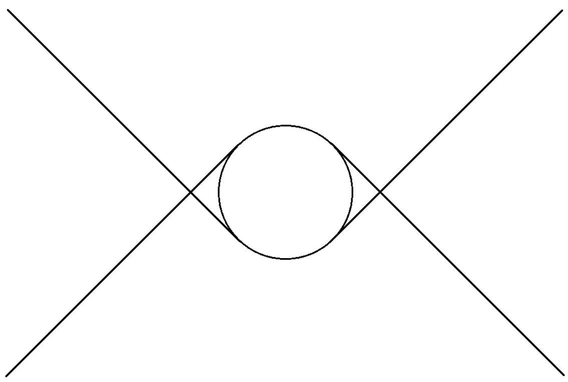
HISTORY AND PHILOSOPHY OF SCIENCE

MASTER'S THESIS

Spacelike and Timelike Non-Locality

Author:
Sam Michael Rijken

Supervisors:
Dr. G. Bacciagaluppi
Prof. dr. F. A. Muller



31 January 2019

Abstract

In this thesis, I analyze the concepts of spacelike and timelike non-locality in the context of Bell's conceptual framework and the EPR-type thought experiment. After defining locality in this context, I argue in favor of a Shimonian subdivision of independence assumptions about parameters of an EPR-type thought experiment and I argue against alternative subdivisions. I then discuss direct timelike analogues of spacelike locality that have recently been introduced in the literature on quantum foundations. I argue that there are indeed good reasons for considering the possibilities for timelike non-locality, but that the formal similarity between spacelike and timelike locality in the context of Bell's conceptual framework is misleading. I conclude that spacelike and timelike non-locality are strongly disanalogous.

Contents

1 Introduction

1.1 General introduction

At a conference held near Copenhagen in 1936 on the philosophical consequences of the infamous two-slit experiment, with notable attendees such as Niels Bohr and Henrik Casimir, the Danish philosopher Jørgen Jørgensen supposedly exclaimed after an ontology-heavy discussion: "One can, damn it, not reduce the whole of philosophy to a screen with two holes" [6, p. 55]. Jørgensen's worry is very reasonable: how do we justify having a single, seemingly absurd but poorly-understood experimental result overthrow most of two thousand years' worth of philosophy? Perhaps to Jørgensen's relief, however, we are not reducing the whole of philosophy to a single experiment. There are many such absurd experiments, eerily devoid of concrete interpretation, whose results flew right in the face of early twentieth-century common sense and yet were predicted to astounding accuracy by the novel 'quantum theory'. The entire body of quantum theory lacked—and still lacks—an unambiguous and consistent interpretation of what the theory tells us that the quantum world is really like. There has never even been full agreement about what the correct formalism and dynamical laws should be.

Standard textbook quantum mechanics is not a genuine physical theory, it is a recipe for making experimental predictions. For many physicists this leaves nothing to be desired. But for philosophers of physics, who attempt to interpret physical theories in order to formulate a story about what the world is really like, quantum theory is a minefield. Most conventional approaches to reaching general consensus about a single quantum theory and interpretation thereof have been laid out for decades. Proponents of each interpretation appear deadlocked, facing each other in mutual disagreement over basic metaphysical intuitions about determinism, locality, ontology, etcetera, with little prospect of reaching consensus in the near future. Pure philosophical analysis will not solve the problem, neither will pure experimental or theoretical physics. All this has led to a remarkable emphasis on a few basic, empirically accessible absurd quantum phenomena, which are being philosophically scrutinized as a proxy for the whole of quantum theory. The two-slit experiment is such a phenomenon, and its main philosophical problem is 'the measurement problem': (roughly) why does quantum theory describe an entire wavefunction—a catalogue of all possible results—arriving at a detector, and yet we see only a single result? Another such phenomenon is the EPR-type experiment, with the main philosophical problem being locality: what is the meaning of this apparent spooky action-at-a-distance that is present in quantum mechanics but forbidden by relativity theory? The problem of locality is the main topic of this thesis, the EPR-type experiment is the main context. The central aim of this thesis is to provide a critical assessment of the current literature on spacelike and timelike non-locality, to shine a well-deserved light on some persistent and some novel loopholes to Bell's theorem and evaluate how these should be interpreted in the context of timelike non-locality, and to assess whether the analogy between spacelike and timelike non-locality is strong.

In Chapter 2, I define John Bell's conception of locality that I adopt and discuss in this thesis. In Section 2.1, I then describe how this conception came to be, and I discuss it more thoroughly in Section 2.2. Finally, I describe the original EPR-type thought experiment in the context of Bell's framework in Section 2.3. I end by highlighting some strengths and weaknesses of the contemporary paradigm of Bell's conception and formalization of locality.

In Section 3, I focus on the most paradigmatic results of the contemporary locality-debate: Bell's theorem and its accompanying Bell-type inequalities. First, in Section 3.1, I formulate some basic assumptions that are necessary for understanding and interpreting Bell's theorem. I emphasize several assumptions that too often remain implicit and I specify some details of the experimental context that are necessary for the analysis in the second half of this thesis. In Section 3.2, I briefly summarize the literature on Bell-type inequalities, which focuses on independence relations between different components of the EPR-type experiment, and I present a new argument against the utility of some of these independence relations.

The second half of this thesis starts with Chapter 4. There is an increasing interest in the possibility of non-local correlations in time over and above non-local correlations in space, and, in association, increasing interest in retrocausal and all-at-once interpretations of quantum mechanics. This has led to the introduction of direct timelike analogues of Bell's theorem and Bell-type inequalities. In Section 4.1, I introduce the proposed definition of timelike locality analogously to spacelike locality and I discuss how the independence relations discussed in the previous sections may be applicable to the timelike analogue of Bell's theorem. Then, in Section 4.2, I review the three (apparent) possibilities for theories that violate the timelike analogue of Bell-type inequalities and are often considered to be loopholes to the regular Bell's Theorem. I discuss each of these thoroughly in the subsequent chapter, but I first conclude Chapter 4 with a discussion of Bell-type inequalities and formally similar inequalities such as Leggett-Garg's and Brukner et al.'s in Section 4.3.

Chapter 5 is concerned with the three main possibilities for timelike non-locality: superdeterminism, retrocausality, and all-at-once. First, in Section 5.1, I critically assess a relatively recent result that argues for the analogy between spacelike and timelike variant of the EPR-type experiment. In Section 5.2, I argue that superdeterministic theories are undesirable and unscientific despite them always being a logical possibility. I share this sentiment with many others, but I hope to add some new weight to this position by showing how superdeterminism violates some of my formulated basic assumptions and my specification of the EPR-type experiment. Section 5.3 is concerned with retrocausality. Here I first discuss a recent explicit argument for retrocausality in the quantum world, the Price Argument, and I argue that this argument is not quite as strong as it appears to be. I emphasize here that retrocausality does not actually violate timelike locality as defined in Chapter 4. Finally, in Section 5.4, I discuss the one interpretation that actually violates such timelike locality: the all-at-once interpretation proposed by Ken Wharton. This interpretation naturally results in a radical departure from standard practice in the philosophy of physics, however, and I argue that we must be very careful about the possible consequences of an interpretation such as all-at-once. I will conclude that all-at-once theories are a very interesting subject of investigation, but that the EPR-type experiment and Bell's theorem are not the main tool that we should use for such investigation. I highlight some intriguing aspects of all-at-once that should be analyzed, but which are beyond the scope of this thesis.

Some of the sections in this thesis are focused on summarizing the existing literature, while others contain more novel arguments and analysis. By presenting this all in the single formalism and treatment of this thesis, I hope to provide a framework for the discussion of locality in quantum theory applicable outside the scope of this thesis.

1.2 Prelude to philosophy of quantum mechanics

As said, quantum mechanics is not a physical theory is the usual sense. It is a mathematical machinery for predicting experimental results, and it does so mainly probabilistically. There is still no consensus about what quantum mechanics tells us that the world is like. In this section, I briefly present the main axioms of quantum theory that we need for this thesis. These I will consider to be the basis of the orthodox interpretation of quantum mechanics commonly referred to as the Copenhagen interpretation. I introduce the problems faced by the Copenhagen interpretation and briefly elaborate on the three main contenders for solving these problems: Bohmian mechanics, collapse theories, and many-worlds.

The axioms that form the basis of quantum mechanics can succinctly be formulated as:[1]

1. Every physical system is associated with a complex Hilbert space.
2. The state of a physical system is described by vectors in Hilbert space, represented by a wavefunction Ψ.
3. Physical quantities are represented by Hermitian operators in the Hilbert space, whose eigenvalues represent possible measurement results.
4. If two wavefunctions Ψ_1 and Ψ_2 represent two states of a system, then a linear combination of these, $\Psi_1 \otimes \Psi_2$, also represents a state of the system.
5. An isolated system evolves unitarily in time according to Schrödinger's equation.
6. Upon measurement of an observable X, the state of the system (the wavefunction) collapses instantaneously into an X-eigenstate corresponding to the eigenvalue observed. Which particular X-eigenstate the system collapses into is predicted only probabilistically according to the Born rule.

There is an obvious tension between axioms 5 and 6. Even before we ask ourselves tough questions such as "what counts as a measurement?", we must already recognize that a quantum mechanical system has two different types of dynamics in two different contexts: and isolated system evolves unitarily in time, but a measured (or, 'observed') system collapses non-unitarily. A fundamental distinction between these two contexts appears arbitrary, and many feel that it has no place in a physical theory (cf. [17]). This, then, is one possible formulation of the infamous *measurement problem* of quantum mechanics: how can we account for the collapse of the wavefunction that occurs upon measurement of a quantum system?

One may reply to this problem in two ways: (i) appeal to the incredible success of quantum theory thusfar and claim nothing needs to be added, or (ii) strive for a different formulation of quantum mechanics by adding mathematical or philosophical baggage often collectively (but inaccurately) referred to as 'hidden variables'. Those who take up option (i) in one way or another I will collectively refer to as adherents of the *Copenhagen interpretation*. Admittedly, however, there is no such thing as a unified Copenhagen interpretation (cf. [39, §5]). All historical actors had their own views on quantum theory, and it does no justice to their philosophical sophistication to consider them as a collective. But whatever their differences are, they do share the sentiment that

[1] Following, for example, [90, p. 4] or [54, p. 3] or [71, §3].

the mathematics of quantum mechanics needs no modification. Those who disagree with this have taken up option (ii) in a multitude of ways. Three of the main alternatives that are relevant for this thesis are discussed below.

The first alternative is *Bohmian mechanics*.[2] In this theory, the wavefunction is not a complete description of a quantum system. The theory postulates that the positions of point-like particles are always well-defined, and that these particles are guided in their motion by a 'guidance equation' that expresses their velocities in terms of the wavefunction [61]. All other observables are determined by (or dependent on) the particles' positions and the measurement context. The collapse of the wavefunction that occurs during measurement is thus not an actual physical 'event' but merely reveals the particle's pre-existing position or some derivative thereof. The evolution of a quantum system is now unitary and fully deterministic, and the measurement problem is solved. A remarkable characteristic of Bohmian mechanics is that it formalizes non-locality explicitly because the position of a particle in a multi-particle system is instantaneously dependent on the position of all other particles through their shared wavefunction. The main problem of Bohmian mechanics is, therefore, that its causal structure appears to require a preferred reference frame and the theory can thus not be made Lorentz invariant.

The second alternative are *collapse theories*, the main contender of which is known as GRW [57]. For collapse theories, the collapse of the wavefunction is an objective physical event. This is achieved by adding two parameters to Schrödinger's equation, one for when and one for where the collapse takes place. There is thus no need for an anthropocentric concept of 'measurement' in the theory, and the measurement problem is again solved. Because both added parameters are probabilistic, the theory is fundamentally indeterministic. The main problem of collapse theories is that it is hard to comprehend its ontology. The three main options here are to simply regard the entire wavefunction as the ontology ($\text{GRW}_\emptyset$), to derive the ontology from the system's mass density (GRW_m), or to only consider the 'flashes' (collapses) themselves as genuinely real (GRW_f).[3] The latter option was proposed intuitively by Bell himself in [16, ch. 22] but only later explicitly formulated by Tumulka in [123], wherein serious steps were taken towards constructing a Lorentz invariant version of GRW_f (although relativistic generalizations of earlier collapse theories long existed—eg. [56, 104]).

The third alternative is *many-worlds*. This interpretation adds no 'hidden variables' to the mathematical formalism of quantum mechanics but instead proposes a radical reinterpretation thereof. The idea here, originally proposed by Everett [52], is, roughly, to take the perspective from within the wavefunction rather than outside it, and to treat every 'possible' measurement result represented by the wavefunction as equally genuinely real. Given the difficulty of comprehending the image of the world that many-worlds suggests, the theory has been (and still is) subject to much misunderstanding. Fortunately, a specific version of the theory has recently been rigorously explored in [128], and one may hope that more consensus will be reached about many-worlds in the near future. Presently, one of the main problems for this theory are the interpretation of the notion of probability—which is central to quantum mechanics, but hard to make sense of in a world where 'every possibility happens'.[4] The main problem of the theory with regards

[2]Originally proposed by Bohm in [21], based on rediscovered ideas by de Broglie from the 1920's about wave-particle duality and pilot-waves (Cf. [14, §2] and the references therein.). See also [36].

[3]See especially [55, Part II] for a most recent elaboration.

[4]Another issue is, for example, the reliance on functionalism for the emergence of classical worlds (cf. [94]). However, some regard this as a good characteristic of the theory while others consider it as a problem.

to this thesis is that it is hard to make sense of Bell's highly classical notion of 'local beables' and his probabilistic definition of local causality in the context of many-worlds, and I therefore do not discuss this interpretation in this thesis.

We can see that all interpretations provide radically different ontological stories—what exists, how it constitutes the outside world, and what role the wavefunction plays therein. It appears thusfar that the reasons for preferring one interpretation over another are mainly subjective, based on one's personal a priori metaphysical commitments, and that there are very few objective criteria for theory choice in the philosophy of quantum mechanics. In the past few decades, this has led to a thriving industry of no-go theorems in the foundations of quantum mechanics, all of which aim to constrain the myriad of ontological possibilities by constructing contradictions that arise when certain plausible assumptions are conjoined. One example is, of course, Bell's theorem, which states that no local ontology can ever reproduce the predictions of quantum theory. A more recent example is the PBR-theorem [110], which roughly states that the wavefunction can not represent mere information about underlying reality, but it must have some degree of reality itself: the wavefunction cannot be purely epistemic rather than ontic. Although such no-go theorems are probably the main driving force of progress in the foundations of quantum mechanics, we must realize that the *impossibilities in principle* provided by no-go theorems are obviously valid only for theories that adhere to the assumptions made in the construction of the no-go theorem (cf. [75]). Very often (and at least for these two examples), one such basic assumption is classical one-way directed causality. Because the exotic interpretations discussed in this thesis such as retrocausality and all-at-once violate this assumption, contrary to the more 'traditional' interpretations such as Bohmian mechanics and collapse theories, they are naturally impervious to much of the standard consequences of such no-go theorems and we may entertain their possibility even though it appears that many of the ontological possibilities for the foundations of quantum mechanics have been severely restricted. But, of course, this does not mean that such exotic interpretations are completely immune to the ontological problems in quantum mechanics. To the contrary, especially for such interpretations we must have a very clear idea about what the restrictions on the possible ontologies are. In this thesis, we will be concerned with Bell's conception of ontology in terms of *local beables*, which has proved to be very useful as it has become a central concept in the foundations of quantum mechanics. As we will see in Section 2.2, however, this is a rather classical conception of what exists, and I will voice some concerns about Bell's conceptions of ontology and causality in Section 2.4. But I am getting ahead of myself. Let us start at the beginning.

2 Defining locality

2.1 A history of locality

Locality is often—if not always—defined as the negation of action-at-a-distance: whatever action or causality is, it doesn't happen instantaneously between spacelike separated regions. This was already a problem right at the heyday of the mechanized worldview. Descartes' world was governed only through action-by-contact, and a widespread philosophical sentiment was that action-by-contact was the way to go. But only a few decades later Newton came along with a law of gravity that obeyed no such restriction. We can recognize the discomfort of the philosophical community with Newton's dynamics in the famous Leibniz-Clarke debate at the beginning of the eighteenth century, in the fourth letter by Leibniz [7, p. 27]:

> It is also a supernatural thing that bodies should attract one another at a distance without any intermediate means, and that a body should circulate without shooting off along a tangent, though nothing hinders it from doing that. For these effects can't be explained by the nature of things.

When the nature and behavior of things is considered to be limited to action-by-contact, then the interaction between heavy bodies lacking intermediate means cannot be explained.

Fast forward two hundred years to the beginning of the twentieth century, and we see physical theories with a built-in restriction on the transmission speed of action. The propagation speed of the electric and magnetic fields in Maxwell's electrodynamics can be read off to be the speed of light.[5] Einstein's special theory of relativity then integrated this into a general system with dynamical laws that are valid in any reference frame and in which the speed of light is an absolute limit on the propagation of action, and this laid the cornerstone for his dynamics described in the general theory of relativity. Besides these monumental achievements, Einstein also helped to lay the foundations for quantum theory with, amongst other things, his Nobel prize-winning work on the photo-electric effect. But his role in quantum mechanics would indeed 'provide some of the great ironies of twentieth-century physics' [105, p. 195], for Einstein rejected the dominant view of the implications of quantum mechanics on basis of an oft-quoted sentiment similar to that of Leibniz described above [47]:

> If one asks what, irrespective of quantum mechanics, is characteristic of the world of ideas of physics, one is first of all struck by the following: the concepts of physics relate to a real outside world, that is, ideas are established relating to things such as bodies, fields, etc., which claim a 'real existence' that is independent of the perceiving subject—ideas which, on the other hand, have been brought into as secure a relationship as possible with the sense-data. It is further characteristic of these physical objects that they are thought of as arranged in a space-time continuum. An essential aspect of this arrangement of things in physics is that they lay claim, at a certain time, to an existence independent of one another, provided these objects 'are situated in different parts of space'. Unless one makes this kind of assumption about the independence of the existence (the 'being-thus') of objects which are far apart from

[5] Cf. [101, §1.2-1.3] for an accessible derivation.

> one another in space which stems in the first place from everyday thinking—physical thinking in the familiar sense would not be possible. It is also hard to see any way of formulating and testing the laws of physics unless one makes a clear distinction of this kind. [...]
>
> The following idea characterizes the relative independence of objects far apart in space (A and B): external influence on A has no direct influence on B; this is known as the 'principle of contiguity', which is used consistently only in the field theory. If this axiom were to be completely abolished, the idea of the existence of (quasi-) enclosed systems, and thereby the postulation of laws which can be checked empirically in the accepted sense, would become impossible.

Again we see here locality defined as the negation of action-at-a-distance in an axiom called 'the principle of contiguity'.[6] (We also see explicit mention of the observer-independent reality of things and the necessity of arranging these in a space-time continuum, but this will be discussed in Section 3.1.) Already in 1909 Einstein used an example to illustrate his intuitions on the conflict between this principle of contiguity and the quantum-mechanical description of light before and after measurement [14, §7.2]. More significantly, at the famous 1927 Solvay conference, Einstein invoked several thought experiments to argue that quantum mechanics violated this very basic requirement for the world of ideas of physics—the principle of contiguity—and, therefore, that quantum mechanics had to be incomplete. One version of such thought experiments, now known as 'Einstein's boxes' due to a later reformulation by de Broglie, can be summarized as follows (cf. [98]).

Consider a particle enclosed in a box described by a wavefunction Ψ that is spread out over the entire box. The usual interpretation of this description is that the particle is somewhat 'everywhere' in the box, and only localizes itself at a certain location x with probability $|\Psi_x|^2$ when a measurement of its position is made. Now, before we do such a measurement, imagine cutting the box in half without disturbing the wavefunction and carrying both halves of the box away to opposite ends of the Earth. If we now open one half-box and perform a position measurement and do indeed find a localized particle, this immediately implies that the other half-box cannot contain the particle, even though there is half a wavefunction in that part of the box. If we reject the idea that a measurement at one half-box can instantaneously influence the other half-box according to the principle of contiguity, then the natural consequence would be to conclude that the particle was is one half-box all along and that the probabilistic description provided by the wavefunction Ψ is incomplete. Therefore, quantum mechanics is incomplete.

But Einstein's arguments were neither well received nor well understood. The nature of the discussion at and following the 1927 Solvay conference can be accurately characterized as confused and ambiguous and it was clear to none what precisely the arguments of the others were in the debate on the completeness of quantum mechanics. The first major shift in this debate came with the publication of the 1935 paper by Einstein, Podolsky, and Rosen (EPR) [48]. Here, finally an argument was explicitly presented that quantum mechanics must be incomplete if one assumed the principle of contiguity to hold, and EPR anticipated that quantum mechanics would be replaced by a fully contiguous theory. The paper immediately received critical review from most prominent physicists working on

[6] Wiseman [137, slide 15] holds that 'local action' is a better translation of the German '*Nahewirkung*' than 'contiguity'. I agree to a certain extent, but I choose to remain congruent with the bulk of the literature and use 'contiguity' instead.

quantum mechanics (most notably by Bohr [22], cf. [11]) and still continues to do so today: the structure of the thought experiment presented by EPR is still paradigmatic for contemporary discussions about the non-contiguity of quantum mechanics (including the model discussed in this thesis). Sadly, as Einstein lamented, the conflated language of the EPR paper—which was written by Podolsky and not by Einstein—focused the attention on EPR's debatable 'reality criterion' and away from the quintessential part of the argument: contiguity.

Later, in his autobiographical notes, Einstein stated that his theorem established that one of three statements must be false [46]: (i) the statistics of quantum mechanics are complete, (ii) there is an independent reality of distant things, (iii) locality (the principle of contiguity) is true. It has been argued that Einstein's theorem should thus be called the completeness-reality-locality theorem [138]. Einstein fully endorsed (ii) and (iii), and thus found his theorem to imply that the statistics of quantum mechanics were incomplete. That is to say, at the very least, the framework of quantum mechanics should be supplemented by additional 'hidden' variables that are to restore contiguity and still retrieve (or possibly improve on) the statistics of quantum mechanics. Einstein himself went on to search for a deeper local theory underlying both quantum mechanics and general relativity (and it is even argued that Einstein would consider abandoning the field concept—so essential for general relativity—in favor of locality [69, p. 199]).

The next big step in the locality debate was taken by Bohm in the 1950s, who did two important things of relevance for this discussion: he reformulated the EPR thought experiment with mathematical clarity and made it empirically accessible [20, §22.16-18], and he presented a fully functioning, deterministic, non-contiguous interpretation of quantum mechanics based on rediscovered ideas from de Broglie [21]. The latter contribution, now often referred to as Bohmian mechanics, made, for the first time, the non-contiguity of quantum mechanics explicit in its mathematics while at the same time providing an unambiguous ontology: a point-particle with a continuously well-defined position is guided in its motion by a wavefunction that evolves according to Schrödinger's equation. The non-contiguity is explicit because the position of a particle in a multi-particle system is instantaneously dependent on the position of all the other particles (somewhat analogously to how heavy bodies are instantaneously dependent on each other's positions in Newtonian mechanics). And yet, despite Bohm's successful incorporation of non-contiguity explicitly into the mathematics of quantum mechanics, it was still not entirely clear whether *all* interpretations of quantum mechanics must necessarily be non-contiguous. This changed when Bell introduced his conceptual framework and accompanying theorem a decade later.

2.2 After Bell: local causality

In his 1964 paper, Bell proved that, given some reasonable assumptions and barring prima facie undesirable loopholes (Section 5), no contiguous theory at all can reproduce the statistics of quantum mechanics in the EPR thought experiment [16, ch. 2]. In doing so, he seemingly shut the door on any Einsteinian attempt of restoring contiguity in physics and strengthened the 'apparent incompatibility, at the deepest level, between the two fundamental pillars of contemporary theory': relativity theory demands contiguity but quantum mechanics violates it. Bell's theorem, given its wide scope, has been the central pillar of the locality debate since the sixties, but it has also been subject to much misunderstanding. Evidently it is hard—but not impossible—to identify precisely what

Bell did (cf. [23, 87]). Bell's 'characteristically whimsical style' of writing [119, p. 5] is refreshing and inspiring, but it does not help much in this regard. Luckily, Bell followed up his 1964 paper with a flurry of papers that elaborate on his particular terminology and theorem in various contexts—this collection now constitutes Bell's *Speakable and unspeakable in quantum mechanics*—which I elaborate on below.

Bell noted that the existing terminology of, for example, *observables* in quantum mechanics was mathematically precise, but physically 'rather woolly' [16, ch. 7]. To substitute or support this terminology of observables, Bell introduced the term *beable*. Beables are (a direct translation of) the ontology of a physical theory: a beable is that which a physical theory posits to actually *be* there, instead of it being a mere mathematical convenience. Maxwell's electromagnetism is often used as an example: the electric and magnetic fields **E** and **B** are there—they are beables—while the electric and magnetic potentials V and **A** are not—they are mathematical conveniences.[7] Bell further specified a *local beable* as a beable that can be assigned to some bounded space-time region. We recognize here the characteristics of the world of ideas of physics described by Einstein: the beables of a physical theory claim a real existence that is independent of the perceiving object, and these can be called local beables if they can be thought of as arranged in bounded space-time regions in the space-time continuum of relativity theory.

With local beables in hand, Bell formulates the causal structure in physical theories in terms of *local causality*. First, like Einstein, only intuitively [18, p. 224]:

> 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.

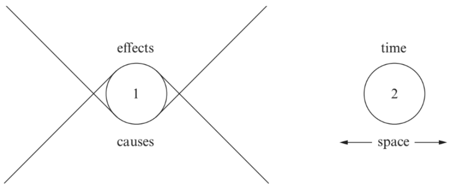
See Figure 1a for his drawing of the situation. Then, Bell formulated this intuition in a way 'sufficiently sharp and clean for mathematics' [18, p. 225]:

> A theory will be said to be locally causal if the probabilities attached to values of local beables in a space-time region 1 are unaltered by specification of values of local beables in a space-like separated region 2, when what happens in the backward light cone of 1 is already sufficiently specified, for example by a full specification of all local beables in a space-time region 3.

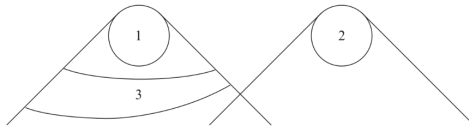
See Figure 1b for his drawing of the situation. Note that there is no mention of cause and effect in this definition, which Bell consciously avoided because those are precisely the type of vague terms that he aimed to remove from physical theories (cf. [17]). Neither is this definition restricted to either deterministic or indeterministic theories: if the 'full specification' of region 3 completely screens off region 1 from region 2, then neither in the irreducibly indeterministic nor in the fully deterministic case can there be a 'locally causal' explanation of possible correlations between region 1 and 2 (cf. [99]).

Bell applied his reasoning to (Bohm's reformulation of) the EPR thought experiment, and used his formulation of local causality to factorize the joint probability of obtaining certain measurement results in region 1 and 2. He then demonstrated that this factorized probability was violated by quantum mechanics. The details of this thought experiment are discussed in the next sections. To repeat, it is often unclear precisely what assumptions go where in Bell's original line of reasoning and the literature on non-locality in quantum mechanics has made tremendous progress since Bell, but it cannot be denied that Bell

[7] There is much to be said about the reality of potentials but this is beyond the scope of this thesis. Cf. [89, 93] for the most recent discussions.



(a) "Space-time location of causes and effects of events in region 1."



(b) "Full specification of what happens in 3 makes events in 2 irrelevant for prediction about 1 in a locally causal theory."

Figure 1: Images and captions from [18, pp. 224-225].

laid the foundation for the present-day conceptual scheme and area of research relating to non-locality in quantum mechanics. Nevertheless, it's fair to say that Bell's *nouvelle cuisine* still finds itself in hell's kitchen.

