily interested in the relation between nonlocality and relativity. Hence we will, in accordance with logical conservatism, retain all other assumptions as far as possible and treat nonlocality as if it were a fact of nature.

[13]We will elaborate on this in Chapters 4 and 5.

Chapter 4

Gisin's Theorem

Nicolas Gisin has offered an argument against the possibility of relativistic hidden variable theories which are empirically adequate [16]. It is our view that this argument touches the core of the tension between nonlocality and relativity. For this reason, we devote the following chapter to a careful study of Gisin's argument. We review Gisin's original argument and identify its two major limitations: the largely implicit character of its assumptions and its restriction to the deterministic case. In the following, we give a clear account of the assumptions involved and generalize Gisin's argument beyond deterministic theories. The most straightforward way to do so is via ontological models. But the ontological models framework turns out to be inadequate as it lacks a natural asymmetric notion of influence which is required to make sense of no-retrocausality, a crucial assumption of Gisin's argument. Causal models, by contrast, can express no-retrocausality without difficulty. We therefore reformulate Gisin's argument as a no-go theorem via the causal model framework and arrive at the following strictly stronger result: there exists no causal model of the Bell scenario that respects (i) weak measurement independence, (ii) no-retrocausality, and (iii) causal Lorentz invariance. The chapter concludes with some clarifying remarks about different notions of relativity.

4.1 Gisin's Argument and its Limitations

In the following, we present a faithful sketch of *Gisin's argument* for the impossibility of covariant hidden variables based on [39, 16]. We will only make slight modifications regarding notation for the sake of consistency. The argument can be summarized as follows:

Consider a Bell scenario from a Lorentz frame $\mathcal{F}_{AB}$ in which Alice performs her experiment before Bob. Gisin argues that in $\mathcal{F}_{AB}$, Alice's outcome cannot depend on Bob's settings while Bob's outcome could nonlocally depend on Alice's settings

$$a = f(x, \lambda)\,, \qquad b = g(x, y, \lambda)\,. \tag{4.1}$$

In a Lorentz frame $\mathcal{F}_{BA}$, where Bob performs his experiment before Alice, the converse is true

$$a' = f'(x', y', \lambda')\,, \qquad b' = g'(y', \lambda')\,. \tag{4.2}$$

Gisin then argues that 'in a covariant nonlocal model Alice's [and Bob's] result should be independent of the reference frame'. But the argument implicitly requires something stronger, namely for all relevant physical quantities to have identical values in both frames, meaning that

$$\alpha = \alpha' \tag{4.3}$$

for the Lorentz frames $\mathcal{F}_{AB}$ and $\mathcal{F}_{BA}$ and $\alpha \in \{a, b, x, y, \lambda\}$[1]. This implies

$$a = f(x, \lambda)\,, \qquad b = g'(y, \lambda)\,. \tag{4.4}$$

which is a Bell-local model of the Bell scenario. But this is ruled out by observed CHSH inequality violations. Hence Gisin concludes that 'there is [sic] no covariant nonlocal models of quantum correlations, not more than local models'.

A natural question to ask is why Gisin, in the frame $\mathcal{F}_{BA}$, allows Alice's outcome to depend on Bob's setting but not also on Bob's outcome, and analogously for Bob's outcome in the frame $\mathcal{F}_{AB}$. The reason is that, in deterministic models, any violation of locality can, without loss of generality, be written as a violation of parameter independence. To see this assume a deterministic model of the Bell scenario. Without

[1] It turns out that one only needs to assume that the operational variables a, b, x, y have the same values in both reference frames in order to conclude Bell locality. This is a very reasonable assumption to make, simply on the grounds of empirical adequacy. Hence, this assumption, which we will sometimes refer to as the Lorentz invariance of values is rather weak. This is a considerable difference to the genuinely probabilistic case. We will come back to this point in Chapter 5.

loss of generality, any nonlocal model can be written as

$$a = h(b, x, y, \lambda)\,, \qquad b = g(x, y, \lambda)\,. \tag{4.5}$$

A further dependence of b on a is ruled out on pain of circularity. Substituting the second equation into the first, we obtain

$$a = h(g(x, y, \lambda), x, y, \lambda) \equiv f(x, y, \lambda)\,, \tag{4.6}$$

which is of the aforementioned form.

Stepping back, we agree that Gisin's argument is certainly very interesting but we are of the opinion that this conclusion overstates the actual strength of the argument as presented above. Gisin's argument rests on some non-trivial assumptions which can be rejected, thereby allowing for nonlocal models to be 'covariant' in Gisin's sense. Furthermore, Gisin's argument exhibits two major limitations:

(i) Its assumptions are not explicitly stated.

(ii) It is restricted to the deterministic case.

Regarding (ii), Gisin actually suggests that this argument can be generalized to probabilistic theories as follows:

> For stochastic models the situation is interesting. If one interprets probabilities in the usual classical way, one may merely add some random variable to turn the model deterministic, as can be done for any probability appearing in classical physics [...]. [16]

We will postpone our discussion of this point to after the assumptions of the argument have been analyzed.

4.2 The Assumptions of Gisin's Argument

4.2.1 Free Experimentation

The first assumption we will discuss concerns some form of free experimentation. In his original argument, Gisin must assume that the hidden variable λ is statistically independent of the settings x and y in order to rule out superdeterministic explanations.

When formulating Gisin's argument in the ontological models framework, the same assumption of statistical independence is required. In the causal models framework, it suffices to assume weak measurement independence (WMI). Here, we will not define these notions again, as they have already been introduced and sufficiently discussed in detail in Chapter 3[2].

4.2.2 No-Retrocausality

The assumption of no-retrocausality is invoked when Gisin argues that, in the frame $\mathcal{F}_{AB}$, Alice's outcome a depends only on her setting x and the hidden variable λ, and similarly that, in the frame $\mathcal{F}_{BA}$, Bob's outcome b depends only on his setting y and λ. The assumption can be stated as the following general physical principle:

> For all spacetime regions $\mathcal{X}$ and $\mathcal{Y}$, and for any local physical quantities X and Y located in $\mathcal{X}$ and $\mathcal{Y}$ respectively, if X is determined before Y in some Lorentz frame, then Y cannot influence X.

A vague aspect of this principle, which must be further explicated, is the notion of influence. More precisely, what it means to say that 'Y cannot influence X'. In Gisin's original argument, the relevant notion of influence is given in terms of functional dependence. In particular, Y does not influence X if

$$x = f(y, z) = f(z)\,, \tag{4.7}$$

where z denotes possible additional dependence. In the ontological and causal models frameworks, the notion of influence must be understood in different ways. In what follows, we will use the following natural interpretations of the statement 'Y cannot influence X':

(i) $p(x \mid y, z) = p(x \mid z)$, where z represents possible additional dependence;

(ii) Y is not a cause of X, meaning, $Y \notin \mathrm{Pa}(X)$.

This gives rise to the following *probabilistic* formulation of no-retrocausality:

Definition 4.1 (Probabilistic No-Retrocausality)**.** For all spacetime regions $\mathcal{X}$ and $\mathcal{Y}$, and for any local physical quantities X and Y located in $\mathcal{X}$ and $\mathcal{Y}$ respectively, if X

[2]In the following we will always, anticipating certain ways of extending Gisin's argument beyond determinism, provide formulations of Gisin's assumptions for both ontological and causal models.

is determined before Y in some Lorentz frame, then

$$p(x \mid y, z) = p(x \mid z), \tag{4.8}$$

where z denotes possible additional dependence.

For causal models, *no-retrocausality* (NRC) simply prohibits causal influences from the future to the past in all reference frames:

Definition 4.2 (No-Retrocausality)**.** For all spacetime regions $\mathcal{X}$ and $\mathcal{Y}$, and for any local physical quantities X and Y located in $\mathcal{X}$ and $\mathcal{Y}$ respectively, if X is determined before Y in some Lorentz frame, then Y is not a cause of X.

We remark that NRC implies no-backward causality (NBC), while the converse is not true. NRC forbids $Y \in \mathrm{Pa}(X)$ whenever X is determined before Y in *some* Lorentz frame. By contrast, NBC forbids $Y \in \mathrm{Pa}(X)$ only if X lies in the past light cone of Y, and hence only if X is determined before Y in *all* Lorentz frames.

As shown in Chapter 3, NBC together with WMI implies, by Bell's theorem, that there exists no empirically adequate causal model that also respects local causality (LC). Therefore, the strictly stronger assumption of NRC, together with WMI, likewise guarantees that any empirically adequate causal model must violate LC.

4.2.3 Lorentz Invariance

The final assumption of the argument is motivated by the theory of special relativity and is explicitly mentioned by Gisin:

> In this note I like to ask whether such additional nonlocal variables could be covariant in the sense of relativistic time-order invariant predictions, that is invariant under a velocity-boost that changes the time ordering of events.

We will call this condition Lorentz invariance[3]. As mentioned in our discussion of Gisin's original argument, one needs to assume that for both Lorentz frames $\mathcal{F}_{AB}$

[3]We choose this terminology following a remark of Bell, which we already cited in the introduction of this thesis:

> Those paradoxes are simply disposed of by the 1952 theory of Bohm, leaving as the question, the question of Lorentz invariance. So one of my missions in life is to get people to see that if they want to talk about the problems of quantum mechanics — the real problems of quantum mechanics — they must be talking about Lorentz invariance. [59]

and $\mathcal{F}_{BA}$ the values of all relevant physical quantities are the same. This idea can be expressed as the following very general principle:

> The objective description of physical phenomena does not depend on the choice of Lorentz frame.

Here, we have to specify what exactly constitutes an objective description of physical phenomena. This is obviously dependent on the underlying physical theory or framework under consideration. We propose the following for ontological and causal models respectively:

(i) The probability distribution over all physical quantities.

(ii) The causal structure.

This leads us to the following definitions of Lorentz invariance in the case of ontological and causal models:

Definition 4.3 (Probabilistic Lorentz Invariance)**.** The probability distribution over all physical quantities does not depend on the choice of Lorentz frame.

Definition 4.4 (Causal Lorentz Invariance)**.** The causal structure does not depend on the choice of Lorentz frame.

Causal Lorentz invariance will be denoted by (CLI). It turns out that the following strictly weaker principle also suffices for ontological models:

Definition 4.5 (Probabilistic Frame Compatibility)**.** For all physical quantities X and Y, if $(X \perp\!\!\!\perp Y \mid Z)$ holds in some Lorentz frame, then $(X \perp\!\!\!\perp Y \mid Z)$ holds in all Lorentz frames, where the random variable Z represents possible additional dependence.

Probabilistic frame compatibility is strictly weaker than probabilistic Lorentz invariance. While the conditional independence relations between random variables may be frame independent, the numerical values of the underlying probability distributions can, in principle, nevertheless depend on the Lorentz frame under consideration. By contrast, if the joint probability distribution over all physical quantities is the same in all Lorentz frames, then the same conditional independence relations must hold in every frame.

4.3 Unsuccessful Generalizations of Gisin's Argument

4.3.1 The Need for a Probabilistic Generalization

As mentioned above, one major limitation of Gisin's argument is that it applies only to the deterministic case. This is reminiscent of Bell's original theorem, which was likewise first formulated for deterministic models [3] and only later extended to probabilistic theories [60]. It is therefore natural to ask whether Gisin's argument can likewise be generalized beyond determinism.

In the following, we discuss two natural strategies for generalizing Gisin's argument and explain why both fail for principled reasons. The first, proposed by Gisin himself, appeals to the fact that every probabilistic theory can be simulated by a deterministic one. A version of the second strategy has been put forward in [61] and proceeds via something akin to the ontological models framework. We will argue that neither strategy provides a genuine generalization of Gisin's argument. Their respective failures clarify why a further extension, using causal models, is needed.

4.3.2 Deterministic Simulation

Before examining the ontological models formulation of Gisin's argument, we will first clarify why Gisin's simpler proposal is not workable. Gisin claims that his original argument, which applies to the deterministic case can be straightforwardly extended to the probabilistic case by 'merely add[ing] some random variable[s]'. This procedure amounts to the dilation of a deterministic theory.

The reason this generalization via dilation does not go through is simple. It is a truth of probability theory that the predictions of any probabilistic theory can be simulated by a deterministic theory. But this does not mean that the probabilistic theory and its deterministic simulation share the same properties.

The following example makes this point clear. Assume that some genuinely probabilistic theory T is nonlocal and violates outcome independence while respecting parameter independence. As we have seen before, any deterministic theory must violate parameter independence. Hence, any deterministic simulation of T must also violate parameter independence. This is a clear-cut example of empirically equivalent theories which have very different properties, one respects parameter independence and one violates it.

In general, it is simply insufficient to argue that just because a probabilistic theory can be simulated deterministically, the theory in question must exhibit the same properties as its simulation, like for example being able to jointly respect all of the assumptions of Gisin's argument. Such a conclusion must be established on its own merits. In the case of Gisin's argument, the question is whether there exists a sensible dilation of its assumptions. In the following subsection we will argue that this is not the case.

4.3.3 Ontological Models

We have seen that a straightforward generalization of the argument via dilation, as advocated by Gisin, is not possible. A second rather natural way of proceeding is to use ontological models and the definitions given in the previous section to generalize Gisin's argument[4]. This leads to the following theorem:

Theorem 4.6 (Probabilistic Gisin)**.** *There is no ontological model of the Bell scenario that accounts for the operational statistics and satisfies:*

(i) Statistical Independence,

(ii) Probabilistic No-Retrocausality,

(iii) Probabilistic Frame Compatibility.

Proof. Assume an ontological model that accounts for the observed statistics in a Bell scenario and satisfies (i)-(iii). By the definition of conditional probability and the (i) we obtain

$$p(a, b \mid x, y) = \int_{\boldsymbol{\Lambda}} p(a, b \mid x, y, \lambda)\, p(\lambda \mid x, y) \tag{4.9}$$

$$= \int_{\boldsymbol{\Lambda}} p(a \mid b, x, y, \lambda)\, p(b \mid x, y, \lambda)\, p(\lambda)\,. \tag{4.10}$$

We consider a frame $\mathcal{F}_{AB}$ in which Alice's experiment takes place before Bob's; meaning that in $\mathcal{F}_{AB}$, Alice's outcome a is ascertained before Bob chooses his measurement setting y. Hence, due to probabilistic no-retrocausality, we find

$$p(a \mid b, x, y, \lambda) = p(a \mid x, \lambda)\,. \tag{4.11}$$

[4]As briefly mentioned above, this has been done before in [61] in a rather muddled way.

In a frame $\mathcal{F}_{BA}$ in which Bob's experiment takes place before Alice's, probabilistic no-retrocausality implies that b cannot depend on x

$$p(b \mid x, y, \lambda) = p(b \mid y, \lambda) \tag{4.12}$$

Now, by probabilistic frame compatibility, these independence relations must hold in all frames. Hence, we conclude that $p(a, b \mid x, y)$ is Bell-local:

$$\begin{aligned} p(a, b \mid x, y) &= \int_{\mathbf{\Lambda}} p(a \mid b, x, y, \lambda)\, p(b \mid x, y, \lambda)\, p(\lambda) && (4.13)\\ &= \int_{\mathbf{\Lambda}} p(a \mid x, \lambda)\, p(b \mid y, \lambda)\, p(\lambda)\,. && (4.14) \end{aligned}$$

□

This reformulation of the argument using ontological models appears to be the sought-after generalization of Gisin's original argument. But, as so often, things are not as they seem. The problem with the ontological models formulation is that ontological models do not have the resources to naturally express a relevant and asymmetric notion of influence, which is needed to properly capture the principle of no-retrocausality.

We have argued that, within the language of ontological models, the most natural way to formulate the claim 'Y does not influence X' is as $(X \perp\!\!\!\perp Y \mid Z)$ for some, possibly trivial, additional dependence Z. The problem with this way of understanding the claim 'Y does not influence X' is that it also implies 'X does not influence Y'. This is a direct consequence of the symmetry of statistical independence.

But then probabilistic no-retrocausality implies, by the symmetry of statistical independence, that the future cannot depend on the past because the past cannot depend on the future. This is clearly implausible, as this would imply that for any given Lorentz frame $\mathcal{F}$, each moment of time is statistically independent of every other. Hence, for the ontological models formulation of the no-go theorem, there is an easy way out: simply deny the assumption of probabilistic no-retrocausality, as it is not a plausible principle. Any no-go result that rests on even one implausible assumption is itself uninteresting. Hence, to generalize Gisin's theorem beyond the deterministic case, the ontological models framework is not suitable, as it lacks a natural asymmetric notion of influence. This gives rise to the need for causal modeling in the context of Gisin's argument, which we will explore in the following section.

4.4 Gisin's Theorem

Finally, we cast the argument in what we consider to be its strongest form, as a theorem about the possibility of causal explanation. When we talk about *Gisin's theorem* from now on, we will refer to the causal version unless specified otherwise. We will offer two proofs of this result: a direct proof following Gisin's original argument, which is somewhat involved and very similar to the proof of the causal formulation of Bell's theorem. The second proof is much simpler and uses primarily conceptual analysis together with the causal formulation of Bell's theorem provided in Chapter 3.

Theorem 4.7 (Causal Gisin)**.** *There is no causal model of the Bell scenario that accounts for the operational statistics and satisfies:*

(i) Weak Measurement Independence,

(ii) No-Retrocausality,

(iii) Causal Lorentz Invariance.

Proof. We begin by assuming the most general causal explanation of the Bell scenario.

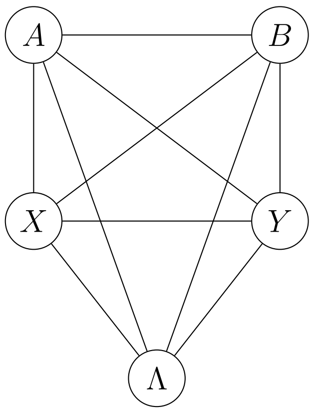


Figure 4.1: The most permissive generalized causal structure compatible with the operational Bell scenario.

We assume (i)–(iii). Applying (i) yields:

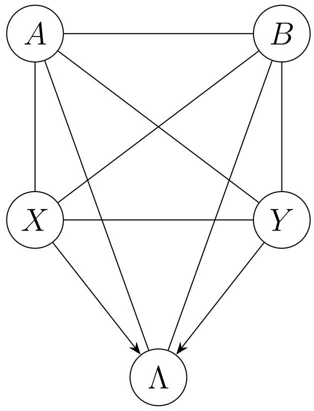


Figure 4.2: The generalized causal structure constrained by WMI.