2.3 The EPR thought experiment

Bell's retelling of Bohm's reformulation of EPR's formulation of Einstein's thought experiment goes along the following lines. Consider a pair of spin-1/2 particles prepared in the singlet state. These two particles 1 and 2 are sent off into opposite directions towards measurement apparatuses where measurements are made of their spins σ_1 and σ_2, respectively, according to a certain measurement settings a and b. If measurement of the component $\sigma_1 \cdot a$ yields the result $+1$, then, according to quantum mechanics, measurement of $\sigma_2 \cdot b$ when $b = a$ must necessarily yield the result -1 and vice versa. Now, since the two measurements were made at spacelike separated locations, neither the choice of measurement setting nor the result of the measurement at one location can influence the measurement result at the other location. But since we can predict with certainty the result of measuring any chosen component of σ_2 after measuring the same component of σ_1, it apparently follows that the result of such measurement must be predetermined. Since quantum mechanics does not provide a description of this predetermination, we may aim for a more complete specification of the state by means of unknown parameters λ from the common past of both particles. Such was the first reformulation of the problem by Bell. But, as he himself later also acknowledged, we do not need to aim for pre-determination, because this might perhaps be too strong a requirement (and indeed we now know that this is the case). We could simply aim for a more complete, local description of how the measurement results came to be, whether it be pre-determined or not, or deterministic or stochastic. This reduces the assumptions that we have to make to some assumption of (observer-independent) *realism* and some form of *locality*. Essentially, both these assumptions are included in Bell's conception of local beables.

Now, let such more complete description be effected by means of those parameters λ, whose precise nature is irrelevant except for the requirement that it consists of local beables.[8] The probability for a measurement result A after measuring $\sigma_1 \cdot a$ should now be completely determined by a and λ, and the probability for result B after measuring $\sigma_2 \cdot b$ should be completely determined by b and λ. This means, in particular, that the probability for result A for particle 1 does not depend on the setting b and certainly not

[8]Note that Bell did not require that the nature of the hidden variables described by the parameters λ are that of local beables. I disagree with Bell: his definition of *local causality* holds if and only if we are concerned only with local beables. I will elaborate on this view in the rest of this thesis.

on the result B, nor does the probability for B depend on a or A. This means that the probabilities for attaining results A and B are:

$$\begin{aligned} P(A|a, b, B, \lambda) &= P(A|a, \lambda), \\ P(B|a, b, A, \lambda) &= P(B|b, \lambda), \end{aligned} \tag{2.1}$$

which may take on all values $0 \leq P \leq 1$ for indeterministic theories and may be limited to only values 0 and 1 for deterministic theories. This relation, equation 2 in [16, ch. 7] (in my notation), Bell used as the earliest formulation of his local causality. Now, Bell states (most clearly in [18]), if we take the standard rule for the joint probability of results A and B: $P(A, B|a, b, \lambda) = P(A|B, a, b, \lambda)P(B|a, b, \lambda)$ and we invoke local causality as defined in Equation (2.1), we may declare redundant certain of the variables upon which we conditionalize in this expression to obtain:[9]

$$P(A, B|a, b, \lambda) = P(A|a, \lambda) \cdot P(B|b, \lambda). \tag{2.2}$$

This *factorized* equation tells us that A and B have no dependence on one another, nor on the remote settings, but only on their own local setting and the complete description λ.

Now, it can be shown that quantum mechanics does not obey the factorized Equation (2.2): if we repeat the EPR experiment described above many times, we can define an expectation value for the product of results A and B as

$$E(a, b) := \int_\Lambda \sum_{A,B} AB \cdot P(A, B|a, b, \lambda)\rho(\lambda|a, b)d\lambda, \tag{2.3}$$

where $\rho(\lambda|a, b)$ is a probability distribution for the hidden variables λ. We now make two important assumptions: (i) λ is not in any way influenced by the variables a and b, so that we may write $\rho(\lambda|a, b) = \rho(\lambda)$, and (ii) the probability distribution $\rho(\lambda)$ is normalized, so that $\int \rho(\lambda)d\lambda = 1$. From this, we can derive the CHSH-inequality [28] (see Appendix A for a full derivation):

$$|E(a, b) + E(a, b') + E(a', b) - E(a', b')| \leq 2, \tag{2.4}$$

which is violated by quantum mechanics, where the value of this expression can approach $2\sqrt{2}$. Thus, quantum mechanics cannot be embedded in a locally causal theory.

2.4 Concluding remarks

Today, the philosophy of physics community is firmly situated in the contemporary paradigm centered around Bell's conceptual framework.[10] The beauty of Bell's concep-

[9] It is sometimes argued that Bell found it sufficiently satisfactory to declare the distant measurement result B redundant from purely intuitive reasons (cf. [114, ch. 3, footnote 15]) because Bell's earlier definitions of *local causality* (as in equation 2 in [16, ch. 7]) does not very explicitly include it its formulation as well. As we have seen, however, Bell defined local causality more broadly in terms of light cones [17], from which reasoning the redundancy of the distant result B is obvious and the derivation of Eq. (2.2) is clear. The important lesson is that Bell considered space-time regions that included both a measurement setting and a result, and declared them jointly redundant. The explicit distinction between 'influence' on the probabilities by the settings and the results came later, mainly through the works of Jarrett and Shimony, as we will see in Section 3.

[10] To illustrate, consider the recently established *John Bell Institute for the Foundations of Physics* (www.johnbellinstitute.org).

tual framework is, firstly, that it framed the locality problem in a largely interpretation-independent manner and thus allows us to do philosophy of quantum mechanics with a very wide scope. This is contrary to other central problems of quantum mechanics, such as the measurement problem, which mostly demand a strongly interpretation-dependent treatment. As all the problems in quantum mechanics are fundamentally interconnected, one may hope that Bell's framework may then assist us in the search for a solution to the other problems. Another advantage is that the ontology roughly represented by beables is very intuitive, especially when considered in the highly abstract context of quantum theory. It has, to use a term Schrödinger often employed, *Anschaulichkeit* (cf. [38]). I endorse the Bellian framework, as is evident by the focus thereon in this thesis, but I do consider it more of a coarse-graining proxy approach that may guide us in future research rather than as 'actual ontology'. There are several reasons for treating Bell's framework with caution, some of which are not acknowledged explicitly often enough in the literature.

The first reason of caution is that, to use Bell's words against himself, a beable itself is physically (or philosophically) quite precise, but mathematically rather woolly. It is not straightforward to see how to extract a beable from a wavefunction, especially not a local beable. Let's say that an isolated, wavefunction localized to a bounded region of space-time is a local beable. Then we must immediately acknowledge that such wavefunctions are very rare because most unrestricted wavefunctions rapidly spread out over all space and so quickly lose candidacy for being a beable.[11] Similarly, it is not clear whether overlapping wavefunctions can be beables, however small the overlap may be. One could try to define local beables with more mathematical precision (see [62, fn. 29]), but I believe one will quickly find oneself working again in a specific interpretation-dependent context rather than the broader Bellian framework.

Relatedly, Bohm's mathematical representation of the EPR-thought experiment that was the original target of Bell's framework contains no notion of space or time, whilst the very problem this framework aims to deal with is *locality.* This problem is best formulated by Muller in [95, p. 242, references omitted]:

> *Lacuna 1:*
>
> Whereas the notions space and time figure prominently in the foundational-philosophical investigations of locality, and whereas the notions of space and time are necessarily presupposed in the very idea of *locality*, space and time are absent from the Bohm singlet.
>
> ...
>
> To fill Lacuna 1 means (i) to give a description of the Bell-experiment where Galilean space-time is part of the mathematical pasta, if only in order to express mathematically that the spin measurements are performed simultaneously in different regions of space, (ii) to formulate a locality condition of a manifestly spatio-temporal nature, and (iii) to prove Bell's Theorem in space-time.
>
> *Lacuna 2:*
>
> All descriptions of the Bell-experiment are in the frame-work of orthodox quantum mechanics, which is in fact a *Galilean-covariant* theory and therefore not supposed to satisfy a locality condition at all. On the contrary, it should

[11] Nor does the beable-framework appear to be really applicable to the universe as a whole [42].

> *violate* locality generically just as the *actio in distans* potentials of Galilean-covariant classical mechanics violate locality generically.

Interestingly, there is a long tradition of formulating Bell's locality and Bell-type inequalities in Minkowski space-time and quantum field theories—eg. [26, 45, 77, 95, 111, 112], to name a few, and by Bell himself. But in the contemporary literature on violations of Bell-type inequalities, where one mainly aims to do *causal* analysis in the tradition following Shimony (as we will see in the next section) these concrete space-time formalizations of Bell-type inequalities are largely missing. The most popular current formalism for such analysis is the basic probabilistic formalism that I adopted in this thesis as well, which has no reference to space or time at all. One could see this as a weakness because the mathematics is ambiguous with regards to the very problem that it aims to capture, or one could see it as a strength in the sense that the Bellian framework is applicable in a very broad context—as long as one keeps in mind that the formalism is not very precise. With regards to timelike non-locality, however, as we will see, it is precisely this ambiguity and the lack of concrete formulation in space-time that allows one to entertain Bell-type non-locality in time rather than in space: there is nothing in the formalism that stops me from doing so. A reasonable worry here is that such attempts fallaciously exploit the ambiguity in the Bellian formalism, and that, since my objections against such attempts in Section 5 are different, here lies a whole new class of objections to these attempts.

To build on this a bit, we must realize that the problem of locality is strongly interwoven with the phenomenon of entanglement, but that the Bohm singlet that was the original subject of Bell's theorem is only a very special case of entanglement. It is, for example, only bipartite, whilst entanglement can exist between three or more particles (as in the GHZ scenario [37]). Moreover, entanglement comes in degrees and can exist between different degrees of freedom. Especially in the pure metaphysics literature on entanglement, one regularly comes across 'all-or-nothing' metaphysical definitions of entanglement based solely on the Bohm singlet or on Bell's theorem that do not take into account all of the aspects of entanglement described above (cf. [74]). We must always be careful that we don't draw strong conclusions from Bell's theorem that do not hold outside this very special scenario.

A different reason for caution is with regards to Bell's definition of (local) causality. As said in Section 2.2, Bell's definition of causality does not include the 'vague terms' such as cause and effect themselves, nor is it restricted to deterministic or indeterministic theories. Bell achieves this by adopting a probabilistic definition of causality: causes raise the probability of their effects. Indeed, most Bell-type analysis is in terms of probabilities. But probabilistic definitions of causality are (metaphysically) quite contentious as one can quickly come up with examples of causes that don't raise the probability of their effects and non-causes that do (see [60] for a recent discussion). Recently, efforts are made to supplement this probabilistic causality in the context of Bell's theorem with graphical causal models (eg. [139, 119]). This point will briefly return in Section 5.4. It is useful to already here mention that especially retrocausalists adopt a different conception of causality, alongside Bell's probabilistic framework. They use an *interventionist* account [109, p. 7763]: a variable X is a cause of a variable Y if and only if a free intervention on X makes a difference to Y. As I said, I still endorse the Bellian framework. I believe that these problems faced by Bell's probabilistic causality are relatively harmless if we keep in mind that this framework is not as precise as we might want it to be. The lack of metaphysical rigor here is not a no-go for the Bellian framework, but it does indicate that we must always be wary not to draw too strong conclusions about causality from considerations about probability too hastily.

Another reason for caution is with regards to the probabilistic formalism itself. There are mainly two competing notations for Bellian conditional probabilities, one with a conditionalization stroke such as $P(A|a,b)$ and one with subscripts such as $P_{a,b}(A)$. There is a difference between the two, and this difference will become relevant in various parts of this thesis, so we need to understand it. The difference has been stated most clearly by Butterfield in [26, p. 117-118], which I quote at length:

> [By] having [$\mathrm{prob}_{i,j}(x,y)$] contain i and j in the subscript, rather than after a conditionalization stroke, as in $\mathrm{prob}(x,y|i,j)$, we are assuming that there are many probability spaces (each with the four events $\langle +1,+1\rangle, \langle +1,-1\rangle, \langle -1,+1\rangle, \langle -1,-1\rangle$), each labeled by its settings $\langle i,j\rangle$, rather than one big probability space, whose events would be all the various quadruples $\langle x,y,i,j\rangle$. Mathematically, there is little difference between these approaches. The big space is, of course, partitioned into subspaces, labeled by $\langle i,j\rangle$, each obtained by conditionalizing on its value $\langle i,j\rangle$. And given the many spaces, labeled by the settings, it is a mathematical triviality to build a single space whose conditional probabilities $\mathrm{prob}(x,y|i,j)$, $\mathrm{prob}(x|i,j)$, etc., are the given $\mathrm{prob}_{i,j}(x,y)$, $\mathrm{prob}_{i,j}(x)$, etc. (The big space we build is not unique: we can assign each label $\langle i,j\rangle$ whatever probability we like, provided they sum to one.) But there is a conceptual difference between the approaches. With many spaces, one is not committed to apparatus settings, i and j, having a probability; with the big space, one is. I do not think that every proposition or event has a probability; and since one thinks of apparatus settings as decided by the experiment, they seem quite good candidates for not having a probability. In any case, it seems to be worth signaling where one can avoid the assumption that some proposition has a probability. Accordingly, I prefer to use many spaces and the above notation. Nevertheless, for the sake of uniformity with the other papers in this volume, I am happy to use '$\mathrm{prob}(x,y|i,j)$' etc.: we shall never have to calculate suspicious probabilities like $\mathrm{prob}(i,j)$, or $\mathrm{prob}(i,j|x,y)$.

We will see that my objection raised to certain independence assumptions in Section 3.2.3 is precisely that we have to calculate suspicious probabilities for measurement settings.[12] Personally, I still choose to work with the conditionalization stroke for three reasons: I find it much easier to read, the bulk of the literature that I used for this thesis uses the same notation,[13] and, like Butterfield, I find myself justified in using this notation because I explicitly acknowledge the difference between the two notations, thus avoid the pitfalls, and do not commit myself to an assumption that some propositions about measurement settings have probabilities by emphasizing the role of measurement settings simply as boundary conditions for the model. This last point will become more apparent in the next section.

[12] Although it will not be discussed much further in this thesis, I do need to mention here a recent event. A strongly related problem of whether or not there exists some 'unconditional probability for hidden variables' was the cause of a feud between Maudlin [88] and Leifer [80] that followed an interesting paper by Leifer and Pusey [81]. One might suspect that such feuds may be avoided if there was less ambiguity in Bell's probabilistic formalism.

[13] At the Foundation2018 conference in Utrecht, Wiseman's presentation came directly after mine, and we were both happy to see that we were using the same notation for the same topic.

3 Spacelike non-locality

3.1 Basic assumptions

Consider the representation of the EPR-type thought experiment in Figure 2. Nearly every paper that discusses EPR-type thought experiments provides a similar figure, and I do not intend to reinvent the wheel, but there is one important difference with this figure from the rest: in the worldlines of the measurement apparatuses I explicitly introduce the measurement settings (temporally) previous to the measurement interaction and the measurement result after the measurement interaction. Often in the literature the measurement settings, apparatus, and outcome are all represented in the same space-time region. I separate these to show clearly that the measurement setting is a boundary condition for the model, but also because this temporal order of setting–interaction–result reverses when we invoke symmetry arguments in Section 5.3. Because 'EPR-type thought experiment' is quite a mouthful, I henceforth refer to this experimental setup as simply 'the model'.

If we recast the description of the experiment from Section 2.3 in terms of the model, the description is as follows. Consider a quantum system λ composed of two entangled, indistinguishable particles I and II. These particles are sent off in opposite directions towards measurement apparatuses μ_a and μ_b that are situated at space-time regions V_a and V_b respectively. At these locations, measurements on I and II are performed according to measurement settings a and b, yielding results A and B, respectively. The variable λ is the complete specification of the state of the composite system, and, similarly, μ_a and μ_b are complete specifications of the states of the measurement apparatuses.[14] The possible values for all these variables are defined to be: $A \in \{-1, 1\}, B \in \{-1, 1\}, a \in \{r, r'\}, b \in \{r, r''\}, \mu_a \in M_a, \mu_b \in M_b, \lambda \in \Lambda$, where M_a, M_b, Λ are the state spaces for μ_a, μ_b, λ.

Before we continue the discussion of the model and the independence assumptions that lead to Bell's factorized probability function, let us evaluate the basic assumptions that have been made during the formulation of the model above and which loopholes for avoiding the consequences of Bell's proof these assumptions already exclude. Many such lists have been made and none is exhaustive (eg. [27, 129, 137]), but here I list broadly the assumptions that are necessary for formulating and interpreting Bell's theorem. The last two assumptions are those that the loopholes to Bell's theorem discussed in Section 5 violate.

Observer-independent reality: *There exists an observer-independent reality that is, to a certain extent, described by a physical theory.*

This excludes: (i) quietism or instrumentalism about physical theories, for which questions like 'does quantum mechanics tell us there is non-locality in the world' should either not be asked or are unworthy of being answered. (ii) Observer-dependent realism, such as Bohr's interpretation of quantum mechanics based on his concept of complementarity (cf. [70]). (iii) Solipsism and philosophical skepticism

[14] Seevinck [114, p. 49], following Clifton et al. [29] and Butterfield [26], contra Bell [16, ch. 7] argues that the specification of λ, μ_a, μ_b should be the complete state of the entire causal history of the systems in question. Whilst I acknowledge that this specification is more general than mere specification of the state at one instant in time, I side with Bell in stating that the specific nature of λ, μ_a, μ_b is empirically inaccessible and therefore irrelevant for the purpose of the model—as long as we may assume that their distributions are properly normalized.

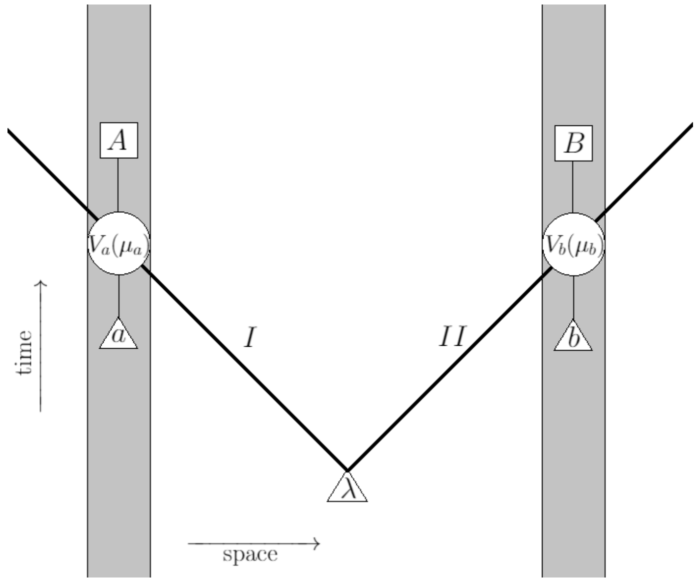


Figure 2: Representation of the EPR-type thought experiment in space-time with worldlines of the particles I and II and measurement apparatuses μ_a and μ_b. Triangles indicate that the measurement settings a and b and the state λ are boundary conditions for the experiment, squares indicate that the measurement results A and B are empirically accessible results of the experiment.

about the existence of an external world, where any statements about properties of the external world are nonsensical (cf. [100]).

Space-time continuum: *The things in the external world are arranged in a space-time continuum.*

This is a necessary assumption for the formulation of local causality. Because the quantum mechanical description takes place in a (generally higher-than-three-dimensional) Hilbert space, this assumption demands that we project the predictions of whatever interpretation of quantum mechanics we invoke onto a four-dimensional space-time continuum. This excludes statements about the locality of a theory in configuration space: if we cannot talk about space-time, we cannot talk about local causality.

Local beables: *The observer-independent things can be assigned to some bounded region of the space-time continuum.*

This is Bell's definition of local beables. It distinguishes between beables and mere mathematical conveniences, and restricts the beables to a bounded region of space-time. This excludes both beables existing in the entirety of space-time and single (non-contiguous) beables distributed over disjunct regions of space-time.

Local causality: *The causes for local beables lie entirely in their backwards lightcones, and the effects of local beables lie entirely in their forwards lightcones.*

The probabilities attached to values of local beables in a space-time region 1 are unaltered by specification of values of local beables in a disjunct space-time region 2, when what happens in region 1 is already sufficiently described by a full specification of all local beables in a space-time region 3 that completely screens off the backward lightcone of region 1 from (the backwards lightcone of) region 2, see Figure 1b.

Unique outcomes: *Every single run of the experiment gives only one set of unique results.*

This excludes the many-worlds interpretation of quantum mechanics, Many-worlds is usually excluded from the non-locality discussion for a few other reasons as well (eg. it is hard to formulate Bellian local causality in many-worlds), but I find this the easiest way. It is clear from the literature that the locality-debate takes on quite a different shape in the context of many-worlds, and this thesis is not the place for it.

Independent boundary conditions: *The specific values of boundary conditions are independent of each other and of any other variable of the model.*

This (symmetrically) excludes dependence of the measurement settings on each other, and dependence of the measurement settings on the hidden variables and vice versa.

Model completeness: *The space-time continuum and the local beables described by the model in Figure 2 are, from the perspective of the variables in the model, the entirety of reality.*

This ensures the independence of the model, so that we do not smuggle in some external factors not originally described by the model to explain features in the model. The assumption is made to ensure the simple status of 'boundary conditions' for the measurement settings and the quantum state. This excludes: (i) postulating a common cause for both the values of the measurement settings in a single run of the experiment, in accordance with the previous assumption, and (ii) smuggling (the free will of) a human experimenter into the model to account for the specific values of the measurement settings. To use a familiar phrase: from the perspective of the model, this model is everything that is the case.

In Section 2, we have seen that the first four of these assumptions, *Observer-independent reality, Space-time continuum, Local beables*, and *Local causality* are necessary assumptions that we need to make before we are even able to formulate Bell's theorem and therefore to understand what non-locality entails in this context. The assumption of *Unique outcomes* is presented in this thesis in order to explicitly exclude many-worlds from this discussion. Because the locality debate (and its philosophical and metaphysical consequences) transforms fundamentally when considered in the context of many-worlds, this is beyond the scope of this thesis.[15]

This list is by no means complete, but it is sufficient for the analysis in this thesis. Another explicit assumption could concern, for example, counterfactual definiteness: the model is well-defined for all possible combinations of boundary conditions and variables. There have recently been attempts to blame the violation of Bell-type inequalities solely on the violation of such counterfactual definiteness, but the sting seems to be out the argument (see [19, 84, 65, 85] and [76]). I omit such assumptions because they are not necessary for the arguments in this thesis.

[15] See, for example, [122] for a recent argument that non-locality does not exist in many-worlds. See further [126] and the references therein. In my opinion, non-locality in the sense of Bell appears impossible (or very hard) to define unambiguously in many-worlds—its meaning changes when considered in many-worlds—and I am therefore usually unimpressed by arguments along these lines.

The two assumptions that are relevant for this thesis are *Independent boundary conditions* and *Model completeness*. The latter, *Model completeness*, is violated by the superdeterministic proposals that I discuss and argue against in 5.2. *Independent boundary conditions* is the assumption that is violated by every loophole (including retrocausal and all-at-once) that I discuss in Section 5 in one way or another. It is essentially the strongest form of what is commonly referred to as 'statistical independence' or 'source independence'. As we will see in the next section, it is the very first assumption that we need to make in order to derive the factorized equation (2.2) that Bell used as his definition of local causality, and it is needed again in the derivation of Bell-type inequalities (see Appendix A).

3.2 Spacelike factorizability

3.2.1 Independence assumptions

If we want to declare redundant certain variables upon which we conditionalize in our probability functions in order to derive Bell's factorized equation (2.2), we must consider and constrain their possible (inter)dependencies. Seevinck [114] has done so explicitly in the clearest way, and therefore I mainly follow his line of thought in this section.[16] We start by acknowledging that in any specific measurement we have no knowledge of the precise states of λ, μ_a, μ_b, but that we can only constrain the possible probability distributions to depend backwards lightcones according to the principle of local causality. The first thing we state is that the distribution of λ is statistically independent of the other boundary conditions—the measurement settings a and b—and formalize this as **Source Independence**:

$$\text{SI:} \qquad \rho(\lambda|a,b) = \rho(\lambda|a',b) = \rho(\lambda|a,b') = \rho(\lambda|a',b') = \rho(\lambda). \tag{3.1}$$

This assumption relates to the basic assumptions of *Independent measurement settings* and *Model completeness*: the measurement setting a and b and the state λ are boundary conditions *for* the model, they cannot be further specified by anything *in* the model.

The values of μ_a and μ_b, however, are not boundary conditions for the model. Their distributions may thus depend on the boundary conditions in their own backwards light cones. Firstly, this implies that these distributions are dependent on their measurement settings but not on the setting of their counterparts. This is formalized as **Apparatus Locality**:

$$\begin{aligned} \text{AL:} \qquad &\rho(\mu_a|a,b,\lambda) = \rho(\mu_a|a,b',\lambda) = \rho(\mu_a|a,\lambda), \\ &\rho(\mu_b|a,b,\lambda) = \rho(\mu_b|a',b,\lambda) = \rho(\mu_b|b,\lambda). \end{aligned} \tag{3.2}$$

Secondly, these distributions are also independent of each other. This allows for factorization of their joint distribution with an assumption called **Apparatus Factorizability**:

$$\text{AF:} \qquad \rho(\mu_a,\mu_b|a,b,\lambda) = \rho(\mu_a|a,b,\lambda)\rho(\mu_b|a,b,\lambda). \tag{3.3}$$

[16] I adopt his notation with some small deviations. Firstly, I reversed Seevinck's choice of a, b for measurement outcome and A, B for measurement setting to stay more in line with Bell and the papers discussed in Section 5. Secondly, probabilities conditioned on every variable of the model, including measurement apparatus states, are denoted by a bold and capital $\mathbf{P}$, while probabilities averaged over all possible measurement apparatus states are denoted by a smaller P.

The conjunction of AL and AF then gives **Total Apparatus Factorizability**:

$$\text{TAF:} \qquad \rho(\mu_a, \mu_b|a, b, \lambda) = \rho(\mu_a|a, \lambda)\rho(\mu_b|b, \lambda). \tag{3.4}$$

Now if we consider that our first condition, Source Independence, implies that $\rho(\lambda, \mu_a, \mu_b|a, b)$ $= \rho(\mu_a, \mu_b|a, b, \lambda)\rho(\lambda)$, and we take the conjunctions of this with TAF, we obtain the grand total **Independence of Systems and Apparatuses**:

$$\text{ISA:} \qquad \rho(\mu_a, \mu_b, \lambda|a, b) = \rho(\mu_a|a, \lambda)\rho(\mu_b|b, \lambda)\rho(\lambda). \tag{3.5}$$

ISA formalizes the independence of λ from the settings a and b, the independence of μ_a from b and μ_b, and the independence of μ_b from a and μ_a. It is useful to already note here that ISA reduces to SI when we average over all possible measurement states μ_a and μ_b with $\int_{M_a \times M_b} \rho(\mu_a|a, \lambda)\rho(\mu_b|b, \lambda)d\mu_a d\mu_b = 1$.