Because $A \in J^{+}(X)$ and $B \in J^{+}(Y)$, in all reference frames X is determined before A, and Y before B. NRC therefore implies, independently of Lorentz frame:

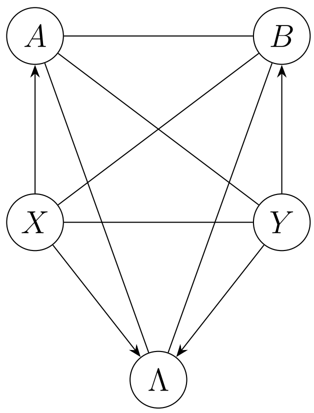


Figure 4.3: The generalized causal structure further constrained by NRC.

We will now consider a reference frame $\mathcal{F}_{AB}$ in which Alice carries out her experiment before Bob. This is always possible due to the spacelike separation of the two parties. NRC then allows only the following causal relations:

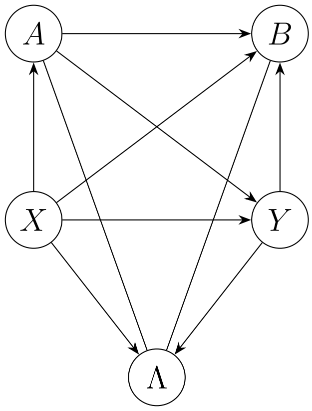


Figure 4.4: The causal structure constrained by NRC in $\mathcal{F}_{AB}$.

We now consider a frame $\mathcal{F}_{BA}$ in which Bob performs his experiment before Alice. Hence, by NRC, we obtain:

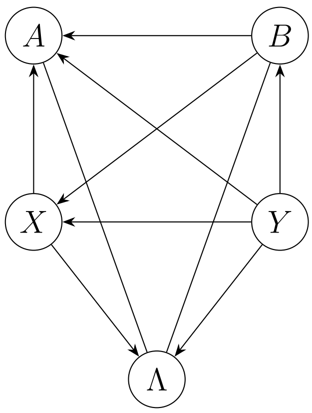


Figure 4.5: The causal structure constrained by NRC in $\mathcal{F}_{BA}$.

By CLI, we obtain the following allowed causal structure, which is given by only those causal connections that are allowed in both frames:

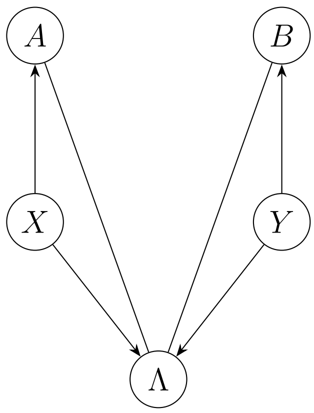


Figure 4.6: The causal structure constrained by CLI.

We notice that this is exactly the same generalized causal structure that we arrived at in our proof of the causal formulation of Bell's theorem. By a completely analogous argument, one can now show that this causal structure, when one assumes CLI and NRC, leads to a Bell-local structure. Again, the settings have no causes. If we can show that no causal connection between the two wings of the experiment is possible, we can conclude Bell locality by lemma 3.15, which is in conflict with the observed predictions.

There are four possible ways for there to be a causal connection between Alice's and Bob's sides:

(1) $X \to \Lambda \to B$

(2) $A \to \Lambda \to B$

(3) $Y \to \Lambda \to A$

(4) $B \to \Lambda \to A$

We will now show that (1) is impossible given (i)–(iii). We proceed by indirect proof and assume that X is a cause of Λ and Λ is a cause of B:

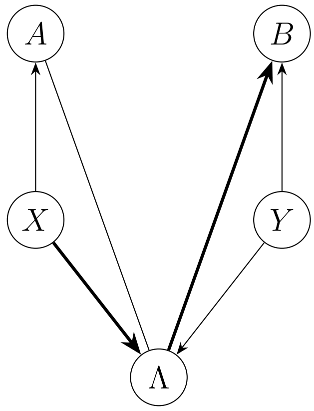


Figure 4.7: Causal structure assumed for an indirect proof.

There are three possibilities regarding the spacetime relations between $\mathcal{X}$ and $\mathcal{R}_\Lambda$, the spacetime region in which the relevant part of Λ is located:

(I) $\mathcal{R}_\Lambda \subseteq J^-(\mathcal{X})$.

(II) $\mathcal{R}_\Lambda \subseteq J^+(\mathcal{X})$.

(III) $\mathcal{X}$ and $\mathcal{R}_\Lambda$ are space like separated.

We will examine (I)–(III) and show that they all lead to a contradiction, thereby establishing that the directed path $X \to \Lambda \to B$ is impossible given (i)–(iii).

The case of (I) directly contradicts NRC, as it would imply that $t_{\mathcal{R}_\Lambda} < t_{\mathcal{X}}$ in all frames.

Now, we consider (II), most involved case. Because $\mathcal{R}_\Lambda \subseteq J^+(\mathcal{X})$ and $\mathcal{X}$ is spacelike separated from $\mathcal{B}$, either $\mathcal{R}_\Lambda \subseteq J^+(\mathcal{B})$ or $\mathcal{R}_\Lambda$ is spacelike separated $\mathcal{B}$. The case of $\mathcal{R}_\Lambda \subseteq J^+(\mathcal{B})$ is directly ruled out by NRC, analogous to (I). We therefore consider the case of spacelike separation between $\mathcal{R}_\Lambda$ and $\mathcal{B}$. Due to their spacelike separation, there exists a reference frame $\mathcal{F}_{B\Lambda}$ in which $t_B < t_\Lambda$. In this frame, Λ cannot cause B due to NRC. By CLI, this implies that Λ cannot cause B in any frame. Hence, there can be no causal connection between Alice's and Bob's wings in this case.

There still remains possibility (III), the case of spacelike separation between $\mathcal{X}$ and $\mathcal{R}_\Lambda$. If they are spacelike separated, there exists a frame $\mathcal{F}_{\Lambda X}$ in which $t_\Lambda < t_X$. Hence, by NRC, X cannot be a cause of Λ in $\mathcal{F}_{\Lambda X}$. By CLI, this implies that X cannot cause Λ in any frame. Hence, in this case as well, there can be no causal connection between Alice's and Bob's sides.

As in the case of the causal formulation of Bell's theorem, this proof can be adapted to block options. Hence, there is no causal connection between the two sides of the experiment. This, together with the fact that the settings cannot have causal parents, implies by lemma 3.15, that any causal explanation of the Bell scenario compatible with (i)–(iii) must be Bell-local. □

As mentioned above, there is another, more conceptual, way of inferring the causal formulation of Gisin's theorem. The idea of this alternative proof is the following: first, we show that the assumptions of Gisin's theorem imply the assumptions of the causal version of Bell's theorem. Then, by Bell's theorem, we conclude that there can be no empirically adequate causal model that satisfies the assumptions of Gisin's theorem. We therefore establish the following lemma:

Lemma 4.8. *The following implications hold:*

(i) No-retrocausality implies no backward causality.

(ii) No-retrocausality and causal Lorentz invariance imply local causality.

Proof. The truth of statement (i) can be easily seen. NBC states that X cannot cause Y if $\mathcal{Y}$ lies in the past light cone of $\mathcal{X}$. This is equivalent to:

(1) For all random variables X and Y, X cannot cause Y if $t_X > t_Y$ in all frames.

NRC states:

(2) For all random variables X and Y, X cannot cause Y if $t_X > t_Y$ in any frame.

This way of spelling out the two principles makes it clear that NRC is strictly stronger than NBC, as it forbids a causal influence from X to Y in all cases in which the same is true for NBC, and in additional cases.

To show (ii), assume NRC and CLI for a conditional proof. Consider arbitrary local physical quantities, represented by the random variables X and Y, located in spacelike separated spacetime regions $\mathcal{X}$ and $\mathcal{Y}$. We now show that there cannot be a causal connection between X and Y by way of indirect proof. Assume, for contradiction, that X is a cause of Y in some Lorentz frame $\mathcal{F}$. Because $\mathcal{X}$ and $\mathcal{Y}$ are spacelike separated, there exists a Lorentz frame $\mathcal{F}_{YX}$ in which Y is determined before X. By CLI, if X is a cause of Y in some frame, then it is a cause of Y in all Lorentz frames, and hence also in $\mathcal{F}_{YX}$. However, by NRC, X cannot be a cause of Y in $\mathcal{F}_{YX}$, since Y is determined before X in that frame. This contradiction shows that X cannot cause Y. Since X and Y were arbitrary spacelike separated local quantities, this establishes LC. □

Because WMI, NBC, and LC jointly imply Bell locality, the same follows from WMI, NRC, and CLI. However, Bell locality is violated experimentally. Therefore, WMI, NRC, and CLI cannot jointly hold. This is exactly the content of the causal formulation of Gisin's theorem.

In summary, the causal formulation of Gisin's theorem extends Gisin's original insight beyond the deterministic case to genuinely probabilistic theories. Unlike the ontological models formulation, this result is non-trivial, since NRC is, at least prima facie, a plausible physical condition. This is possible because the notion of independence available in causal models is not symmetric. If X is not a cause of Y, it is possible that Y is a cause of X.

4.5 The Notion of Relativity

We have claimed before that Gisin's theorem is able to provide a rigorous formulation of the conflict between non-locality and relativity. Nonlocality enters indirectly via the assumptions of NRC and WMI. Any empirically adequate causal model that satisfies these conditions must, by Bell's theorem, be nonlocal. Relativity enters via CLI, which enforces that causal structure is independent of the choice of Lorentz frame.

But there are other possible notions of relativity which have been proposed. They can be loosely sorted into three main types, based on

(i) Locality,

(ii) Lorentz Invariance,

(iii) Spacetime Structure.

Locality-based notions include conditions such as relativistic causality, local causality, probabilistic local causality, Bell locality, parameter independence, or no-signaling. Furthermore, theory-dependent notions such as propagation locality or interaction locality may fall in this camp [18]. The general idea, which can be interpreted in many different ways, seems to be this: special relativity prohibits faster-than-light influences. Different ways of understanding what it means for something to influence something else give rise to the different conceptions of relativity listed above. The problem with these conceptions of relativity as locality listed above is twofold: either they are too permissive, like for example no-signaling, which is respected by relativistic versions of Bohmian mechanics, widely considered to be not genuinely relativistic, or they are incompatible with most realist accounts, such as relativistic causality, local causality, or probabilistic local causality; or simply with experiment, as is the case for Bell locality.

There are different ways in which Lorentz-invariance-based approaches can be formulated. A well-known conception of Lorentz invariance can be given via the transformation properties of dynamically allowed histories. Physical theories are thought to consist of a set $\mathcal{K}$ of kinematically allowed histories and a subset $\mathcal{D} \subseteq \mathcal{K}$ of dynamically allowed histories. A physical theory is Lorentz-invariant if and only if any Lorentz transformation Λ maps dynamically allowed histories to dynamically allowed histories, meaning that $\Lambda\mathcal{D} = \mathcal{D}$ for all $\Lambda \in O(1,3)$ [62, 18]. Our discussion of this definition is primarily based on Struyve [62].

The problem with this way of making sense of Lorentz invariance is that it allows physical theories to be made Lorentz-invariant in a trivial way. To see this, consider physical theory $\mathcal{T}$ which is not Lorentz invariant. We now formulate a new, Lorentz invariant theory $\mathcal{T}^*$ based on and empirically equivalent to $\mathcal{T}$. The trick is to make reference frames part of $\mathcal{T}^*$'s kinematically allowed histories, $h^* = (\mathcal{F}, h) \in \mathcal{K}^*$, where $\mathcal{F}$ is a Lorentz frame and h is a kinematically allowed history in $\mathcal{T}$.

For each frame $\mathcal{F}$, let $\mathcal{D}_\mathcal{F}$ denote the set of histories that satisfy the equations of motion of $\mathcal{T}$ when expressed in that frame. The set of dynamically allowed histories of the new theory $\mathcal{T}^*$ is then defined as $\mathcal{D}^* = \{(\mathcal{F}, h) \mid h \in \mathcal{D}_\mathcal{F}\}$. Lorentz transformations act simultaneously on both components of the history according to $\Lambda h^* = (\Lambda\mathcal{F}, \Lambda h)$. Since the equations of motion have the same form in every frame related by a Lorentz transformation, it follows that

$$(\mathcal{F}, h) \in \mathcal{D}^* \quad \Rightarrow \quad (\Lambda\mathcal{F}, \Lambda h) \in \mathcal{D}^* \,. \tag{4.15}$$

Hence the set $\mathcal{D}^*$ is closed under Lorentz transformations, and $\mathcal{T}^*$ is Lorentz-invariant.

Other approaches to relativity based on invariance or symmetry considerations are Anderson's criterion in terms of absolute objects and independent Lorentz invariance of subsystems. We will not expand on these notions here, but instead refer the reader to Struyve's treatment [62]. He concludes that all of these definitions are also not stringent enough, because some Bohmian models that are generally not considered to be seriously Lorentz-invariant comply with them [62].

The remaining approach to relativity is based on spacetime structure. Minkowski spacetime $\mathbb{M}$ can be defined as $\mathbb{M} = (\mathbb{R}^4, \eta)$, with η being the Minkowski metric.

Maudlin, for example, advocates roughly the following definition of a genuinely relativistic theory: a physical theory is relativistic if and only if it can be formulated by appealing only to the existence of relativistic spacetime structure [63]. If additional spacetime structure, such as a preferred foliation, is adopted, the theory in question is not considered to be genuinely relativistic. This notion of relativity is fairly clear-cut and (rightly) excludes the relativistic versions of Bohmian mechanics which were compatible with the previously discussed invariance-based definitions. Hence, the prohibition of additional, non-relativistic spacetime structure seems to be a necessary part of the common understanding of relativity.

We now briefly comment on the relation between (i)–(iii). As argued above, locality-based approaches to the notion of relativity seem untenable due to Bell's theorem. Interestingly, Tumulka notes that for local theories, Lorentz invariance seems to be coextensional with 'relativity', while the same is not true in non-local theories:

> This example shows why for a general theory, Lorentz invariance is a very weak property that does not capture what we mean by 'relativistic'. For a local theory, in contrast, I cannot think of a case in which the theory is Lorentz-invariant but would not be regarded as relativistic. So it seems that for local theories, Lorentz invariance is a good criterion for being relativistic. [18]

The presence of non-local influences seems to complicate the question of what makes a theory relativistic considerably. As described above, a natural general strategy for making a non-local theory comply with this notion of Lorentz invariance invokes additional spacetime structure. This conflicts with another important aspect of relativity. But, as will be discussed in the following chapter, there are ways of reconciling this notion of Lorentz invariance without postulating additional spacetime structure, even for non-local physical theories. We consider both of these properties to be at least necessary conditions for any serious notion of relativity. We further note that this condition is implied by the principle of relativity and hence motivated by the ontic symmetry principle, as argued in Chapter 2.

Our notion of causal Lorentz invariance is slightly orthogonal to the usual notions of relativity discussed here. Because causal structure is not transparently connected to spacetime structure, it is unclear how it is supposed to transform under the action of the Lorentz group. In the following, we will advocate for the Lorentz invariance of causal connections, which says that the skeleton of the causal structure, but not necessarily the direction of the arrows, must be independent of the Lorentz frame.

Our general approach to the questions discussed in this section can be roughly characterized as follows: we are not truly interested in terminological disputes or in finding the *true* definition of relativity. Instead, we argue that the relevant notions should reveal themselves by playing important roles in productive arguments, such as those of Bell, Gisin, etc. Furthermore, we are motivated by logical conservatism, to preserve as much of the spirit of relativity as possible given violations of local causality.

Chapter 5

Relativistic Explanations of the Bell Scenario

This chapter builds on Gisin's theorem and considers what kinds of nonlocal and genuinely relativistic explanations of the Bell scenario are possible. First, the arguments developed in the previous chapter are summarized. In light of these results, abandoning causal Lorentz invariance in favor of no-retrocausality is advocated as a response to Gisin's theorem. The reason is that arguments, such as the problem of causal paradoxes, which have already been raised against backward causality, also threaten retrocausal accounts. The most natural way of defusing these objections is to adopt a preferred foliation of spacetime. However, this move appears to violate the spirit of relativity even more severely than abandoning causal Lorentz invariance. In order to preserve as much of causal Lorentz invariance as possible, we instead retain the Lorentz invariance of causal connections. It is argued that this way of responding to Gisin's theorem is incompatible with determinism. A uniqueness theorem is then established showing that there exists only one effectively causal and empirically adequate explanation of the Bell scenario that satisfies (i) strong measurement independence, (ii) no-retrocausality, and (iii) Lorentz invariance of causal connections. The properties of this unique causal explanation are discussed. Finally, Tumulka's relativistic GRW model with flash ontology is introduced. It is shown that its description of the Bell scenario agrees with the unique causal explanation derived earlier.

5.1 The Case for Frame Dependence

5.1.1 The State of Play

Because we have, throughout this thesis, accepted nonlocality in response to Bell's theorem, we should, by logical conservatism, also accept weak measurement independence. Hence, Gisin's theorem forces us to choose between no-retrocausality and causal Lorentz invariance (CLI). In the following, we will consider upholding no-retrocausality (NRC) while rejecting CLI. By giving up CLI we allow for different reference frames to exhibit different causal structures. We will call this feature the *frame dependence* of causal structure. This stance seems prima facie to be in conflict with our stated goal of understanding the interplay between nonlocality and relativity, as CLI appears to capture the relevant notion of relativity. The remainder of this section will be devoted to showing that this impression is mistaken.

5.1.2 Troubles with Retrocausality

The reason why we advocate rejecting CLI in order to preserve NRC is twofold. First, we have already provided arguments in favor of no-backward causality (NBC), which also apply in the case of NRC. In Chapter 3, we pointed out that allowing backward causation undermines the usual conception of causality according to which causes temporally precede their effects. Furthermore, allowing backward causation creates the possibility of causal loops, which may lead to logical paradoxes. These arguments apply directly to NRC as well.

Second, in order to avoid the problems associated with retrocausality, one may introduce a preferred frame $\mathcal{F}^*$ and define the dynamics of the theory in question with reference to $\mathcal{F}^*$. This allows one to treat all retrocausal connections that may arise as merely apparent, since in $\mathcal{F}^*$ causal influences propagate only forward in time. Because the dynamical story, as seen from $\mathcal{F}^*$, is free from retrocausality and hence from causal paradox, the latter also holds for arbitrary reference frames. This strategy is, for example, employed in relativistic generalizations of Bohmian mechanics [7, 6].

The problem with this strategy is that the introduction of a preferred reference frame entails additional non-lorentz-invariant spacetime structure, thereby contradicting Einstein's principle of relativity. This principle, as argued in Chapter 2, can be motivated by plausible heuristics such as the ontic symmetry principle (OSP).

Hence, while accepting retrocausality allows us to retain CLI, it may lead us to invoke additional non-relativistic spacetime structure. This tension with the principle of relativity undermines the initial motivation for accepting retrocausal influences. If one were to find a physically plausible way of implementing retrocausality without the use of a preferred frame or similar spacetime structure in order to account for CHSH violations, this would significantly alter the dialectical situation. At present, no such approach is known to us.

Furthermore, as we will see in the following, the rejection of CLI does not have to be wholesale. It is possible to preserve much of the motivation underlying CLI without generating a contradiction via Gisin's theorem by instead accepting the Lorentz invariance of causal connections.

5.1.3 Lorentz Invariance of Causal Connection

CLI states that the underlying causal structure is identical in all Lorentz frames. In the spirit of logical conservatism, we want to preserve as much of the content of CLI. If we give up too much, our study of nonlocal and relativistic explanations of the Bell scenario becomes rather pointless.

A natural relaxation of CLI is a condition which we will refer to as *Lorentz Invariance of Causal Connections* (LICC). It can be stated as follows:

Definition 5.1 (Lorentz Invariance of Causal Connection)**.** If X and Y are directly causally connected in some Lorentz frame $\mathcal{F}$ then X and Y are directly causally connected in all Lorentz frames.

Here, 'X and Y are directly causally connected' simply means that either X is a cause of Y or *vice versa*. This condition allows the direction of arrows to depend on the choice of Lorentz frame while enforcing that the skeleton of the causal structure must remain invariant.

Formally, while CLI states that $G_{\mathcal{F}} = G_{\mathcal{F}'}$ for all Lorentz frames $\mathcal{F}$ and $\mathcal{F}'$, where $G_{\mathcal{F}}$ is the causal graph in $\mathcal{F}$, LICC only claims that $\mathrm{skel}(G_{\mathcal{F}}) = \mathrm{skel}(G_{\mathcal{F}'})$, where $\mathrm{skel}(G)$ denotes the skeleton of the directed graph G. When discussing frame-dependent causal structures it is often necessary, in order to characterize them completely, to consider them in multiple frames of reference. This will become relevant in the following sections, when we will show that one cannot easily strengthen LICC any further without running into serious problems.

5.1.4 A Note on Determinism

We will now discuss why, in the case of deterministic models, adopting a deterministic variant of LICC is not of much help. This further clarifies that the deterministic case is actually qualitatively different from the probabilistic case. Even though it is possible to reject CLI and hold on to LICC in response to Gisin's theorem in the case of probabilistic theories, one cannot do so in the deterministic case. This implies that, given the relevant assumptions, the most natural way of responding to Gisin's theorem entails indeterminism.

To see this, recall Gisin's original argument for the deterministic case. Again, consider the most general deterministic model of the Bell scenario in the frames $\mathcal{F}_{AB}$ and $\mathcal{F}_{BA}$ that respects statistical independence and NRC:

$$a = f(x, \lambda)\,, \qquad b = g(x, y, \lambda)\,. \tag{5.1}$$

$$a' = f'(x', y', \lambda')\,, \qquad b' = g'(y', \lambda')\,. \tag{5.2}$$

Consider a single run of the experiment. Then, as a matter of consistency with the operationally accessible settings and outcomes, we must have that $a = a'$, and similar for all additionally primed and unprimed operational variables:

$$a = f(x, \lambda)\,, \qquad b = g(x, y, \lambda)\,. \tag{5.3}$$

$$a = f'(x, y, \lambda')\,, \qquad b = g'(y, \lambda')\,. \tag{5.4}$$

It directly follows that

$$a = f(x, \lambda)\,, \qquad b = g'(y, \lambda')\,. \tag{5.5}$$

This is already a Bell-local model, since λ' cannot functionally depend on x, as this would violate NRC[1]. Bell locality becomes even more manifest by defining $\tilde{\lambda} = (\lambda, \lambda')$, which leads to

$$a = \tilde{f}(x, \tilde{\lambda}) \equiv f(x, \lambda)\,, \qquad b = \tilde{g}(y, \tilde{\lambda}) \equiv g'(y, \lambda')\,. \tag{5.6}$$

Hence, in the deterministic case Gisin's argument leaves no room for giving up the Lorentz invariance of values, since doing so is in conflict with the empirical adequacy of the model. Furthermore, it is entirely unclear what a sufficiently weakened version of the Lorentz Invariance of values, analogous to LICC, would even be. This implies that the proposed way of responding to Gisin's theorem by giving up (causal) Lorentz

[1] This is because if $\lambda' = h(x')$, then $b' = g'(y', \lambda') = g'(y', h(x')) \equiv k(y', x')$. But this is in violation of NRC as b' is determined before x' in $\mathcal{F}_{BA}$.

invariance and holding on to the Lorentz invariance of connections does not go through in the deterministic case.

5.2 Characterizing Frame Dependence

5.2.1 Causal Discovery and Effective Causal Models

The problem of causal discovery concerns which underlying causal structures are compatible with a given probability distribution over operationally accessible variables. To render this task tractable, the inferred causal connections and latent variables are subjected to naturalness constraints. An example of this is provided by Wood and Spekkens [26], who point out that one need not consider causal structures wherein a latent variable merely acts as a mediator between operational variables. This is because any probability distribution over the operational variables can be accounted for by a model that simply posits a direct causal influence. Furthermore, latent variables that are only common effects of operational variables also leave the distribution over the latter unchanged. Hence, they conclude that '[l]atent variables have nontrivial consequences for the observed distribution only when they act as common causes of the observed variables' [26].

In the following we are interested in finding nonlocal and relativistic causal explanations of the Bell scenario. Not every possible causal structure compatible with the operational statistics and the physical requirements we impose is explanatory in the required sense. Consider, for example, a causal model with a latent common effect. Such an additional hidden variable does nothing to explain the observable behavior in the Bell scenario. Hence, we will restrict ourselves to causal models which do not exhibit unnecessary structure. Such causal models will be referred to as *effective causal models* and can be defined as follows:

Definition 5.2 (Effective Causal Model)**.** An effective causal model is any causal model with a causal structure that satisfies:

(i) No latent nodes which are only mediators between operational variables.

(ii) No latent nodes which are only common effects of operational variables.

Equivalently, effective causal models only allow for additional latent nodes if they serve as common causes of operationally accessible variables.

5.2.2 Towards a Uniqueness Theorem

In the following, we prove a *uniqueness theorem* for causal explanations of the Bell scenario. It states that, given a small set of natural assumptions, which we briefly motivate below, exactly one causal structure is compatible with the operational statistics observed in Bell experiments.

As a response to Gisin's theorem, we opted to give up CLI in favor of NRC while remaining committed to the strictly weaker LICC. We further assume strong measurement independence (SMI). Although this assumption is strictly stronger than the condition of WMI employed so far, it still follows from the principle of free experimentation. Finally, we restrict attention to effective causal models, since we are only interested in causal structures that actually play an explanatory role.

Regarding the operational statistics, we take the distribution $p(a, b \mid x, y)$ observed in the Bell scenario to satisfy $2 < S_{\text{CHSH}} \leq 4$, meaning that the correlations violate Bell locality while respecting the no-signaling constraints. This is in agreement with the results of experimental Bell tests discussed in Chapter 3. We are now in a position to state the central result of this thesis:

Theorem 5.3 (Uniqueness)**.** *There exists exactly one effective causal model that reproduces the operational statistics in a Bell scenario and satisfies:*

(i) Strong Measurement Independence,

(ii) No-Retrocausality,

(iii) Lorentz Invariance of Causal Connections.

The model is given by the pair of frame-dependent causal structures shown below:

Figure 5.1: The unique effective causal structure compatible with Bell-violating, no-signaling statistics under assumptions (i)–(iii).

Proof. We proceed in two steps. First, we show that for effective causal models, under the assumptions (i)–(iii), no additional causal connections are compatible. Second, we use the fact that the operational statistics observed in the Bell scenario violate Bell locality and respect no-signaling to show that all of the arrows in Figure 5.1 must correspond to an actual causal influence. It follows that the structure shown in Figure 5.1 is the unique effective causal model satisfying (i)–(iii) that reproduces the operational statistics of the Bell scenario.

Assume the most permissive generalized causal structure of the Bell scenario consisting only of operationally accessible quantities. Because we do not assume that the causal structure is frame independent, we have to consider the causal structure in multiple frames in order to characterize it completely and allow for all possible time orderings. Because of the spacetime locations of the operational quantities in the Bell scenario, it is sufficient to consider two frames $\mathcal{F}_{AB}$ and $\mathcal{F}_{BA}$. In $\mathcal{F}_{AB}$, Alice performs her experiment before Bob, and vice versa in $\mathcal{F}_{BA}$:

Figure 5.2: The most permissible generalized causal structure of the operational Bell scenario.

By (i) and (ii), the causal structure is restricted to

Figure 5.3: The generalized causal structure constrained by (i) and (ii).

Because there is no direct causal connection between X and B in $\mathcal{F}_{BA}$ and no direct causal connection between Y and A in $\mathcal{F}_{AB}$, (iii) forbids these connections in all frames.

Figure 5.4: The generalized causal structure further constrained by (iii).

Because we only consider effective causal models, any latent random variable Λ must act as a common cause. SMI in only allows Λ to be a common cause of A and B. Hence, the most permissible structure compatible with (i)–(iii) is given by the following diagram:

Figure 5.5: The most permissive effective generalized causal structure compatible with (i)–(iii)

Note that in order for Λ to serve as a common cause of A and B in all frames, Λ must lie in the past light cone of both A and B. This means that we do not gain anything by including additional reference frames in our analysis.

We will now show that the arrows between the operational quantities are thick, meaning that the operational statistics enforce the existence of the relevant causal connections. We have to consider

(1) $A \to B$ in $\mathcal{F}_{AB}$,

(2) $B \to A$ in $\mathcal{F}_{BA}$,

(3) $X \to A$,

(4) $Y \to B$.

In both (1) and (2), if we were to remove the arrows, then there would be no path leading from the other wing of the experiment to A and B, respectively. Because both of the settings have no parents, Lemma 3.10 implies that the operational statistics must be Bell-local. Hence, the arrows connecting A and B must be thick.

In the case of (3), the removal of the arrow would imply that $(A, B \perp\!\!\!\perp X \mid Y, \Lambda)$ via the Markov condition. By the fact that $(X, Y \perp\!\!\!\perp \Lambda)$ we obtain $(A, B \perp\!\!\!\perp X \mid Y)$:

$$p(a, b \mid x, y) = \int_{\mathbf{\Lambda}} \mathrm{d}\lambda \, p(a, b \mid x, y, \lambda) \, p(\lambda \mid x, y) \tag{5.7}$$

$$= \int_{\mathbf{\Lambda}} \mathrm{d}\lambda \, p(a, b \mid y, \lambda) \, p(\lambda) \tag{5.8}$$

$$= p(a, b \mid y) \,. \tag{5.9}$$

Hence,

$$E(x, y) = \sum_{a,b} ab \, p(a, b \mid x, y) = \sum_{a,b} ab \, p(a, b \mid y) = E(x) \,, \tag{5.10}$$

which implies

$$|S_{\mathrm{CHSH}}| = |E(0) + E(0) + E(1) - E(1)| = |2E(0)| \leq 2 \,. \tag{5.11}$$

The case of (4) is completely analogous. Hence, removing any of the arrows between the operational variables leads to a Bell-local model, which contradicts the operational statistics. Now we consider the remaining arrows connecting the common cause Λ to A and B. We assume that both arrows are missing. Hence, we obtain the following joint probability in the frame $\mathcal{F}_{AB}$:

$$p(a, b, x, y) = p(x) \, p(y) \, p(a \mid x) \, p(b \mid a, y) \,. \tag{5.12}$$

When evaluating the marginal $p(b \mid x, y)$ using the joint probability derived from the causal structure in $\mathcal{F}_{AB}$, we obtain

$$p(b \mid x, y) = \sum_{a} p(a \mid x) \, p(b \mid a, y) \,. \tag{5.13}$$

We have assumed that $p(a, b \mid x, y)$ respects the operational statistics recorded in actual experiments. This means that the operational statistics violate Bell locality but must also respect the no-signaling constraints. From no-signaling it follows that

$$\begin{aligned} 0 &= p(b \mid x=0, y) - p(b \mid x=1, y) && (5.14)\\ &= \sum_a \left(p(a \mid x=0) - p(a \mid x=1)\right) p(b \mid a, y) && (5.15)\\ &= (p(a=1 \mid x=0) - p(a=1 \mid x=1))(p(b \mid a=1, y) - p(b \mid a=-1, y)) \,. && (5.16) \end{aligned}$$

Hence,

$$p(a=1 \mid x=0) - p(a=1 \mid x=1) = 0 \vee p(b \mid a=1, y) - p(b \mid a=-1, y) = 0 \,. \quad (5.17)$$

The first case implies

$$p(a \mid x) = p(a) \,. \quad (5.18)$$

Hence,

$$\begin{aligned} E(x,y) &= \sum_{a,b} ab\, p(a \mid x)\, p(b \mid a, y) && (5.19)\\ &= \sum_{a,b} ab\, p(a)\, p(b \mid a, y) = E(y) \,. && (5.20) \end{aligned}$$

This again leads to

$$|S_{\mathrm{CHSH}}| \leq 2 \,. \quad (5.21)$$

In the second case we have

$$p(b \mid a, y) = p(b \mid y) \,. \quad (5.22)$$

Hence,

$$p(a, b \mid x, y) = p(a \mid x)\, p(b \mid a, y) = p(a \mid x)\, p(b \mid y) \,, \quad (5.23)$$

which is a trivially Bell-local behavior. The same is true in the frame $\mathcal{F}_{BA}$. Hence, we must have a non-trivially acting hidden variable Λ in order to reproduce the operational statistics. This implies that Λ must be a common cause of A and B.

In the case of a common cause we can reproduce the operational statistics, as the causal structure is sufficiently permissive to simulate the PR box, or any no-signaling behavior for that matter. To see this, consider the following construction in the case of $\mathcal{F}_{AB}$. An analogous construction can be given for $\mathcal{F}_{BA}$. Here we will, for the sake of simplicity, take $a, b \in \{0, 1\}$ instead of $\{-1, 1\}$. Let the common cause Λ contain a uniformly distributed random bit U. Define

$$a = x \oplus u \,, \qquad b = a \oplus (a \oplus u) y \,, \quad (5.24)$$

as allowed by the causal structure in $\mathcal{F}_{AB}$. This assignment satisfies the defining condition of the PR box

$$a \oplus b = a \oplus (a \oplus (a \oplus u)y) = (a \oplus u)y = ((x \oplus u) \oplus u)y = xy\,. \tag{5.25}$$

It is easy to see that the relevant marginals are also satisfied,

$$p(a \mid x) = p(b \mid y) = \frac{1}{2}\,. \tag{5.26}$$

The model is therefore no-signaling. It follows that this causal structure can simulate a PR box and hence achieve $|S_{\mathrm{CHSH}}| = 4$. Since the model can reach $|S_{\mathrm{CHSH}}| = 4$, it can realize any value in the interval $[0, 4]$ by convex mixing with the completely random behavior $p(a, b \mid x, y) = \frac{1}{4}$.

To see this, let Λ additionally contain another random bit V with $p(v = 0) = p \in [0, 1]$. Now, if $v = 0$, implement the PR strategy stated above. If instead $v = 1$, implement the completely random behavior. The resulting CHSH value is

$$|S_{\mathrm{CHSH}}| = 4p \in [0, 4]\,, \tag{5.27}$$

where the Tsirelson bound is obtained by setting $p = \frac{1}{\sqrt{2}}$.

Hence, the causal structure given in Figure 5.1 is the unique effective causal model which respects (i)–(iii) and is able to reproduce the operational statistics □

5.2.3 Properties of the Unique Causal Explanation

In the following we characterize several interesting properties of the unique causal structure derived in the previous subsection. We distinguish between causal properties, which concern the structure of the causal graph itself, and probabilistic properties, which concern the probability distributions compatible with this structure. For convenience, we reproduce the relevant causal structure below:

Figure 5.6: The unique effective causal structure compatible with the relevant requirements.

The following causal properties seem to be especially noteworthy: First, and unsurprisingly so, both causal structures violate local causality (LC). Second, and equally unsurprising, is the fact that the causal structure can differ between frames. But the way in which CLI is violated is interesting. In agreement with LICC the skeleton of both graphs are identical. Only the direction of the allowed nonlocal influence between A and B depends on the choice of reference frame. Due to NRC, whenever A is determined before B, A is taken to be the cause of B, and vice versa. This suggests a very natural way to extend this kind of reasoning beyond the Bell scenario which is jointly implied by LICC and NRC: Given two spacelike separated but directly causally connected nodes X_1 and X_2, and denote the times at which they are determined as t_1 and t_2 respectively. Whenever $t_1 < t_2$, X_1 causes X_2 and vice versa.

Another causal property is the agreement between the causal explanation given above and the ontic symmetry principle (OSP). Recall that the OSP forces symmetries in the operational statistics to be preserved on the ontological level. As discussed in Chapter 2, the operational statistics $p(a, b \mid x, y)$ are symmetric under swap of parties. Hence, the causal structure which underlies the Bell scenario should also respect this symmetry. Taken individually, each of the two causal structures violates this symmetry as in $\mathcal{F}_{AB}$ Alice's outcome A is a cause of Bob's outcome B while in $\mathcal{F}_{BA}$ the converse is true. If one instead considers both structures together, the causal explanation respects the symmetry in question. This is why this nonlocal explanation does not single out of a preferred frame.

It is interesting to note that if one, instead of giving up CLI, rejects NRC, there exists only one effective causal model compatible with SMI and the OSP. This is exactly given by the structure used as an example for a causal explanation of the Bell scenario which satisfies the OSP in Chapter 2:

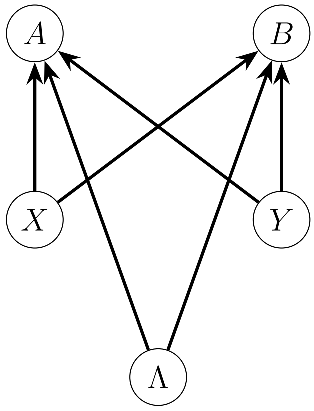


Figure 5.7: The only effective causal explanation of the Bell scenario which respects SMI, CLI and the OSP.