Now that we have constrained the dependencies of the distributions with TAF, we can formalize probability assignments for the measurement outcomes A and B conditional on all $a, b, \mu_a, \mu_b, \lambda$. We start by considering the probabilities for certain measurement results A and B individually, and again use the fact that they may only depend on the variables in their backwards lightcone. The probability for A may thus only depend on a, μ_a, and λ, and the probability for B may only depend on b, μ_b, and λ. Seevinck, following Clifton et al. [29], calls this assumption **Outcome Locality**:

$$\begin{aligned} \text{OL:} \qquad \mathbf{P}(A|a, b, \mu_a, \mu_b, \lambda) &= \mathbf{P}(A|a, \mu_a, \lambda), \\ \mathbf{P}(B|a, b, \mu_a, \mu_b, \lambda) &= \mathbf{P}(B|b, \mu_b, \lambda). \end{aligned} \tag{3.6}$$

The outcomes A and B themselves also lie outside each other's light cone, and their joint probability may thus be factorized analogously to (3.2) in the assumption called **Outcome Factorizability**:

$$\text{OF:} \qquad \mathbf{P}(A, B|a, b, \mu_a, \mu_b, \lambda) = \mathbf{P}(A|a, b, \mu_a, \mu_b, \lambda)\mathbf{P}(B|a, b, \mu_a, \mu_b, \lambda). \tag{3.7}$$

The conditions OL and OF are identical to Jarrett's conditions 'Locality' and 'Completeness', respectively, as introduced in his influential paper 'On the Physical Significance of the Locality Conditions in the Bell Arguments' [72]. The conjunction of OL and OF gives, for the joint probability, the assumption called **Total Factorizability**:

$$\text{TF:} \qquad \mathbf{P}(A, B|a, b, \mu_a, \mu_b, \lambda) = \mathbf{P}(A|a, \mu_a, \lambda)\mathbf{P}(B|b, \mu_b, \lambda). \tag{3.8}$$

TF is explicitly dependent on the states of the measurement apparatuses μ_a and μ_b, but our insufficient knowledge of these states complicates any calculation of surface probabilities $P(A)$ and $P(B)$. Indeed, we only have empirical access to the conditional probabilities $P(A|B, a, b)$ and $P(B|A, a, b)$. Following Bell [16, ch. 7], it is often argued that the states of the measurement apparatuses should not be included in the probabilities, but that these probabilities should instead be averaged over all possible $\mu_a \in M_a$ and $\mu_b \in M_b$. We cannot know the state of the measurement apparatus in single experiment procedure, and are thus somewhat forced to assume that the measurement settings a and b and the state λ are the only relevant variables for the outcomes A and B. (Cf. Shimony [116, p. 226], Clifton et al. [29, p. 161], and Seevinck [114, §3.2.4].) Seevinck justly notes that we cannot simply include the μ_a and μ_b in the state λ by redefining $\lambda' = (\lambda, \mu_a, \mu_b)$ because

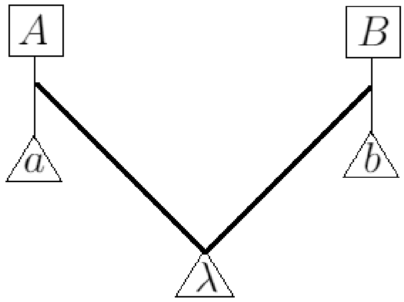


Figure 3: Remaining variables in the toy model after rendering the measurement apparatus states irrelevant by averaging over all their possible states.

then the measurement apparatuses become dependent on the state of their counterpart measurement apparatus, which lie outside their past light cones and thus violates local causality.

Now, although we can derive Bell-type inequalities with the conjunction of Outcome Locality and Outcome Factorizability (Locality and Completeness, respectively, in Jarrett's terminology),[17] the bulk of the contemporary literature on Bell-type inequalities is solely concerned with the μ-averaged probabilities. We define the new probability functions—here specifically for the joint probability—and explicitly include this averaging in the assumptions, call it **Apparatus Averaging**:

$$P(A,B|a,b,\lambda) := \int_{M_a \times M_b} \mathbf{P}(A,B|a,b,\mu_a,\mu_b,\lambda)\rho(\mu_a,\mu_b|a,b,\lambda)d\mu_a d\mu_b, \tag{3.9}$$

and similarly for other probabilities. For the sake of clarity, we can represent the connections between the remaining variables as in Figure 3. I discuss the independence assumptions for the μ-averaged probability function in the next section.

3.2.2 Shimony factorizability

The μ-averaged analogues of OL and OF were first introduced by Shimony [117]. The analogue of OL he called **Parameter Independence**:

$$\text{PI:} \qquad \begin{aligned} P(A|a,b,\lambda) &= P(A|a,\lambda), \\ P(B|a,b,\lambda) &= P(B|b,\lambda), \end{aligned} \tag{3.10}$$

and the analogue of OF he called **Outcome Independence**:[18]

$$\text{OI:} \qquad \begin{aligned} P(A|B,a,b,\lambda) &= P(A|a,b,\lambda), \\ P(B|A,a,b,\lambda) &= P(B|a,b,\lambda). \end{aligned} \tag{3.11}$$

The analogy between the OL–OF and PI–OI independence assumptions is obvious, but it should be noted here that the two sets of independence assumptions are logically in-

[17] See, for example, [12, Lecture 8, Part II, p. 4] for a derivation. I take this opportunity to emphasize my gratitude for Bacciagaluppi's excellent education and for his supervision of this thesis. He taught me nearly everything I know about the foundations of quantum mechanics.

[18] The original notation of OI is $P(A,B|a,b,\lambda) = P(A|a,b,\lambda)P(B|a,b,\lambda)$, but I present it in the form of (3.11) because this is easier to compare with Maudlin's assumptions in the next section.

dependent: both sets of independence assumptions can be derived from the assumption of local causality and Bell-type inequalities can in turn be independently derived from the conjunction of either set of assumptions.[17] Moreover, PI can be violated whilst OL holds and vice versa, and similarly OI can be violated whilst OF holds and vice versa.[19] Violation of one set of assumptions does not inform us about the possible violation of the other set of assumptions, but a violation of either set of assumptions certainly informs us that local causality is violated. One must therefore distinguish explicitly between (the physical meaning of) Jarrett's assumptions and Shimony's assumptions, I will return to this point in the concluding remarks of this section.

The conjunction of PI and OI gives the μ-averaged analogue of TF, called simply **Factorizability**:

$$\text{F:} \qquad P(A,B|a,b,\lambda) = P(A|a,\lambda)P(B|b,\lambda). \tag{3.12}$$

We have now obtained Bell's factorized Equation (2.2) with explicit formulation of all necessary assumptions accompanying the definition of a locally causal theory and the details of the model.

Thus, to repeat: if you have a locally causal theory, you can derive Factorizability (3.12). If you have Factorizability, then you can derive CHSH-inequalities (Appendix A). CHSH-inequalities are violated by all quantum mechanics (properly understood in a locally causal space-time context). Ergo, quantum mechanics is never a locally causal theory.

There are now two options: either we acknowledge this and accept that quantum mechanics implies that its peculiar non-locality (understood as a violation of Bell's local causality) is simply a fact of the world, or we deny one of the basic assumptions of the model and argue that there is no non-locality in the conventional sense. The latter option is discussed in Section 5. Regarding the former, the power of Bell's theorem is that this is mandatory for any standard realist interpretation of quantum mechanics that aims to be applicable in a space-time context. Bohmian mechanics and collapse theories are standard examples, and they violate Parameter Independence and Outcome Independence, respectively, thus manifesting the non-locality by violating either one of Shimony's independence assumptions. Advocates of either realist interpretation then face the task of constructing a Lorentz-invariant formulation of the theory—this usually involves introducing a preferred reference frame (that is conveniently phenomenologically shrouded from observation)—and, accordingly, constructing the theory so that non-locality is manifest and yet superluminal signaling remains impossible. It is also a simple consequence—and this appears to be most difficult to accept for many physicists and philosophers alike—that one must generally make peace with the idea of non-locality. Perhaps one may console oneself with the thought that this non-locality is 'discriminating' in the sense that, in Maudlin's words [86, p. 22], the causal influence does not (instantaneously) propagate outwards in every direction but that it is a private connection only between the two (or more) particles in question, and that it may thus lend itself more easily to the 'postulation of laws which can be checked empirically in the accepted sense' and avoid the empirical and metaphysical anarchy of which Einstein was so apprehensive.[20]

Before I discuss Maudlin's independence assumptions in the next section, some words about the aim of Shimony's independence assumptions need to be said. Whereas a fail-

[19] See [12, Lecture 8, Part II, p. 4] for demonstrations.

[20] This is not the case for certain existing results claiming 'entanglement in time' discussed in Section 4.8, which indicates that those results are not concerned with the same type of non-locality as we are concerned with here.

ure of Bell's original Factorizability (3.12) equation could tell you that there was some correlation between the spacelike separated measurement events as a whole, it could not assist in the analysis of what aspect of each measurement was the culprit of the non-local correlations. Shimony's independence assumptions can do this: they give you the feeling of putting the finger on the sore spot of non-locality of either the measurement setting or the measurement outcome. And, because it seems that different interpretations of quantum mechanics violate different independence relations, we get the idea that these independence assumptions lead us forward in the analysis of the possibly-causal links in the non-locality of quantum mechanics. But inferring a causal link from the violation of an already suspicious probability function is (metaphysically) contentious, and it forces us to carefully re-evaluate what precisely we may infer from these probabilities, if anything at all. One thing we need to take into consideration is that the literature is in disagreement about whether or not OI and PI refer to the same type of causal intervention. Whereas PI clearly refers to an action (of introducing a measurement setting to the system), OI does not an *interventionist* feel to it. I aim to take no stance on this debate in this thesis, and I try to sidestep it as follows.

For the case of Shimony's independence assumptions—in contrast with Maudlin's, as we will see—I shall make the following statement: insofar as we are able to infer a causal link from the violation of a probability function at all, we are certainly only able to do so if the probabilities that we use can be reasonably interpreted as ontic probabilities. Ontic probabilities, in these cases, are those probabilities that we compute with definite values for the variables upon which we conditionalize. So, for example, to be able to interpret the probabilities in the relation $P(A|a, b, \lambda) = P(A|a, \lambda)$ as ontic probabilities, we need to input definite values for a, b, and λ. If, on the contrary, we cannot input definite values and instead need to average over all possible values for a specific measurement setting, then this integration move should be interpreted as a move motivated by ignorance, and the resulting probabilities can only be reasonably interpreted as epistemic probabilities. And epistemic probabilities surely do not provide us with the same feeling of putting the finger on a causal link as ontic probabilities do. This difference between ontic and epistemic probabilities in the context of Bell-type inequalities will become more clear in the next section, where I argue that Maudlin's independence assumptions can only be reasonably interpreted in terms of epistemic probabilities, and can thus not assist us much in the search for a causal link.

3.2.3 Maudlin factorizability

In his seminal *Quantum Non-locality and Relativity* [86], Maudlin carefully analyzes the non-locality in quantum mechanics in the context of different interpretations and their possibilities of reconciliation with relativity theory (and what this would require). His conclusions may be summarized as follows. Violation of the CHSH-inequality does not require superluminal matter or energy transport, nor does it entail the possibility of superluminal signaling. It does, however, require superluminal causal connections and it can only be accomplished if there is superluminal information transmission.[21] Barring the usual loopholes, of course. Standard realist interpretations that violate CHSH-inequalities

[21] According to Maudlin, this is always the case, both when Outcome Independence and when Parameter Independence are violated. But, as I said in the previous section, Maudlin's stance on this topic is not universally shared. If one holds instead that a violation of Outcome Independence does not imply any causal link in the same way as that Parameter Independence does, then Maudlin's conclusion that violations of CHSH-inequalities always require superluminal causal connections obviously does not follow.

can possibly be made Lorentz-invariant, but it seems that this would require either backwards causation or hyperplane dependence. It could even suggest the existence of a single preferred reference frame shrouded from observation (and thus the failure of Lorentz invariance). See [86, Ch. 10] for further discussion of these matters, which are outside the scope of this thesis.

Besides discussing the reconciliation of quantum non-locality and relativity, Maudlin argues that we cannot experimentally test whether it is specifically failure of Outcome Independence or of Parameter Independence that leads to the violation of the CHSH-inequality. Indeed, he states, the entire discussion of 'which assumption fails' is rather arbitrary because the conjunction of Outcome Independence and Parameter Independence is not the only set of assumptions logically equivalent to Factorizability. We have already seen this is the case in the previous section with respect to Jarrett's independence assumptions, but Maudlin has introduced another alternative set of assumptions that are logically independent from the sets we already know. Before I discuss Maudlin's alternative set of assumptions, it must be said that Seevinck [114, §3.2.6] too argues firmly against such possible 'experimental metaphysics'. Seevinck does so on philosophical grounds because, indeed, 'the relationship between on the one hand OI and PI and other the other hand spacelike causation, signaling and relativity are much more subtle than is acknowledged in most of the experimental metaphysics projects', as Maudlin similarly argues throughout his book, but also on mathematical grounds. Consider, for example, the hilarious non-local analogue of PI where the measurement outcomes may only depend on the settings of their counterparts, **Anti-Parameter Independence**:

$$\text{API:} \qquad \begin{aligned} P(A|a,b,\lambda) &= P(A|b,\lambda), \\ P(B|a,b,\lambda) &= P(B|a,\lambda). \end{aligned} \tag{3.13}$$

The conjunction of API with Outcome Independence (which is the same in both the local and the non-local case) leads to **Non-local Factorizability**:

$$\text{NF:} \qquad P(A,B|a,b,\lambda) = P(A|b,\lambda)P(B|a,\lambda), \tag{3.14}$$

from which we also get a CHSH-inequality using the standard derivation [114, §3.3]. This already means that Parameter Independence is not necessary for the derivation of CHSH-inequalities (which we already knew), and, consequently, it appears to imply that we cannot test experimentally whether it holds or fails. But there are two problems here. The first problem is that we can easily dismiss API as a useful tool for analysis on intuitive (and even empirical) grounds. The probabilities on the right-hand side of the equations of (3.13) specify measurement results that are only dependent on distant measurement settings (and λ). This implies, for example, that Alice cannot obtain a result if Bob has not selected a measurement setting. That this is not the case can be easily verified by an EPR-type experiment with only one active wing.[22] Clearly, Anti-Parameter Independence is just a joke.

But the second problem is even more pressing. Logically independent sets of independence assumptions, which can be independently derived from local causality and from which Factorizability can be independently derived, are, to repeat the obvious, *logically in-*

[22] It should be noted that Shimony's original paper is one of the very few papers that actually define independence relations for three possible measurement settings (instead of the usual two): two possibilities for opposite measurement settings, and one possibility for *no measurement setting* and therefore no experiment [117, p. 187].

dependent. A violation of either of Shimony's assumptions does not imply anything about a violation of either of Jarrett's assumptions. Nor does a violation of Anti-Parameter Independence inform us about violations in Shimony's or Jarrett's assumptions. Likewise, whatever alternative set of assumptions Maudlin proposes, it cannot imply anything about possible violations in Shimony's and Jarrett's independence assumptions. The 'which assumption fails'–discussion with regards to Shimony's assumptions is completely impervious to any conclusions about an alternative set of assumptions, and so Maudlin's argument can not possibly go through. All that Maudlin can do is argue that his alternative set of assumptions is better or more useful in the search for a causal link than Shimony's or Jarrett's sets. But I will now argue that this is certainly not the case, and I will return to the point of the logically independent sets of assumptions in the concluding remarks of this section.

So, consider the alternative set of assumptions proposed by Maudlin [86, p. 87], which I call **Maudlin Parameter Independence**:

$$\text{MPI:} \qquad \begin{aligned} P(A|a,b,B,\lambda) &= P(A|a,B,\lambda), \\ P(B|a,b,A,\lambda) &= P(B|b,A,\lambda), \end{aligned} \tag{3.15}$$

and **Maudlin Outcome Independence**:

$$\text{MOI:} \qquad \begin{aligned} P(A|a,B,\lambda) &= P(A|a,\lambda), \\ P(B|b,A,\lambda) &= P(B|b,\lambda). \end{aligned} \tag{3.16}$$

And this is what Maudlin has to say about these [86, p. 87]:

> [MOI] says that the probability assigned to a given outcome on one side is not changed if one knows the result (for example, passed or absorbed) but not the setting on the other side. [MPI] says that the probability of a given result on one side, assuming one already knows the result on the other, is not changed if one knows also the distant setting. One might very well call [MOI] "outcome independence" and [MPI] "parameter independence", since [MOI] concerns conditionalizing on the distant outcome and [MPI] on the distant setting. But now the verdict for a theory can be changed: orthodox QM violates [MPI] but not [MOI]. Factorization is no more correctly seen as the sum of outcome independence and parameter independence than as the sum of [MOI] and [MPI]. And, as Bell saw, it is the failure of factorization alone which gives us the most startling result: no account can be given according to which all of the causes of every event lie in its past light cone.

Maudlin states that the conjunction of MPI and MOI also leads to Factorizability. This is true, and it is intuitively very obvious which variables can be declared redundant on the basis of the assumptions accompanying local causality needed to derive Factorizability, but mathematically it is not entirely straightforward. Seevinck has demonstrated that we need some extra steps in the accurate mathematical derivation of Factorizability from the conjunction of MPI and MOI, I have included his proof in Appendix B.[23] But Maudlin's main argument for the arbitrariness of the 'which assumption fails'–discussion is that the standard predictions of quantum mechanics violate Outcome Independence when one assumes PI and OI, but they violate Maudlin's Parameter Independence when one assumes

[23] Ironically, we must prove that Parameter Independence holds.

MPI and MOI. (And then Maudlin still has to argue that his set of assumptions is better.) However, this argument doesn't go through because there is a crucial difference between the way we can interpret (most charitably) Shimony's independence assumptions PI and OI and Maudlin's independence assumptions MPI and MOI.

The difference is this. At least since Shimony, we use these probability functions such as Factorizability and, specifically, OI and PI because they give us an idea that we may be putting our finger on a causal link. With the distinction between OI and PI we can now say, for example, that Bohmian mechanics violates PI while collapse theories violate OI. We are able to do so because we feel confident in interpreting the probability relations of OI and PI—even after all the assumptions, simplifications and mathematical sloppiness discussed above—as representing something about the world and not merely as representing ignorance: the probabilities can be reasonably interpreted as ontic probabilities, not merely as epistemic probabilities. Indeed, we do the same with Factorizability: precisely because we take Factorizability seriously as representing something about the world do we consider that a violation of Factorizability tells us something about an occurrence in the outside world that needs explanation. But, as I mentioned in the previous section, we can only justifiably interpret the probabilities as ontic probabilities if we have complete knowledge of all the values of the variables that we use to calculate the probabilities with. This is always possible for both OI and PI, but it is never possible for MOI and MPI.

To see that this is the case, consider, for example, the first equations from OI and PI: $P(A|B,a,b,\lambda) = P(A|a,b,\lambda)$ and $P(A|a,b,\lambda) = P(A|a,\lambda)$, respectively. We can readily evaluate these equations for the EPR-type experiment in orthodox quantum mechanics by simply using arbitrary measurement settings and outcomes for the variables upon which we conditionalize, and choosing λ to be Ψ for the singlet state (see [114, §3.7.2] for a full evaluation). This then provides us with well-defined probabilities for the measurement result A for each possible combination of settings a and b and distant results B. But now take the first equations from MOI and MPI: $P(A|B,a,b,\lambda) = P(A|B,a,\lambda)$ and $P(A|B,a,\lambda) = P(A|a,\lambda)$, respectively. Intuitively, there is already something suspicious about the probability $P(A|B,a,\lambda)$: what distant measurement result does B refer to without reference to any distant setting b? Quantum mechanics only provides values for probabilities for measurement outcomes given a certain choice of measurement settings, and it turns out that MOI and MPI cannot yet be evaluated in QM similarly to OI and PI because we need to make an extra assumption about the distant measurement setting b that is not needed for evaluating OI and PI. This is demonstrated by Seevinck in [114, §3.7.2], who shows that in the evaluation of MOI and MPI we are stuck with an unknown distribution $\rho(a,b,\lambda)$.[24] Given that we did already assume that the distribution $\rho(\lambda)$ is independent of the other boundary conditions with SI (3.1), the extra assumption that we need to make is that the distribution of the settings a and b are independent of each other: $\rho(a,b) = \rho(a)\rho(b)$.

Informally, for the case of a single measurement procedure, we can interpret this as an averaging procedure over all possible measurement settings b. We specify the possible values for b and state that all values are equiprobable with the above assumption so that we can evaluate the probability $P(A|B,a,\lambda)$ in quantum mechanics. But this means that we can no longer reasonably interpret the resulting probability as an ontic probability. Because we have manually inserted a distribution of possible distant settings b, we have

[24] We must evaluate MOI and MPI with $P(A|B,a,\lambda) = \frac{\int_{\Lambda_b} P(A,B|a,b,\lambda)\rho(a,b,\lambda)db}{\int_{\Lambda_b}\int_{\Lambda_A} P(A,B|a,b,\lambda)\rho(a,b,\lambda)dbdA}$.

limited ourselves to a discussion of epistemic probabilities. So, if we desire to use violations of independence relations such as PI–OI or MPI–MOI to identify a causal link, we can only reasonably do so with OI and PI and we can definitely not do so with MOI and MPI because only OI and PI can be interpreted as ontic probabilities.[25]

Now, it must be said that Maudlin agrees with this critique.[26] In reply, firstly, Maudlin stated that the aim of the relevant section of his book was to argue against the common sentiment that violations of PI are really indications of a serious conflict with fundamental Lorentz invariance whilst violations of OI would allow for 'peaceful coexistence'. There is much to be said about the potentially peaceful coexistence between spontaneous collapse theories (which violate OI) and special relativity, but this is beyond the scope of this thesis.[27] But this is rather irrelevant, because, as I already argued, Maudlin's independence assumptions do not imply anything about PI–OI. They certainly do not argue against the common sentiment that violations of PI are really indications of a serious conflict with fundamental Lorentz invariance whilst violations of OI allow for peaceful coexistence. Especially since the probabilities in Maudlin's assumptions can only be interpreted epistemically whilst those in Shimony's can be interpreted ontically. Secondly, Maudlin understandably pointed out that what we really need are concrete, ontologically clear theories to analyze their non-locality, instead of having to use these independence relations as proxies. This is hard to disagree with, although it must be noted that the scope of analysis by using OI and PI is obviously broader than that of a single theory—which is sometimes an advantage—and that I believe that the case for using OI and PI has been made stronger by my argument against their counterparts MOI and MPI.

3.3 Concluding remarks

In the previous sections, I discussed three different sets of assumptions: Jarret's OL and OF, Shimony's OI and PI, and Maudlin's MOI and MPI. All three sets are independently derivable from Bell's definition of local causality, and from all three sets can Bell-type inequalities independently be derived.

Bell originally derived Factorizability directly from his definition of local causality, and with Factorizability he derived his inequalities. He then demonstrated that his inequalities are violated by quantum mechanics, and, consequently, concluded that local causality must therefore be violated. Although this is apparently enough to demonstrate that local causality is violated (barring the usual loopholes), it does not inform us much about what individual aspect of the model is responsible for this violation.

Jarrett's independence assumptions OL and OF were great improvements in this regard, as they explicitly distinguished between the contribution of two independence assumptions to the violation of Total Factorizability (3.8). I argued that we are searching for probabilities that we can interpret ontically, and since Jarrett's independence assumptions are most complete they appear to be the best candidate for this task. However,

[25]In a sense, the only reasonable conclusion we may draw from a violation of MOI and MPI is whether the distributions of a and b are independent of each other. But we have already assumed that is this the case with SI (Eq. (3.1) states that $\rho(a, b, \lambda)$ factorizes), so we have learned nothing new.

[26]Private communication.

[27]Although I must admit that the case for fundamentally relativistic collapse theories appears more promising than the case for, say, fundamentally relativistic Bohmian mechanics. See, for example, Myrvold's direct reply to Maudlin's arguments in [96] (but compare with his own challenge to such endeavours in [97]). See also the more recent [8, 15, 125]. Notably, Butterfield [26, p. 118] once explicitly wrote "I believe in peaceful co-existence." Upon recent inquiry in private communication, he confirmed to me that he remains convinced of peaceful coexistence in the sense that we should keep looking for a Lorentz invariant-ish interpretation of quantum mechanics.

Total Factorizability and Jarrett's independence assumptions include variables for the complete states of measurement apparatuses μ_a and μ_b, which we can never specify completely. It is impossible to concretely evaluate these assumptions individually, and most of the contemporary literature on Bell-type inequalities is therefore solely concerned with the μ-averaged probabilities.

These probability functions are known as Shimony's OI and PI. Given the μ-averaging procedure, there is an obvious but inevitable epistemic aspect to these independence relations. But they are the best we can do in our search for obvious culprits when Factorizability is violated, and especially for Shimony's original purpose of using his independence assumptions to investigate controllable non-locality. Shimony's independence relations are certainly easier to interpret than Jarrett's, and, as I argued, they can even reasonably be interpreted ontically. I argued that this is not the case for Maudlin's other alternative set of independence assumptions, MOI and MPI, which can only be interpreted epistemically (on top of the inevitable epistemic aspect of the μ-averaging). In Figure 4, we see the three different ways of deriving Bell-type inequalities from the assumption of local causality. The violation of Bell-type inequalities implies the violation of local causality, but we see that one may reconstruct the violation in three different ways. I argued that Shimony's independence assumptions are most suited for our purpose of finding accessible independence relations which provide probabilities can be interpreted ontically. That is to say, from which we might conclude (in at least some cases) spacelike non-local causal influence. The time has now come to investigate timelike non-locality.

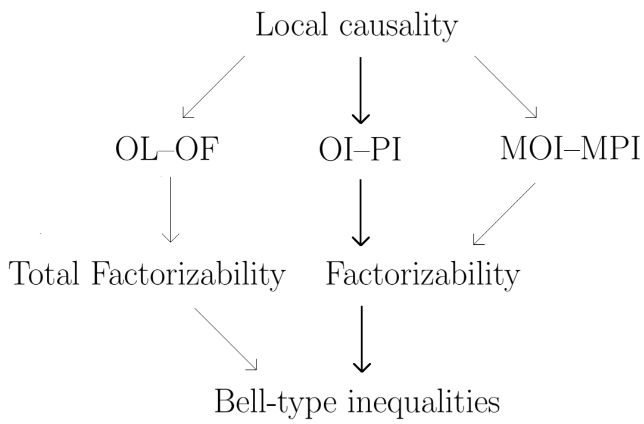


Figure 4: Three different ways of deriving Bell-type inequalities.

4 Timelike non-locality

4.1 Timelike factorizability

The first clear exposition of the case for timelike non-locality—or, rather, the case against timelike locality—is by Adlam in a paper [2] that is an elaboration of one half of a talk (eg. [1]) wherein she argues that both 'temporal locality'[28] and 'objective chance' are (intertwined) concepts that have crept in the practise of physics without any intrinsic reason for this. The succint and clear introduction to her paper deserves to be quoted in full [2, p. 1]:

> Since the discovery of Bell's theorem, the physics community has broadly come to take seriously the possibility that the universe might contain physical processes which are spatially nonlocal. However, there has been no such revolution with regard to "temporal locality", i.e., the assumption that the probabilities attached to the outcomes of a measurement performed at a given time depend only on the state of the world at that time. Indeed, temporal locality remains almost ubiquitous in the way that scientists think about science and about what constitutes a reasonable scientific hypothesis. An assumption so widespread and yet so infrequently justified is in serious danger of becoming a dogma. While it is true that temporal locality has previously been recognised as problematic by parts of the physics community, we argue that this recognition is not sufficiently widespread and that the assumption is actively limiting progress in the field of quantum foundations. In this article, we investigate the origins of this way of thinking about physics, arguing that it has become dominant for historical and pragmatic reasons rather than good scientific ones. We then explain why temporal locality is in tension with relativity, and review some recent results which cast doubt on the status of temporal locality in modern physics.

This is quite a lot to unpack. In this section, I first discuss Adlam's definition of timelike non-locality and elaborate upon it. I then briefly consider her examples of and qualitative arguments for timelike non-locality. Finally, I show why this result is manifestly different from the familiar temporal Bell-type inequalities such as the Legget-Garg inequalities [79] or the results of Brukner et al. [25]. I end this section with some general remarks and conclusions, including a discussion of a recent concrete result by Brierly et al. [24] that might be considered as a serious technical result in favor of timelike non-locality.

Adlam starts by using the Factorization equation (3.12) for Bell's theorem as the definition of spacelike locality and then defines timelike locality analogously [2, p.2]:

> Suppose that two observers, Alice and Bob, perform measurements on a shared physical system. At some time t_a, Alice performs a measurement with measurement setting a and at some time $t_a + \delta$ she obtains a measurement outcome A; likewise, at some time t_b, Bob performs a measurement with measurement setting b and at some time $t_b + \delta$ he obtains a measurement outcome B. Let λ_{t_a} be the state of the world at time t_a and let λ_{t_b} be the state of the world at time t_b.