To see this, first note that by giving up NRC as a response to Gisin's theorem, one can preserve CLI and therefore specify the causal structure independently of the choice of Lorentz frame. Furthermore, SMI rules out any arrows going into X and Y. The OSP then forces Alice's side to mirror Bob's. This rules out any direct causal influence between A and B, since such an influence would introduce a causal cycle. Hence, the only remaining nonlocal arrows are $X \to B$ and $Y \to A$. Removing only one of these arrows would violate the OSP, removing both would render the model Bell-local. Similarly for the arrows $X \to A$ and $Y \to B$. By assuming effective causal models and SMI, Λ can only act as a common cause of A and B. It can be shown that Λ is required to ensure the compatibility with the no-signaling constraints. Hence, only the causal structure pictured above remains viable given the aforementioned assumptions. We note that this structure, in contrast to the one identified by the uniqueness theorem, does not immediately suggest a generalization beyond the Bell scenario.

We will begin our discussion of the probabilistic properties by noting that these types of explanation require genuine indeterminism. This is because the framework of deterministic models is too narrow to allow for frame-dependent causal influence while remaining empirically adequate. We will defer to our discussion above on this point.

The nonlocal causal structure of this explanation gives rise to the following independence relations: in $\mathcal{F}_{AB}$, the Markov condition implies $(A \perp\!\!\!\perp Y \mid X, \Lambda)$ while B might still depend on X given Λ through mediation by A. In $\mathcal{F}_{BA}$ it follows that $(B \perp\!\!\!\perp X \mid Y, \Lambda)$, while A might still depend on Y given Λ through mediation by B. In both classes of frames, outcome independence generically fails. If one assumes that probabilities must be Lorentz-invariant, it follows that the joint probability distributions of both structures must be identical. Hence, PI must hold generally and the violation of locality must be due to a failure of OI.

But one quickly notices the following disturbing fact: The causal structures in $\mathcal{F}_{AB}$ and $\mathcal{F}_{BA}$ do not allow for the same joint probability distribution $p(a,b,x,y,\lambda)$ if one assumes empirically adequate operationally statistics. If one forces the structures in $\mathcal{F}_{AB}$ and $\mathcal{F}_{BA}$ to generate the same ontic distribution, that distribution must be Bell-local.

Theorem 5.4. *If the causal structures, shown in Figure 5.6, give rise to the same joint probability $p(a,b,x,y,\lambda)$ in $\mathcal{F}_{AB}$ and $\mathcal{F}_{BA}$, with a,b,x,y binary variables, then the operational statistics $p(a,b \mid x,y)$ must be Bell-local.*

Proof. By the Markov condition we obtain in frame $\mathcal{F}_{AB}$ and $\mathcal{F}_{BA}$ respectively

$$p(a,b,x,y,\lambda) = p(x)\,p(y)\,p(\lambda)\,p(a \mid x,\lambda)\,p(b \mid a,y,\lambda)\,, \tag{5.28}$$

$$p(a,b,x,y,\lambda) = p(x)\,p(y)\,p(\lambda)\,p(a \mid b,x,\lambda)\,p(b \mid y,\lambda)\,. \tag{5.29}$$

We assume the joint probabilities in both frames are identical. This implies the following condition

$$\frac{p(a \mid b,x,\lambda)}{p(a \mid x,\lambda)} = \frac{p(b \mid a,y,\lambda)}{p(b \mid y,\lambda)}\,. \tag{5.30}$$

Because the left-hand side does not depend on y, the right-hand side must also be independent of y.

$$\frac{p(b \mid a,y=0,\lambda)}{p(b \mid y=0,\lambda)} = \frac{p(b \mid a,y=1,\lambda)}{p(b \mid y=1,\lambda)}\,. \tag{5.31}$$

For $b = 1$ we obtain

$$\frac{p(1 \mid a,y=0,\lambda)}{p(1 \mid y=0,\lambda)} = \frac{p(1 \mid a,y=1,\lambda)}{p(1 \mid y=1,\lambda)}\,. \tag{5.32}$$

Analogously, we set $b = -1$. By making use of normalization we find that

$$\frac{1-p(b=1 \mid a,y=0,\lambda)}{1-p(b=1 \mid y=0,\lambda)} = \frac{1-p(b=1 \mid a,y=1,\lambda)}{1-p(b=1 \mid y=1,\lambda)}\,. \tag{5.33}$$

It should be noted that we have implicitly assumed that the relevant denominators are non-zero. Furthermore, we have as usual assumed that $a,b \in \{-1,1\}$ and $x,y \in \{0,1\}$. In the following we will use the conventions stated below

$$\alpha_y = p(b=1 \mid a,y,\lambda)\,, \qquad \beta_y = p(b=1 \mid y,\lambda)\,. \tag{5.34}$$

Hence, we have the following system of equations

$$\alpha_0\beta_1 = \alpha_1\beta_0\,, \tag{5.35}$$

$$\alpha_0 + \beta_1 - \alpha_0\beta_1 = \alpha_1 + \beta_0 - \alpha_1\beta_0\,. \tag{5.36}$$

By using the first equation the second equation reduces to

$$\alpha_1 = \alpha_0 + \beta_1 - \beta_0\,. \tag{5.37}$$

By eliminating α_1 from the first equation we obtain

$$\alpha_0\beta_1 = (\alpha_0 + \beta_1 - \beta_0)\beta_0\,, \tag{5.38}$$

which is equivalent to

$$(\alpha_0 - \beta_0)(\beta_1 - \beta_0) = 0 \tag{5.39}$$

Hence either

$$\alpha_0 = \beta_0 \qquad \vee \qquad \beta_1 = \beta_0\,. \tag{5.40}$$

In the first case

$$\alpha_1 - \alpha_0 = \beta_1 - \beta_0\,, \tag{5.41}$$

implies

$$\alpha_1 = \beta_1\,, \tag{5.42}$$

and therefore

$$\alpha_y = \beta_y\,. \tag{5.43}$$

This puts the following restrictions on the conditional probabilities

$$p(b \mid a, y, \lambda) = p(b \mid y, \lambda)\,, \tag{5.44}$$

which guarantees Bell-local operational statistics

$$p(a, b \mid x, y) = \int_{\mathbf{\Lambda}} \mathrm{d}\lambda\, p(\lambda)\, p(a \mid x, \lambda)\, p(b \mid a, y, \lambda) = \int_{\mathbf{\Lambda}} \mathrm{d}\lambda\, p(\lambda)\, p(a \mid x, \lambda)\, p(b \mid y, \lambda)\,. \tag{5.45}$$

In the second case we have

$$\beta_0 = \beta_1\,, \tag{5.46}$$

and hence

$$p(b \mid y, \lambda) = p(b \mid \lambda)\,. \tag{5.47}$$

This also leads to Bell-local statistics as

$$E(x,y) = \sum_{a,b} ab\, p(a,b \mid x,y) \tag{5.48}$$

$$= \sum_{a,b} ab \int_{\mathbf{\Lambda}} \mathrm{d}\lambda\, p(\lambda)\, p(a \mid b,x,\lambda)\, p(b \mid y,\lambda) \tag{5.49}$$

$$= \sum_{a,b} ab \int_{\mathbf{\Lambda}} \mathrm{d}\lambda\, p(\lambda)\, p(a \mid b,x,\lambda)\, p(b \mid \lambda) \equiv E(x)\,, \tag{5.50}$$

which forces $|S_{\text{CHSH}}| \leq 2$:

$$|S_{\text{CHSH}}| = |E(0) + E(0) + E(1) - E(1)| = |2E(0)| \leq 2\,. \tag{5.51}$$

This concludes the proof. □

This result seems to contradict the fact that rGRWf provides a Lorentz-invariant probability law which (approximately) recovers the predictions of quantum theory. In the subsequent section we analyze the Bell scenario through the lens of rGRWf and explain why this result does not apply to that model. The most important properties identified above can be summarized as follows:

(i) Causal Connections are independent of reference frame.

(ii) Causal direction dependent on relative time order prohibits retrocausality.

(iii) No preferred reference frame.

(iv) Genuine indeterminism.

(v) Outcome independence fails.

(vi) Parameter independence holds.

In the following, we will see how rGRWf concretely realizes these properties.

5.3 The Bell Scenario in rGRWf

5.3.1 The rGRWf Model

In this subsection we briefly review the developments that led to the rGRWf model and explain how it reconciles nonlocality with relativity. The technical aspects of our exposition mostly follow Tumulka's presentation [64], while many of the conceptual pointers are taken from Bell [65, 66] and Maudlin [67]. For a more in-depth treatment of rGRWf we refer to [8, 64, 68, 18].

The original Ghirardi-Rimini-Weber (GRW) model was introduced to provide a precise account of when wave function collapse occurs. In orthodox quantum theory the wave function is said to collapse when a measurement takes place. The difficulties associated, with this way of speaking, usually referred to as the *measurement problem*, have often been noted, perhaps most vividly by Bell:

> It would seem that the theory is exclusively concerned about 'results of measurement', and has nothing to say about anything else. What exactly qualifies some physical systems to play the role of 'measurer'? Was the wavefunction of the world waiting to jump for thousands of millions of years until a single celled living creature appeared? Or did it have to wait a little longer, for some better qualified system ... with a Ph.D.? If the theory is to apply to anything but highly idealized laboratory operations, are we not obliged to admit that more or less 'measurement-like' processes are going on more or less all the time, more or less everywhere? Do we not have jumping then all the time? [66]

GRW theory provides a concrete answer to the questions raised by Bell. In the GRW model each particle undergoes spontaneous localization events that occur randomly. The central idea behind GRW-type models is that the Schrödinger equation does not provide a complete description of the dynamics of quantum systems. While unitary evolution still governs the dynamics between collapse events, the overall dynamics also includes additional non-linear and genuinely indeterministic localization processes.

GRW models can be supplemented with different *primitive ontologies*, i.e., variables describing matter in spacetime. An example of a primitive ontology is the particle ontology of classical mechanics. Another example is the matter density ontology, where the matter content is given by a continuous distribution $m(\vec{x}, t)$. In the following we will consider GRW models with a flash ontology, which we will characterize shortly. The reason for supplementing the abstract scheme outlined above with a primitive ontology is the following: we want our theories to describe the physical world. In particular, they should give an account of ordinary material objects in spacetime. If a theory lacks a primitive ontology, the connection between its mathematical formalism and the behavior of ordinary material objects often remains opaque. Introducing a primitive ontology makes explicit how ordinary material objects are constituted and thereby clarifies how the predictions of the theory relate to the macroscopic world [64].

In the somewhat baroque *flash ontology*, first suggested by Bell [66], matter is described by discrete flashes, meaning elementary events represented by space-time points. In this view, ordinary physical matter consists of many of these flashes in spacetime. In GRW models with flash ontology (GRWf) a flash $X = (\vec{x}, t)$ exists at the center of any wave function collapse event.

The mathematical structure of non-relativistic GRWf for N distinguishable, non-interacting particles can be sketched as follows [64][2]: For N distinguishable particles it is useful to introduce N types of flashes, one for each particle. Initially, the wave function ψ at time t_0 is given, which is a normalized vector in a Hilbert space $\mathcal{H} = \bigotimes_{i=1}^{N} \mathcal{H}_i$. Furthermore, seed flashes for each flash type $\{X_{i,0}\}_{i=1}^{N}$ are also given. For each flash type there exists a flash rate operator $\Lambda_i(\vec{x})$. The flash rate for particle type i at $\vec{x} \in \mathbb{R}^3$, i.e. the probability per unit time that a flash of type i occurs at $\vec{x}$, is given by

$$\langle\psi|\, \Lambda_i(\vec{x})\, |\psi\rangle \; . \tag{5.52}$$

In GRWf, the flash rate operators $\Lambda_i(\vec{x})$ act as multiplication by Gaussians in the position basis:

$$\Lambda_i(\vec{x})\psi(\vec{x}_1, \ldots, \vec{x}_N) = \frac{1}{(2\pi\sigma^2)^{3/2}\tau} e^{-\frac{(\vec{x}-\vec{x}_i)^2}{2\sigma^2}} \psi(\vec{x}_1, \ldots, \vec{x}_N)\,, \tag{5.53}$$

where σ is the width of the localization and τ the average time between two collapses of the same type. Typical values for these parameters are $\sigma \approx 10^{-7}$ meters and $\tau \approx 10^{8}$ years. The Hamiltonian $H = \sum_{i=1}^{N} H_i$ defines the time evolution between collapse events.

We will denote a sequence of events $x_0, \ldots, x_n$ in spacetime as $f = (x_0, \ldots, x_n)$. The collapse operator $C_{x'}(x)$ which encodes the evolution from a previous collapse event x' to the next collapse event x is given by

$$C_{x'}(x) = \Lambda(\vec{x})^{\frac{1}{2}} W_{t-t'} \tag{5.54}$$

with

$$W_{t-t'} = \begin{cases} e^{-iH(t-t') - \frac{1}{2}\int \mathrm{d}^3x\, \Lambda(\vec{x})\, t-t'} & \text{if } t - t' \geq 0 \\ 0 & \text{else} \end{cases} \; . \tag{5.55}$$

An important property of the collapse operator $C_{x'}(x)$ is that

$$\int \mathrm{d}^4x\, C_{x'}(x)^* C_{x'}(x) = \mathbb{1}\,. \tag{5.56}$$

[2]We will restrict ourselves to distinguishable and non-interacting particles as they are sufficient to model the Bell scenario.

A sequence f of n collapses will be denoted by $C(f)$ and is given by

$$C(f) = C_{x_{n-1}}(x_n) \dots C_{x_0}(x_1)\,. \tag{5.57}$$

The generalization to N particle types is straightforward. For N non-interacting systems with different flash types, the joint distribution of the first n_i flashes of each type i is given by

$$p\Big(X_{i,k} \in \mathrm{d}^4 x_{i,k} : i \leq N,\ k \leq n_i\Big) = \left\| \left(\bigotimes_{i=1}^{N} C_{f_i} \right) \psi \right\|^2 \mathrm{d}f\,, \tag{5.58}$$

where $X_{i,k} = (\vec{x}_{i,k}, t_{i,k})$ denotes the k-th flash of type i, and

$$\mathrm{d}f = \prod_{i=1}^{N} \prod_{k=1}^{n_i} \mathrm{d}^4 x_{i,k}\,. \tag{5.59}$$

A very important property of GRWf is its relative time translation invariance which can be understood as a limit of Lorentz invariance [65]. To see this consider the one-dimensional Lorentz transformations in natural units

$$x' = \gamma(x - vt)\,, \qquad t' = \gamma(t - vx)\,. \tag{5.60}$$

For small v and large $|x|$ we have

$$vx \approx \mp a\,, \qquad \gamma \approx 1\,, \tag{5.61}$$

and hence the relevant Lorentz transformations reduce to

$$x' \approx x\,, \qquad t' \approx t \pm a\,. \tag{5.62}$$

This means that for very widely separated systems Lorentz invariance implies relative time invariance, which can be shown to hold in GRWf [66]. This led Bell to the following conclusion about the possibility of a relativistic generalization of GRWf:

> [...] I am particularly struck by the fact that the model is as Lorentz invariant as it could be in the nonrelativistic version. It takes away the ground of my fear that any exact formulation of quantum mechanics must conflict with fundamental Lorentz invariance. [66]

While the relative time translation invariance of GRWf is certainly encouraging in this respect, a fully relativistic extension of GRWf was only worked out by Tumulka in

2006 [8].

In order to generalize GRWf to the relativistic case, such a generalization must only make use of structures inherent to Minkowski spacetime. The classical spacetime structure is invoked at two points in the non-relativistic model: The probability of collapses is specified with respect to a universal time t and the Gaussian of the collapse operator $\Lambda(\vec{x})$ is defined on a simultaneity surface, again with respect to the universal time [67].

Tumulka realized that structures native to Minkowski spacetime $\mathbb{M}$ together with a set of seed flashes can function sufficiently similar to universal time in non-relativistic GRWf theories. Consider a seed flash X_0. We want the next flash X_1 to lie inside its future lightcone $J^+(X_0)$. Every such flash occurs at some proper time from X_0. Surfaces of constant proper time form future hyperboloids $\mathbb{H}(X_0, X_1)$ which foliate the interior of $J^+(X_0)$. Because X_1 must lie in $J^+(X_0)$, this foliation plays a role analogous to the foliation provided by the universal time t in the non-relativistic GRWf model. This enables us, given a seed flash and an initial wave function, to define probabilities for the flash occurring on any given part of a chosen hyperboloid using the relativistic Dirac equation. Iterating this process yields the probability for any finite sequence of flashes. The generalization to N distinguishable particles without interaction is straightforward.

After this introduction to the main ideas of rGRWf we will specify some of its most pertinent mathematical details [64]. Given two spacelike 3-surfaces Σ_1 and Σ_2 in $\mathbb{M}$, let

$$U_{\Sigma_1}^{\Sigma_2} : L^2(\Sigma_1) \to L^2(\Sigma_2)\,, \tag{5.63}$$

denote the unitary evolution of the wave function from Σ_1 to Σ_2, which is defined by the Dirac equation

$$(i\gamma_i^\mu \partial_{i,\mu} - m_i)\, \psi(x_1, \ldots, x_N) = 0\,. \tag{5.64}$$

Let the flash rate operator $\Lambda_\Sigma(x)$ on a spacelike 3-surface Σ be given by

$$\Lambda_\Sigma(x)\psi(y) = \frac{\mathcal{N}}{\tau} \exp\left(-\frac{\text{s-dist}_\Sigma^2(x,y)}{2\sigma^2}\right) \psi(y)\,, \tag{5.65}$$

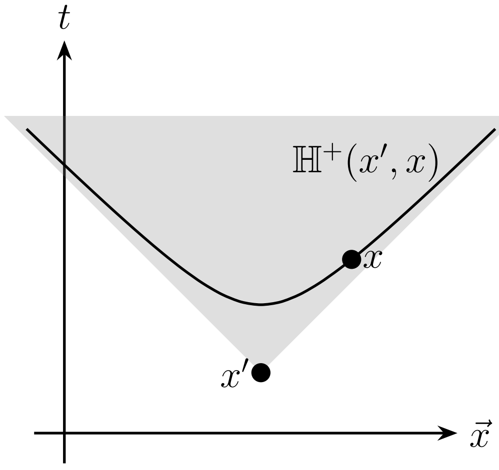
with normalization

$$\int_\Sigma \mathrm{d}^3x \Lambda_\Sigma(x) = \frac{1}{\tau} \tag{5.66}$$

In the following, we will take Σ to be the future hyperboloid based at the previous flash x'

$$\Sigma \equiv \mathbb{H}^+(x', x) = \left\{ y \in J^+(x') \mid |y - x'| = |x - x'| \right\}\,, \tag{5.67}$$

as shown in the diagram given below:



The collapse operator can then be specified as follows:

$$C_{x'}(x) = \mathbb{1}_{x\in J^+(x')}\, e^{\frac{-|x'-x|}{2\tau}}\, U_{\Sigma}^{\Sigma_0} \Lambda_{\Sigma}(x)^{\frac{1}{2}} U_{\Sigma_0}^{\Sigma}\,. \tag{5.68}$$

The factor $\mathbb{1}_{x\in J^+(x')}$ ensures that flashes of the same type are timelike separated. The factor $e^{\frac{-|x'-x|}{2\tau}}$ expresses the exponential distribution of the proper waiting time. The remaining terms are the Heisenberg evolution of the square root of the collapse rate operator $\Lambda_\Sigma(x)$. Again, initial conditions consisting of a seed flash of every type and an initial wave function on a spacelike 3-surface must be specified. The joint distribution of the first n_i flashes of each type i is then, in complete analogy to the non-relativistic case, given by

$$p\Big(X_{i,k} \in \mathrm{d}^4 x_{i,k} : i \leq N,\ k \leq n_i\Big) = \left\| \left(\bigotimes_{i=1}^{N} C_{f_i} \right) \psi \right\|^2 \mathrm{d}f\,. \tag{5.69}$$

This concludes the definition of the rGRWf model, which is manifestly Lorentz-invariant since it is formulated entirely in terms of relativistic spacetime structure.

5.3.2 A Comparison between rGRWf and Previous Results

In the following we give a brief account of the description of the Bell scenario in rGRWf. We will then go on to compare the properties of this description with those recovered in the previous section.

Consider an rGRWf model of two distinguishable, non-interacting particles of type A and B. The Bell scenario is realized by two spacelike separated Stern–Gerlach experiments which can each be oriented in one of two different directions, corresponding to the settings $x, y \in \{0, 1\}$. The Bell scenario is then governed by a global Hamiltonian

$$H = H_A(x) + H_B(y)\,, \tag{5.70}$$

where $H_A(x)$ is the local Hamiltonian of Alice's set-up, which depends on her measurement setting x, while $H_B(y)$ is the local Hamiltonian of Bob's set-up, which depends on his measurement setting y. Hence, we do not have to explicitly include the measurement apparatus or the observers in the model.

We are interested in the first flashes of type A and B, respectively. The local Hamiltonians are such that for Alice's first flash there are four possible spacetime regions. If the apparatus is oriented according to $x = 0$, the flash X_A is found either in the spacetime region $\mathcal{A}_{0,-1}$ or $\mathcal{A}_{0,1}$, depending on whether the spin is determined to be down or up. If the apparatus is oriented according to $x = 1$, X_A is found either in $\mathcal{A}_{1,-1}$ or $\mathcal{A}_{1,1}$. Hence,

$$X_A \in \{\mathcal{A}_{x,a} \mid x \in \{0,1\}, a \in \{-1,1\}\} = \{\mathcal{A}_{0,-1}, \mathcal{A}_{0,1}, \mathcal{A}_{1,-1}, \mathcal{A}_{1,1}\}\,. \tag{5.71}$$

Similarly,

$$X_B \in \{\mathcal{B}_{0,-1}, \mathcal{B}_{0,1}, \mathcal{B}_{1,-1}, \mathcal{B}_{1,1}\}\,. \tag{5.72}$$

The settings, and hence the corresponding local Hamiltonians, are freely chosen at random with

$$p(x) = p(y) = \frac{1}{2}\,. \tag{5.73}$$

Furthermore, the initial wave function ψ_0 is in a singlet state and two seed flashes $X_{A,0}$ and $X_{B,0}$ are specified. In the following, they will serve as the common cause λ, meaning that

$$\lambda \equiv (\psi_0, X_{A,0}, X_{B,0})\,. \tag{5.74}$$

Using the formal machinery outlined in the previous subsection, it is now possible to derive a conditional probability $p(X_A, X_B \mid x, y, \lambda)$ which is independent of the reference frame:

$$p(X_A, X_B \mid x, y, \lambda) = \left\| \Big(C_{X_{A,0}}(X_A, x) \otimes C_{X_{B,0}}(X_B, y) \Big)\, \psi_0 \right\|^2\,. \tag{5.75}$$

It is well known that this distribution violates outcome independence while respecting parameter independence [17, 19]. To see this consider the following marginal

$$p(X_A \mid x, y, \lambda) = \int \mathrm{d}^4 X_B \left\| \Big(C_{X_{A,0}}(X_A, x) \otimes C_{X_{B,0}}(X_B, y) \Big) \psi_0 \right\|^2 \tag{5.76}$$

$$= \langle \psi_0 | \, C_{X_{A,0}}(X_A, x)^* C_{X_{A,0}}(X_A, x) \otimes \int \mathrm{d}^4 X_B \, C_{X_{B,0}}(X_B, y)^* C_{X_{B,0}}(X_B, y) \, |\psi_0\rangle \tag{5.77}$$

$$= \langle \psi_0 | \, C_{X_{A,0}}(X_A, x)^* C_{X_{A,0}}(X_A, x) \otimes \mathbb{1}_B \, |\psi_0\rangle \tag{5.78}$$

$$= p(X_A \mid x, \lambda) \, . \tag{5.79}$$

By completely analogous reasoning one can show that

$$p(X_B \mid x, y, \lambda) = p(X_B \mid y, \lambda) \, . \tag{5.80}$$

Since rGRWf reproduces the Bell correlations while satisfying parameter independence, outcome independence must be violated:

$$p(X_A \mid X_B, x, y, \lambda) \neq p(X_A \mid x, y, \lambda) \, , \tag{5.81}$$

Furthermore, it is easy to see that the following independence relations must be satisfied:

$$p(X_A \mid X_B, x, y, \lambda) = p(X_A \mid X_B, x, \lambda) \, , \tag{5.82}$$

$$p(X_B \mid X_A, x, y, \lambda) = p(X_B \mid X_A, y, \lambda) \, . \tag{5.83}$$

This is because the locations of the flashes carry the information about their respective settings. If, for example, $X_A \in \mathcal{A}_{0,-1}$, then it must be the case that $x = 0$. Hence, if X_B is specified, conditioning on y provides no additional information about X_A, and similarly for the converse case.

This enables a factorization of the joint probability as required by the unique causal explanation derived above:

$$p(X_A, X_B \mid x, y, \lambda) = p(X_A \mid X_B, x, y, \lambda) \, p(X_B \mid x, y, \lambda) \tag{5.84}$$

$$= p(X_A \mid X_B, x, \lambda) \, p(X_B \mid y, \lambda) \, , \tag{5.85}$$

$$p(X_A, X_B \mid x, y, \lambda) = p(X_B \mid X_A, x, y, \lambda) \, p(X_A \mid x, y, \lambda) \tag{5.86}$$

$$= p(X_A \mid x, \lambda) \, p(X_B \mid X_A, y, \lambda) \, . \tag{5.87}$$

By applying parameter independence we obtain the joint distribution generated by the unique causal explanation

$$p(X_A, X_B, x, y, \lambda) = p(\lambda)\, p(x)\, p(y)\, p(X_A \mid X_B, x, \lambda)\, p(X_B \mid y, \lambda) \tag{5.88}$$
$$= p(\lambda)\, p(x)\, p(y)\, p(X_A \mid x, \lambda)\, p(X_B \mid X_A, y, \lambda)\,. \tag{5.89}$$

At first glance, this factorization seems to imply that the statistics $p(X_A, X_B \mid x, y)$ must be Bell-local by Theorem 5.4. But this result rests on the assumption that the outcomes are binary random variables. Both X_A and X_B can each take one of four possible values. Hence, the argument does not go through and $p(X_A, X_B \mid x, y)$ can violate Bell locality. Furthermore, in agreement with the unique causal explanation of the Bell scenario, rGRWf does not single out a preferred frame or any other non-relativistic spacetime structure and is genuinely indeterministic.

The rGRWf model does not itself possess a native notion of causal structure but it can be naturally supplemented with one. We have already seen that rGRWf allows for a factorization compatible with the unique causal explanation.

Now consider the Bell scenario in the frame $\mathcal{F}_{AB}$. In $\mathcal{F}_{AB}$ the spacetime location of X_A is only influenced by the orientation of Alice's Stern–Gerlach apparatus specified by the setting x, and by initial data λ if one prohibits retrocausal influences. Bob's flash X_B is influenced by X_A via the updated wave function, which can be inferred from the location of X_A together with ψ_0, and by the orientation y of his apparatus. Hence, it is natural in $\mathcal{F}_{AB}$ to describe the Bell scenario via the following causal structure:

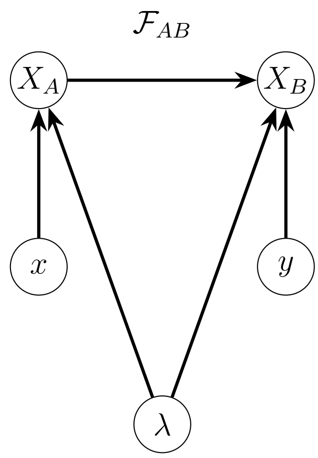


Figure 5.8: The natural causal explanation of the Bell scenario in rGRWf, as seen from $\mathcal{F}_{AB}$

An analogous causal structure can be recovered in the case of $\mathcal{F}_{BA}$. This is exactly the causal explanation recovered by the uniqueness theorem given earlier. Hence, if one chooses to interpret rGRWf in a causal way, its description of the Bell scenario most naturally involves exactly the causal structure that is the subject of our uniqueness theorem. The question whether one should interpret rGRWf in a causal way will be briefly discussed in the conclusion. In summary, rGRWf offers a description of the Bell scenario which, rather unsurprisingly, either exhibits or is compatible with the properties of the unique causal explanation.

Chapter 6

Conclusion

The central concern of this thesis has been the tension between nonlocality and special relativity. More concretely, we have investigated what constraints are necessarily imposed upon theories which are both nonlocal and genuinely relativistic.

Our investigation of this conflict began with an exploration of nonlocality. This led us to provide a rigorous formulation of the EPRB argument for n parties, showing that any outcome-independent ontological model of the EPRB scenario that reproduces perfect (anti-)correlations must be locally deterministic. We then provided a derivation of Bell's theorem from purely causal principles, demonstrating that no empirically adequate causal model of the Bell scenario can jointly satisfy (i) weak measurement independence, (ii) local causality, and (iii) no-backward causality.

To clarify the tension with relativity, we presented a generalization of Gisin's argument beyond deterministic theories using causal models. This result shows that no empirically adequate causal model of the Bell scenario can jointly satisfy (i) weak measurement independence, (ii) causal Lorentz invariance, and (iii) no-retrocausality.

Our discussion of possible responses to Gisin's theorem led to a uniqueness result identifying the only effectively causal and empirically adequate explanation of the Bell scenario that satisfies (i) strong measurement independence, (ii) Lorentz invariance of causal connections, and (iii) no-retrocausality. Finally, this unique causal explanation of the Bell scenario was compared with the account provided by rGRWf and found to be in agreement with it. Taken together, these results clarify the structure of the tension between nonlocality and relativity. We now turn to several general lessons that can be extracted from the analysis presented in this thesis.

If one adopts a broadly realist account, one must accept violations of faithfulness. Wood and Spekkens [26] have shown that any empirically adequate causal explanation of the Bell scenario must involve fine-tuning, while Ying et al. [28] have demonstrated that quantum, and even more general, causal models cannot preserve faithfulness in the Wigner's friend scenario. Hence, if one wishes to provide a causal account of the Bell scenario, there is little reason to prefer non-classical causal models over classical ones, since the former cannot generally restore faithfulness. Among the possible causal explanations of the Bell correlations, we have argued that nonlocality provides the most reasonable option. The failure of faithfulness also has methodological consequences. In particular, it implies that the Leibnizian methodological principle must be abandoned, since this principle entails faithfulness [25]. We have argued that the ontic symmetry principle can serve many of the same purposes without prohibiting fine-tuning.

We have seen that, in order to adequately express Gisin's argument, one requires a general framework that provides an asymmetric notion of causal influence. Gisin's original formulation relied on the relation of functional dependence, thereby restricting the result to deterministic models. Because statistical independence is symmetric, Gisin's argument cannot be naturally formulated within the framework of ontological models. By contrast, causal models offer an asymmetric notion of influence and allow Gisin's argument to be extended naturally to probabilistic theories.

As discussed above, the assumptions underlying Gisin's theorem are weak measurement independence (WMI), no-retrocausality (NRC), and causal Lorentz invariance (CLI). Because NRC implies no-backward causality (NBC), the conjunction of WMI and NBC entails, by Bell's theorem, the existence of nonlocal causal influences. However, if the underlying causal structure is Lorentz-invariant, the relativity of simultaneity implies that such nonlocal influences correspond to retrocausal influences in some reference frames. But this is prohibited by NRC. This contradiction forces us to abandon at least one of the aforementioned assumptions.

We have advocated abandoning CLI in favor of NRC as a response to this dilemma. The reason is that objections raised against backward causality, such as the threat of causal paradoxes, also apply to retrocausal explanations. The most natural way of avoiding these difficulties is to posit a preferred frame of reference. But doing so seems to contradict the spirit of relativity even more severely than abandoning CLI. To preserve as much as possible of the motivation behind CLI, we retain the Lorentz invariance of causal connections. As we have shown, this assumption naturally leads to the uniqueness theorem for causal explanations of the Bell scenario.

The uniqueness theorem shows that the space of reasonable causal explanations of the Bell scenario that are both nonlocal and meaningfully relativistic is quite small. The properties of the unique causal explanation include violations of local causality (LC) and CLI, genuine indeterminism, the absence of a preferred frame, Lorentz-invariant causal connections, frame dependence of the direction of causal influence consistent with NRC, and certain independence relations in the joint ontic probability distribution. The account of the Bell scenario provided by rGRWf is in overall agreement with these features, even though the theory itself does not possess a native notion of causal influence.

Some authors have argued that the very notion of causation ceases to be well-defined in rGRWf. Recent examples of this view are provided by Allori and Tumulka respectively:

> In fact, in GRW-type theories the distinction between cause and effect as we commonly understand it seems to vanish, so that this principle does not even apply. [...] [B]ecause of their stochasticity, in GRW-type theories cause and effect can be thought of as symmetrical: nothing can be regarded as the cause and the other as the effect in an obvious way. [19]

> Now the claim is that the theory does not have to specify which of the two ways nature uses and that nature does not have to use either of the two ways. [...] Thus, there is no need for a direction of influence; it is enough if the fundamental physical theory prescribes the joint distribution. [18]

Allori argues that, due to the apparent symmetry of cause and effect in rGRWf, the distinction between cause and effect breaks down. Tumulka goes further by suggesting that there is no need to specify a direction of influence and presumably no need for causal explanation at the level of fundamental physics.

In response to Allori's concern, our account makes clear that one does not have to forsake causation entirely. By allowing the direction of causal influence to depend on the choice of Lorentz frame in order to comply with NRC, one can maintain a frame-indexed asymmetry of causal influence.

We agree with Tumulka that causal notions are not necessary ingredients for the formulation of fundamental physical theories. However, the framework of causal models, and the additional structure it supplies over and above probabilities, has proven well suited for understanding the relation between nonlocality and relativity. Bell's theorem can be clarified as a result about the limits of causal explanation. Gisin's theorem, and especially the uniqueness theorem, cannot naturally be stated in strictly statistical language.

Furthermore, there is independent theoretical support for a more prominent role of causality in fundamental physics. Spekkens points out that in physical theories different choices of kinematics can be accommodated, without changing their empirical predictions, by modifying the dynamics and vice versa. He concludes that 'the distinction between kinematics and dynamics [. . .] is purely conventional and should be abandoned' [69]. He instead proposes that the notion of causal structure may serve as a promising candidate for a more fundamental characterization of physical reality[1].

Tumulka's remark quoted above seems to suggest some sort of eternalist view. According to eternalism, sometimes called the block universe view, all events in spacetime exist on equal terms. Nothing ever truly comes into being or ceases to exist. This position is often motivated by an appeal to special relativity. More explicitly, a well-known argument due to Rietdijk and Putnam shows that, given the relativity of simultaneity and plausible assumptions about shared truth values between observers, events in the future of any given observer must already be determinate [71, 72]. This seems to contradict the possibility of genuinely indeterministic events[2].

One should note that eternalism is not inconsistent with the form of dynamical indeterminism exhibited by rGRWf. While it is true under eternalism that the spacetime location of a flash X is determinate even before the flash occurs, rGRWf may simply be unable to tell us where X will be located given our current knowledge, and can only predict its location probabilistically. Hence, ontic determinacy does not imply deterministic predictability. Although eternalism is not logically incompatible with the indeterminism exhibited by rGRWf, it sits somewhat uneasily with the dynamics of rGRWf, in which flashes are generated sequentially according to probabilistic laws.

Moreover, the argument for eternalism from relativity hinted at above makes use of an assumption referred to by Del Santo and Gisin as *present reality*. It states that any two distant observers at relative rest attribute the same truth values to empirical propositions about events lying on the same plane of simultaneity in their rest frame [73]. Del Santo and Gisin reject this principle and instead advocate the relativity of indeterminacy, according to which there are regions of spacetime to which different inertial observers attribute different probabilities for the outcomes of experiments. Only in the overlap of their future light cones, when their predictions become empirically comparable, must they agree [73]. Del Santo and Gisin thereby reconcile the relativity of simultaneity with an open future.

This reconciliation between an open future and relativity mirrors the reconciliation between nonlocality and relativity advocated in this thesis. The former is achieved by relativizing indeterminacy, while the latter is achieved by relativizing the direction of

[1]For a complementary perspective on this topic see, for example, Adlam [70].
[2]For a presentation of the argument following Del Santo and Gisin [73], see Appendix B.

causal influence[3]. Both provide a similar sort of answer to the question 'How does nature do it?' [75]: Differently in different frames yet in a way that preserves overall agreement.

We hope that the hints revealed by the arguments of Del Santo and Gisin, Bell's theorem, Gisin's theorem, and especially the uniqueness theorem may prove useful in orienting the search for future physics. While rGRWf coheres remarkably well with many of these results, the general structure of GRW-type dynamics, which consist of unitary evolution supplemented by random wave-function collapses, appears somewhat ad hoc and lacks a convincing physical motivation. Furthermore, although rGRWf has been extended to include indistinguishable particles and interactions [68], a formulation capable of recovering quantum field theory has yet to be developed. Similarly to Bohmian mechanics, we therefore tentatively regard rGRWf and other relativistic collapse theories as important proofs of principle rather than as the final story of microscopic reality. This naturally suggests promising directions for future research. In this spirit, we conclude with the following remark by Bell:

> However that may be, long may Louis de Broglie continue to inspire those who suspect that what is proved by impossibility proofs is lack of imagination. [14]

[3]It is interesting that both realist and antirealist approaches to quantum theory may lead to a relativity of aspects of physical reality that are generally assumed to be objective. Realist approaches, by arguments such as those discussed in this thesis or those provided by Del Santo and Gisin [73], may be forced to accept the relativity of causal direction and/or the relativity of indeterminacy. By contrast, antirealist approaches may lead to relativity, or more precisely observer dependence, of facts about the world, as suggested by arguments based on Wigner's friend scenarios [74, 53, 28].