[28] Again, I prefer 'timelike', to distinguish it from the more familiar temporal non-locality of Leggett–Garg etc., see Section 4.3

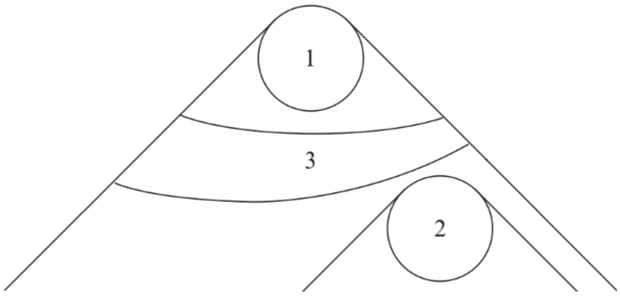


Figure 5: Full specification of what happens in 3 makes events in 2 irrelevant for prediction about 1 in a timelike local theory.

The definition of timelike locality is then the Timelike Factorized Equation (TLF):

$$\text{TLF:} \qquad P(A, B|a, b, \lambda_{t_a}, \lambda_{t_b}) = P(A|a, \lambda_{t_a}) \cdot P(B|b, \lambda_{t_b}). \tag{4.1}$$

Adlam stresses that the central idea of this definition is that “all influences on a measurement outcome would be mediated by the state of the world immediately prior to the measurement” [2, p. 2]. This definition thus implies that there is ‘timelike action at a distance’ if TLF (4.1) does not hold: there is timelike non-locality if the state of the world immediately prior to a measurement cannot fully mediate all influences on that measurement. Since ‘the state of the world’ is a rather vague expression, it might be helpful to define TLF with the same imagery and terminology as Bell originally did for his local causality. Recall Bell’s definition, which, as we may recognize now, may apply to both the spacelike and timelike case:

> A theory will be said to be locally causal if the probabilities attached to values of local beables in a space-time region 1 are unaltered by specification of values of local beables in a space-like separated region 2, when what happens in the backward light cone of 1 is already sufficiently specified, for example by a full specification of all local beables in a space-time region 3.

Bell emphasizes that region 2 is space-like separated from region 1, but there is thusfar little that prohibits us from adjusting his original Figure 1b into a timelike analogue, as presented in Figure 5. Here too, the full specification of values of local beables in region 3 should completely screen off any influence from region 2 to region 3.[29]

Some remarks need to made before we can continue. First, Adlam’s definition of timelike locality cited above explicitly mentions the temporal difference between the time where the measurement setting is introduced and the time where the measurement result is obtained, just like I emphasized in Section 3.1. This is obviously important in the case of timelike non-locality because our definition of λ has now been specified from Bell’s ‘complete specification of hidden variables at any arbitrary time after the source preparation and before the measurements’ to Adlam’s ‘state of the world at time t_i,’ where t_i is the time of a measurement with setting i. This difference raises the question whether the λ’s in Factorization (3.12) and TLF (4.1) are even comparable. I argue that they are: although Adlam’s ‘state of the world’ λ_{t_i} appears to include the description of the measurement apparatus—and thus TLF appears to be some hybrid between Jarrett’s

[29] Indeed, as Adlam also mentions, we can see with this definition that both spacelike and timelike locality are essentially included in Bell’s conceptual framework.

and Shimony's independence assumptions—but, as the examples in Sections 5 will show, we are only concerned with the quantum state immediately prior to the measurement and (again) not so much with the measurement apparatuses.

Now, we can straightforwardly apply Shimony's analysis of independence assumptions for Factorization (3.12) to the definition of timelike locality TLF (4.1), which provides us with the timelike analogue of Outcome Independence (TOI):

$$\begin{aligned} \text{TOI:} \qquad & P(A|B,a,b,\lambda_{t_a},\lambda_{t_b}) = P(A|a,b,\lambda_{t_a},\lambda_{t_b}), \\ & P(B|A,a,b,\lambda_{t_a},\lambda_{t_b}) = P(B|a,b,\lambda_{t_a},\lambda_{t_b}), \end{aligned} \tag{4.2}$$

and we are tempted to immediately formulate the timelike analogue of Parameter Independence. However, since we have an extra conditional variable for our probability, this condition would be doing too much work on its own, and so we must dissect it further. A natural suggestion would be first an assumption of Timelike State Independence (TSI):

$$\begin{aligned} \text{TSI:} \qquad & P(A|a,b,\lambda_{t_a},\lambda_{t_b}) = P(A|a,b,\lambda_{t_a}), \\ & P(B|a,b,\lambda_{t_a},\lambda_{t_b}) = P(B|a,b,\lambda_{t_b}), \end{aligned} \tag{4.3}$$

and only then the adequate form of Timelike Parameter Independence (TPI):

$$\begin{aligned} \text{TPI:} \qquad & P(A|a,b,\lambda_{t_a}) = P(A|a,\lambda_{t_a}), \\ & P(B|a,b,\lambda_{t_b}) = P(B|b,\lambda_{t_b}). \end{aligned} \tag{4.4}$$

Starting from TLF (4.1), therefore, we declare redundant first the distant measurement outcome, then the distant state, and finally the distant parameter. Note that we run into the same criticism as I provided against Maudlin's independence assumptions if we declare redundant first the distant parameter and then the distant state: if we input no specific value of the distant measurement setting, we must average over all possible distant measurement settings to specify the distribution of the distant state, whereby we limit ourselves to mere epistemic probabilities.

But there are a few problems with the timelike factorized equation (4.1) as a definition for timelike locality. The first problem is that, as we will see throughout the rest of this thesis, the above Shimonian subdivision of TLF into three different independence relations is not as useful as it is in the spacelike case. For whereas in the spacelike case we can demonstrate straightforwardly that, for example, Bohmian mechanics violates Parameter Independence while collapse theories violate Outcome Indepenence, this is not quite the case for the timelike independence relations. Here we seem to run into a triviality problem: by construction, none of the three independence relations are violated by retrocausal theories whilst all three independence relations are violated by an all-at-once theory. This triviality problem forces us to reconsider the usefulness of a Shimonian analysis for timelike non-locality, and it can then be interpreted as weakening the analogy between spacelike and timelike Bell-type inequalities even more. Of course, we need show that this apparent triviality problem is indeed a problem by quantitatively analyzing concrete retrocausal and all-at-once theories (the latter which are quite sparse), which I have unfortunately not done (yet), but I hope it will become clear in Section 5 that this triviality problem is quite self-evident from purely conceptual analysis of these theories.

The second problem is perhaps even more pressing. It appears that TLF is not particularly suited to represent the violation of Source Independence (3.1), $\rho(\lambda|a,b) = \rho(\lambda)$, the violation of which is the loophole to the conclusion of spacelike non-locality in the regular

EPR-type experiment. And it was this possible violation of Source Independence in the spacelike EPR-type experiment that was the original motivation for considering retrocausal theories, and, subsequently, all-at-once. Indeed, the formulation of a factorizable joint probability is not enough to derive a Bell-type inequality, whether spacelike or timelike. We need an assumption about the independence between the boundary conditions as well, along the lines of:

$$\text{SIT:} \qquad \rho(\lambda_{t_a}, \lambda_{t_b}|a, b) = \rho(\lambda_{t_a}|a)\rho(\lambda_{t_b}|b). \tag{4.5}$$

But emphasis on the violation of such a a timelike analogue of Source Independence then takes away the significance of Adlam's definition of timelike non-locality. It can perhaps be argued that a violation of timelike factorizability (4.1) also includes the possible violation of Source Independence because the boundary conditions described by Source Independence now lie on on temporally distant 'states of the world' and are thus included in the λ_{t_i}'s. But then it must be admitted that violation of timelike factorizability (4.1) hardly puts the finger on the sore spot of the violation that is the violation of Source Independence, and we need to do more work in order to understand the reason why timelike factorizability is violated.

Moreover, the possibly interesting violation of Source Independence as a timelike non-local influence in the regular spacelike EPR-type experiment is not adequately captured at all by defining timelike locality analogously to spacelike locality. Whether or not TLF by itself can satisfactorily represent timelike non-locality for a *timelike* EPR-type experiment (we will come across several of these in Section 5), it cannot account for the timelike non-locality in a *spacelike* EPR-type experiment that is the violation of Source Independence any better than Factorizability (3.12) can. And because it is Source Independence that is violated by the 'loopholes' to Bell's theorem discussed in Section 5, it is again the question of how relevant this definition and subdivision of timelike non-locality can actually be for evaluating theories that are candidates for timelike non-locality. But before these loopholes can be considered, we must understand in more detail Adlam's motivation for taking timelike non-locality very seriously.

4.2 Non-Markovian theories

Adlam discusses three different ways in which physical theories might be timelike non-local: theories with non-Markovian laws, retrocausal theories, and all-at-once theories. In this section, I shall elaborate on Adlam's presentation of theories with non-Markovian laws, and I briefly introduce her views on retrocausal and all-at-once theories, which I discuss thoroughly in Sections 5.3 and 5.4, respectively.

For a theory with dynamical *Markovian* laws, the future evolution of a physical state can be fully determined solely from the present state, without needing to add any past states or other information about the present state's history. In the context of Bell's theorem, it is said that the *Markov property* can be informally defined as demanding that the Timelike Factorized Equation holds, with the added requirement that the region in between relevant space-time regions used to 'shield off' causal influence is "sufficiently thick" [62, §6]. Formally, the Markov property for probability calculus states that 'only the most recent conditioning matters' [58, p. 60]:

$$P(\lambda_{t_i}, t_i|\lambda_{t_0}, t_0; \lambda_{t_1}, t_1; ...; \lambda_{t_{i-1}}, t_{i-1}) = P(\lambda_{t_i}, t_i|\lambda_{t_{i-1}}, t_{i-1}), \tag{4.6}$$

where λ_{t_i} is the state of the world at time t_i.[30] Gillespie [58] calls this the 'past-forgetting property' of a Markovian process. For our model, the Markov property for a certain result B that depends only on the state of the world immediately prior to the measurement and adheres to parameter independence and outcome independence would be, for example:

$$P(B; \lambda_{t_b}, t_b | A; a; b; \lambda_{t_0}, t_0; \lambda_{t_1}, t_1; ...; \lambda_{t_{b-1}}, t_{b-1}) = P(B; \lambda_{t_b}, t_b | b; \lambda_{t_{b-1}}, t_{b-1}). \quad (4.7)$$

Adlam states that we have good reason to be cautious about the Markov property in the context of quantum theories for two different reasons. The first reason is with regard to recent results about the exponential growth of the number of variables with respect to physical size in ontological Markovian theories of quantum mechanics (especially [66, 68, 91]). Ontological models are all those quantum mechanical theories that aim to add an underlying reality to operational quantum mechanics by means of the parameters λ (cf. [67, §II]), and thus includes Bohmian mechanics. The result is stated most clearly by Montina in [91, p. 10]:

> [T]he main criticism against the known ontological models, such as Bohmian mechanics, is indeed the exponential growth of the ontic-space dimension with physical size. We have shown rigorously that this feature is unavoidable in the framework of causal Markovian theories. Thus our result seems to put another nail in the hidden-variable coffin and to imply that realistic interpretations do not provide a practical advantage for the study of quantum systems. ... In order to avoid the exponential growth of the number of ontic variables, we have one possibility: to discard some hypotheses of the theorem. In our opinion, the Markovian property is the only one that may be sacrificed. More drastically, we could discard the causality hypothesis. It is interesting to observe that the Bell theorem and the Lorentz invariance seem to suggest the same conclusion. The Bell theorem establishes that an ontological theory of quantum mechanics cannot be local, and relativity implies that a nonlocal theory is also noncausal. We suspect that realism in quantum mechanics with nonpathological consequences for the ontic-space dimension could be possible with noncausal rules for evaluating correlations among events.

It can be argued that the aim of a realistic interpretation is not to provide a practical advantage over the bare-bones quantum mechanical formalism but to add a realistic interpretation at the cost of some practical advantage, but a theory should of course be workable. In any case, if we aim to avoid exponential growth of the ontic-space dimension with respect to physical size, Montina states that we are left with a dilemma: either sacrifice the Markovian property of the theory or sacrifice its traditional causal structure. See [127] for a recent proposal of a Bohmian approach to non-Markovian Schrödinger-type equations, but compare with [92, p. 1], which notes that, in certain contexts, "[n]on-Markovian equations lead to a loss of analytical expressions as well as larger computational times." We may thus consider whether obtaining smaller ontic-space dimension at the cost of loss of analytical expressions as well as larger computational times is actually retaining a practical advantage.[31] Montina's other horn of the dilemma, to sacrifice causal structure,

[30] This definition is only apparently time-asymmetric. See [9, p. 44] for a general definition of the Markov property and a demonstration of its time symmetry.

[31] Perhaps the case for dropping the Markovian property can be made stronger if there is some physical motivation for doing so. This might be the case in Derakhshani's [40, p. 136], where, in his Nelsonian

is interesting because it appears to be precisely what all-at-once theories are suggesting. This will be discussed in Section 5.4.

There is a second reason for seriously considering non-Markovian theories. After questioning the status of the Markov property with regard to the arguments presented above, Adlam continues with an interesting observation [2, p. 9]: according to the reversible dynamics of unitary quantum mechanics, information about the past—stored in the wavefunction—never gets lost but just gets spread out into the environment due to decoherence. Much of this information will be stored in wavefunctions of entangled systems which are irreducible global properties of non-individual systems, and the information thus ends up stored in a state that is not the state of a particular individual system. So, Adlam continues [2, p. 9]:

> [If] the formalism tells us that no information about the past is ever lost (except possibly in a measurement process) and also that most of this information usually cannot be attached to any single system or any particular physical location, then are we really saying anything particularly meaningful when we assert that the information is nonetheless all stored in the present state of the world? Under these circumstances, it is certainly more ontologically economically and arguably also more natural to say simply that measurement results at the present time depend *directly* on the history of the system, without any need for mediation via a nebulous state-like entity.

Although I sympathize with such considerations to a certain extent, I must contest that, at least intuitively, it does feel like we are saying something meaningful when we assert that the information determining the next state of the world is all stored in the present state of the world: we can still claim there is a traditional (dynamical and asymmetric in time) causal connection between consecutive states of the world, albeit epistemically shrouded. It does not yet appear that the bargain is good enough here to trade off our intuitive causal picture for an all-at-once acausal—thus apparently timelike non-local—picture of the world based on consideration about the Markov property alone. We need more. But we must also note at this point that the discussion about the Markov property in the context of quantum theories has now become rather disconnected from the original context of EPR-type experiments from which we first encountered it, and which is the central topic of this thesis. We have seen that violation of the Markov property is a serious possibility for timelike non-local theories, but we must draw this section to a close so as not to lose sight of the central aims of this thesis.[32]

Adlam goes on to discuss the recent developments regarding the retrocausal and all-at-once type approaches, for they might add sufficient reason to seriously consider timelike non-locality. She begins by noting the interesting paper by Leifer and Pusey [81], which builds on the Price Argument discussed in Section 5.3 that argues retrocausality must be present in quantum mechanics under the assumption of a certain type of time symmetry. If the concept of retrocausality is invoked as a way to save either or both timelike and spacelike locality, however, Adlam sees this as a retrogade step. Indeed, the aforementioned arguments for retrocausality in quantum mechanics obey the definition of timelike

(stochastic) mechanics, the ether-like background interacts with the ontology to generate Markovian laws, but such Markovian interaction appears to be coarse-graining because interaction with an ether would more naturally be non-Markovian.

[32] See further [82] for a thorough review of the Markov property in quantum theories.

locality (4.1).[33] Moreover, they strongly appear to be arguing for timelike locality because the few explicitly retrocausal quantum mechanics theories (Cramer's transactional interpretation [34, 35] and Aharonov et al.'s two-time interpretation [4]) depend fundamentally on mediating states that carry information through time—in both directions but still contiguously.

Adlam ends her discussion here with the more intuitive point that such contiguous retrocausal theories appear to be involved in a very 'finely tuned balancing act', because, in such theories, the backwards evolving states must be compatible with the forwards evolving states and vice versa [2, p. 10]. Such balancing quite naturally results in an image of the world where the past solves the future and the future solves the past, leading towards some global interdependence of all states. Adlam concludes [2, p. 10]: "In this picture, the assumption of temporal locality begins to seem highly artificial, and talk of a 'backwards-evolving state' or 'influences that travel back in time' look like rhetorical devices designed to preserve the appearance of temporal locality in a theory whose underlying structure is really temporally nonlocal." Thus we arrive at all-at-once, Adlam concludes.

4.3 Temporal Bell-type inequalities and timelike non-locality

In order to understand genuine timelike non-locality, we need to distinguish this concept (leading to timelike Bell-type inequalities) from the family of *temporal* Bell-type inequalities on which there is a rich body of literature, and of which the Leggett-Garg inequalities [79] are the main result. As may be evident, herein lies the main reason for me to use the terms *timelike* and *spacelike* for discussions involving genuine non-locality, as opposed to *temporal* (and *spatial)* for discussions disconnected from the locality debate. Adlam [2, p. 15–16] dismisses these types of inequalities because they all appear to implicitly assume some form of locality in time and therefore exclude the possibility of timelike non-locality by construction, and she argues that we must look for a different type of result that could provide a strong argument in favor of timelike non-locality.

One might contest to this dismissal that we do the same for spacelike Bell-type inequalities: we construct spacelike Bell-type inequalities by assuming spacelike locality, and from a violation of the inequalities can we then conclude a violation of spacelike locality. But there is a difference between these two cases that will be relevant throughout in this section: the assumption of local causality completely rules out local influence for spacelike Bell-type inequalities, but local causality can not rule out causal influence in the temporal case. Violations of temporal Bell-type inequalities appear to simply allow for an explanation in terms of local causality, referred to as the *communication loophole*, and so we need a stronger result over and above the mere violation of temporal Bell-type inequalities. In this section, I will point towards possible results of this kind for several types of temporal Bell-type inequalities.

To begin our discussion of temporal Bell-type inequalities, we must realize that there are two similar types of temporal Bell-type inequalities that nonetheless describe two different experimental situations: the original Leggett-Garg-type inequalities [79] that

[33] The Price Argument is not really formal, but we will see in Section 5.3 that the timelike locality is evident. The Leifer-Pusey result is formal, however, and it includes an explicit assumption of timelike locality in the form of λ-*mediation* [81, p. 10]: "The ontic state λ mediates any remaining correlation between the preparation and the measurement." And they go on to elaborate on why timelike locality is essentially already included in the definition of an 'ontic state' [81, p. 10]: "The λ-mediation assumption is really just part of the definition of what we mean by an ontic state. The properties of the system [immediately prior to a measurement, I may add] are supposed to be the cause of the correlation between preparation and measurement."

are concerned with two-time correlations of a *single* observable measured with respect to three or more different times, and newer inequalities (eg. [25, 53]) that do consider two different observables measured at two different times, which is more in line with the spacelike Bell-type inequalities. Both of these types of inequalities need to be considered separately.

Let's start with the Leggett-Garg inequalities. These inequalities were originally proposed to test whether the macroscopic world (when extrapolated up from quantum mechanics) adheres to some of our classical intuitions about the world, which are represented with the following assumptions. (1) *Macroscopic Realism*: A macroscopic system that has two or more distinct possible states is at any time in one of those well-defined, pre-measurement-existing states. (2) *Non-invasive Measurability*: We can measure this value without disturbing the system's state or subsequent dynamics. The Leggett-Garg scenario then describes successive measurements on a single system with a single, dichotomous observable $Q = \pm 1$ at times $t_1 < t_2 < t_3 < t_4$, which enables us to derive Leggett-Garg type inequalities using four different two-time correlation functions of this observable $\langle Q(t_i)Q(t_j)\rangle$:[34]

$$\text{LG:} \qquad |\langle Q(t_1)Q(t_2)\rangle + \langle Q(t_2)Q(t_3)\rangle + \langle Q(t_1)Q(t_4)\rangle - \langle Q(t_2)Q(t_4)\rangle| \leq 2, \tag{4.8}$$

which is violated by quantum mechanics. This inequality is obviously structurally similar to the Bell-type (CHSH) inequality (2.4), especially considering the fact that the assumption of Non-invasive Measurability plays essentially the same role for the temporal scenario as the assumption of Locality does for the spacelike scenario. But, whereas it is often concluded from violation of spacelike Bell-type inequalities that the phenomenon of entanglement leads to a failure of Locality, for these Leggett-Garg-type inequalities it is most often concluded that the non-classical invasiveness of a measurement on a quantum state causes a violation of Non-invasive Measurability.[35] There is thus no entanglement at play in the Leggett-Garg inequalities, which is usually a necessary requirement for spacelike non-locality, and we are already less enticed to consider the temporal Leggett-Garg-type inequalities being perfectly analogous to spacelike Bell-type inequalities.[36]

Adlam dismisses the relevance of the Leggett-Garg-type inequalities on different grounds, however. She points to the fact that the assumptions (1) and (2) already imply that "by definition the measurements references in the Leggett-Garg inequalities are only allowed to depend on the state of the world at the time of the measurement, which makes temporal locality irrelevant: whether or not the world is [timelike] non-local in general, for this specific type of measurement there is no freedom for the measurement result to depend on anything other than the present state of the world" [1, p. 15]. These inequalities therefore appear to exclude the type of timelike non-locality that we are interested in. Remarkably, this point is clarified somewhat with an assumption that has recently been made explicit (by Leggett himself, amongst others) in addition to the two previous assumptions: (3) *Induction*: the outcome of a measurement on the system cannot be affected by what will

[34] See, for example, [141, §II] for a derivation that uses the same notation as I do in Appendix A and which indicates clearly where the assumptions are invoked in the derivation. I use four different times here to strengthen the analogy with the spacelike CHSH-inequality.

[35] Indeed, Bohmian mechanics is a theory that demonstrates this invasiveness explicitly for the temporal Leggett-Garg-type inequalities (in the form of contextuality) just as it demonstrates the non-locality explicitly for the spacelike Bell-type inequalities, see [10].

[36] See [83] for further arguments along these lines about the methodological differences between the two types of inequalities.

or will not be measured on it later.[37] 'Induction' might be somewhat of a misnomer, because we are here assuming that there are no retrocausal influences: this is exactly the same assumption as is often referred to as 'no retrocausality' in the spacelike Bell-type inequalities literature and which I excluded with one part of 'Independent boundary conditions'. Consider the words of Emary et al. on the assumption of Induction [49, p. 3]:

> Whilst the derivation of the [Leggett-Garg inequalities] certainly relies on assumption [3], so does much of our understanding of the natural world. As this assumption reflects such basic notions about causality and the arrow of time, it has remained unchallenged in discussions of the source of [Legget-Garg inequality] violation (but see [78] for a word of caution on this point).

And Leggett, the reference cited in the final sentence, states [78, p. 4]:

> [T]he most impervious to challenge would seem at first sight to be [Induction], in the sense that once we give up our 'common sense' notions concerning the 'arrow of time' it seems very difficult to continue to do physics at all in the mold to which we have been accustomed. Actually, in this writer's opinion it is entirely possible, indeed probable, that the next major revolution in physics will force us to do just that, but unfortunately it is in the nature of scientific revolutions that their content is difficult or impossible to forecast, so speculation along these lines seems pointless at this time.

Although Leggett is apprehensive about exploring this retrocausal loophole himself, we can recognize here a similar trend of increasing interest in retrocausality as has been the case for spacelike Bell-type inequalities in the past couple of decades. But, once again, retrocausal-yet-contiguous causal influence is not the type of timelike non-locality that we are looking for. We want something stronger, more non-local, preferably involving entanglement.

Interestingly, there appears to be a recent result with regard to these Leggett-Garg inequalities of the form that Adlam is looking for. Bacciagaluppi [10, §4] discusses a recent result by Dzhafarov and Kujala [43, 44] where the signalling (the communication loophole) is taken into account in Leggett-Garg-type inequalities. This apparently allows for distinction between violations of Leggett-Garg-type inequalities as a result of mere signalling (violation of marginal selectivity) or as a result of non-locality *in time* (non-local contextuality between consecutive measurements). I do not fully understand the results of Dzhafarov and Kujala, and I do not feel qualified to comment on their implications, but it does appear that these results could provide an argument for genuine timelike non-locality and in the very least reignite the discussion of the role of contextuality and entanglement in the Legget-Garg-type inequalities. [38]

Besides the original Leggett-Garg inequalities, there exists a second family of temporal Bell-type inequalities commonly referred to as 'entanglement in time'. This appears quite promising for our current discussion, and I take Brukner et al.'s [25] as the focal point of a brief discussion. We start again with the same three assumptions (1)–(3) as we made for the Leggett-Garg inequality (and note that Brukner et al. gave the latter, Induction, the

[37] I borrow use the formulation of [49, p. 3]. See also the references therein, especially Leggett's own [78].

[38] See also Dieks et al.'s [41] for a very early result (it predates the original Leggett-Garg inequality) of contextuality in temporal Bell-type inequalities.

more pertinent title *Locality in time*). In contrast to the Leggett-Garg inequality, however, we consider now two dichotomous observables instead of one. Consider two consecutive measurements on a single system, one dichotomous measurement at time t_1 and one at time $t_2 > t_1$. At both measurements, the experimenter has a choice of two measurement settings, a or a' and b or b', leading to observables $A = \pm 1$ and $B = \pm 1$, respectively. We can now derive a temporal Bell-type inequality [25, p. 2] for 'entanglement-in-time':

$$\text{EIT:} \qquad |E(a,b) + E(a,b') + E(a',b) - E(a',b')| \leq 2, \tag{4.9}$$

which is again violated by quantum mechanics. The equivalence between the regular Bell-type inequality and this temporal inequality led Brukner et al. to consider that that the two cases are perfectly analogous. And indeed, it appears that the analogy is rather strong for this bipartite system. However, as Emary et al. have argued [49, p. 6], it is not complete because "in the extension to multi-partite entanglement, spatial entanglement is known to be monogamous [73], but the temporal version was found to be polygamous." It is unclear to me whether this has any implication for our present discussion, however.

Luckily, Adlam dismisses the relevance of these types of inequalities on the same grounds as before:

> Here the derivation depends on temporal locality and also "realism", defined as the assumption that measurement results are determined by hidden properties that the particles carry prior to and independently of observation; one assumes that the state of the world at the time of measurement would be expected to include these "hidden properties" and, thus under this assumption measurement results can depend only on the state of the world at the time of the measurement, so the auxhiliary assumption already implies temporal locality, in the sense in which we have used the term.

Therefore, as Adlam makes clear, we must look for a plausible argument in favor of genuine timelike non-locality in the sense in which we have used the term that does not assume timelike locality by construction. She finds this in Brierly et al's [24], who also comment on the 'entanglement in time' inequalities discussed above [24, p. 2]:

> [I]dentifying a violation of temporal Bell inequalities with so-called "entanglement in time" [25] is not always accurate. Indeed, forward signaling of classical information in the sequential scenario leaves a kind of *communication loophole*: correlation functions which would be considered nonclassical in the spatial setting can be simulated by classical protocols that use a classical communication channel with capacity given by the Holevo bound of the corresponding quantum particle.

Brierly et al. then go on to demonstrate a violation of the Holevo bound for a timelike GHZ scenario (the upper limit of how much information can be stored in a quantum system, here constructed as $\log_2 m$ for a GHZ state with m subsystems) for a sufficiently large amount of measurement steps and number of settings per observer. This result implies two things for our present discussion. Firstly, it confirms that violation of temporal Bell-type inequalities does not straightforwardly imply timelike non-locality. Secondly, however, it also confirms that, at least for certain measurement set-ups where the communication loophole is taken explicitly into account, it is possible to violate this loophole in the sense that more information is transmitted through time than is allowed classically.[39] And here Adlam finds her concrete result in favor of genuine timelike non-locality.