Appendix A

Proofs

A.1 Generalization of Theorem 3.3

To generalize the EPRB argument to n parties we first need to define the notion of an n-party EPRB scenario:

Definition A.1 (Generalized EPRB scenario)**.** The generalized EPRB scenario is an operational situation with no inputs and n outputs $a_1, \ldots, a_n \in \{-1, 1\}$ which are pairwise either perfectly correlated or perfectly anti-correlated, meaning either

$$p(a_i, a_j) = \mu_{ij}\, \delta_{a_i,-1}\delta_{a_j,-1} + (1 - \mu_{ij})\, \delta_{a_i,1}\delta_{a_j,1}\,, \tag{A.1}$$

or

$$p(a_i, a_j) = \mu_{ij}\, \delta_{a_i,-1}\delta_{a_j,1} + (1 - \mu_{ij})\, \delta_{a_i,1}\delta_{a_j,-1}\,, \tag{A.2}$$

for all outcomes $a_i, a_j \in \{-1, 1\}$ with $\mu_{ij} \in [0, 1]$.

Using this definition, the argument given in Chapter 2 can be extended in the following way.

Theorem A.2. *Any outcome-independent ontological model of the generalized EPRB scenario is locally deterministic. That is, for almost all $\lambda \in \mathbf{\Lambda}$, the conditional probabilities satisfy*

$$p(a_i \mid \lambda) = \delta_{a_i, a_i(\lambda)} \in \{0, 1\}\,, \tag{A.3}$$

for all $a_i \in \{-1, 1\}$ and all $i \in \{1, \ldots, n\}$.

Proof. We begin by defining a linear functional, which we will call the generalized

EPRB functional,

$$S_{\mathrm{EPRB}n} = \sum_{1\le i<j\le n} \sum_{a_i,a_j} a_i a_j \begin{cases} p(a_i,a_j) & \text{if } a_i, a_j \text{ are correlated,} \\ -p(a_i,a_j) & \text{if } a_i, a_j \text{ are anti-correlated} \end{cases} . \tag{A.4}$$

In the generalized EPRB scenario, each pair (i,j) contributes 1 to $S_{\mathrm{EPRB}n}$, so

$$S_{\mathrm{EPRB}n} = \sum_{1\le i<j\le n} 1 = \frac{n(n-1)}{2} . \tag{A.5}$$

This is also the largest possible value of $S_{\mathrm{EPRB}n}$ attainable on valid probability distributions, since for each pair (i,j) one has

$$\left| \sum_{a_i,a_j} a_i a_j \, p(a_i,a_j) \right| \le 1 . \tag{A.6}$$

Now assume an outcome independent ontological model of the generalized EPRB scenario, which is of the form

$$p(a_1,\dots,a_n) = \int_{\mathbf{\Lambda}} \mathrm{d}\lambda \, p(\lambda) \prod_{i=1}^{n} p(a_i \mid \lambda) . \tag{A.7}$$

The marginal probabilities $p(a_i,a_j)$ are given by

$$p(a_i,a_j) = \sum_{\{a_k \mid k\neq i,j\}} p(a_1,\dots,a_n) , \tag{A.8}$$

where the sum runs over all outcomes a_k with $k \neq i,j$. Substituting the ontological model description, we obtain

$$\begin{aligned} p(a_i,a_j) &= \sum_{\{a_k \mid k\neq i,j\}} \int_{\mathbf{\Lambda}} \mathrm{d}\lambda \, p(\lambda) \prod_{l=1}^{n} p(a_l \mid \lambda) && \text{(A.9)} \\ &= \int_{\mathbf{\Lambda}} \mathrm{d}\lambda \, p(\lambda) \, p(a_i \mid \lambda) \, p(a_j \mid \lambda) . && \text{(A.10)} \end{aligned}$$

We evaluate $S_{\mathrm{EPRB}n}$ in the ontological model and obtain

$$S_{\mathrm{EPRB}n} = \sum_{1\le i<j\le n} \sum_{a_i,a_j} a_i a_j \begin{cases} \int_{\mathbf{\Lambda}} \mathrm{d}\lambda \, p(\lambda) \, p(a_i \mid \lambda) \, p(a_j \mid \lambda) & \text{if } a_i, a_j \text{ are correlated,} \\ -\int_{\mathbf{\Lambda}} \mathrm{d}\lambda \, p(\lambda) \, p(a_i \mid \lambda) \, p(a_j \mid \lambda) & \text{if } a_i, a_j \text{ are anti-correlated} \end{cases} \tag{A.11}$$

$$= \sum_{1\le i<j\le n} \int_{\mathbf{\Lambda}} \mathrm{d}\lambda \, p(\lambda) \, f_{ij}(\lambda) , \tag{A.12}$$

where, for each pair (i,j), we defined the function f_{ij} as follows

$$f_{ij}(\lambda) \equiv \sum_{a_i,a_j} a_i a_j \begin{cases} p(a_i \mid \lambda)\, p(a_j \mid \lambda) & \text{if } a_i, a_j \text{ are correlated,} \\ -p(a_i \mid \lambda)\, p(a_j \mid \lambda) & \text{if } a_i, a_j \text{ are anti-correlated} \end{cases} . \tag{A.13}$$

We first consider the case where a_i and a_j are correlated. For convenience, set

$$x = p(a_i = 1 \mid \lambda)\,, \qquad y = p(a_j = 1 \mid \lambda)\,. \tag{A.14}$$

Then

$$f_{ij}(\lambda) = xy - x(1-y) - (1-x)y + (1-x)(1-y) \tag{A.15}$$

$$= (1-2x)(1-2y)\,. \tag{A.16}$$

Since $x, y \in [0,1]$, it follows that $1-2x$, $1-2y \in [-1,1]$, and hence

$$-1 \leq f_{ij}(\lambda) \leq 1\,. \tag{A.17}$$

The maximal value $f_{ij}(\lambda) = 1$ occurs if and only if $(x,y) \in \{(0,0),(1,1)\}$, that is, only in the locally deterministic case where $p(a_i \mid \lambda) = p(a_j \mid \lambda) \in \{0,1\}$. Similarly, when a_i and a_j are anti-correlated, the extra minus sign yields

$$f_{ij}(\lambda) = -(1-2x)(1-2y)\,, \tag{A.18}$$

so again $-1 \leq f_{ij}(\lambda) \leq 1$, and the maximal value $f_{ij}(\lambda) = 1$ occurs if and only if $(x,y) \in \{(0,1),(1,0)\}$, that is, only in the deterministic case where $p(a_i \mid \lambda), p(a_j \mid \lambda) \in \{0,1\}$ with $p(a_i \mid \lambda) \neq p(a_j \mid \lambda)$. In all cases we therefore have, for each pair (i,j) and all $\lambda \in \mathbf{\Lambda}$,

$$f_{ij}(\lambda) \leq 1\,. \tag{A.19}$$

It follows that

$$S_{\mathrm{EPRB}n} = \sum_{1\leq i<j\leq n} \int_{\mathbf{\Lambda}} \mathrm{d}\lambda\, p(\lambda)\, f_{ij}(\lambda) \tag{A.20}$$

$$\leq \sum_{1\leq i<j\leq n} \int_{\mathbf{\Lambda}} \mathrm{d}\lambda\, p(\lambda) \tag{A.21}$$

$$= \frac{n(n-1)}{2}\,. \tag{A.22}$$

But in the generalized EPRB scenario we have already seen that

$$S_{\mathrm{EPRB}n} = \frac{n(n-1)}{2}\,. \tag{A.23}$$

Thus the inequality must be saturated, and hence for each pair (i, j) we must have

$$\int_{\mathbf{\Lambda}} \mathrm{d}\lambda\, p(\lambda)\, f_{ij}(\lambda) = 1\,. \tag{A.24}$$

Since $f_{ij}(\lambda) \leq 1$ for all $\lambda \in \mathbf{\Lambda}$ and $\int_{\mathbf{\Lambda}} \mathrm{d}\lambda\, p(\lambda) = 1$ with $p(\lambda) \geq 0$, we must have

$$f_{ij}(\lambda) = 1 \tag{A.25}$$

for almost all $\lambda \in \mathbf{\Lambda}$ and for all pairs (i, j). As shown above, this is possible only if, for almost all λ, the conditional probabilities $p(a_i \mid \lambda)$ and $p(a_j \mid \lambda)$ are deterministic for all $i, j \in \{1, \ldots, n\}$. Hence, for almost all $\lambda \in \mathbf{\Lambda}$ and all $i \in \{1, \ldots, n\}$,

$$p(a_i \mid \lambda) \in \{0, 1\}\,, \tag{A.26}$$

which amounts to local determinism. □

A.2 Lemma 3.17

The local set $\mathcal{L}$ is defined as

Definition A.3 (Local Set)**.**

$$\mathcal{L} = \left\{ p(a, b \mid x, y) \,\middle|\, p(a, b \mid x, y) = \int_{\Lambda} \mathrm{d}\lambda\, p(a \mid x, \lambda)\, p(b \mid y, \lambda)\, p(\lambda) \right\}\,, \tag{A.27}$$

where $p(a, b \mid x, y)$ is a valid probability.

It can be shown that $\mathcal{L}$ is a polytope with the local deterministic behaviors as its extremal points:

Theorem A.4. *The local set $\mathcal{L}$ is a polytope whose extremal points are the local deterministic behaviors, given by*

$$p(a, b \mid x, y) = \delta_{a,a(x)}\, \delta_{b,b(y)}\,. \tag{A.28}$$

where $a(x), b(y) \in \{-1, 1\}$ are deterministic response functions depending only on the settings x and y, respectively.

Proof. We know from Fine's theorem that any $p(a, b \mid x, y) \in \mathcal{L}$ can be expressed as a convex combination of local deterministic behaviors. Hence, only local deterministic behaviors can be extremal points of $\mathcal{L}$. Assume for an indirect proof, that a local deterministic behavior $\delta_{a,a(x)}\,\delta_{b,b(y)}$ can be written as a proper convex combination of local deterministic behaviors:

$$\delta_{a,a(x)}\,\delta_{b,b(y)} = \mu\,\delta_{a,a_1(x)}\,\delta_{b,b_1(y)} + (1-\mu)\,\delta_{a,a_2(x)}\,\delta_{b,b_2(y)}\,, \tag{A.29}$$

with $\delta_{a,a_1(x)}\,\delta_{b,b_1(y)} \neq \delta_{a,a_2(x)}\,\delta_{b,b_2(y)}$ and $0 < \mu < 1$. This is only possible if the two behaviors differ for at least one pair of settings $(\tilde{x}, \tilde{y})$. For these settings, there are outcomes $(a', b') \neq (a'', b'')$ such that $\delta_{a',a_1(\tilde{x})}\,\delta_{b',b_1(\tilde{y})} = 1$ and $\delta_{a'',a_2(\tilde{x})}\,\delta_{b'',b_2(\tilde{y})} = 1$. This implies that $\delta_{a',a(\tilde{x})}\,\delta_{b',b(\tilde{y})} = \mu$ and $\delta_{a'',a(\tilde{x})}\,\delta_{b'',b(\tilde{y})} = 1-\mu$. But this is impossible as $\delta_{a,a(x)}\,\delta_{b,b(y)}$ is a local deterministic behavior and can only take values in $\{0, 1\}$. Hence, we have reached a contradiction and conclude by indirect proof that $\mathcal{L}$ is a polytope with the local deterministic behaviors as its extremal points. □

Theorem A.4 can be used to prove that Bell-local behaviors must satisfy the CHSH inequality.

Lemma A.5 (CHSH)**.** *Any behavior $p(a, b \mid x, y)$ that is Bell-local satisfies the CHSH inequality:*

$$|S_{\text{CHSH}}| = |E(0,0) + E(0,1) + E(1,0) - E(1,1)| \leq 2\,. \tag{A.30}$$

Proof. Because $E(0,0) + E(0,1) + E(1,0) - E(1,1)$ is linear in the probabilities, it takes its maximal value over $\mathcal{L}$ at the extremal points of $\mathcal{L}$, which are the local deterministic behaviors. In the case of deterministic behaviors we find:

$$E(x,y) = \sum_{a,b} ab\,\delta_{a,a(x)}\,\delta_{b,b(y)} = a(x)b(y) \tag{A.31}$$

Hence, the CHSH functional becomes:

$$\begin{aligned} S_{\text{CHSH}} &= E(0,0) + E(0,1) + E(1,0) - E(1,1) && \text{(A.32)}\\ &= a(0)b(0) + a(0)b(1) + a(1)b(0) - a(1)b(1) && \text{(A.33)}\\ &= a(0)(b(0) + b(1)) + a(1)(b(0) - b(1))\,. && \text{(A.34)} \end{aligned}$$

Now because $b(0)$ and $b(1)$ must both take values in {-1, 1}, we either have that $b(0) = b(1)$ or that $b(0) = -b(1)$. In the first case we obtain:

$$|a(0)(b(0) + b(1)) + a(1)(b(0) - b(1))| = |2a(0)b(0)| = 2\,. \tag{A.35}$$

Similarly for the second case:

$$|a(0)(b(0)+b(1))+a(1)(b(0)-b(1))| = |-2a(1)b(0)| = 2\,. \tag{A.36}$$

We conclude that the maximal value of the absolute value of the CHSH functional over $\mathcal{L}$ is 2. Hence, for any arbitrary $p(a,b \mid x,y) \in \mathcal{L}$, the CHSH inequality must hold:

$$|E(0,0)+E(0,1)+E(1,0)-E(1,1)| \leq 2\,. \tag{A.37}$$

□

A.3 Theorem 3.19

Theorem A.6 (Tsirelson)**.** *The maximal violation of the CHSH inequality achievable by quantum theory is* $2\sqrt{2}$.

Proof. We begin by assuming that the statistics $p(a,b \mid x,y)$ obtained in the Bell scenario, are explained by quantum theory. Hence, there exists a Hilbert space $\mathcal{H}$ and a quantum state $|\psi\rangle \in \mathcal{H}$ that describe the Bell scenario. Furthermore, there must exist two binary observables $E_{a|x}, F_{b|y} \in \mathcal{B}(\mathcal{H})$, with eigenvalues in $\{1,-1\}$, which satisfy $E_{a|x} = E_{a|x}^2 = E_{a|x}^\dagger$ and analogously for $F_{b|y}$. Because of the spacelike separation of the wings of the experiment, the commutator of the two observables must vanish: $[E_{a|x},\, F_{b|y}] = 0$. The statistics can then be expressed as follows:

$$p(a,b \mid x,y) = \langle\psi|\, E_{a|x}\, F_{b|y}\, |\psi\rangle\ . \tag{A.38}$$

We will now define two operators A_x and B_y:

$$A_x = E_{1|x} - E_{-1|x}\,, \tag{A.39}$$

$$B_y = F_{1|y} - F_{-1|y}\,. \tag{A.40}$$

It is easy to verify that they satisfy the following relations:

$$A_x^2 = B_y^2 = \mathbb{1}\,, \tag{A.41}$$

$$\langle\psi|\, A_x\, B_y\, |\psi\rangle = E(x,y)\,, \tag{A.42}$$

where E is the correlation function. We will now go on to define three additional operators, namely Z, Z_1, and Z_2 in terms of A_x and B_y:

$$Z \equiv Z_1^2 + Z_2^2 \tag{A.43}$$

$$\equiv \left(\frac{A_0 + A_1}{\sqrt{2}} - B_0\right)^2 + \left(\frac{A_0 - A_1}{\sqrt{2}} - B_1\right)^2 \tag{A.44}$$

$$= 4\mathbb{1} - \sqrt{2}(A_0B_0 + A_0B_1 + A_1B_0 - A_1B_1)\,. \tag{A.45}$$

Due to this definition, Z must be positive semi-definite:

$$\langle\psi| Z |\psi\rangle = \langle\psi| Z_1^2 |\psi\rangle + \langle\psi| Z_2^2 |\psi\rangle \geq 0\,. \tag{A.46}$$

Furthermore we find:

$$\langle\psi| Z |\psi\rangle = \langle\psi| 4\mathbb{1} |\psi\rangle - \langle\psi| \sqrt{2}(A_0B_0 + A_0B_1 + A_1B_0 - A_1B_1) |\psi\rangle \tag{A.47}$$

$$= 4 - \sqrt{2}(E(0,0) + E(0,1) + E(1,0) - E(1,1))\,. \tag{A.48}$$

From this, via the positive semi-definiteness of Z, it follows that:

$$4 - \sqrt{2}(E(0,0) + E(0,1) + E(1,0) - E(1,1)) \geq 0\,. \tag{A.49}$$

This yields the following upper bound on S_{CHSH}

$$S_{\text{CHSH}} = E(0,0) + E(0,1) + E(1,0) - E(1,1) \leq 2\sqrt{2}\,. \tag{A.50}$$

The lower bound can be obtained by an analogous argument, which establishes the Tsirelson bound:

$$|S_{\text{CHSH}}| = |E(0,0) + E(0,1) + E(1,0) - E(1,1)| \leq 2\sqrt{2}\,. \tag{A.51}$$

$\square$

Appendix B

Additional Material

B.1 Von Neumann's Impossibility Proof

In the following, we aim to give a reconstruction of von Neumann's well-known impossibility argument [11], mostly following Ballentine [76]. Here, we will restrict our considerations to finite-dimensional Hilbert spaces. In what follows, we will always take $\mathcal{H}$ to be a finite-dimensional complex Hilbert space with $\dim(\mathcal{H}) \geq 2$ and let

$$\mathcal{A} = \{A \in \mathcal{L}(\mathcal{H}) \mid A = A^\dagger\}\,, \tag{B.1}$$

be the real vector space of Hermitian operators.

Definition B.1 (Density operator)**.** A *density operator* on a Hilbert space $\mathcal{H}$ is a linear operator $\rho \in \mathcal{L}(\mathcal{H})$ satisfying:

(i) Hermiticity: $\rho = \rho^\dagger$.

(ii) Positivity: $\rho \geq 0$.

(iii) Normalization: $\mathrm{Tr}(\rho) = 1$.