So what may we conclude from these different types of temporal Bell-type inequalities with regard to timelike non-locality? On the one hand, formal similarities between spacelike Bell-type inequalities and temporal variants thereof certainly do not directly imply that there is similar non-locality in both cases. The biggest and most obvious difference is that we can rule out all communication between spacelike separated measurements but that we must allow communication between temporally consecutive measurements, at least forward in time. We must therefore be very careful not to conclude timelike non-locality from mere formal similarity between different types of inequalities. But, on the other hand, it appears that at least for several cases there are interesting results with regard to timelike non-locality when we take the communication loophole into account. As said, it might be possible to demonstrate that there is genuine non-local contextuality between consecutive measurements in a Leggett-Garg scenario, indicating timelike non-locality. It is also possible to demonstrate a violation of the limit to classical information transmission in a timelike GHZ-scenario for a sufficiently large amount of measurement steps and number of settings per observer. So, although the results that argue in favor of genuine timelike non-locality are sparce and easy to conflate with possibilities for retrocausality or simply situations that allow for locally causal explanations, they do exist and we must take them seriously.

4.4 Concluding remarks

In this section, I discussed timelike locality as defined analogously to Bell's definition of spacelike locality in terms of Factorizability. I reviewed Adlam's motivation for seriously considering the possibilities for timelike non-locality. Philosophically, and largely independently of Bell's conceptual framework, there are at least some good reasons to take timelike non-locality seriously. Indeed, as Adlam emphasized, the assumption of timelike locality is so widespread and yet so infrequently justified that it is in serious danger of becoming a dogma. This is already an intriguing observation worthy of further analysis.

I discussed two types of recent results that possibly argue in favor of timelike non-locality. Firstly, there are good reasons to consider the possibility of non-Markovian theories. The current main ontological interpretations of quantum mechanics such as Bohmian mechanics suffer from exponential growth of their ontic space dimension with physical size, and it appears that this problem may be remedied by giving up on either the Markov property or causal structure. The Markov property is also analyzed reasonably well using Bell's conceptual framework. Bell's intuitive idea of 'screening off' influence on an event by specifying a sufficiently thick slice of its backwards light-cone is a reasonable description of the definition of the Markov property, and the definition of the Markov property is easily combined with Shimony's OI and PI. But more in-depth analysis of the possibilities for non-Markovian theories quickly lead us away from the framework of the EPR-type experiment and Bell's terminology, and are therefore beyond the scope of this thesis. But they are certainly interesting and should be studied further.

The second type of result possibly in favor of timelike non-locality is with regards to the more familiar temporal Bell-type inequalities. Like Adlam, I distinguished two types of such temporal inequalities, both of which describe manifestly different measurement

[39]This relates to (the violation of) the relatively recent concept of 'information causality', which is a generalization of no-signaling conditions. Cf. [102, 103].

scenario's: Leggett-Garg-type inequalities and Bruckner et al.'s 'entanglement-in-time'-type inequalities. Here, however, the analogy between spacelike and timelike non-locality becomes rather superficial and misleading when using Bell's conceptual framework. Although the different types of inequalities (spacelike, timelike, and temporal) share strong resemblance formally, their violations allow for entirely different types of explanations. Contrary to the violation of spacelike Bell-type inequalities, the violation of a temporal Bell-type inequality allows for a perfectly locally causal explanation of its correlations because the two wings of the EPR-type experiment are allowed to communicate classically (but asymmetrically in time). This is referred to as the *communication loophole*. This is the case for the 'entanglement-in-time' inequalities, and so their name is rather misleading as they don't refer to genuine non-locality. We therefore need a stronger result over and above mere violation of these inequalities to argue in favor of genuine timelike non-locality. Adlam found this in the recent result by Brierly et al. [24], which considered a certain GHZ scenario and found that, for sufficiently large amounts of measurement steps and number of settings per observer, more information is transmitted than is allowed classically (by the communication loophole). This does point us in the direction of genuine timelike non-locality. With regard to the other type of temporal inequalities, whereas Adlam dismissed the relevance of Leggett-Garg-type inequalities for timelike non-locality entirely, I briefly discussed a recent result [10] where it is argued that, when taking the communication loophole into account in the Leggett-Garg inequalities, there is the possibility of distinguishing between violations of the inequalities as a result of mere signalling or as a result of genuine non-local contextuality in time. Here, too, I find a result that we might use to argue in favor of genuine timelike non-locality.

What is now clear, however, is that the formal similarity between Bell's spacelike inequalities and the temporal inequalities discussed above is misleading. We cannot straightforwardly conclude genuine timelike non-locality from the violation of a temporal inequality, but we must search for results that take a locally causal explanation into account. It is remarkable that there are indeed such results, but Bell's conceptual framework did not really assist us in finding and understanding them.

Now, turning back to the definition of timelike locality discussed at the beginning of this section. It appears that we should take timelike non-locality seriously, but there are several more problems with the analogy between spacelike and timelike non-locality besides the problems I just discussed. Firstly, our analysis of timelike non-locality does not appear to benefit from a Shimonian subdivision into independence relation as well as spacelike non-locality does. It seems that, when we analyze potentially timelike non-local theories, we will run into a triviality problem: by construction, either a theory will not violate any independence assumption (such as retrocausality) or it will violate all independence assumptions (such as all-at-once). Especially given the arguments in favor of Shimonian analysis that I provided in the previous section, this triviality problem makes the analogy between spacelike and timelike non-locality even more troublesome. Secondly, whether or not timelike factorizability can adequately capture timelike non-locality for a *timelike* EPR-type experiment, it cannot properly account for the timelike non-locality in a *spacelike* EPR-type experiment that is the violation of Source Independence. It is now time, therefore, to discuss the types of theories that violate Source Independence, after which I will return to the question of whether defining timelike locality analogously to spacelike locality (in terms of factorizability) is useful or misleading.

5 Violating 'Source Independence': three proposals

In this section, I discuss three different proposals for theories that violate Source Independence and are thus possible candidates for being timelike non-local. We will see that only the last proposal, all-at-once, may perhaps genuinely be timelike non-local, albeit not in a causal sense. First, however, I critically assess a recent argument in favor of defining timelike locality analogously to spacelike locality using Bell's conceptual framework. Although the argument has three authors, I refer to it as 'the Evans trilemma' because Evans is its main author, and the positions of the other two authors, Price and Wharton, are discussed in Sections 5.3 and 5.4, respectively.

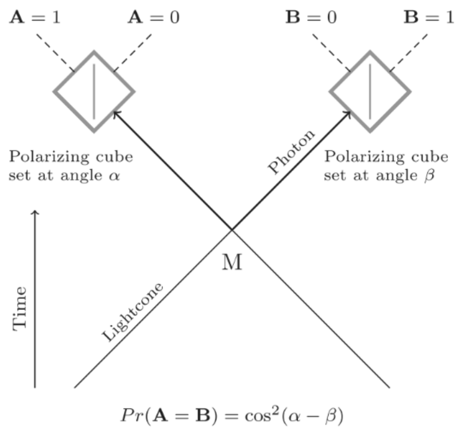
5.1 The Evans trilemma

Consider the abstract of the 'New Slant on the EPR-Bell Experiment' by Evans, Price and Wharton [51, p. 297]:

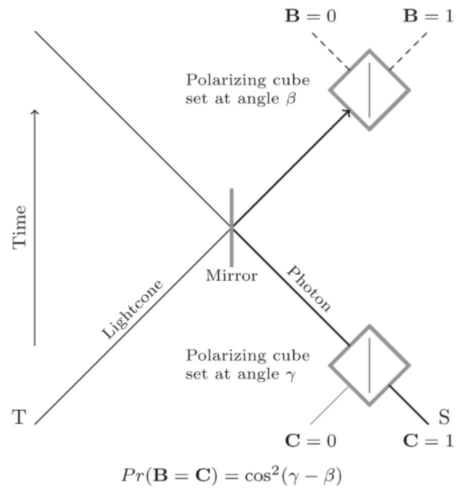
> [T]he correlations characteristic of Einstein–Podolsky–Rosen (EPR)–Bell (EPRB) experiments also arise in familiar cases elsewhere in quantum mechanics (QM), where the two measurements involved are timelike rather than spacelike separated; and in which the correlations are usually assumed to have a local causal explanation, requiring no action-at-a-distance (AAD). It is interesting to ask how this is possible, in the light on Bell's Theorem. We investigate this question, and present two options. Either (i) the new cases are nonlocal too, in which case AAD is more widespread in QM than has previously been appreciated (and does not depend on entanglement, as usually construed); or (ii) the means of avoiding AAD in the new cases extends in a natural way to EPRB, removing AAD in these cases too. There is a third option, viz., that the new cases are strongly disanalogous to EPRB. But this option requires an argument, so far missing, that the physical world breaks the symmetries which otherwise support the analogy. In the absence of such an argument, the orthodox combination of views—action-at-a-distance in EPRB, but local causality in its timelike analogue—is less well established than it is usually assumed to be.

In [51], Evans et al. turn the standard EPR-type experiment (with photons) sideways, so that the two (polarization) measurements now have a timelike separation rather than the usual spacelike separation. Correlations in these cases are not normally thought to be a manifestation of non-local causality as there appears to be a perfectly normal local causal explanation: the two timelike separated measurements are in each others light cone. My main aim is now to critically evaluate the analogy between spacelike and timelike non-locality that Evans et al. are arguing for. I argue that one of the possible paths for Evans et al.'s symmetry requirements is probably wrong, whilst the other is at least problematic.

If you truly subscribe to spacelike non-locality, then you should only take up positions (i) and (iii). But either option comes at a price. Option (i) conflicts strongly with with the by now relatively standard view of non-locality. Moreover, the AAD in the SEPRB situation appears not to depend on any entanglement, which is a necessary criterion for Bell's theorem. Option (iii) amounts to the rejection of one or more seemingly attractive symmetries that support the analogy between EPRB and SEPRB.



(a) EPRB (with photons).



(b) Sideways EPRB (SEPRB).

Figure 6: Comparing the regular and time-symmetric EPR-type experiment. Images and captions from [51, p. 302-303].

Evans et al. themselves argue in favor of either (i) or (ii). Wharton is the main proponent of (i) and Price the main proponent of (ii). I shall argue strongly in favor of (iii) and voice my concerns about (i) and (ii).

5.1.1 The experiments

In Figure 6a, we see a regular space-time representation of a set-up of the EPR-type experiment, performed with a pair of photons. The two photons, described by an entangled state of zero total polarization, are emitted from space-time location M into opposite directions towards spacelike separated (ideal) polarizing cubes set at angles α and β. These angles are boundary conditions for the model, chosen by the experimenters—say, Alice and Bob, respectively. When arriving at the polarizing cubes, the photons can either be transmitted ($A = 1, B = 1$) or reflected ($A = 0, B = 0$), the result probabilistically depending on their polarization before entering the polarizing cubes (which is the same for every run of an EPR-type experiment).[40] In this particular set-up, the sting of Bell's theorem is that the joint probability of the same result is known to be $P(A = B) = \cos^2(\alpha - \beta)$ (instead of simply $P(A = B) = 1/2$), violating Bell's inequality when one considers suitable combinations of settings. This indicates that this result can not be replicated with the addition of any *local* hidden variables at M (which therefore do not depend on the distant—or future—settings α and β).

In Figure 6b, we see the same set-up repeated, but shifted sideways in the space-time diagram. The distance between the two polarizing cubes is now timelike rather than spacelike. The rest of the set-up appears identical. However, there are some peculiar changes with respect to the standard EPRB experiment. Firstly, we are now considering a single-photon experiment: a photon enters through the polarizing cube with setting γ, gets reflected at a mirror at location M[41], and subsequently passes through the second

[40] A and B are the measurement results, represented by Boolean variables in [51]. I omit the original bold capitals to remain consistent with the rest of this thesis.

[41] This reflection at the mirror is 'for heuristic purposes only'. The results should be the same if the photon is not reflected but simply sent towards a distant polarizing cube. The reflection at the mirror,

polarizing cube with setting β. Secondly, the measurement 'result' C, which should be the analogue of the result A in Figure 6a, is no longer a measurement result but rather an input choice made by the experimenter—let's call him Charlie. This means that, if Charlie chooses $C = 1$ and an arbitrary setting γ, then the probability for Bob of obtaining result $B = 1$ is given by $\cos^2(\beta - \gamma)$. Similarly, if Charlie chooses $C = 0$ and arbitrary γ, Bob's result of $B = 0$ is $\cos^2(\beta - \gamma)$. Therefore, the total joint probability for results B and C is again $P(B = C) = \cos^2(\beta - \gamma)$. On first sight, therefore, the results of the regular EPRB-experiment and the sideways EPRB-experiment (SEPRB) are analogous, and their implications for Bell's Theorem—which uses only this joint probability—should be the same. I describe the three possibilities for interpreting this result below, after raising some preliminary concerns about the analogy between the EPRB-experiment and SEPRB-experiment itself.

It appears that we must be cautious about the adequacy of the analogy for several reasons. The most obvious reason is again the communication loophole discussed in Section 4.8: the SEPRB-experiment allows for a perfectly locally causal explanation whilst the EPRB-experiment does not. After all, the two different polarizing cubes at C and B lie in each others light cone and can therefore causally influence each other. (More specifically, Charlie's procedure should be able to influence Bob's procedure). This does imply, however, that the two set-ups, EPRB and SEPRB, are more disanalogous than is apparent by the probability calculus. This is one of the options in Evans et al.'s trilemma (the one that I shall be arguing for), but it requires arguments against the symmetry considerations that Evans et al. employ in order to justify the analogy between the spacelike and the timelike experiment. These will be discussed below. The second reason of caution is with respect to the interpretation of this probability calculus. The two joint probabilities $P(A = B) = \cos^2(\alpha - \beta)$ and $P(B = C) = \cos^2(\beta - \gamma)$ may be structurally similar, but there is an important difference between them: the conditional variable A is a genuine measurement result (with a probability distribution) whilst the conditional variable C is a boundary condition for the model, chosen by Charlie (without a probability distribution). This makes it quite dubious to interpret these two probabilities $P(A = B)$ and $P(B = C)$ equally, as the former is the joint probability of two measurement results whilst the latter is the joint probability of a measurement result corresponding with a boundary condition.[42] The third and last reason of caution is concerned with null results for the SEPRB-experiment. In the regular EPRB-experiment, there are no null results possible because all possible combinations of measurement results for Alice and Bob are valid for analysis. This is not the case for the SEPRB-experiment: if Charlie inputs a photon with a certain polarization γ_{in} through $C = 1$ (transmission) and sets his polarizing cube to setting $\gamma = \gamma_{in} + \pi/2$, then the photon will not pass Charlie's polarizing cube, and Bob will subsequently obtain a null result on his side of the experiment. To avoid this situation, Evans et al. mention that the polarization of the incoming photon at Charlie's polarizing cube is irrelevant as long as we 'consider a null result at B to be no experiment at all' [51, p. 303]. This point may or may not have bearing on the validity of the symmetry arguments—time-symmetry in this case, I elaborate on this point in Section 5.3—but it certainly raises a problem for the analogy between the EPRB and

however, allows us to imagine the photon being absorbed by and re-emitted from the mirror, which, in effect, allows us to consider the photons before and after the mirror reflection as two different photons. This consideration also makes the action symmetry that will be discussed below easier to interpret. This was pointed out to me by Evans in private communication.

[42] Note that this objection appears similar to my objection to Maudlin's independence assumptions in Section 3.2.3.

SEPRB experiments.

5.1.2 The symmetry arguments

With respect to the present analysis, Evans et al. mention that there are two crucial assumptions necessary for the derivation of Bell's Inequality. The first is *Independence*, which states that all possible hidden variables of the two-photon system are statistically independent of the measurement settings at the polarizer cubes. The second is *Locality*, which states that variables in our model may only be dependent on the variables and boundary conditions in their past light cone. In the formalism adopted in this thesis, assuming Independence implies assuming Source Independence (3.1), and assuming Locality amounts to assuming the validity of Factorizability (3.12)—which was Bell's definition of what he called *local causality*.[43]

If correlations violate Bell's Inequality, this would imply that at least one of the assumptions of Independence or Locality is violated. For the EPRB experiment, Locality is usually considered to be violated, naturally resulting in a conclusion of 'non-locality' (I aim to avoid the more ambiguous term 'action-at-a-distance'). But for the SEPRB experiment, the natural interpretation is that Independence fails. In order to do so, however, Evans et al. state that we are effectively considering the photon (or additional hidden variables) as a Bellian *beable* that can carry the information of measurement settings—that will violate Independence—from Charlie's polarizer to Bob's. They state that such beables provide a *Locality Model* for the SEPRB experiment. The argument by Evans et al. is now that if (i) there exists a Locality Model for the SEPRB experiment, and (ii) the ontology of the experiment—the beables—respect certain symmetries that can transform the SEPRB into the EPRB experiment, then we must accept that there exists a Locality Model for the EPRB case too. A model that, of course, would violate Independence retrocausally. If one finds this conclusion unacceptable, then one is now forced to argue that either there does not exist a Locality Model for the SEPRB case or the symmetry arguments do not hold.

Although Evans et al. describe two different lines of reasoning for applying symmetry arguments to transpose EPRB into SEPRB, both lines require the same two necessary symmetry arguments: *Time symmetry* and *Action symmetry*. I will now briefly introduce the two symmetry considerations (I will elaborate on them in Sections 5.3 and 5.4, respectively) and sketch the line of reasoning for transposing EPRB into SEPRB.

One line of reasoning provided by Evans et al. is based on highly intuitive reasoning about *Time symmetry*. I have mentioned that, if we assume there to be a Locality Model for the SEPRB case, it is natural (or simply mandatory if we choose to work with Bellian terminology) that there exists a beable that carries the information about Charlie's setting towards Bob, so that the Independence assumption fails. Evans et al. now state that, "unless the Locality Model in question is time-asymmetric, it also yields a Locality Model for EPRB." [51, p. 311] They raise the following question: if there exists a beable that carries information for Charlie to Bob, is there also a beable that carries information from Bob to Charlie? They justly state that, if one considers the regular quantum state ψ as the only beables travelling between Charlie and Bob, and if we assume the projection postulate as we do in bare-bones quantum mechanics, then the answer is obviously 'No, there is no beable carrying information from Bob to Charlie'. This means that such

[43] It may be noted here that this definition of locality based on intuitively 'screening off' causal influence by specifying a segment of a variable's past light cone applies equally well to the spacelike and the timelike case. To see this, compare Figure 1b with Figure 5.

an interpretation is time-asymmetric. Evans et al. now state that time-symmetry can be restored by introducing a second wave function carrying information from Bob to Charlie[44] or by using second-order wave equations constrained by both past and future interaction.[45] These cases could be interpreted as suggesting beables travelling from Bob back to Charlie.

The point that Evans et al. try to make here is that a time-symmetric Locality Model for the SEPRB experiment must necessarily contain beables both from Charlie to Bob and from Bob to Charlie [51, p. 311]. Moreover, once we have these two types of beables travelling forwards and backwards in time to provide the measurement apparatuses with information about the distant apparatus' settings, thus violating Independence, we can fairly straightforwardly imagine that such beables can also provide a Locality Model (and violate Independence) for the regular EPRB experiment. However, whilst this latter statement is true,[46] the former statement is false. Even if we take the wavefunction ψ itself seriously as a beable—which is already problematic for reasons alluded to in Section 1.2—then only those quantum mechanical theories with a collapsing wave function are actually time asymmetric. Bohmian mechanics is a simple counterexample to the claim that time symmetric quantum mechanics must necessarily involve multiple beables travelling in opposite directions of time. In fact, any quantum mechanical theory without a collapsing wavefunction is perfectly time symmetric. It is only the projection postulate that makes quantum mechanics time asymmetric.

The second line of reasoning from Evans et al. is perhaps more sophisticated. It does include intuitive reasoning with regards to time symmetry similarly to the first line of reasoning discussed above, but it starts with another symmetry: the *Action symmetry*. They begin by noting that the two experiments, EPRB and SEPRB, span bounded regions of space-time with identical electromagnetic action (in quantum electrodynamics). This means that, in terms of the Feynman path-integral, we may write the joint probability as:

$$P(A_{t_0}, A_{t_f}) = \left| \int e^{iS(A)/\hbar} \mathcal{D}A \right|^2, \tag{5.1}$$

where A is the electromagnetic 4-potential, and the integral is over all field configurations consistent with the initial boundary A_{t_0} and the final boundary A_{t_f}. The derivation of and physical motivation for Equation (5.1) is rigorously discussed in Wharton, Miller, and Price's [136]. I will return to this in Section 5.4, but for the present discussion it suffices to briefly introduce the action symmetry principle that Evans et al. aim to exploit. The action symmetry is most explicitly introduced in [136, p. 528]:

> **Feynman Integral Symmetry Hypothesis (FISH):** For any two experiments with an action duality (a well-defined spacetime transformation that maps the classical Lagrangian density of one experiment onto the classical Lagrangian density of the other), any realistic ontology must also map between the two experiments under the same spacetime transformation.

[44] This is simply the two-time interpretation [3] of the Aharonov and Vaidman's two-state vector formalism [4], see Section 5.3.

[45] This refers to Wharton's novel interpretation of the Klein-Gordon equation [130], which can arguably be considered a precursor to Wharton's emphasis on (past and future) boundary conditions and his Lagrangian quantum mechanics and accompanying all-at-once interpretation. This will be discussed in Section 5.4.

[46] It was suggested already in the 1950s by Costa de Beauregard [32], cf. [109]. This concept is now known as the 'Parisian zigzag'. See Section 5.3.

The idea is thus that we should be able to transform a Locality Model for the SEPRB into a Locality Model for EPRB because the electromagnetic action is preserved when the geometry of the experimental set-up is permuted. In the words of Evans et al.: "this action-preserving permutation of the *geometry* (from that of SEPRB to that of EPRB) provides a simple QED-inspired template for a corresponding permutation of the *ontology*" [51, p. 308]. They explicitly mention that one should come up with reasons for thinking that the ontology of SEPRB and EPRB does not reflect the symmetry of the mathematics if one disagrees with the hypothesis cited above.

But the application of an action symmetry principle alone is not enough to satisfy all symmetries because a spatial symmetry remains violated. The intuitive Locality Model for the SEPRB experiment has a one-way causality, where the beables travel only from Charlie to Bob. If we permute this experiment to an EPRB experiment, then the ontology suggested by only the action symmetry does violate Independence, but it does so spatially asymmetrically: there is only a beable travelling from Alice to Bob, not the other way around. In order to recover the spatial symmetry for the EPRB experiment, Evans et al. suggest, we must treat the future settings α and β on an equal footing, leading to beables travelling both from Alice to Bob and vice versa. This, in turn, implies that the SEPRB experiment too requires beables travelling in both directions. Before I continue with the Evans trilemma, I must note that now something feels suspicious: we started our reasoning with the assumption of a Locality Model for the SEPRB experiment (which has one-way causality) and we end up with SEPRB with both-way causality. Moreover, if we insist on one-way causality, we could imagine turning this argument upside down into a peculiar argument in favor of spatially asymmetric causality in the regular EPRB: the line of reasoning would imply that only one wing of the EPRB experiment is endowed with (non-local or zig-zag) causal influence over the other wing, while the other wing is merely a passive receiver of causal influence. Such spatially asymmetric causality is quite possibly more undesirable than both retrocausality or symmetric non-local causality.

5.1.3 Discussion

If we accept either of the two symmetry-based strategies for transforming a Locality Model for SEPRB into a Locality Model for EPRB, we are confronted with the Evans trilemma introduced above. To repeat:

(i) There is no Locality Model for either EPRB or SEPRB because both experiments exhibit non-locality.

(ii) There is a Locality Model for both EPRB and SEPRB because the symmetry arguments hold.

(iii) There is a Locality Model for SEPRB but not for EPRB because the symmetry arguments do not hold.

In [51], Evans et al. aim to take no specific stance on which option is preferable. We will recognize in the remainder of Section 5, however, that Price takes up (ii) and Wharton takes up (i). I find myself choosing option (iii), mainly for the reasons mentioned in Section 5.1.1. I believe the two experiments, EPRB and SEPRB, are disanalogous not for reasons relating to the underlying ontology but because the construction of the experiments is not analogous. When switching from EPRB to SEPRB, the measurement 'result' is not a result anymore but a setting instead—which already indicates that we are concerned with a different, disanalogous experiment—and, consequently, the interpretation of the probability calculus becomes rather suspect because we change from a joint probability of two measurement results to a joint probability of a measurement setting and a result.

Regardless, Evans et al. discuss a few common replies to the trilemma. First, they justly remark that both instrumentalists about quantum mechanics and those who subscribe to many-worlds naturally sidestep the trilemma by avoiding the 'naive' ontology of Bellian beables that has been used in the construction of the experiments and arguments above. Like Evans et al., I grant the intrumentalists and many-worldists their way out, and repeat that both interpretations fall outside the scope of this thesis.

The second reply that Evans et al. discuss is with regard to time symmetry. It can be claimed, they state, that 'the time-asymmetry of a Locality Model for SEPRB with C-beables but no B-beables does not need to be fundamental, so that such a model does not conflict with the assumed time-symmetry of the fundamental ontology that we're after' [51, p. 315]. And they then reply that this can be claimed for classical electrodynamics (which yields the same probability ratio's, as we will see again in Section 5.3) but not for the case of a single photon, where the time-symmetry is manifestly different. But, as I discussed above, it is not entirely clear what Evans et al. precisely mean with 'time symmetry'. They cannot mean that the causal description of the experiment should be time symmetric, because this would already smuggle in retrocausality and so make the whole argument circular. Apparently, neither do they mean the familiar time symmetry in the sense that the natural laws work the same when time is reversed, $F : t \mapsto -t$, because such time symmetry is clearly obeyed by the interpretation of quantum mechanics without collapse of the wavefunction—Bohmian mechanics being the main candidate in the context of this thesis. But if they then instead insist on some time symmetry of solutions, then the requirement is much too strong. Such time symmetry has never been required of a physical theory for it to be considered time symmetric. Evans et al. aim to strengthen their point by appealing to the Price Argument that is discussed in Section 5.3, but there I make the same point that the time symmetry requirement is either too strong or too vague. Moreover, it is also acknowledged by Evans et al. that Bohmian mechanics is perfectly immune to these arguments (for it has no wave function collapse). Price himself does often mention that the absence of an argument for retrocausality in the context of Bohmian mechanics is no argument for the absence of retrocausality in the context of Bohmian mechanics [106, p. 81], however, but this is hardly an argument in the present discussion.

Finally, Evans et al. discuss two more possible objections to their trilemma in favor of option (iii): one concerning Maudlin's argument of inconsistency in retrocausal theories and one concerning 'free variables'. I agree with Evans et al. that both objections are unsatisfactory, although I do so perhaps for different reasons. Regarding the first, Maudlin's opposition against retrocausality is well-known [86, p. 201] (see further [86, ch. 7]):

> If the course of present events depend on the future and the shape of the future is in part determined by the present then there must be some structure which guarantees the existence of a coherent mutual adjustment of all the free variables [...] Any theory with both backwards and forwards causation cannot have such a structure. Data along a single hypersurface do not suffice to fix the immediate future since that in turn may be affected by its own future. The metaphysical picture of the past generating the future must be abandoned, and along with it the mathematical tractability of local theories.

I side with Evans et al. in holding that Maudlin is perhaps too strongly committed to this metaphysical picture of the past generating the future [51, fn. 22], which obviously excludes retrocausality in physics from the start.

The second objection, concerning 'free variables', exploits the persistent red herring of 'free willed experimenters' that have been an explicit subject of discussion in the context of EPR-type experiments at least since Bell (see [16, ch. 12] or [18, p. 229]). The rejection of the assumption of Independence not via retrocausal influence of the 'freely chosen' measurement settings on the initial quantum state but, reversely, by claiming that the settings were not chosen freely but are instead determined by the initial quantum state via local interactions is an experimental loophole that has been closed more and more over the past couple of decades [59]. But it is always possible to claim that the measurement settings and the initial quantum state are correlated by some common past. When a common cause in the distant past determines both the measurement setting and the initial quantum state, there is nothing mysterious about correlations between the settings and quantum state (and they may even be accounted for purely locally). But explaining away correlations like this is often referred to as such an 'unscientific' proposal that nearly all philosophers and physicists feel justified in dismissing it entirely. Moreover, such proposals are an argument against Bell's theorem entirely, and, consequently, against non-locality in both the regular EPRB and the new SEPRB case. They are not an argument against the Evans trilemma itself, because this trilemma presupposes the validity of Bell's theorem in the regular EPRB experiment. I will now discuss this proposal in more detail.

5.2 Proposal 1: Superdeterminism

The contemporary bogeyman of the non-locality debate is superdeterminism. This position holds that, given a single run of the experiment, the measurement settings a and b and the state λ are not independent boundary conditions but they are instead all three determined by some common cause in their distant past. Accordingly, the relevant correlation is not the two-body correlation between the measurement outcomes A and B that is the subject of the Bell-type inequalities, but instead a three-body correlation between all boundary conditions a, b, and λ. The idea is that their spacelike non-local correlations should be explained by locally causal dependence on their common past. It must be said that there is no timelike causal dependence as there is, perhaps, in retrocausal theories. But I treat this option here because retrocausal and superdeterministic theories are often dismissed together with a single assumption similar to *Independent boundary conditions*. The main proponent of this position is 't Hooft [120], whose recent paper on this matter [121] caused quite a stir on social media.[47] 't Hooft especially argues against the (very misleading) usual description of *Independent boundary conditions* as requiring the assumption that the measurement settings are chosen by experimenters with some philosophical notion of free will.

To understand where this position deviates from the usual treatment of the model, we need to understand some of the terms often employed by 't Hooft (cf. [120, §14.3]) and others. According to 't Hooft, 'determinism' (in a causally closed universe) means that all physical phenomena are direct consequences of physical laws that do not leave anything to chance, and 'superdeterminism' means that the experimenters too behave in accordance with the same laws. With regards to this distinction between determinism

[47] On 9 September 2017, Tim Maudlin posted a direct reaction to this paper on his Facebook page, starting with: "Here is my futile attempt to prevent more of this."

and superdeterminism, it is often replied that regular determinism already means that the experimenters and their choices of measurement are fully determined by their causal past: a more standard definition of determinism is that future evolution of a system is uniquely specified by data along a suitable spacelike hypersurface in combination with dynamical laws. Even if we let some pseudo-randomizer like a lottery-machine make the choice for us, this too is fully determined in regular determinism, albeit unpredictable. But I quite agree with the usefulness of the term 'superdeterminism': what is 'super' about superdeterminism is that it appeals to something outside of the descriptive power of the model: it drags the experimenter in by the hairs and demands the model to account for their actions.

't Hooft, apparently being a non-compatibilist, holds that superdeterminism is incompatible with philosophical free will. But the assumption of *Independent boundary conditions* is often very unfortunately named *Free Will* or *Free Choice* or the like, and this diverges our attention from the necessity of indepenent boundary conditions to the apparent fallacious anthropocentrism of involving philosophical free will in a physical theory. The usage of 'free willed experimenters' can be traced back to the original EPR paper, but it takes on explicit form in Bell's papers (especially [16, ch. 12]). Bell emphasized that this assumption was necessary for scientific practice, and he denied on several occasions that he was talking about philosophical free will.[48] Bell really meant to say that independent boundary conditions were necessary for the building and evaluating physical models, not that it required philosophical free will. But he did not do this clearly enough, presumably, and this helped philosophical free will become the biggest red herring of the foundations of quantum mechanics, as is evident from the numerous scattered mentions of free will across the literature, and, especially, the Free Will Theorem [30, 31] and dismissive replies thereto (eg. [63, 124, 140]).[49]

In the context of this thesis, superdeterminism amounts to extending the model by replacing the boundary conditions a, b, λ by a new boundary condition that is their common cause, call it Ω. See Figure 7. This would allow for factorizability of the joint probability distributions necessary for the derivation of CHSH-inequalities:

$$\rho(a, b, \lambda|\Omega) = \rho(a|\Omega)\rho(b|\Omega)\rho(\lambda|\Omega). \tag{5.2}$$

This violates first and foremost the assumption of *Model completeness* by replacing the formerly-boundary conditions by a new boundary condition Ω and even by introducing experimenters that may have free will (and then denying the latter). It is usually acknowledged that it is indeed a logical possibility in principle that there is a common source that forces the measurement settings to take up the specific value that provides results which confirm the predictions of quantum mechanics. It is also evident, however, that the common cause is always unknowable in practice and that we should therefore not appeal to it in our explanation of the correlation. So a common objection to this violation of *Model completeness* is this: systematic correlations have always been explained by a direct causal influence or by an in practice specifiable common cause. If we now start

[48] But also consider Bell's comment that [16, p. 101]: "Here I would entertain the hypothesis that experimenters have free will. But according to CHS it would not be permissible for me to justify the assumption of free variables 'by relying on a metaphysics which has not been proved and which may well be false'. Disgrace indeed, to be caught in a metaphysical position! But it seems to me that in this matter I am just pursuing my profession of theoretical physics."

[49] Here I want to stress that, even if the Free Will Theorem does prove anything new, it can be convincingly shown that the topic of philosophical free will is logically disjunct from the Free Will Theorem [13].

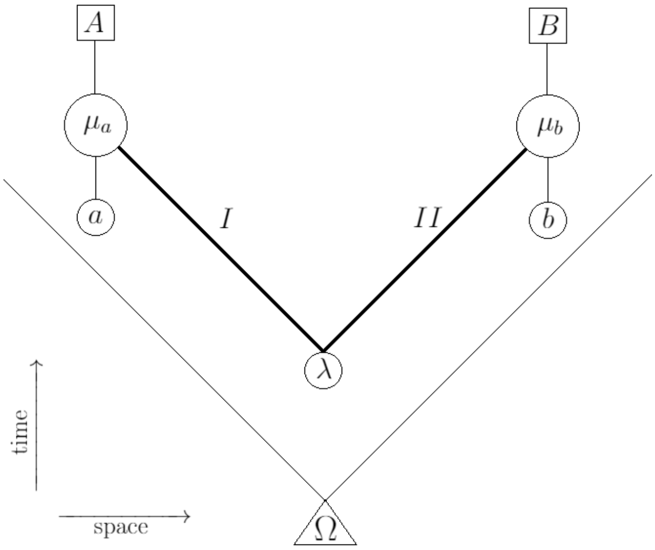


Figure 7: The superdeterministic common source dependence represented as replacing the previously-boundary conditions a, b, λ with a new boundary condition Ω.

explaining systematic correlations not by direct causal relations but by postulating an in practice unknowable common source, then it seems we cannot trust any knowledge based on empirical research in the history of science.

Superdeterminism also similarly violates the *Independent boundary conditions* assumption: there is a systematic bias caused by the common past that forces the measurement settings to take up only those values that would confirm the predictions of quantum mechanics and not those that would contradict the predictions. Understandably, this is dismissed as too 'conspirational'. Again, there is nothing conspirational about determinism or superdeterminism in itself, there is only something conspirational about using determinism to explain away correlations that demand a scientifically reasonable causal explanation. This objection may be formulated in two ways. Firstly, this type of explanation would suggest that the laws of physics that we humans come up with are entirely contingent on the specific composition of an initial state, and that there is nothing that enables us to say anything about the relation of these laws to the external world. It suggests that the laws of quantum mechanics as applied to the EPR-type thought experiment are not fundamental in any sense but merely a mirage, a result of some very specifically fine-tuned initial state that makes us believe the laws of quantum mechanics hold good over and over again. And this fine-tuning would be much more severe than the familiar fine-tuning of, for example, the laws of thermodynamics and their necessary Past Hypothesis. This led some authors such as Maudlin to suggest that superdeterminism should be called hyperfine-tuning instead. I prefer superdeterminism, as it immediately refers to the 'super'-extension of the toy model, indicating a violation of *Model completeness* (but let us not get hung up on semantics). The second formulation of this objection is a bit more quantitative, and was clearly formulated by Bacciagaluppi [13, slide 28]: the strangeness of the correlations lies not in the statement that "my free choices are predetermined, but that there is the same dependence between the settings and the hidden variables whether the choice of settings is by me, or by Ruward, or by a random number generator, or by anything else (Bell's favourite example were the Swiss lotto machines)!"

Because superdeterminist and retrocausal approaches to the EPR-type experiment both avoid the conclusion of non-locality by violating Source Independence (3.1), it is often said that retrocausality and superdeterminism are essentially the same, or that retrocausality is superdeterminism is disguise. Since superdeterminism is the black sheep in the philosophy of quantum mechanics, advocates of retrocausality have replied to this comparison with objections similar to those I presented in this section (cf. [108, 109]). As the next section is concerned with retrocausality, it is useful to already here summarize the retrocausalist's objections and to distinguish their approach from superdeterminism. The objections, mainly stated explicitly in [109, §6], are as follows. Firstly, retrocausality accepts and enforces the experimenter's free choice because this is a 'standard presupposition of all experimental science'. Relatedly, this notion of 'free choice' enables the retrocausalist to employ a standard *interventionist* account of causality to construct a story of what is happening in an experimental context such as the EPR-type experiment.[50] But although these objections are sound, they do tend to make one focus on a particular understanding of causality (which is different from a probabilistic account of causality). The best objection in my opinion, therefore, is the following point made by Price and Wharton [109, p. 7763]:

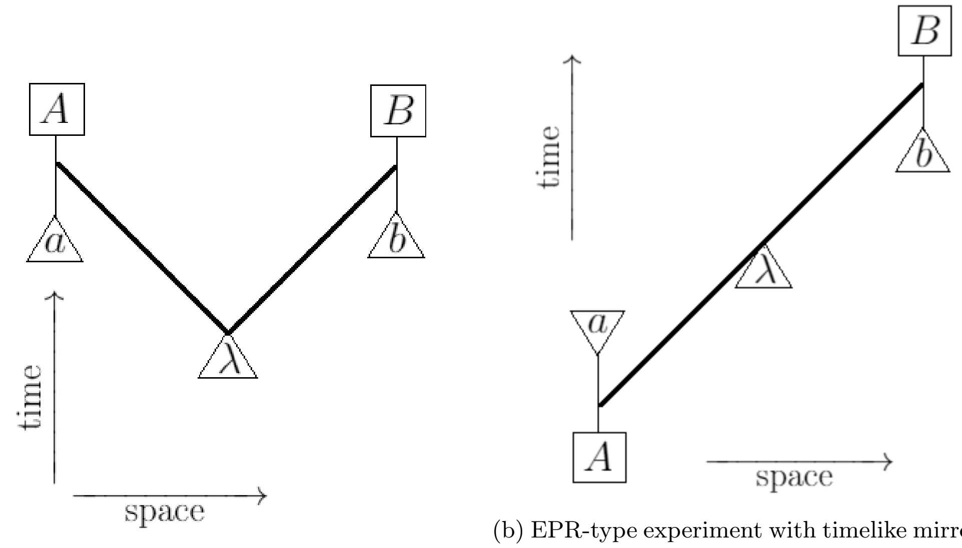
> The retrocausal model introduces no strange new hidden variables to control measurement settings. To the extent that it proposes new hidden variables, they are internal to the model (and, at least in the version referred to in these paragraphs, of a familiar kind ...)

We can now recognize that this objection points precisely to the unwarranted 'super'-extension of the model by the superdeterminist, violating the basic assumption of *Model completeness*. By making this assumption explicit at the start of this thesis in Section 3.1, I hope to have given this assumption enough oomph to make it a satisfactory objection to superdeterminism. I will now discuss the retrocausal proposal in detail.

5.3 Proposal 2: Retrocausality

Retrocausality in quantum mechanics, originally proposed by Costa de Beauregard (eg. [32, 33]) in the form of the 'Parisian zigzag', has been undergoing a minor resurgence in recent years. Retrocausal dynamics have been worked into concrete interpretations of quantum mechanics during the end of the twentieth century in, for example, Cramer's Transactional Interpretation [34, 35] and the Two-Time Interpretation of [3] Aharonov and Vaidman's Two State Vector Formalism [4]. The idea of allowing retrocausality in the sense that, in the EPR model, the state λ is dependent on measurement settings a and b is a persistent loophole to the conclusion of spacelike non-locality from Bell's Theorem, and it continues to be a tempting loophole even to more recent no-go theorems such as the PBR Theorem [110] but only quite recently has an explicit argument been proposed that the demand of both quantum-like discreteness and time-symmetry at the most fundamental level of physics requires some form of retrocausality. This argument, originally proposed by Price [106] and elaborated upon by Price and Wharton [107, 108, 109], has evidently gained some traction. Price's argument is regularly discussed by authors such as Evans [50, 51] and Adlam [2], and it was built upon in a recent result by Leifer and Pusey [81], which consequently sparked debate between Maudlin [88] and Leifer [80] (where

[50] I briefly mentioned this interventionism in Section 2.4: "a variable X is a cause of a variable Y if and only if a free intervention on X makes a difference to Y" [109, p. 7763].



(a) Regular EPR-type experiment.

(b) EPR-type experiment with timelike mirrored left-hand side of the experiment, resulting in a time-symmetric set-up.

Figure 8: Comparing the regular and time-symmetric EPR-type experiment.

the cause of their feud was precisely the ambiguity in the Bellian probability formalism, cf. footnote 12). Because Price and Wharton have presently developed rather divergent views of the implications of retrocausality, I discuss Price's original argument in this section (and I call it the Price Argument even though Wharton was a co-author for many of the relevant papers) and I discuss Wharton's apparent current point of view in the next section.[51] Again, we already know that retrocausal theories are not actually timelike non-local in the sense that we are looking for in this thesis. But since retrocausal theories are intimately connected to genuinely timelike non-local theories such as all-at-once—as we have seen in Section 4—understanding these proposals too will help us in better understanding timelike non-locality in general.

To get an intuitive idea of what the Price Argument amounts to in the context of the EPR-type experiment, see Figure 8. To obtain Figure 8b from 8a, we have temporally inverted everything on the left-hand side of the regular EPR-type experiment. The set-up is now completely (geometrically) time-symmetric. But two things immediately stand out: the role of λ in 8b is now unclear and needs to be specified, and the temporal order of measurement setting a and result A has flipped with comparison to the regular EPR-type experiment. Both these issues will be dealt with by specifying the details of the experiment in the Price Argument, but we must already acknowledge that we are here concerned with an experimental context that is manifestly different from the more familiar temporal EPR-type experiments discussed in Section 4. I provide here a brief summary of my critique before I discuss the Price Argument in detail.

My reformulation of the Price Argument consists of four assumptions that together imply retrocausality: *realism* + *discreteness* + *time-symmetry* + *null-avoidance* ⇒ *retrocausality*. *Realism* states that our ontology consists only of local beables. This is a rea-

[51] In private discussion, Derakhshani explained to me that Price might personally subscribe as much to 'all-at-once' as Wharton. I have not read this anywhere in print, however, while Price does have numerous papers (and a book) arguing for dynamical retrocausality, and so I find myself justified in representing him as advocating this position.

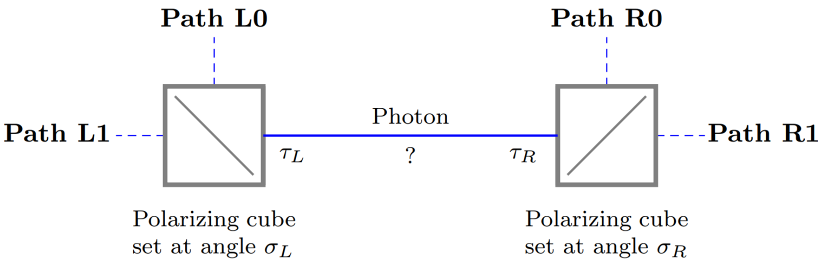


Figure 9: Experimental set-up for the Price Argument. Image from [107, p. 12].

sonable assumption, but it is violated by, for example, positing non-local beables (which would already pose problems for conclusions about spacelike locality from Bell's theorem). *Discreteness* states that we allow no superpositions of our local beables: the (wavefunctions of the) photons in the experiment enter and exit the measurement apparatuses through a single channel only. This assumption is naturally violated by any interpretation of QM with a non-collapsing wavefunction such as Bohmian mechanics. *Time-symmetry* holds that the ontological description of the situation should remain the same if the arrow of time is reversed. This assumption is violated by all theories with dynamical asymmetry such as collapse theories. *Null-avoidance* states that we only consider those cases in which a photon travels through both apparatuses successfully, no null results are allowed on either side. With a time-symmetric flavor, this implies that we both *pre-select* passing photons on one side (say, the left-hand polarizer) and *post-select* on the other side. Even given that the Price Argument already does not appear to have much bearing on standard practice in foundations of quantum mechanics because all previous assumptions are naturally violated by most interpretations of QM, *null-avoidance* is also doing quite some implicit work in the argument. At the end of this section, I will argue that taking *null-avoidance* into account reduces the scope of the Price Argument even further.

5.3.1 Price Argument: classical

Price usually presents his experiment first for the case of classical electromagnetism, where only assumptions of *time-symmetry* and *realism* are required, and only then for the quantum case. The only difference between the classical and the quantum case is an extra assumption called *discreteness*. In doing so, he demonstrates the remarkable work done by this latter assumption. Interestingly, only Price's original paper [106] makes explicit use of these assumptions and formulates the Price Argument in the logical form of: *realism* + *time-symmetry* + *discreteness* ⇒ *retrocausality*. The rest of his (and Wharton's) papers on the subject [107, 108, 109] present a more qualitative form of the argument and the implications thereof without explicit use of these assumptions. I present the argument with explicit use of the assumptions to highlight the work that they do and to hopefully make my critique more evident.

Consider the experimental set-up of Figure 9 and imagine for the moment an experiment where we use only the right-hand polarizer. We send a beam of light with polarization τ_R and intensity I_0 to an ideal polarizing cube set at angle σ_R. Classical electromagnetism predicts that the beam of light with split into two output beams: a transmission beam R1 with intensity $I_{R1} = I_0 \cos^2(\tau_R - \sigma_R)$ and polarization $\tau_R = \sigma_R$, and a reflection beam R0 with intensity $I_{R0} = I_0 \sin^2(\tau_R - \sigma_R)$ and linear polarization

$\tau_R = \sigma_R + \pi/2$. This is a familiar experiment. But now consider the time-reversed situation, where we use only the left-hand polarizer of Figure 9. In this case, the two beams L1 and L0 are combined into a single beam in the polarizing cube set at angle σ_L. Clearly, the two beams will combine into a single beam of polarization τ_L with intensity I_0 only if the input beam of path L1 has intensity $I_{L1} = I_0 \cos^2(\tau_L - \sigma_L)$ and polarization $\tau_L = \sigma_L$ and the input beam of path L0 has intensity $I_{L0} = I_0 \sin^2(\tau_L - \sigma_L)$ and polarization $\tau_L = \sigma_L + \pi/2$ (and the two beams are exactly in phase).

We now combine the two polarizing experiments described above into the full single experiment as presented in Figure 9: two beams of light coming from paths L0 and L1 are combined into a single beam at the left-hand polarizer, and this beam travels to the right-hand polarizer, where it is split into the two beams R0 and R1 with the conditions described above. Now we consider the case where our usual experimenters, Alice and Bob, have control only over the polarizer settings σ_L and σ_R, respectively, just as they do in the regular EPR-type experiment. Let's ask ourselves what type of control Alice and Bob have over the intermediate polarization τ of the beam in between the two polarizers. Clearly, Bob has no control over τ by adjusting σ_R, because a shift in σ_R will only result in a change of intensities I_{R1} and I_{R0}. The case for Alice is a bit harder to imagine because we know neither the intensity nor the polarization of the incoming beams of light via L1 and L0. If these were all constant, then Alice evidently has some control over τ because she can adjust her setting σ_L so that she combines the two beams coming from L0 and L1 with different ratios of intensity, resulting in a different intermediate polarization.

However—or, rather, therefore—this situation is not time-symmetric: we allow the intensities I_{R1} and I_{R0} to vary while we hold I_{L1} and I_{L0} constant. This means that we have a manifestly different experiment if we flip the arrow of time and the two beams now meet at the right-hand polarizer and split at the left-hand polarizer. To enforce time-symmetry, therefore, Price introduces a Demon that controls the inputs over L1 and L0 and who knows the setting σ_L that Alice has chosen, in exactly the same (time-reversed) way as nature 'knows' the setting σ_R and controls the variable outputs at R1 and R0 on the right-hand side of the experiment. Of course, the Demon is only there to stimulate our intuition: its role is only to make sure that time-symmetry is enforced in the sense that the physical situation before the left-hand polarizer is the same as we would usually consider the situation to be after the right-hand polarizer. In fact, it will turn out that the work that the 'Demon' is doing in the argument is a combination of the assumptions of *discreteness* and *null-avoidance*. I will elaborate on this point in the discussion and leave the heuristic Demon in place for now. Now Alice has no control over the intermediate polarization τ if she only has control over σ_L because the Demon can now produce any intermediate polarization τ that it desires by choosing appropriate inputs for L1 and L0. In this time-symmetric classical electromagnetic experiment, therefore, neither Alice nor Bob have any control over the intermediate polarization τ. This changes, however, if we consider the quantum equivalent where Alice and Bob perform the same experiment on a single photon instead of a beam (that can be classically described).

5.3.2 Price Argument: quantum

To simulate the quantum equivalent of the full experiment described above, we consider for each run of the experiment just a single photon instead of a beam of photons and we make the assumption of *discreteness*: the single photon enters and leaves the polarizers only on one channel or other, no superpositions are allowed. As Price mentions, the factors

$\cos^2(\tau_R - \sigma_R)$ and $\cos^2(\tau_L - \sigma_L)$ now represent probabilities rather than intensities. For example, a photon with polarization τ_R travelling toward the right-hand polarizer has a probability $\cos^2(\tau_R - \sigma_R)$ of being detected on the R1 output and a probability of $\sin^2(\tau_R - \sigma_R)$ of being detected on the R0 output. The interpretation of the probabilities for the left-hand polarizer is a little more subtle [106, p. 77]: we can only state, for example, that a photon emerging from the left-hand polarizer with polarization τ_L has a probability $\cos^2(\tau_L - \sigma_L)$ of having entered the polarizer from the L1 channel if we explicitly state that we have no knowledge of which input channel is chosen and thus treat both options L0 and L1 with equal probability (ie. we avoid the cases where the photon source lies only on the L1 channel and our probability factors do not make any sense).

Interestingly, Alice now does have some control over the intermediate polarization of the photon when it exits the left-hand side polarizer. For if Alice chooses an arbitrary setting σ_L and the Demon is obliged to present her with a photon that passes through the polarizer, the Demon's only choices are to input either a photon via path L1, which will then emerge from the polarizer with polarization τ_L, or input a photon with via path L0, in which case it will emerge with polarization $\tau_L + \pi/2$. Alice therefore has control of τ_L by controlling the setting σ_L, forcing it to being either $\tau_L = \sigma_L$ or $\tau_L = \sigma_L + \pi/2$. This control is a direct result of the assumption of *discreteness* in combination with *null-avoidance*. If we now consider the right-hand side of the experiment, and we want to enforce time-symmetry in the sense that τ_R relates to σ_R, R1 and R0 in exactly the same way as τ_L relates to σ_L, L1 and L0 as described above, then Bob must have the same control over τ_R by adjusting σ_R as Alice does over τ_L by adjusting σ_L. Here we find our retrocausality: Bob should have some retrocausal control over τ_R by only shifting σ_R. However, we have made two important assumptions in this last step: (i) we assume what Price calls *realism*, which means that both the photon exiting the left-hand side polarizer with polarization τ_L is a genuine element of reality—it is a local beable—and similarly for the photon of the right-hand side polarizer with polarization τ_R (it remains to be discussed whether τ_L and τ_R are describing the same photon), and (ii) we consider only the photons that successfully pass the left-hand side polarizer and, similarly for the right-hand polarizer, we demand that τ_R is such that the photon always passes the polarizer to either R1 or R0 (which always happens); no null results are allowed on either side. As said, I call this last assumption *null-avoidance*, and I stress here that it is often only assumed *en passant* by Price and others.

It is important to note that, analogously to the non-locality-without-signaling of the standard EPR-type experiment, the retrocausality in this experiment cannot be used to send signals from one polarizer to the other (cf. [107, p. 13]). Both Alice and Bob have some control over their respective photon's polarization, but not enough to send signals with these polarizations. If we run the experiment from left to right, for example, the polarization of the incoming photon can always vary by an angle of $\pi/2$ from what we would want it to be.

5.3.3 Discussion

To summarize the quantum case, we have considered a single photon undergoing an experiment as presented in Figure 9 from left to right. We first assumed *discreteness*: the single photon enters and leaves the polarizers only on one channel or other, no superpositions are allowed. We then assumed *realism*: the photon exiting the left-hand side polarizer with polarization τ_L is a local beable, and so is the photon leaving the right-hand side polarizer with polarization τ_R. Then, *time-symmetry*: τ_R relates to σ_R, R1 and R0 in exactly the same way as τ_L relates to σ_L, L1 and L0. More informally this assumption implies that the experiment should be exactly the same when performed from right to left. And, finally, *null-avoidance*: we pre-select and post-select the cases where the settings σ_L and σ_R and photon polarizations τ_L and τ_R are such that the photon can successfully pass both polarizers in either time direction. The conjunction of these four assumption thus implies retrocausality, and, when compared to the classical case, it seems that *discreteness* is doing the most work. Because Price only explicitly mentions the first three assumptions, he naturally only discusses the options for dropping any of these.

Dropping *realism* appears most undesirable. It is most clearly an option for the instrumentalist about quantum mechanics, but this position is rather uninteresting in the context of this thesis. There is, however, a more subtle way of dropping *realism*. The natural interpretation of *realism* is that both τ_L and τ_R describe Bell-type beables restricted to their own bounded space-time region, albeit with τ_L travelling forwards in time and τ_R travelling backwards in time. This is the ontology suggested by the Two-Time Interpretation of the Two State Vector Formalism [3] (about which it is rather convincingly argued that it does not solve the measurement problem [113]). One could object thereto that this is not ontologically parsimonious, and instead propose a single beable, say τ_L, that 'knows' both settings σ_L and σ_R [106, footnote 13]. I note here that this is essentially an all-at-once proposal and I repeat Adlam's remark from Section 4.2 that using the terminology of local Bellian *beables* in such an all-at-once context appears too artificial. But if we insist, then we face the problem that (dynamical) causality—and, consequently, retrocausality—loses its meaning in an all-at-once context.

Now consider dropping *discreteness*, which is the assumption that is apparently doing the most heavy lifting in the Price Argument. We have seen that, if we allow photons to input partially through both channels L1 and L0, neither Alice nor Bob have any control over the intermediate polarization τ and there is no argument for retrocausality. In the classical experiment, this unproblematically amounts to inputting beams with different intensities and polarizations through both channels. In the quantum case, it would amount to inputting a superposition of the photon being in both channels. Price mentions that this is possible for proponents of interpretations of quantum mechanics with an ontological view of the wave function (without objective collapse)—Bohmian mechanics again being the obvious candidate in the context of this thesis. Since Bohmian mechanics has a non-collapsing wave function (and yet manages to produce discrete measurement results), it naturally violates the assumption of *discreteness* and so the Price Argument for retrocausality does not go through. We may note that, as Price stresses, the fact that this argument doesn't work does not imply that there can be no retrocausality in Bohmian mechanics. Indeed, a toy model by Goldstein and Tumulka suggests that 'Parisian zigzag' retrocausality in a Bohmian framework can provide a Lorentz-invariant explanation of the EPR-type experiment, albeit with an apparently different purpose than Price (but the difference is not quite clear to me) [64, p. 561]: "Whereas Price seeks to exploit backwards causation to avoid non-locality, we use it to *achieve* non-locality in a Lorentz invariant way." It seems that the Bohmian might retain Lorentz-invariance by allowing retrocausality, but that (spacelike) non-locality will remain manifest in Bohmian mechanics (which is unsurprising considering it is built into the dynamics). The real trouble for the Price Argument here is that dropping *discreteness* actually amounts to adopting a realist interpretation of quantum mechanics with a non-collapsing wavefunction. This is a very natural option, and so the Price Argument becomes a bit of a moot point. We could even interpret the entire Price Argument upside down: because collapsing-wavefunction theories are covered by the Price Argument, and this would then imply retrocausality, this can be interpreted as an argument against collapsing-wavefunction theories for the majority of physicists that don't like retrocausality. But it is unlikely that this is what Price had in mind with his argument.