We will present von Neumann's argument via two connected theorems. The first shows that if a functional $E : \mathcal{A} \to \mathbb{R}$ satisfies certain conditions, then there exists a unique density operator such that $E(A)$ is given by the Born rule. We will make use of an assumption which is not included in von Neumann's original presentation but simplifies the argument. We assume that $E(\mathbb{1}) = 1$, a condition referred to as the normalization of E.

Theorem B.2 (Von Neumann's Representation Theorem)**.** *Let $E : \mathcal{A} \to \mathbb{R}$ be a functional satisfying for all $A, B \in \mathcal{A}$:*

(i) Linearity: For all $a, b \in \mathbb{R}$, $E(aA + bB) = aE(A) + bE(B)$.

(ii) Positivity: If $A \geq 0$, then $E(A) \geq 0$.

(iii) Normalization: $E(\mathbb{1}) = 1$.

Then there exists a unique density operator ρ such that for all $A \in \mathcal{A}$,

$$E(A) = \mathrm{Tr}(\rho A)\,. \tag{B.2}$$

Proof. Since $\mathcal{A}$ is a finite-dimensional real Hilbert space under the Hilbert-Schmidt inner product

$$\langle A, B\rangle = \mathrm{Tr}(AB)\,, \tag{B.3}$$

the Riesz-Fréchet representation theorem guarantees that for every linear functional E there exists a unique operator $\rho \in \mathcal{A}$ such that

$$E(A) = \mathrm{Tr}(\rho A)\,, \tag{B.4}$$

for all $A \in \mathcal{A}$. We now show that ρ is a density operator: Hermiticity of ρ follows immediately from $\rho \in \mathcal{A}$. To show positivity of ρ, let $\phi \in \mathcal{H}$ be any unit vector. The rank-one projector $P_\phi = |\phi\rangle\langle\phi|$ satisfies $P_\phi \geq 0$, so by (ii),

$$0 \leq E(P_\phi) = \mathrm{Tr}(\rho P_\phi) = \langle\phi|\rho|\phi\rangle\,. \tag{B.5}$$

Since this holds for all unit vectors ϕ, we conclude that $\rho \geq 0$. Finally, using (iii), we deduce normalization of ρ as follows:

$$\mathrm{Tr}(\rho) = \mathrm{Tr}(\rho\mathbb{1}) = E(\mathbb{1}) = 1\,. \tag{B.6}$$

Thus ρ satisfies all defining properties of a density operator, and hence

$$E(A) = \mathrm{Tr}(\rho A)\,, \tag{B.7}$$

for all $A \in \mathcal{A}$, where ρ is a unique density operator. □

We now introduce the notion of dispersion-free functionals which is needed for the second part of von Neumann's argument:

Definition B.3. A functional $E : \mathcal{A} \to \mathbb{R}$ is *dispersion-free* if for every $A \in \mathcal{A}$,

$$E(A^2) = E(A)^2 \,. \tag{B.8}$$

This allows us to state von Neumann's impossibility proof:

Theorem B.4 (Von Neumann's impossibility theorem)**.** *There exists no dispersion-free functional $E : \mathcal{A} \to \mathbb{R}$ satisfying (i)–(iii).*

Proof. For an indirect proof, assume that there exists a dispersion-free functional E. By Theorem B.2, assumptions (i)–(iii) imply that there exists a density operator ρ such that

$$E(A) = \mathrm{Tr}(\rho A) \,, \tag{B.9}$$

for all observables A. We now show that this leads to a contradiction by setting A equal to $P_\phi = |\phi\rangle \langle\phi|$ with $\langle\phi|\phi\rangle = 1$. We find

$$E(P_\phi) = \mathrm{Tr}(\rho P_\phi) = \langle\phi| \rho |\phi\rangle \,, \tag{B.10}$$

which implies by the dispersion-freeness of E that

$$E(P_\phi^2) = \langle\phi| \rho |\phi\rangle = (\langle\phi| \rho |\phi\rangle)^2 = E(P_\phi)^2 \,. \tag{B.11}$$

Hence, $\langle\phi| \rho |\phi\rangle \in \{0, 1\}$. This is only possible for arbitrary $|\phi\rangle \in \mathcal{H}$ if $\rho = 0$ or $\rho = \mathbb{1}$. But both of these cases contradict the requirement of ρ being a valid density operator:

$$\mathrm{Tr}(0) = 0 \neq 1 \,, \tag{B.12}$$

$$\mathrm{Tr}(\mathbb{1}) = \dim(\mathcal{H}) \geq 2 \neq 1 \,. \tag{B.13}$$

Therefore, no dispersion-free functional E satisfying (i)–(iii) exists. □

Up to this point, the discussion has been entirely formal. The functional E has been treated as an abstract map on the algebra of observables, and von Neumann's representation theorem shows that, under assumptions (i)–(iii), any such functional must be of the form

$$E(A) = \mathrm{Tr}(\rho A) \,. \tag{B.14}$$

Using this result, von Neumann's impossibility theorem then establishes that no dispersion-free functional E satisfying (i)–(iii) can exist.

To connect this result to the hidden-variable question, one must regard $E(A)$ as representing the expectation value of the observable A in some underlying physical state. In particular, the position that a hidden variable theory should be dispersion-free, and that the expectation value functional E should satisfy the assumptions (i)–(iii) is ruled out by von Neumann's result. While this result is mathematically correct and conceptually interesting, as it recovers the Born rule, its relevance to the hidden variable question hinges on the plausibility of the relevant assumptions.

Bell pointed out that the same result can be deduced from strictly weaker assumptions, without making use of positivity and replacing normalization with $E(\mathbb{1}) \neq 0$, in a more elementary way. This argument from strictly weaker premises will make it easier to discuss the options available to proponents of hidden variable theories in response to these impossibility proofs. The following proof is a more formal version of arguments that can be found in [77, 78] with an explicit generalization to Hilbert spaces with $\dim(\mathcal{H}) > 2$.

Theorem B.5 (Bell's Impossibility Theorem)**.** *There exists no dispersion-free functional $E : \mathcal{A} \to \mathbb{R}$ satisfying for all $A, B \in \mathcal{A}$:*

(i) Linearity: For all $A, B \in \mathcal{A}$ and $a, b \in \mathbb{R}$, $E(aA + bB) = aE(A) + bE(B)$.

(ii) Non-Triviality: $E(\mathbb{1}) \neq 0$.

Proof. First consider the case $\dim(\mathcal{H}) = 2$. Let $\sigma_x, \sigma_y \in \mathcal{A}$ denote the Pauli matrices, which satisfy

$$\sigma_x^2 = \sigma_y^2 = \mathbb{1}\,, \qquad \{\sigma_x, \sigma_y\} = 0\,. \tag{B.15}$$

Set $c = E(\mathbb{1})$. Dispersion-freeness together with non-triviality implies

$$E(\sigma_i)^2 = E(\sigma_i^2) = E(\mathbb{1}) = c \neq 0\,. \tag{B.16}$$

If $c < 0$, then the equation $E(\sigma_i)^2 = c$ has no real solution, contradicting the assumption that E is real-valued. We conclude that $c > 0$. Hence,

$$E(\sigma_i) = \pm\sqrt{c}\,, \tag{B.17}$$

for $i \in \{x, y\}$. By linearity,

$$E(\sigma_x + \sigma_y) = E(\sigma_x) + E(\sigma_y) \in \{-2\sqrt{c}, 0, 2\sqrt{c}\}\,. \tag{B.18}$$

But

$$(\sigma_x + \sigma_y)^2 = \sigma_x^2 + \sigma_y^2 + \sigma_x\sigma_y + \sigma_y\sigma_x \tag{B.19}$$
$$= 2\,\mathbb{1}\,, \tag{B.20}$$

and therefore, by linearity and non-triviality,

$$E\left((\sigma_x + \sigma_y)^2\right) = E(2\mathbb{1}) = 2E(\mathbb{1}) = 2c\,. \tag{B.21}$$

Dispersion-freeness now yields

$$E(\sigma_x + \sigma_y)^2 = 2c\,, \tag{B.22}$$

and hence

$$E(\sigma_x + \sigma_y) = \pm\sqrt{2c}\,. \tag{B.23}$$

This contradicts the earlier result that $E(\sigma_x + \sigma_y) \in \{-2\sqrt{c}, 0, 2\sqrt{c}\}$. Hence there can exist no such functional E in the case where $\dim(\mathcal{H}) = 2$. Now let $\mathcal{H}$ be any Hilbert space with $\dim(\mathcal{H}) = n > 2$. We will construct a pair of operators X, Y with the following properties:

$$X^2 = Y^2 = Q\,, \qquad \{X, Y\} = 0\,, \tag{B.24}$$

with $E(Q) \neq 0$. This enables us to run the same argument as in the two-dimensional case. Let $\mathbb{1} = \sum_{k=1}^{n} P_k$ be a decomposition of the identity into orthogonal rank-one projectors. We claim that one can always choose two projectors P_i and P_j such that

$$E(P_i + P_j) \neq 0\,. \tag{B.25}$$

To see this, assume for contradiction that

$$E(P_i + P_j) = 0\,, \tag{B.26}$$

for all pairs $i < j$. Then the sum over all such pairs must also vanish:

$$\begin{aligned}
0 &= \sum_{i<j} E(P_i + P_j) && \text{(B.27)}\\
&= \sum_{i<j} \Big(E(P_i) + E(P_j)\Big) && \text{(B.28)}\\
&= (n-1) \sum_{k=1}^{n} E(P_k) && \text{(B.29)}\\
&= (n-1) E\left(\sum_{k=1}^{n} P_k\right) && \text{(B.30)}\\
&= (n-1) E(\mathbb{1}) \neq 0\,, && \text{(B.31)}
\end{aligned}$$

which is a contradiction. Hence, there exist orthogonal rank-one projectors P_i and P_j

such that

$$E(P_i + P_j) \neq 0\,. \tag{B.32}$$

Define $Q = P_i + P_j$ and set $S = \mathrm{Ran}(Q)$. Since P_i, P_j are orthogonal rank-one projectors, Q is a rank-two projector onto S and $\dim(S) = 2$. The observables X, Y can then be specified as follows:

$$X = \sigma_x \oplus 0_{S^\perp}\,, \tag{B.33}$$

$$Y = \sigma_y \oplus 0_{S^\perp}\,. \tag{B.34}$$

These operators satisfy the desired properties as can easily be checked:

$$X^2 = \sigma_x^2 \oplus 0_{S^\perp}^2 = \mathbb{1}_S \oplus 0_{S^\perp} = P_i + P_j = Q\,, \tag{B.35}$$

and analogously for Y. Furthermore

$$\begin{aligned}
\{X, Y\} &= XY + YX && \text{(B.36)}\\
&= (\sigma_x \oplus 0_{S^\perp})(\sigma_y \oplus 0_{S^\perp}) + (\sigma_y \oplus 0_{S^\perp})(\sigma_x \oplus 0_{S^\perp}) && \text{(B.37)}\\
&= \{\sigma_x, \sigma_y\} \oplus 0_{S^\perp} && \text{(B.38)}\\
&= 0\,. && \text{(B.39)}
\end{aligned}$$

Since linearity and dispersion-freeness apply to all observables in $\mathcal{A}$, the same contradiction as above follows. Therefore, no dispersion-free functional E satisfying both (i) and (ii) exists for any Hilbert space with $\dim(\mathcal{H}) \geq 2$. □

Naturally, two questions arise: first, the systematic question of how a proponent of a realist view of quantum phenomena should respond to von Neumann's result. Second, the historical question of what role the result has played in the discussion of the hidden variable question. The second question will not concern us here. We instead refer to [79] for a contemporary discussion of these issues.

With regard to the first question, it is clear that von Neumann's and Bell's impossibility proofs do not rule out hidden variables *tout court*. Bell's argument makes it clear that if one wants the expectation value E to be dispersion-free for non-commuting observables, then one has to reject linearity. This is because under the assumption of positivity of E, meaning that $A \geq 0$ implies $E(A) \geq 0$, the non-triviality condition cannot be rejected while keeping linearity without forcing $E(A) = 0$ for arbitrary $A \in \mathcal{A}$[1]. This can be seen by the following indirect argument: Assume

[1]It should be emphasized that positivity is not used in Bell's impossibility proof itself, but only

non-triviality is false and hence $E(\mathbb{1}) = 0$. For any projector P, it holds that

$$P \leq \mathbb{1}\,, \tag{B.40}$$

and therefore by linearity and positivity of E we have

$$0 \leq E(P) \leq E(\mathbb{1}) = 0\,, \tag{B.41}$$

which implies that $E(P) = 0$. For any $A \in \mathcal{A}$, this finally yields via its spectral decomposition and linearity of E that

$$E(A) = \sum_i a_i E(P_i) = 0\,, \tag{B.42}$$

Hence, if one posits the existence of a dispersion-free expectation value functional E for all observables, E must violate linearity. Bell gives the following explanation of why it might be sensible to give up on the linearity of the expectation value:

> At first sight the required additivity of expectation values seems very reasonable, and it is rather the non-additivity of allowed values (eigenvalues) which requires explanation. Of course the explanation is well known: A measurement of a sum of noncommuting observables cannot be made by combining trivially the results of separate observations on the two terms — it requires a quite distinct experiment. [...] But this explanation of the nonadditivity of allowed values also established the nontriviality of the additivity of expectation values. The latter is a quite peculiar property of quantum mechanical states, not to be expected a priori. There is no reason to demand it individually of the hypothetical dispersion-free states, whose function it is to reproduce the measurable peculiarities of quantum mechanics when averaged over. [77]

Bell's critique of overambitious interpretations of von Neumann's result amounts to the claim that we should not expect that in hidden variables theories expectation values for experimentally incompatible physical quantities to be linearly additive.

in the following argument showing that rejecting non-triviality while maintaining linearity leads to absurdity.

B.2 On the Localizability of Latent Causes

One possible objection to the causal formulation of Bell's theorem provided in Chapter 3 is to reject the implicit assumption that at least part of the latent variable Λ can be localized in spacetime. This blocks our derivations of Bell's theorem and also Gisin's theorem. It is necessary to distinguish three degrees of localizability, or lack thereof:

(i) Λ is defined on a bounded spacetime region.

(ii) Λ is defined on an unbounded spacetime region.

(iii) Λ is not defined on any spacetime region, bounded or otherwise.

In case (i), Λ is a local physical quantity. This is not the case in (ii). Consider for example the electric field $\vec{E}$ which is defined on a spacelike hypersurface. But even here it is possible to define local physical quantities in terms of $\vec{E}$, by restricting oneself to bounded spacetime regions. The value of $\vec{E}$ at $x = (\vec{x}, t)$ provides an example. For both (i) and (ii), the arguments provided in Chapter 3 go through.

The situation is different in the case of (iii). Here, Λ is no longer a physical quantity defined on spacetime, which makes it in general impossible to define local physical quantities restricting Λ to a bounded spacetime volume. An example of this is the wave function which is not defined on physical space but configuration space. In this case, LC does not suffice, since it only prohibits direct causal relations between spacelike separated local physical quantities. As a result, our derivation of Bell's theorem from causal principles is no longer valid.

But by using Lemma 3.10, a similar argument from slightly stronger assumptions to the same conclusion can be made. To this end, we consider the following strengthened version of local causality, which we will refer to as *strong local causality*.

Definition B.6 (Strong Local Causality)**.** For all spacetime regions $\mathcal{X}$, $\mathcal{Y}$ and any local physical quantities X and Y in $\mathcal{X}$ and $\mathcal{Y}$ respectively, if $\mathcal{X}$ and $\mathcal{Y}$ are spacelike separated, then there exists no directed path between X and Y.

Strong LC, in contrast to the weaker version we originally considered, does not merely prohibit spacelike separated variables from being each other's causal parents. Instead, it prohibits even indirect causal connections between spacelike separated local physical quantities. Hence, the following alternative formulation of Bell's theorem can be given:

Theorem B.7. *There is no causal model of the Bell scenario that accounts for the operational statistics and satisfies:*

(i) Weak Measurement Independence,

(ii) Strong Local Causality,

(iii) No-Backward Causality.

Proof. We begin by assuming the most permissive causal structure underlying the Bell scenario. By (ii), there can be no directed path from X and/or A to Y and/or B. Similarly, there can be no directed path from Y and/or B to A and/or X. Assumption (i) excludes Λ as the cause of X and Y. Using (iii), we further conclude that A cannot be the cause of X; analogously, B cannot be the cause of Y. Hence, the settings X and Y have no causal parents. These are exactly the conditions of Lemma 3.10. Hence, (i)–(iii) jointly imply Bell locality. But every Bell-local behavior satisfies the CHSH inequality by Lemma 3.17. Since experiments violate the CHSH inequality, the operational statistics cannot be Bell-local. This contradicts our initial assumption. Thus we conclude that no causal model satisfying (i)–(iii) can account for the operational statistics of the Bell scenario. □

In the case of Gisin's theorem, the situation is a little more subtle. For the sake of concreteness, consider a wave function of two particles $\psi(x_1, x_2, t)$. While it does not make sense to specify ψ at a location, as it is defined on configuration space, one can meaningfully specify ψ at a time. Hence, it is possible to imagine extensions of no-retrocausality that still apply to delocalized quantities such as ψ. But the proof given in Chapter 4 further relies on considerations which attribute spacetime relations, such as spacelike separation, between Λ and the operational quantities. Hence, it does not seem possible to save the proof by extending no-retrocausality (NRC) to delocalized quantities. However, in a similar way as in the case of LC, one can strengthen NRC so as to obtain a result very similar in spirit to Gisin's theorem.

Definition B.8 (Strong No-Retrocausality)**.** For all spacetime regions $\mathcal{X}$ and $\mathcal{Y}$, and for any local physical quantities X and Y located in $\mathcal{X}$ and $\mathcal{Y}$ respectively, if X is determined before Y in some Lorentz frame, then there exists no directed path from Y to X.

Using this definition and weak measurement independence (WMI), the following analog to Gisin's theorem can be obtained:

Theorem B.9. *There is no causal model of the Bell scenario that accounts for the operational statistics and satisfies:*

(i) Weak Measurement Independence,

(ii) Strong No-Retrocausality,

(iii) Causal Lorentz Invariance.

Proof. Assume the most permissive causal explanation of the Bell scenario. WMI and strong NRC imply that X and Y have no parent. In $\mathcal{F}_{AB}$, due to strong NRC, there can be no directed path from Y or B to A. Similarly, in $\mathcal{F}_{BA}$, all directed paths from X or A to B are prohibited. By Causal Lorentz invariance, we conclude that there can be no directed path from Y or B to A and/or from X or A to B. Hence, Lemma 3.10 applies and we have shown that (i)–(iii) jointly imply Bell locality. But every Bell-local behavior satisfies the CHSH inequality by Lemma 3.17. Since experiments violate the CHSH inequality, the operational statistics cannot be Bell-local. This contradicts our initial assumption. Thus we conclude that no causal model satisfying (i)–(iii) can account for the operational statistics of the Bell scenario. □

These considerations show that the localizability of the latent cause Λ is relevant but not central for the main results of this thesis.