There is another problem with the assumptions of *discreteness*, however. I just stated that violating *discreteness* in the case of classical electromagnetism unproblematically amounts to inputting beams with different intensities and polarizations through both channels, just like allowing the input of a superposition for the quantum case. But if the Price Argument goes through until now, and this assumption of *discreteness* is necessary, then the argument would also go through for the classical case in addition to the assumption that we only input our beam of photons through one channel. But this is quite suspicious: the Price Argument is now an argument for retrocausality in classical electromagnetism, but only given that we input our beam of photons through one channel, and the retrocausality disappears completely when we allow input over both channels. Surely an argument for a type of causality should not depend fundamentally on the direction from which your measurement object is sent towards your measurement set-up. It appears that the argument for retrocausality in quantum mechanics is more pressing because in quantum mechanics the assumption of discreteness is much more natural and appears to be forced upon us. But, as I just argued above, this is not the case for quantum mechanics either, because the assumption of *discreteness* is naturally violated by an interpretation of quantum mechanics with a non-collapsing wavefunction.

Moving on, violating *time-symmetry* is perhaps the most natural option to avoid the conclusion of retrocausality because the title of Price's original paper 'Does time-symmetry imply retrocausality? How the quantum world says "Maybe"?' leads one to think that time-symmetry is doing the heavy lifting in this thought experiment. But there are at least two different ways to drop time-symmetry in the Price Argument: (i) adopt an interpretation of quantum mechanics with dynamical asymmetry in the theory itself, or (ii) remove the Demon that we had introduced to enforce the time-symmetry of the experiment (which essentially amounts to violating *null-avoidance*). Option (i) applies to the Copenhagen interpretation, where the time-symmetry of the dynamics of the Schrödinger equation is broken by a measurement,[52] and to collapse theories, where the dynamics are inherently time-asymmetric because evolution of the Schrödinger equation is made non-unitary. Option (ii), removing *null-avoidance* requires some more thought.

In order to enforce one aspect of the time-symmetry of the experiment, Price introduced a heuristic Demon imposed with the task of providing useful photons to the left-hand polarizer. I called this *null-avoidance*: we consider only those cases where a photon successfully passes through the left-hand polarizer via either L0 or L1 just like the photon reaching the right-hand polarizer will always be either transmitted via R1 or reflected via R0. So, let us then drop this assumption for a moment, and consider a photon source that inputs a photon with random polarization τ_{in} via either path L0 or L1 with equal probability. If Alice sets her polarizer to arbitrary setting σ_L, then we know that only half of the incoming photons will successfully pass through Alice's polarizer.

[52] Although, admittedly, the Copenhagen interpretation is not a consistent realist interpretation of quantum mechanics and its adherents would not care much for the Price Argument to begin with.

To see this, consider a photon coming in via L1 with polarization τ_{in}. The probability of the photon transmitting are $P_{L1_{result}} = \cos^2(\tau_{in} - \sigma_L)$ and the probability of the photon not transmitting are $P_{L1_{null}} = \sin^2(\tau_{in} - \sigma_L)$. Likewise for a photon coming in via L0 with polarization τ_{in}: $P_{L0_{result}} = \sin^2(\tau_{in} - \sigma_L)$ and $P_{L0_{null}} = \cos^2(\tau_{in} - \sigma_L)$. Properly normalized, the total probability for Alice having a photon successfully pass her polarizer, these add up to $P_{result} = P_{null} = 1/2 * (\cos^2(\tau_{in} - \sigma_L) + \sin^2(\tau_{in} - \sigma_L)) = 1/2$. Here, Alice has exactly the same control over the intermediate polarization as before: by switching her polarization direction σ_L she controls the intermediate polarization $\tau_L = \sigma_L$ modulo $\pi/2$ for half of the cases, while for the other half she obtains a null result.

We could interpret this in two ways. We could perhaps insist that the photons arriving at the left-hand side detector already had a well-defined polarization (described by some hidden variables). In this case, all control that Alice has (for the same experiment repeated many times) is simply selecting the ensemble of experiments that just so happened to involve photons with polarizations $\tau_{in} = \sigma_L$ modulo $\pi/2$. But now this control is fully due to pre-selection, and Alice has no no causal influence over the photons' intermediate polarizations. If we then again enforce time-symmetry in the sense that τ_R relates to σ_R, R1 and R0 in exactly the same way as τ_L relates to σ_L, L1 and L0 as described above, then Bob must have the same control over τ_R by adjusting σ_R as Alice does over τ_L by adjusting σ_L. But Alice's control is only due to pre-selection and therefore Bob's result is only due to post-selection. And now both Alice's and Bob's control is fully due to *null-avoidance* and there is nothing retrocausal about it.

If we acknowledge, on the other hand, that the photons did not have a well-defined polarization before entering Alice's polarizer, then Alice does indeed appear to have some causal influence over the intermediate polarization. And, therefore, Bob should have some retrocausal influence over this intermediate polarization. That is, of course, if we do not accept any of the multiple ways to naturally violate the assumptions of the Price Argument that I discussed above. And still there is now only retrocausality for half of the runs of the experiment, while for the other half Alice (and Bob) will obtain a null result. In my opinion, the case for retrocausality in quantum mechanics is not very strong.

And even if the case for retrocausality was strong, recall that I mentioned in Section 4 how Adlam argues that thinking through all the way the resulting balancing act of forwards and backwards contiguous causal influences naturally results in a theory where the future solves the past and the past solves the future simultaneously. Such a theory has massive interdependence throughout all regions of spacetime and its underlying (causal (?)) structure is thus really temporally non-local. Thus we arrive at all-at-once. But, although retrocausal theories might have been the first path of arriving at all-at-once, it does not seem necessary to arrive at all-at-once theories via consideration of retrocausal theories. Indeed, we do not need to arrive at all-at-once via considerations about causality at all but we can, in hindsight, instead arrive at all-at-once via considerations about the way we usually do physics and how there might be alternatives thereto.

5.4 Proposal 3: All-at-once

In this thesis, I discussed many different subjects relating to the EPR-type experiment. There are very good reasons to take the spacelike non-locality suggested by violations of Bell's inequalities very seriously, but there are also possibilities for avoiding such conclusions. One such possibility is violation of Source Independence. I discussed two different options, superdeterminism and retrocausality, for violating Source Independence in the previous sections and I argued that the arguments for neither are particularly convincing. But there is another option, which I alluded to several times throughout this thesis: Wharton's all-at-once approach, which takes both spacelike non-locality *and* the violation of Source Independence seriously. We have seen in Section 4 that Adlam strongly advocates taking this option seriously because it is a good candidate for a time-like non-local theory. In this section, I briefly discuss the intriguing proposal by Wharton [130, 131, 132, 133, 134, 135, 136]. Wharton's proposal reaches so deep into every corner of the foundations of physics, however, that any in-depth discussion and critical analysis will immediately take us far away from the context of the EPR-type thought experiment and therefore outside the scope of this thesis. I am not qualified to critically assess this proposal, but I do feel that it is worthy of more attention that it currently gets. Therefore, in this section I aim to present the central ideas of an all-at-once interpretation and to indicate how an all-at-once interpretation apparently ties together many of the concepts that I discussed in this thesis, specifically the violation of Source Independence, the distinction between ontic and epistemic probabilities, and the relation between causality and non-locality. But first, we must take a step back and reconsider the way we usually do physics.

5.4.1 Newtonian and Lagrangian schemas

Wharton often begins his exposition of all-at-once by presenting two different schemas for doing physics and doing physics to interpret the world: the Newtonian schema and the Lagrangian schema.[53] The Newtonian schema can be characterized as follows [132, p. 191]: "we first map our knowledge of the physical world onto some mathematical state, then use dynamical laws to transform that state into a new state, and finally map the resulting (computed) state back onto the physical world." And then we assume that this is also how the universe 'works': it takes the 'present' state of the world and uses dynamical laws to calculate what will come after, and so drive itself through time by generating the future one moment at a time. Indeed, this (Newtonian) schema is so standard for the practice of physics and it is so intuitive that only recently have we begun to realize that the assumption that the universe acts according to the way we humans perceive it and calculate it is inherently anthropocentric. As Wharton states [133, p. 187]: "Now there's one last anthropocentric attitude that needs to go, the idea that the computations we perform are the same computations performed by the universe, the idea that the universe is as 'in the dark' about the future as we are ourselves." So what is the alternative?

The alternative is the Lagrangian schema, and we have already come across it in Section 5.1 in the form of the *Action symmetry* principle. Using a Lagrangian formalism is common practice in many fields of physics (ie. Feynman's path integral formulation of quantum mechanics and Fermat's principle), but it is usually considered as being merely a useful alternative to the 'actual' dynamical description of the world according to the Newtonian schema. In the Lagrangian schema, we again map our knowledge of the physical world onto some mathematical state. But instead of specifying only one instance of time and then applying dynamical laws, we specify the spacetime parameters (the boundary conditions) both at the beginning and the end of the spacetime region in question, and we then apply some extremization principle to solve the intermediate spacetime *all at once*. In doing so, the values of all parameters (both the boundary and the intermediate parameters) become statistically dependent on each other. You cannot

[53] The name 'Newtonian Schema' comes from Smolin [118].

change the value of one parameter without that change affecting the entire spacetime region in question. The resulting picture of the world is that of a static block universe where all parameters are interconnected, and where the dynamical asymmetric-in-time description thereof is merely an approximation.[54]

5.4.2 Probability and the Independence Fallacy

To use the Lagrangian formalism as an alternative to a dynamical formalism is one thing, but to consider it as perhaps more fundamental or more in accordance with the actual physical world than the Newtonian dynamical description is another. It immediately conflicts with many of our deepest intuitions about the world. In a Lagrangian world, for example, there is no dynamics, nor is there a fundamental flow of time. Another such intuition discussed by Wharton that is particularly relevant in the context of this thesis is the Independence Fallacy [132, p.197]: the idea that probabilities ascribed to a subsystem be independent of external geometry. More specifically, the idea that probabilities ascribed to a past or current state of the world are independent of future states.

From a standard dynamical point of view where the current state of the world generates the future state of the world, the future state is, in some sense, obviously dependent on the present state but the present state is independent of the future state. This is precisely what Bell aimed to capture with his definition of local causality: the probability of a certain measurement result is determined only by the parameters in its past light cone. Likewise, as described by the assumption of Source Independence in the context of the EPR-type experiment, the distribution of the parameter λ is independent of the future measurement settings a and b. This assumption has not been recognized as particularly problematic before the advent of quantum mechanics, but in this thesis I already discussed serious reasons for violating it. The two previously discussed options for violating the assumption of Source Independence did so dynamically: superdeterminism uses a common

[54] There is an interesting parallel between this view of physics and an existing discossion in mathematics, beautifully put into words by Shapiro in his *Philosophy of Mathematics* which I quote here to stimulate our intuition about this subject [115, p. 181]

> I mentioned a gap between the practice of mathematics and its current philosophical and semantic formulations. Mathematicians speak and write as if dynamic operations and constructions are being performed: they draw lines, they move figures, they make choices, they apply functions, they form sets. Taken literally, this language presupposes that mathematicians envision creating their objects, moving them around, and transforming them. This manner of speaking goes back to antiquity. Euclid's Elements contains statements that express the capabilities of mathematicians to effect geometrical construction or, in other words, the potentialities of geometrical objects to be created or affected by mathematicians. One of the postulates is "Given any two points, to draw a straight line between them." Taken literally, assertions like this are statements of permission, or of what combinations of moves are possible.
>
> In contrast to the dynamic picture, the traditional Platonist holds that the subject matter of mathematics is an independent, static realm. Accordingly, the practice of mathematics does not change the universe of mathematics. In a deep, metaphysical sense, the universe cannot be affected by operations, constructions, or any other human activity, because the mathematical realm is eternal and immutable. There can be no permission to operate on such a domain.
>
> To belabor the obvious, then, the traditional Platonist does not take dynamic language literally. Euclid's Elements also contains static language, statements about existing geometrical objects. One postulate is "All right angles are equal." For the Platonist, this is the philosophically correct way to speak.

Perhaps, if the fundamental relation between the Newtonian and the Lagrangian schema gains more widespread attention, we may draw inspiration from this similar discussion in mathematics that has been going on since the inception of mathematics.

cause in the distant past of all three parameters and retrocausality uses a dynamical account of causality coming from the future measurement settings back to λ.

But according to the all-at-once approach, there is a fact of the matter what the future measurement settings are, and it would be a serious mistake to insist that parameters λ are independent of those. But this changes everything about the way we can interpret the probability relations that I discussed throughout this thesis. Especially when we try to understand specifically in what way the states (and their probabilities) are dependent upon each other. When comparing Shimony's independence relations to Maudlin's in Section 3.2, I was looking for probabilities that can be interpreted ontically. That is to say, from our dynamical perspective, a change in these probabilities represents an actual change in the world. In the EPR-type experiment, we go as far as to say that a change in probabilities represents a certain non-local causality: updating our joint probability for measurement results A and B with, for example, Alice's measurement setting or result changes the probabilities we ascribe to Bob's possible results. We say that there might be some genuine non-local causality going on in such cases, at least in the case where Shimony's Parameter Independence is violated. This is the violation of Bell's local causality, this is spacelike non-locality.

According to Wharton, from an all-at-once perspective, however, considering that the entirety of spacetime (including the future) is already set in stone, all such probabilities merely represent our incomplete knowledge of the future measurement settings and results. That is to say, all probabilities that we can calculate are epistemic probabilities. In fact, this ties in nicely with Adlam's conviction that there is no such thing as 'objective chance', and one may suspect that the personal dismissal of objective chance and the openness to consider an all-at-once interpretation are two sides of the same coin. From this perspective, there are no objectively chancey events in the world because there is only one actual world and it is already a fact of the matter what the world looks like throughout its entire spacetime. All probabilities are epistemic probabilities. To illustrate this point further, compare again Maudlin's independence assumptions to Shimony's. I argued that especially Maudlin's $P(A|B, a, \lambda)$ is suspicious because we have to average over all possible settings b in order to evaluate it—a move I interpreted as one motivated by ignorance of a distant setting. But now we are not much better off with any other probability function that gives us non-trivial probabilities. The resulting probabilities change when we first evaluate Shimony's $P(A|a, \lambda)$, then add the information about the distant setting with $P(A|a, b, \lambda)$, and then the result with $P(A|B, a, b, \lambda)$, but this tells us nothing about a possibly non-local causal link between any of these parameters but it merely tells us that we are updating our knowledge of the total measurement situation until we reach full knowledge when we obtain a trivial probability.

Similarly, the distribution of the hidden variables $\rho(\lambda|a, b)$ cannot simply be assumed to be independent of the future measurement settings a and b. We cannot assume that Source Independence holds because this is precisely the Independence Fallacy. Bell's theorem has ruled out that the addition of any local hidden variables will make a locally causal explanation of the correlations in an EPR-type experiment possible again. We have already seen two loopholes to this conclusion with superdeterminism and retrocausality. But the all-at-once interpretation does not contradict Bell's theorem because the hidden variables are not local beables in the sense of Bell anymore: by construction, in an all-at-once interpretation every parameter is non-locally dependent on the values of all other parameters in the spacetime region in question, and so to insist that a parameter that can be assigned to a bounded spacetime region is a local beable in this context would be to focus on the wrong aspect of this parameter. The value of the parameter is non-locally dependent on the values of all other relevant parameters of the spacetime region that you want to describe. But it is not a familiar dynamical asymmetric-in-time dependence, and so it is not quite a causal dependence in the usual sense—certainly not in the sense intended by Shimony. This then raises the question whether the Bellian framework of local causality is particularly suited for capturing the essence of the non-locality present in all-at-once. But before I discuss the relation between Bellian local causality and non-locality in all-at-once, I need to discuss an alternative way of interpreting probabilities in all-at-once.[55]

If we just consider the probabilities in all-at-once theories to about our knowledge of the actual world, then we could say that there are only epistemic probabilities in all-at-once. But then there isn't anything lawlike about these probabilities either that we can connect to the world, which seems undesirable because a main goal of physics is generally about finding lawlike descriptions of possibilities, not just about describing the actual world. So we could interpret the probabilities in all-at-once as probabilities over possible worlds, with the events kept fixed with respect to *all* boundary conditions. Then these same probabilities I have used in my examples above can be interpreted as describing some counterfactual, lawlike dependence over possible worlds. In a sense, then, they are ontic probabilities. But certainly not in the same sense as the way I interpreted the probabilities of Shimony's independence assumptions in Section 3.2, which described a dynamically causal dependence that does not depend on the semantics of possible worlds. And introducing counterfactuals and possible worlds into our discussion opens up a Pandora's box that is better left closed for now. The nature of probabilities in all-at-once is an intriguing subject that deserves to be studied further. But the EPR-type model and Bell's conceptual framework are certainly not the right tools for this task, and so it lies beyond the scope of this thesis.

5.4.3 Causality and non-locality

I just discussed two possible ways in which we may interpret probabilities in all-at-once. Either probabilities are all epistemic, in which case we cannot claim in any way that we are describing lawlike dependencies or dynamical causal relations between different events in the actual world. Or, alternatively, probabilities describe lawlike dependencies between events in possible worlds, in which case we may interpret the probabilities as ontic (in some sense). But neither interpretation of probabilities is particularly useful in the context of Bell's conceptual framework of local beables and local causality that I used in this thesis. By construction, the localized values of parameters are non-locally dependent on every other parameter, and so it makes little sense to speak of these parameters as local beables. Likewise, it makes little sense to speak about local causality in a theory where there is no dynamics at all. This then begs the question whether calling an all-at-once theory 'non-local' in the sense of Bell is not just misleading, as there are very few aspects of Bell's conceptual framework that apply to all-at-once theories. Moreover, if we consider Adlam's timelike locality defined analogously to Bell's spacelike locality introduced in Section 4, then all-at-once appears to be the only candidate that violates this definition. But it does so by construction again, and it appears to trivial violate all Shimonian independence relations that we may formulate analogously to the spacelike independence relations. If we then also consider that, according to Wharton, a more interesting aspect of all-at-

[55]Suggested to me by Bacciagaluppi.

once is that it identifies the Independence Fallacy as a fallacy—that is, it violates Source Independence—and we recognize that Adlam's definition of timelike locality is not well-suited to describe the failure of Source Independence, we should be tempted to conclude that Adlam's definition of timelike locality defined analogously to Bell's definition of spacelike locality is not as relevant for all-at-once theories as we might want it to. This, then, again weakens the analogy between spacelike and timelike locality.

Of course, to enforce these intuitions described above, we must study all-at-once in the context of actual theories or toy models. For examples, several possibilities for such models that should be analyzed are introduced in Almada et al.'s [5] and Wharton's [135]. In this thesis, however, I have limited myself mostly to conceptual analysis of the analogy between spacelike and timelike non-locality in the context of Bell's conceptual framework. And I have repeatedly argued that the analogy is much weaker than might be apparent from the formal similarities between contemporary approaches to spacelike and timelike non-locality. In whatever way we choose to study explicit all-at-once theories and toy models, we must be hesitant to use Bell's conceptual framework for this analysis because superficial similarities are likely just misleading.

Adlam has made the intriguing point that timelike locality is such a fundamental and unquestioned assumption in physics that it is in serious danger of becoming a dogma. Wharton has made the equally intriguing point that the apparent self-evidence of the Newtonian schema is equally dangerous. But a solution to either or both problems cannot quite be accurately captured in the conceptual framework of Bell. The framework of Bell was meant to describe non-locality in space, and it needs to necessarily assume locality in time to do so. It has helped us to identify many different dynamical interdependencies between the different parameters of the EPR-type experiment over the past couple of decades. But the application of Bell's framework to the analysis of timelike non-locality does not appear equally promising. Loopholes to Bell's theorem that are possible candidates for being timelike non-local such as retrocausal theories do not violate any of the timelike independence relations that we formalized analogously to the spacelike independence relations. The all-at-once interpretation, on the other hand, appears to violate by construction every spacelike and timelike independence relation that we can think of. Moreover, it seems that we cannot conclude any dynamical 'non-local causality' in the sense of Bell from these violations because there is no such thing as dynamics in an all-at-once description of the world. The concept of timelike non-locality and the all-at-once interpretation are highly intriguing and they should be analyzed much further. But to do so using Bell's conceptual framework, we must fundamentally reinterpret many of the concepts that are fundamental to the framework, such as the interpretation of probabilities, dynamical causality, local beables, and local causality. There will then be not much left of Bell's conceptual framework, and the formal similarities between spacelike and timelike locality in the context of an EPR-type experiment will just become increasingly misleading. Timelike non-locality and the all-at-once interpretation should be studied extensively, but the conceptual framework of Bell is not suited for this task.

We have now reached the end of this thesis. It is time to briefly revisit the claims I made throughout this thesis, and to draw some conclusions.

6 Conclusion

In this thesis, I analyzed the concepts of spacelike and timelike non-locality in the context of Bell's conceptual framework and the EPR-type thought experiment. The main aim of this thesis was to shine a well-deserved light on some persistent and some novel loopholes to Bell's theorem, to evaluate how these should be interpreted in the context of timelike non-locality, and to assess whether the analogy between spacelike and timelike non-locality is strong.

In Section 2, I first presented a brief history of locality. I then introduced Bell's conceptual framework and the EPR-type thought experiment, after which I discussed several reasons for treating Bell's framework with caution. Most of these reasons were connected to the ambiguity in Bell's probabilistic formalism: there is, for example, no reference to space and time in the conventional formalism, which opens up the door for ambiguities when one applies the formalism in a timelike context instead of the traditional spacelike context. I argued that one can justify using Bell's formalism as long as one holds these reasons for caution in mind—this is not done often enough.

Then, in Section 3, I emphasized some of the basic assumptions that are necessary for formulating and interpreting Bell's theorem. I highlighted the assumptions of *Independent boundary conditions* and *Model completeness* as especially important, and we have seen these return throughout this thesis. Then, I introduced the contemporary subdivisions of Bell's factorized equation into several independence assumptions à la Jarrett, Shimony, and Maudlin. I argued that the probabilities from the Shimonian subdivision into Parameter Independence and Outcome Independence can be interpreted as ontic probabilities: they tell us something about the world, and from them might we hope to conclude non-local causality. I argued especially against Maudlin's alternative set of independence assumptions by demonstrating that the probabilities in these assumptions can only reasonably be interpreted as epistemic probabilities: they merely represent ignorance of distant events (measurement settings, mostly) and can not assist us in the search for genuine causal links.

I then introduced Adlam's recently proposed definition of timelike locality analogously to Bell's definition of spacelike locality in terms of a factorizability equation, for which I used a timelike variant of Bell's light-cone drawings. I argued that this definition of timelike locality might not be as useful as it appears on first sight because, firstly, it cannot adequately capture the failure of Source Independence that is the violation of timelike locality in a *spacelike* EPR-type experiment, and secondly, nor can the Shimonian subdivision be applied to timelike locality with the same success as it has for spacelike locality. Indeed, it appears that we will run into a triviality problem: either all Shimonian independence assumptions are violated or none are violated. I then went on to discuss some of Adlam's motivations for taking the concept of timelike non-locality very seriously. I discussed the viability of non-Markovian theories, and I concluded that there is an increasingly good case to be made for dropping the Markov property. Then, I distinguished the concept of *timelike* non-locality from two types of more familiar *temporal* Bell-type inequalities: Leggett-Garg inequalities and 'entanglement-in-time' inequalities. Contrary to Adlam, I acknowledged that there might be reasons for concluding genuine timelike non-locality from certain violations of Leggett-Garg inequalities, even when we take locally causal communication between consecutive measurements (the communication loophole) into account. Finally, I discussed the result that Adlam considers to be among the first concrete results in favor of timelike non-locality: it appears that, for certain EPR-type

(GHZ) measurement set-ups and given certain conditions, more information is transmitted than is allowed classically. Here we find a genuine possibility for timelike non-locality in the sense that the state of the world at one time *directly* depends on the state of the world at another time, without this dependence being mediated contiguously through intermediate states of the world. I concluded that, although there are good reasons for taking the concept of timelike non-locality seriously, Bell's conceptual framework and probabilistic formalism might not be the most useful tool for further analysis of this concept. On the contrary, in Bell's framework, the formal similarities between spacelike and timelike non-locality might just be misleading.

In Section 5, the final section of this thesis, I critically evaluated a recent argument by Evans, Price, and Wharton in favor of a direct timelike analogue of the EPR-type experiment. I argued that the symmetry arguments that were employed to transpose the familiar spacelike EPR-type experiment into a timelike analogue thereof were problematic, and that the construction of the spacelike and timelike experiments certainly is not analogous. Then, I discussed some persistent and some novel loopholes to Bell's theorem, all of which violate Source Independence (or, more broadly, Independent boundary conditions). I argued against superdeterminism by arguing that it violates my assumption of *Model completeness*. I argued against the Price Argument for retrocausality in quantum mechanics because every single one of the assumptions is naturally violated by the standard interpretations of quantum mechanics that I introduced in the introduction to this thesis. Moreover, I argued that, if one does insist on the Price Argument, either the scope of the argument is too broad and it applies to classical electromagnetism too or the scope of the argument is so limited (and applies only to half of the runs of the experiment) that the argument is hardly significant. Finally, I discussed the all-at-once interpretation recently introduced by Wharton. This is a highly intriguing proposal and it deserves more attention. It is the only interpretation that might be genuinely timelike non-local according to Adlam's definition of timelike locality. But the proposal is so radical and so fundamentally different from the standard practice of physics, that many of the concepts at play in Bell's conceptual framework lose their meaning: the interpretation of probability, local beables, local causality and, consequently, non-locality in general. I argued, therefore, that Bell's conceptual framework is not particularly useful for the only theory that might have been genuinely timelike non-local when timelike locality is defined analogously to spacelike locality. I now feel justified in concluding that spacelike locality and timelike locality is strongly disanalogous, and that many of the similarities between the two concepts (not in the least the formal similarity between their definitions) are only superficial and, in many cases, certainly misleading.

But the last word on these matters has certainly not been said, and there are many possibilities for future research, especially with respect to the concept of timelike non-locality and all-at-once interpretations. The philosophical concepts at play here need to be fleshed out much further, and they must be studied in the context of concrete theories or at least in toy models, not just on a conceptual level. These ideas are so admirably radical yet justified to the extent that they deserve more attention from the general philosophy of science community. In the words of Adlam, the dogmatic assumption of timelike locality might be actively limiting progress in the field of quantum foundations. But the field of quantum foundations is a minefield, and the apparent analogy between spacelike and timelike locality is dangerous. In this thesis, I aimed to substantiate the claim that spacelike and timelike non-locality are strongly disanalogous. In doing so, I hope I have contributed to genuine progress in the field of quantum foundations.