B.3 The Rietdijk–Putnam Argument

According to eternalism or the block universe view, all events in spacetime exist on equal terms. This position thereby denies the possibility of true becoming and decouples existence from temporal presentness. Eternalism is often motivated by an appeal to special relativity, especially the relativity of simultaneity. Here we will briefly sketch an argument to this effect originally due to Rietdijk and Putnam [71, 72], following the presentation of Del Santo and Gisin [73].

Consider four inertial observers, Alice, Bob, Charlie, and Debbie. Bob is spacelike separated from Alice and at rest in her frame. Charlie and Debbie are at rest relative to each other and move with a constant, subluminal velocity v relative to Alice and Bob. One can then construct the following situation: at time t_2 in Charlie and Debbie's frame, Charlie's worldline intersects Bob's, while Debbie's worldline intersects Alice's at an earlier time $t_1 < t_2$ in Alice's frame. This setup can be represented by the following spacetime diagram:

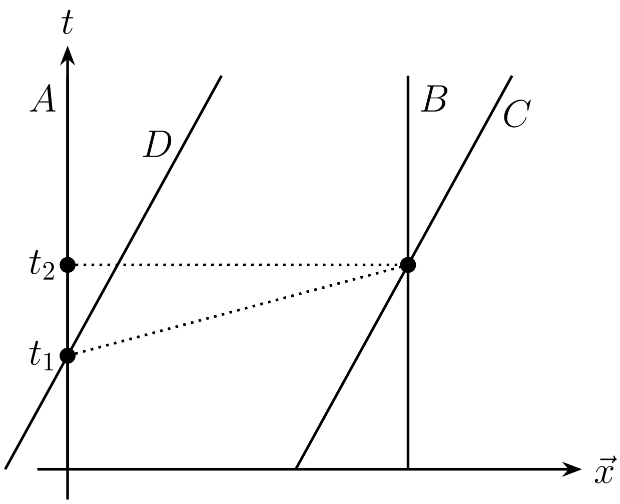


Figure B.1: Spacetime diagram of the setup of the Rietdijk–Putnam argument.

For an indirect proof, assume that Alice performs a genuinely indeterministic experiment which outputs a random bit $a \in \{0, 1\}$ at time t_2 in her frame. Hence, for Alice at any time t with $t < t_2$, the proposition $a = 0$, which we call ϕ, is indeterminate due to the genuinely indeterministic nature of the experiment. We further assume that determinateness satisfies the following intuitive principles:

(i) *Local reality*: Any two observers that locally coincide attribute the same truth values to empirical propositions.

(ii) *Present reality*: Any two distant observers at relative rest attribute the same truth values to empirical propositions about present events (i.e., events lying on the same plane of simultaneity in their rest frame).

By present reality, at time t_2 the truth value of ϕ is determinate for Bob. Because Bob and Charlie coincide at t_2, ϕ must have the same truth value for Charlie by local reality. Again, by present reality, ϕ must already be determinate for Debbie at time t_1, since Debbie lies on the same simultaneity plane as Charlie in their shared rest frame. But Debbie intersects Alice's worldline at t_1, and therefore, by local reality, the value of a must already be determinate for Alice at t_1. This contradicts the assumption that a is the output of a genuinely indeterministic experiment.

Rietdijk and Putnam therefore conclude that special relativity is incompatible with indeterminate events and hence supports an eternalist view of spacetime. Del Santo and Gisin instead opt to reject the assumption of present reality in order to reconcile special relativity with an ontologically open future [73].

Bibliography

[1] Claus Kiefer. *Conceptual Problems in Quantum Gravity and Quantum Cosmology.* `https://arxiv.org/abs/1401.3578`. arXiv:1401.3578 [gr-qc]. 2014.

[2] Karen Crowther. *Why Do We Want a Theory of Quantum Gravity?* Elements in the Philosophy of Physics. Cambridge University Press, 2025.

[3] John S. Bell and Alain Aspect. "On the Einstein–Podolsky–Rosen paradox". In: *Speakable and Unspeakable in Quantum Mechanics: Collected Papers on Quantum Philosophy.* Cambridge University Press, 2004.

[4] Paul Davies. "Interview: John Bell". In: *Omni* (1988), pp. 88–92.

[5] Wayne C. Myrvold. "Relativistic Constraints on Interpretations of Quantum Mechanics". In: *Synthese* 132.3 (2002), pp. 319–346.

[6] Karin Berndl et al. "Nonlocality, Lorentz invariance, and Bohmian quantum theory". In: *Physical Review A* 53.4 (1996), pp. 2062–2073.

[7] Detlef Dürr et al. "Can Bohmian mechanics be made relativistic?" In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 470.2162 (2014).

[8] Roderich Tumulka. "A Relativistic Version of the Ghirardi–Rimini–Weber Model". In: *Journal of Statistical Physics* 125.4 (2006), pp. 821–840.

[9] Niels Bohr. "Can Quantum-Mechanical Description of Physical Reality be Considered Complete?" In: *Physical Review* 48 (1935), pp. 696–702.

[10] Werner Heisenberg. *Physics and Philosophy: The Revolution in Modern Science.* New York: Harper Perennial Modern Thought, 2007.

[11] John von Neumann and Robert T. Beyer. *Mathematical Foundations of Quantum Mechanics: New Edition.* NED - New edition. Princeton University Press, 2018.

[12] David Bohm. "A Suggested Interpretation of the Quantum Theory in Terms of "Hidden" Variables. I". In: *Phys. Rev.* 85.2 (1952), pp. 166–179.

[13] David Bohm. “A Suggested Interpretation of the Quantum Theory in Terms of “Hidden” Variables. II”. In: *Phys. Rev.* 85.2 (1952), pp. 180–193.

[14] John S. Bell and Alain Aspect. “On the impossible pilot wave”. In: *Speakable and Unspeakable in Quantum Mechanics: Collected Papers on Quantum Philosophy*. Cambridge University Press, 2004.

[15] Jon P. Jarrett. “On the Physical Significance of the Locality Conditions in the Bell Arguments”. In: *Noûs* 18.4 (1984), pp. 569–589.

[16] Nicolas Gisin. “Impossibility of covariant deterministic nonlocal hidden-variable extensions of quantum theory”. In: *Physical Review A* 83.2 (2011).

[17] Roderich Tumulka. “Comment on “The Free Will Theorem””. In: *Foundations of Physics* 37.2 (2007), pp. 186–197.

[18] Roderich Tumulka. *Foundations of Quantum Mechanics*. Vol. 1003. Lecture Notes in Physics. Cham, Switzerland: Springer Nature Switzerland AG, 2022.

[19] Valia Allori. “What is It Like to be a Relativistic GRW Theory? Or: Quantum Mechanics and Relativity, Still in Conflict After All These Years”. In: *Foundations of Physics* 52 (July 2022).

[20] Nicholas Harrigan and Robert W. Spekkens. “Einstein, Incompleteness, and the Epistemic View of Quantum States”. In: *Foundations of Physics* 40.2 (2010).

[21] Matthew S. Leifer. “Is the Quantum State Real? An Extended Review of ψ-Ontology Theorems”. In: *Quanta* 3 (2014).

[22] Jonathan Barrett, Robin Lorenz, and Ognyan Oreshkov. *Quantum Causal Models*. 2020. arXiv: `1906.10726 [quant-ph]`. URL: `https://arxiv.org/abs/1906.10726`.

[23] Robert W. Spekkens. *The ontological identity of empirical indiscernibles: Leibniz's methodological principle and its significance in the work of Einstein*. 2019. arXiv: `1909.04628 [physics.hist-ph]`. URL: `https://arxiv.org/abs/1909.04628`.

[24] Gottfried Wilhelm Leibniz and Antoine Arnauld. “Correspondence with Arnauld (1686–1687)”. In: *Philosophical Papers and Letters*. Ed. by Leroy E. Loemker. 2nd. Dordrecht: Reidel, 1969.

[25] Eric G. Cavalcanti. “Classical Causal Models for Bell and Kochen-Specker Inequality Violations Require Fine-Tuning”. In: *Physical Review X* 8.2 (Apr. 2018).

[26] Christopher J. Wood and Robert W. Spekkens. "The lesson of causal discovery algorithms for quantum correlations: causal explanations of Bell-inequality violations require fine-tuning". In: *New Journal of Physics* 17.3 (2015).

[27] Eric G. Cavalcanti and Howard M. Wiseman. "Implications of Local Friendliness Violation for Quantum Causality". In: *Entropy* 23.8 (July 2021), p. 925.

[28] Yìlè Yīng et al. "Relating Wigner's Friend Scenarios to Nonclassical Causal Compatibility, Monogamy Relations, and Fine Tuning". In: *Quantum* 8 (Sept. 2024), p. 1485.

[29] Elie Zahar. "Why did Einstein's Programme supersede Lorentz's?" In: *Method and Appraisal in the Physical Sciences: The Critical Background to Modern Science, 1800–1905*. Ed. by Colin Howson. Cambridge: Cambridge University Press, 1976, pp. 211–276.

[30] Isaac Newton. "406. Newton to Bentley, 25 February 1692/3". In: *The Correspondence of Isaac Newton*. Ed. by Herbert W. Turnball. Cambridge University Press, 1966, pp. 253–256.

[31] Albert Einstein. "Quanten-Mechanik und Wirklichkeit". In: *Dialectica* 2.3–4 (Nov. 1948), pp. 320–324.

[32] John S. Bell and Alain Aspect. "La nouvelle cuisine". In: *Speakable and Unspeakable in Quantum Mechanics: Collected Papers on Quantum Philosophy*. Cambridge University Press, 2004.

[33] Albert Einstein, Boris Podolsky, and Nathan Rosen. "Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?" In: *Physical Review* 47.10 (1935), pp. 777–780.

[34] David Bohm. *Quantum Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1951.

[35] Travis Norsen. "Bell Locality and the Nonlocal Character of Nature". In: *Foundations of Physics Letters* 19.7 (Dec. 2006), pp. 633–655.

[36] Roderich Tumulka. *The Assumptions of Bell's Proof*. 2015. arXiv: `1501.04168 [quant-ph]`. URL: `https://arxiv.org/abs/1501.04168`.

[37] Marek Zukowski and Caslav Brukner. "Quantum non-locality—it ain't necessarily so..." In: *Journal of Physics A: Mathematical and Theoretical* 47.42 (2014).

[38] Travis Norsen. "Against 'Realism'". In: *Foundations of Physics* 37.3 (Feb. 2007), pp. 311–340.

[39] Nicolas Gisin. "Non-realism: Deep Thought or a Soft Option?" In: *Foundations of Physics* 42.1 (Sept. 2010), pp. 80–85.

[40] Federico Laudisa. “How and when did locality become ‘local realism’? A historical and critical analysis (1963–1978)”. In: *Studies in History and Philosophy of Science* 97 (2023), pp. 44–57.

[41] Asher Peres. “Unperformed Experiments Have No Results”. In: *American Journal of Physics* 46.7 (1978), pp. 745–747.

[42] Christopher A. Fuchs. *QBism, the Perimeter of Quantum Bayesianism.* 2010. arXiv: `1003.5209 [quant-ph]`. URL: `https://arxiv.org/abs/1003.5209`.

[43] Anton Zeilinger. *Dance of the Photons: From Einstein to Quantum Teleportation.* New York: Farrar, Straus and Giroux, 2010.

[44] Detlef Dürr and Stefan Teufel. *Bohmian Mechanics: The Physics and Mathematics of Quantum Theory.* Vol. 114. Fundamental Theories of Physics. Berlin Heidelberg: Springer, 2009.

[45] Valerio Scarani. *Bell Nonlocality.* Oxford University Press, 2019.

[46] Sandu Popescu and Daniel Rohrlich. “Quantum Nonlocality as an Axiom”. In: *Foundations of Physics* 24.3 (1994), pp. 379–385.

[47] Boris S. Cirel’son. “Quantum generalizations of Bell’s inequality”. In: *Letters in Mathematical Physics* 4.2 (1980).

[48] B. Hensen et al. “Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometres”. In: *Nature* 526 (2015), pp. 682–686.

[49] M. Giustina et al. “Significant-Loophole-Free Test of Bell’s Theorem with Entangled Photons”. In: *Physical Review Letters* 115.25 (2015).

[50] L. K. Shalm et al. “Strong Loophole-Free Test of Local Realism”. In: *Physical Review Letters* 115.25 (2015).

[51] Lev Vaidman. “Many-Worlds Interpretation of Quantum Mechanics”. In: *The Stanford Encyclopedia of Philosophy.* Ed. by Edward N. Zalta. Fall 2021. Metaphysics Research Lab, Stanford University, 2021.

[52] John-Mark A. Allen et al. “Quantum Common Causes and Quantum Causal Models”. In: *Physical Review X* 7.3 (2017).

[53] Kok-Wei Bong et al. “A strong no-go theorem on the Wigner’s friend paradox”. In: *Nature Physics* 16.12 (Aug. 2020), pp. 1199–1205.

[54] Gerard ’t Hooft. *The Cellular Automaton Interpretation of Quantum Mechanics.* 2015. arXiv: `1405.1548 [quant-ph]`. URL: `https://arxiv.org/abs/1405.1548`.

[55] Gerard 't Hooft. *Free Will in the Theory of Everything*. 2017. arXiv: `1709.02874 [quant-ph]`. URL: `https://arxiv.org/abs/1709.02874`.

[56] Gerard 't Hooft. *An unorthodox view on quantum mechanics*. 2021. arXiv: `2104.03179 [quant-ph]`. URL: `https://arxiv.org/abs/2104.03179`.

[57] John S. Bell and Alain Aspect. "Free variables and local causality". In: *Speakable and Unspeakable in Quantum Mechanics: Collected Papers on Quantum Philosophy*. Cambridge University Press, 2004.

[58] Stephan Bongers et al. "Foundations of Structural Causal Models with Cycles and Latent Variables". In: *Annals of Statistics* 49.5 (2021), pp. 2885–2915.

[59] John S. Bell and Renée Weber. *Interview with John S. Bell*. `https://sites.math.rutgers.edu/~oldstein/bell-weber.html`. Unpublished interview transcript. 1989.

[60] John S. Bell and Alain Aspect. "Bertlmann's socks and the nature of reality". In: *Speakable and Unspeakable in Quantum Mechanics: Collected Papers on Quantum Philosophy*. Cambridge University Press, 2004.

[61] Avi Levy and Meir Hemmo. "Why a Relativistic Quantum Mechanical World Must Be Indeterministic". In: *Quantum, Probability, Logic: The Work and Influence of Itamar Pitowsky*. Ed. by Meir Hemmo and Orly Shenker. Jerusalem Studies in Philosophy and History of Science. Springer, Cham, 2020, pp. 423–447.

[62] Ward Struyve. *Lorentz invariance and quantum mechanics*. 2024. arXiv: `2402.15881 [quant-ph]`. URL: `https://arxiv.org/abs/2402.15881`.

[63] Tim Maudlin. "Non-local correlations in quantum theory: How the trick might be done". In: *Einstein, Relativity and Absolute Simultaneity*. Ed. by William Lane Craig and Quentin Smith. Abingdon: Routledge, 2008, pp. 156–179.

[64] Roderich Tumulka. "Collapse and Relativity". In: *Journal of Physics: Conference Series* 67 (2007), p. 012037. arXiv: `quant-ph/0602208 [quant-ph]`.

[65] John S. Bell. "Towards an Exact Quantum Mechanics". In: *Themes in Contemporary Physics II: Essays in Honour of Julian Schwinger's 70th Birthday*. Ed. by S. Deser and R. J. Finkelstein. Singapore: World Scientific, 1989, pp. 1–26.

[66] John S. Bell and Alain Aspect. "Are there quantum jumps". In: *Speakable and Unspeakable in Quantum Mechanics: Collected Papers on Quantum Philosophy*. Cambridge University Press, 2004.

[67] Tim Maudlin. *Quantum Non-Locality and Relativity: Metaphysical Intimations of Modern Physics*. 3rd ed. Malden, MA: Wiley-Blackwell, 2011.

[68] Roderich Tumulka. “A Relativistic GRW Flash Process with Interaction”. In: *Do Wave Functions Jump?* Springer International Publishing, Oct. 2020, pp. 321–347.

[69] Robert W. Spekkens. “The Paradigm of Kinematics and Dynamics Must Yield to Causal Structure”. In: *Questioning the Foundations of Physics: Which of Our Fundamental Assumptions Are Wrong?* Ed. by Anthony Aguirre, Brendan Foster, and Zeeya Merali. Cham: Springer, 2015, pp. 5–16. arXiv: `1209.0023 [quant-ph]`.

[70] Emily Adlam. *Is There Causation in Fundamental Physics? New Insights from Process Matrices and Quantum Causal Modelling.* 2022. arXiv: `2208.02721 [quant-ph]`. URL: `https://arxiv.org/abs/2208.02721`.

[71] C. W. Rietdijk. “A Rigorous Proof of Determinism Derived from the Special Theory of Relativity”. In: *Philosophy of Science* 33.4 (1966), pp. 341–344.

[72] Hilary Putnam. “Time and Physical Geometry”. In: *The Journal of Philosophy* 64.8 (1967), pp. 240–247.

[73] Flavio Del Santo and Nicolas Gisin. “The Relativity of Indeterminacy”. In: *Entropy* 21.2 (2019), p. 136. DOI: `10.3390/e21020136`.

[74] Časlav Brukner. “A No-Go Theorem for Observer-Independent Facts”. In: *Entropy* 20.5 (2018), p. 350.

[75] Nicolas Gisin. “Quantum Nonlocality: How Does Nature Do It?” In: *Science* 326.5958 (2009), pp. 1357–1358.

[76] Leslie E. Ballentine. “The Statistical Interpretation of Quantum Mechanics”. In: *Reviews of Modern Physics* 42.4 (1970), pp. 358–381.

[77] John S. Bell and Alain Aspect. “On the problem of hidden variables in quantum mechanics”. In: *Speakable and Unspeakable in Quantum Mechanics: Collected Papers on Quantum Philosophy.* Cambridge University Press, 2004.

[78] John S. Bell and Alain Aspect. “Introduction to the hidden-variable question”. In: *Speakable and Unspeakable in Quantum Mechanics: Collected Papers on Quantum Philosophy.* Cambridge University Press, 2004.

[79] Pablo Acuña. “von Neumann’s Theorem Revisited”. In: *Foundations of Physics* 51.3 (2021), p. 73.