A Derivation of the CHSH-inequality

For the same EPR-type experiment repeated many times, define the expectation value for the product of results A and B as:

$$E(a,b) := \int_\Lambda \sum_{A,B} AB \cdot P(A,B|a,b,\lambda)\rho(\lambda|a,b)d\lambda. \tag{A.1}$$

Invoke SI (3.1) to obtain:

$$E(a,b) = \int_\Lambda \sum_{A,B} AB \cdot P(A,B|a,b,\lambda)\rho(\lambda)d\lambda. \tag{A.2}$$

Now invoke Factorizability (3.12) to obtain:

$$\begin{aligned} E(a,b) &= \int_\Lambda \sum_{A,B} AB \cdot P(A|a,\lambda)P(B|b,\lambda)\rho(\lambda)d\lambda \\ &= \int_\Lambda \Big(\sum_A A \cdot P(A|a,\lambda)\Big)\Big(\sum_B B \cdot P(B|b,\lambda)\Big)\rho(\lambda)d\lambda. \end{aligned} \tag{A.3}$$

Redefine $\overline{E}(a,\lambda) := \sum_A A \cdot P(A|a,\lambda)$ and $\overline{E}(b,\lambda) := \sum_B B \cdot P(B|b,\lambda)$ as the average values of A and B, to write:

$$E(a,b) = \int_\Lambda \overline{E}(a,\lambda)\overline{E}(b,\lambda)\rho(\lambda)d\lambda. \tag{A.4}$$

Because $A = \pm 1$ and $B = \pm 1$, their average values must obey $|\overline{E}(a,\lambda)| \leq 1$ and $|\overline{E}(b,\lambda)| \leq 1$. Now consider the combinations of expectation values for settings a, b, and alternative settings a', b':

$$\begin{aligned} E(a,b) + E(a,b') &= \int_\Lambda \Big(\overline{E}(a,\lambda)\overline{E}(b,\lambda) + \overline{E}(a,\lambda)\overline{E}(b',\lambda)\Big)\rho(\lambda)d\lambda \\ &= \int_\Lambda \overline{E}(a,\lambda)\Big(\overline{E}(b,\lambda) + \overline{E}(b',\lambda)\Big)\rho(\lambda)d\lambda. \end{aligned} \tag{A.5}$$

Since $|\overline{E}(a,\lambda)| \leq 1$, we get:

$$|E(a,b) + E(a,b')| \leq \int_\Lambda |\overline{E}(b,\lambda) + \overline{E}(b',\lambda)|\rho(\lambda)d\lambda, \tag{A.6}$$

and, similarly:

$$|E(a',b) - E(a',b')| \leq \int_\Lambda |\overline{E}(b,\lambda) - \overline{E}(b',\lambda)|\rho(\lambda)d\lambda. \tag{A.7}$$

Now we add equations (A.6) and (A.7):

$$\begin{aligned} &|E(a,b) + E(a,b')| + |E(a',b) - E(a',b')| \leq \\ &\int_\Lambda \Big(|\overline{E}(b,\lambda) - \overline{E}(b',\lambda)| + |\overline{E}(b,\lambda) + \overline{E}(b',\lambda)|\Big)\rho(\lambda)d\lambda. \end{aligned} \tag{A.8}$$

Note that $|x-y| + |x+y| \leq 2$ for $|x|, |x'|, |y|, |y'| \leq 1$ and that we had defined $\int_\Lambda \rho(\lambda)d\lambda = 1$. This gives:

$$|E(a,b) + E(a,b')| + |E(a',b) - E(a',b')| \leq 2. \tag{A.9}$$

Finally, use the triangle inequality $|x+y| \leq |x| + |y|$ to obtain the most familiar form of the CHSH-inequality:

$$|E(a,b) + E(a,b') + E(a',b) - E(a',b')| \leq 2. \tag{A.10}$$

B Derivation of Maudlin Factorizability

This derivation follows [114, §3.7]. Start with a general law of conditional probability:

$$P(A, B|a, b, \lambda) = P(A|a, b, B, \lambda)P(B|a, b, \lambda) = P(B|a, b, A, \lambda)P(A|a, b, \lambda). \tag{B.1}$$

Consider the second equality, and apply MOI and MPI to obtain:

$$P(A|a, \lambda)P(B|a, b, \lambda) = P(B|b, \lambda)P(A|a, b, \lambda), \tag{B.2}$$

and, assuming that $P(B|b, \lambda)$ and $P(A|a, \lambda)$ are non-zero, rewrite this as:

$$\frac{P(A|a, b, \lambda)}{P(A|a, \lambda)} = \frac{P(B|a, b, \lambda)}{P(B|b, \lambda)}, \tag{B.3}$$

which holds only if both sides equal 1 and thus if $P(A|a, b, \lambda) = P(A|a, \lambda)$ and $P(B|a, b, \lambda) = P(B|b, \lambda)$. That is, this only holds if PI holds! To demonstrate that this is indeed the case, consider the above derivation for another measurement outcome A' instead of A, which leads to:

$$\frac{P(A'|a, b, \lambda)}{P(A'|a, \lambda)} = \frac{P(B|a, b, \lambda)}{P(B|b, \lambda)}. \tag{B.4}$$

Combine (B.3) and (B.4) to obtain:

$$\frac{P(A|a, b, \lambda)}{P(A|a, \lambda)} = \frac{P(A'|a, b, \lambda)}{P(A'|b, \lambda)}. \tag{B.5}$$

Because we have only dichotomous measurement results A and A', we have:

$$P(A|a, b, \lambda) + P(A'|a, b, \lambda) = P(A|a, \lambda) + P(A'|a, \lambda) = 1. \tag{B.6}$$

Rewrite these as $P(A'|a, b, \lambda) = 1-P(A|a, b, \lambda)$ and $P(A'|a, \lambda) = 1-P(A|a, \lambda)$, and substitute into (B.5) to obtain:

$$\begin{gathered} \frac{P(A|a, b, \lambda)}{P(A|a, \lambda)} = \frac{1 - P(A|a, b, \lambda)}{1 - P(A|a, \lambda)}, \\ P(A|a, b, \lambda) - P(A|a, b, \lambda)P(A|a, \lambda) = P(A|a, \lambda) - P(A|a, \lambda)P(A|a, b, \lambda), \\ P(A|a, b, \lambda) = P(A|a, \lambda). \end{gathered} \tag{B.7}$$

So PI does indeed hold, and we may then rewrite (B.1) as:

$$P(A, B|a, b, \lambda) = P(A|a, \lambda)P(B|b, \lambda), \tag{B.8}$$

which is Factorizability (3.12). We can thus derive Factorizability from the conjunction of MOI, MPI, and the assumptions that $P(B|b, \lambda)$ and $P(A|a, \lambda)$ are non-zero.

References

[1] Adlam, E. A Tale of Two Anachronisms. Presented at the Foundations2018 conference at Utrecht University, 13 July 2018.

[2] Adlam, E. Spooky Action at a Temporal Distance. *Entropy 20*, 41 (2018).

[3] Aharonov, Y., and Gruss, E. Y. Two-time interpretation of quantum mechanics. *Eprint arXiv:quant-ph/0507269* (2005).

[4] Aharonov, Y., and Vaidman, L. The two-state vector formalism: an updated review. *Lecture Notes in Physics 734* (2008), 399–447.

[5] Almada, D., Ch'ng, K., Kintner, S., Morrison, B., and Wharton, K. B. Are Retrocausal Accounts of Entanglement Unnaturally Fine-Tuned? *International Journal of Quantum Foundations 2* (2015), 1–16.

[6] Ananthaswamy, A. *Through two doors at once: the elegant experiment that captures the enigma of our quantum reality.* New York: Dutton, 2018.

[7] Ariew, R., Ed. *G. W. Leibniz and Samuel Clarke: Correspondence.* Cambridge: Hackett Publishing Company, Inc., 2000.

[8] Bacciagaluppi, G. Collapse Theories as Beable Theories. *Manuscrito 33*, 1 (2010), 19–54.

[9] Bacciagaluppi, G. Probability and time symmetry in classical markov processes. In *Probabilities, Causes and Propensities in Physics*, M. Suárez, Ed. Springer Netherlands, Dordrecht, 2011, pp. 41–59.

[10] Bacciagaluppi, G. Leggett-Garg Inequalities, Pilot Waves and Contextuality. *Eprint arXiv:1409.4104v2* (2014).

[11] Bacciagaluppi, G. Did Bohr understand EPR? In *One Hundred Years of the Bohr Atom* (2015), F. Aaserud and H. Kragh, Eds., pp. 377–396.

[12] Bacciagaluppi, G. Lecture notes on Foundations of Quantum Mechanics 2017-2018. *Unpublished* (2017).

[13] Bacciagaluppi, G. The Quantum Mechanical Free Will Theorem. Presented at the Utrecht University Physics Department Day, 22 March 2018.

[14] Bacciagaluppi, G., and Valentini, A. *Quantum Theory at the Crossroads: Reconsidering the 1927 Solvay Conference.* Cambridge University Press, 2009.

[15] Bedingham, D., Dürr, D., Ghirardi, G. C., Goldstein, S., Tumulka, R., and Zanghì., N. Matter Density and Relativistic Models of Wave Function Collapse. *Journal of Statistical Physics 154* (2014), 623–631.

[16] Bell, J. S. *Speakable and unspeakable in quantum mechanics*, first ed. Cambridge University Press, 1987.

[17] Bell, J. S. Against 'measurement'. *Physics World 3*, 8 (1990), 33–40.

[18] Bell, J. S. La Nouvelle Cuisine. In *John S. Bell on the Foundations of Quantum Mechanics*, M. Bell, K. Gottfried, and M. Veltman, Eds. World Scientific Publishing Co., 2001, pp. 216–234.

[19] Blaylock, G. The EPR paradox, Bell's inequality, and the question of locality. *American Journal of Physics 78* (2009), 111–120.

[20] Bohm, D. *Quantum Theory.* New Jersey: Prentice-Hall, Inc., 1951.

[21] Bohm, D. A Suggested Interpretation of the Quantum Theory in Terms of 'Hidden' Variables, I and II. *Physical Review 85*, 2 (1952), 166–193.

[22] Bohr, N. Can Quantum-Mechanical Description of Reality Be Considered Complete? *Physical Review 48* (1935), 696–702.

[23] Bricmont, J. History of Quantum Mechanics or the Comedy of Errors. *Eprint arXiv:1703.00294v1* (2017).

[24] Brierley, S., Kosowski, A., Markiewicz, M., Paterek, T., and Przysiężna, A. Nonclassicality of Temporal relations. *Physical Review Letters 115*, 12 (2015), 120404.

[25] Brukner, C., Taylor, S., Cheung, S., and Vedral, V. Quantum Entanglement in Time. *Eprint arXiv:quant-ph/0402127* (2004).

[26] Butterfield, J. N. A space-time approach to the Bell Inequality. In *Philosophical consequences of quantum theory: reflections of Bell's theorem*, J. Cushing and E. McMullin, Eds. Notre Dame: University of Notre Dame Press, 1989, pp. 114–153.

[27] Butterfield, J. N. Bell's Theorem: What It Takes. *The British Journal for the Philosophy of Science 43*, 1 (1992), 41–83.

[28] Clauser, J. F., Horne, M. A., Shimony, A., and Holt, R. A. Proposed Experiment to Test Local Hidden-Variable Theories. *Physical Review Letters 23*, 15 (1969), 880–884.

[29] Clifton, R. K., Redhead, M. L. G., and Butterfield, J. N. Generalization of the Greenberger-Horne-Zeilinger algebraic proof of nonlocality. *Foundations of Physics 21*, 2 (1991), 149–184.

[30] Conway, J., and Kochen, S. The Free Will Theorem. *Foundations of Physics 36* (2006), 1441–1473.

[31] Conway, J., and Kochen, S. The Strong Free Will Theorem. *Notices of the American Mathematical Society 56* (2009), 226–232.

[32] Costa de Beauregard, O. Une réponse à l'argument dirigé par einstein, podolsky et rosen contre l'interprétation bohrienne des phénomènes quantiques. *Comptes rendus de l'Académie des Sciences 236* (1953), 1632–1634.

[33] Costa de Beauregard, O. Time Symmetry and Interpretation of Quantum Mechanics. *Foundations of Physics 6* (1976), 539–559.

[34] Cramer, J. G. The transactional interpretation of quantum mechanics. *Reviews of Modern Physics 58* (1986), 647–687.

[35] Cramer, J. G. The Transactional Interpretation of Quantum Mechanics and Quantum Nonlocality. *Eprint arXiv:1503.00039v1* (2015).

[36] Cushing, J., Fine, A., and Goldstein, S., Eds. *Bohmian Mechanics and Quantum Theory: An Appraisal.* Dordrecht: Springer, 1996.

[37] Daniel M. Greenberger, Michael A. Horne, A. Z. Going Beyond Bell's Theorem. In *Bell's Theorem, Quantum Theory, and Conceptions of the Universe*, M. Kafatos, Ed. Dordrecht: Kluwer, 1989, pp. 41–59.

[38] de Regt, H. W. Erwin Schrödinger, Anschaulichkeit, and quantum theory. *Studies in History and Philosophy of Modern Physics 28*, 4 (1997), 461–481.

[39] DeBrota, J. B., and Stacey, B. C. FAQBism. *Eprint arXiv:1810.13401v1* (2018).

[40] Derakhshani, M. *Stochastic Mechanics Without Ad Hoc Quantization: Theory and Applications to Semiclassical Gravity.* PhD thesis, Utrecht University, 2018.

[41] Dieks, D., Arnoldus, H., and Nienhuis, G. Sub-poissonian statistics as an experimental test for the contextuality of quantum theory. *Physics Letters 103A*, 1 (1984), 27–31.

[42] Durham, I. T. Bell's Theory of Beables and the Concept of 'Universe'. *Eprint arXiv:1805.02143v1* (2018).

[43] Dzhafarov, E. N., and Kujala, J. V. Probabilistic Contextuality in EPR/Bohm-type Systems with Signaling Allowed. *Eprint arXiv:1406.0243v3* (2014).

[44] Dzhafarov, E. N., and Kujala, J. V. Generalizing Bell-type and Leggett-Garg-type Inequalities to Systems with Signaling. *Eprint arXiv:1406.0243v8* (2015).

[45] Earman, J. Locality, Nonlocality, and Action at a Distance: A Skeptical Review of Some Philosophical Dogmas. In *Kelvin's Baltimore lectures and modern theoretical physics: historical and philosophical perspectives*, R. K. et al., Ed. Cambridge MA: MIT Press, 1987, pp. 449–490.

[46] Einstein. Autobiographical Notes. In *Albert Einstein: Philosopher-Scientist*, P. A. Schilpp, Ed. Evanston: Library of the Living Philosophers, 1949, pp. 1–94.

[47] Einstein, A. Quantum Mechanics and reality. *Dialectica 2* (1948), 320–324.

[48] Einstein, A., Podolsky, B., and Rosen, N. Can Quantum-Mechanical Description of Reality Be Considered Complete? *Physical Review 47* (1935), 777–780.

[49] Emary, C., Lambert, N., and Nori, F. Leggett–Garg inequalities. *Reports on Progress in Physics 77*, 1 (2014), 016001.

[50] Evans, P. W. Quantum Causal Models, Faithfulness, and Retrocausality. *The British Journal for the Philosophy of Science 69*, 3 (2017), 745–774.

[51] Evans, P. W., Price, H., and Wharton, K. B. New Slant on the EPR-Bell Experiment. *The British Journal for the Philosophy of Science 64* (2013), 297–324.

[52] Everett, H. Relative State Formulation of Quantum Mechanics. *Review of Modern Physics 29* (1957), 454–462.

[53] Fritz, T. Quantum correlations in the temporal Clauser–Horne–Shimony–Holt (CHSH) scenario. *New Journal of Physics 12*, 8 (2010), 083055.

[54] Fuchs, C. A. Quantum Foundations in the Light of Quantum Information. *Eprint arXiv:quant-ph/0106166* (2001).

[55] Gao, S., Ed. *Collapse of the Wave Function: Models, Ontology, Origin, and Implications.* Cambridge: Cambridge University Press, 2018.

[56] Ghirardi, G. C., G. R., and Pearle, P. Relativistic dynamic reduction models — general framework and examples. *Foundations of Physics 20*, 11 (1990), 1271–1316.

[57] Ghirardi, G.C., R. A., and Weber, T. Unified dynamics for microscopic and macroscopic systems. *Physical Review D 34* (1986), 470–490.

[58] Gillespie, D. T. *Markov Processes: An Introduction for Physical Scientists*. Amsterdam: Elsevier, 1991.

[59] Giustina, M., et al. Significant-Loophole-Free Test of Bell's Theorem with Entangled Photons. *Physical Review Letters 115*, 25 (2015), 1–7.

[60] Glynn, L. A probabilistic analysis of causation. *The British Journal for the Philosophy of Science 62*, 2 (2011), 343–392.

[61] Goldstein, S. Bohmian mechanics. In *The Stanford Encyclopedia of Philosophy*, E. N. Zalta, Ed., summer 2017 ed. Metaphysics Research Lab, Stanford University, 2017.

[62] Goldstein, S., Norsen, T., Tausk, D. V., and Zanghi, N. Bell's theorem. *Scholarpedia 6*, 10 (2011), 8378.

[63] Goldstein, S., Tausk, D., Tumulka, R., and Zanghì, N. What Does the Free Will Theorem Actually Prove? *Notices of the American Mathematical Society 57* (2010), 1451–1453.

[64] Goldstein, S., and Tumulka, R. Opposite Arrows of Time Can Reconcile Relativity and Nonlocality. *Classical and Quantum Gravity 20* (2003), 557–564.

[65] Griffiths, R. B. EPR, Bell, and quantum locality. *American Journal of Physics 79* (2011), 954–965.

[66] Hardy, L. Quantum ontological excess baggage. *Studies In History and Philosophy of Modern Physics 35*, 2 (2004), 267–276.

[67] Harrigan, N., and Rudolph, T. Ontological models and the interpretation of contextuality. *Eprint arXiv:0709.4266v1* (2007).

[68] Harrigan, N., Rudolph, T., and Aaronson, S. Representing probabilistic data via ontological models. *Eprint arXiv:0709.1149v2* (2008).

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

[70] Iranzo Ribera, N. The Status of Bohr's Complementarity Today: A study of the nature of knowing and being. Master's thesis, Utrecht University, 2017.

[71] Ismael, J. Quantum mechanics. In *The Stanford Encyclopedia of Philosophy*, E. N. Zalta, Ed., spring 2015 ed. Metaphysics Research Lab, Stanford University, 2015.

[72] Jarrett, J. P. On the Physical Significance of the Locality Conditions in the Bell Arguments. *Noûs 18*, 4 (1984), 569–589.

[73] Koashi, M., and Winter, A. Monogamy of quantum entanglement and other correlations. *Phys. Rev. A 69* (2004), 022309.

[74] Ladyman, J., Øystein Linnebo, and Bigaj, T. Entanglement and non-factorizability. *Studies in History and Philosophy of Modern Physics 44*, 3 (2013), 215 – 221.

[75] Laudisa, F. Against the 'no-go' philosophy of quantum mechanics. *European Journal for Philosophy of Science 4*, 1 (2014), 1–17.

[76] Laudisa, F. Counterfactual Reasoning, Realism and Quantum Mechanics: Much Ado About Nothing? *Eprint arXiv:1705.08287v1* (2017).

[77] Leegwater, G., and Muller, F. A. Ranting & Raving about Locality in Quantum Mechanics. Presented at Workshop on History and Philosophy of Physics at Utrecht University, 8 May 2018.

[78] Leggett, A. J. Realism and the physical world. *Reports on Progress in Physics 71*, 2 (2008), 022001.

[79] Leggett, A. J., and Garg, A. Quantum mechanics versus macroscopic realism: Is the flux there when nobody looks? *Physical Review Letters 54* (1985), 857–860.

[80] Leifer, M. S. Time Symmetric Quantum Theory Without Retrocausality? A Reply to Tim Maudlin. *Eprint arXiv:1708.04364v1* (2017).

[81] Leifer, M. S., and Pusey, M. Is a time symmetric interpretation of quantum theory possible without retrocausality? *Proceedings of the Royal Society 473*, 2202 (2017).

[82] Li, L., Hall, M. J. W., and Wiseman, H. M. Concepts of quantum non-Markovianity: A hierarchy. *Physics Reports 759* (2018), 1–51.

[83] Maroney, O. J. E., and Timpson, C. G. Quantum- vs. Macro- Realism: What does the Leggett-Garg Inequality actually test? *Eprint arXiv:1412.6139v1* (2014).

[84] Maudlin, T. What Bell proved: A reply to Blaylock. *American Journal of Physics 78* (2010), 121–125.

[85] Maudlin, T. How Bell reasoned: A reply to Griffiths. *American Journal of Physics 79* (2011), 966–970.

[86] Maudlin, T. *Quantum Non-locality and Relativity*, third ed. Oxford: Wiley-Blackwell, 2011.

[87] Maudlin, T. What Bell did. *Journal of Physics A: Mathematical and Theoretical 47*, 42 (2014), 1–24.

[88] Maudlin, T. Time Symmetric Quantum Theory Without Retrocausality. *Eprint arXiv:1707.08641v1* (2017).

[89] Maudlin, T. Ontological Clarity via Canonical Presentation: Electromagnetism and the Aharonov–Bohm Effect. *Entropy 20*, 6 (2018), 465.

[90] Mehra, J. *The Quantum Principle: its Interpretation and Epistemology.* Dordrecht: Springer, 1974.

[91] Montina, A. Exponential complexity and ontological theories of quantum mechanics. *Physical Review A 77* (2014), 022104.

[92] Mousavi, S. V., and Miret-Artés, S. Quantum-classical transition in dissipative systems through scaled trajectories. *Journal of Physics Communications 2*, 3 (2018), 035029.

[93] Mulder, R. A gauge too far: On the reality of the electromagnetic potential. *Forthcoming* (2018).

[94] Mulder, R. Worldly Patterns: Emergence, Functionalism and Pragmatic Reality in Wallacian quantum mechanics. Master's thesis, Utrecht University, 2018.

[95] Muller, F. A. The Locality Scandal of Quantum Mechanics. In *Language, Quantum, Music*, M. L. D. Chiara et al., Eds. Dordrecht: Kluwer Academic Publisers, 1999, pp. 241–248.

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

[97] Myrvold, W. C. Relativistic Markovian dynamical collapse theories must employ nonstandard degrees of freedom. *Physical Review A 96* (2017), 062116.

[98] Norsen, T. Einstein's Boxes. *American Journal of Physics 73* (2005), 164–176.

[99] Norsen, T. Local Causality and Completeness: Bell vs. Jarrett. *Foundations of Physics 39*, 3 (2009), 273–286.

[100] Norsen, T. Quantum Solipsism and Nonlocality. In *Quantum Nonlocality and Reality: 50 Years of Bell's Theorem*, M. Bell and S. Gao, Eds. Cambridge: Cambridge University Press, 2016, pp. 204–237.

[101] Norsen, T. *Foundations of Quantum Mechanics: An Exploration of the Physical Meaning of Quantum Theory.* Springer International Publishing AG, 2017.

[102] Pawłowski, M., Paterek, T., Kaszlikowski, D., Scarani, V., Winter, A., and Żukowski, M. Information causality as a physical principle. *Nature 461* (2009), 1101–1104.

[103] Pawłowski, M., and Scarani, V. Information Causality. *Eprint arXiv:1112.1142* (2011).

[104] Pearle, P. Toward a Relativistic Theory of Statevector Reduction. In *Sixty-Two Years of Uncertainty*, A. Miller, Ed. New York: Plenum, 1990, pp. 193–214.

[105] Price, H. *Time's Arrow and Archimedes' Point.* Oxford: Oxford University Press, 1996.

[106] Price, H. Does time-symmetry imply retrocausality? How the quantum world says "Maybe"? *Studies in History and Philosophy of Modern Physics 43* (2012), 72–83.

[107] Price, H., and Wharton, K. Dispelling the Quantum Spooks – a Clue that Einstein Missed? *Eprint arXiv:1307.7744* (2013).

[108] Price, H., and Wharton, K. A live alternative to quantum spooks. *Eprint arXiv:1510.06712v2* (2015).

[109] Price, H., and Wharton, K. Disentangling the Quantum World. *Entropy 17* (2015), 7752–7767.

[110] Pusey, M. F., Barrett, J., and Rudolph, T. On the reality of the quantum state. *Nature Physics 8* (2012), 475–478.

[111] Redhead, M. More ado about nothing. *Foundations of Physics 25*, 1 (1995), 123–137.

[112] Redhead, M. L. G., and Wagner, F. Unified Treatment of EPR and Bell Arguments in Algebraic Quantum Field Theory. *Foundations of Physics Letters 11*, 2 (1998), 111–125.

[113] Robertson, K. Can the two-time interpretation of quantum mechanics solve the measurement problem? *Studies in History and Philosophy of Modern Physics 58* (2017), 54–62.

[114] Seevinck, M. P. *Parts and Wholes: An Inquiry into Quantum and Classical Correlations*. PhD thesis, Utrecht University, 2008.

[115] Shapiro, S. *Philosophy of Mathematics: Structure and Ontology*. Oxford: Oxford University Press, 1997.

[116] Shimony, A. Controllable and uncontrollable nonlocality. In *Proceedings of the International Symposium: Foundations of Quantum Mechanics in the light of New Technology*, S. Kamef, Ed. Tokyo: Physical Society of Japan, 1984, pp. 182–203.

[117] Shimony, A. Events and processes in the quantum world. In *Quantum Concepts in Space and Time*, R. Penrose and C. J. Isham, Eds. Oxford: Oxford University Press, 1986, pp. 182–203.

[118] Smolin, L. The unique universe. *Physics World 22*, 6 (2009), 21–26.

[119] Spekkens, R. W. The Paradigm of Kinematics and Dynamics Must Yield to Causal Structure. In *Questioning the Foundations of Physics: Which of Our Fundamental Assumptions Are Wrong?*, A. Aguirre, B. Foster, and Z. Merali, Eds. Springer International Publishing, 2015, pp. 5–16.

[120] 't Hooft, G. *The Cellular Automaton Interpretation of Quantum Mechanics*. Springer International Publishing, 2016.

[121] 't Hooft, G. Free Will in the Theory of Everything. *Eprint arXiv:quant-ph/1709.02874* (2017).

[122] Tipler, F. J. Quantum nonlocality does not exist. *Proceedings of the National Academy of Sciences of the United States of America 111*, 31 (2014), 11281–11286.

[123] Tumulka, R. A Relativistic Version of the Ghirardi–Rimini–Weber Model. *Journal of Statistical Physics 125*, 4 (2006), 821–840.

[124] Tumulka, R. Comment on "the Free Will Theorem". *Foundations of Physics 34* (2007), 186–197.

[125] Tumulka, R. The Point Processes of the GRW Theory of Wave Function Collapse. *Reviews in Mathematical Physics 21* (2009), 155–227.

[126] Vaidman, L. Many-worlds interpretation of quantum mechanics. In *The Stanford Encyclopedia of Philosophy*, E. N. Zalta, Ed., fall 2018 ed. Metaphysics Research Lab, Stanford University, 2018.

[127] Vargas, A. F., Morales-Durán, N., and Bargueño, P. A Bohmian approach to the non-Markovian non-linear Schrödinger–Langevin equation. *Annals of Physics 356* (2015), 498 – 504.

[128] Wallace, D. *The Emergent Multiverse: Quantum Theory According to the Everett Interpretation*. Oxford: Oxford University Press, 2012.

[129] Wessels, L. The Way the World Isn't: What the Bell Theorems Force Us to Give Up. In *Philosophical Consequences of Quantum Theory: Reflections on Bell's Theorem*, J. T. Cushing and E. McMullin, Eds. Notre Dame: University of Notre Dame Press, 1989.

[130] Wharton, K. A Novel Interpretation of the Klein-Gordon Equation. *Foundations of Physics 40* (2010), 313–332.

[131] Wharton, K. Lagrangian-Only Quantum Theory. *Eprint arXiv:1301.7012v1* (2013).

[132] Wharton, K. Quantum States as Ordinary Information. *Information 5* (2014), 190–208.

[133] Wharton, K. The Universe is not a Computer. In *Questioning the Foundations of Physics*, A. Aguirre, B. Foster, and Z. Merali, Eds. Springer International Publishing, 2016, pp. 177–190.

[134] Wharton, K. Towards a realistic parsing of the Feynman path integral. *Quanta 5* (2016), 1–11.

[135] Wharton, K. A New Class of Retrocausal Models. *Entropy 20*, 6 (2018), 410–427.

[136] Wharton, K., Miller, D. J., and Price, H. Action Duality: A Constructive Principle for Quantum Foundations. *Symmetry 3*, 3 (2011), 524–540.

[137] Wiseman, H. M. Three Boos for 'Locality'. Presented at the Foundations2018 conference at Utrecht University, 12 July 2018.

[138] Wiseman, H. M. From Einstein's theorem to Bell's theorem: a history of quantum non-locality. *Contemporary Physics 47*, 2 (2006), 79–88.

[139] Wood, C. J., and Spekkens, R. W. The lesson of causal discovery algorithms for quantum correlations: causal explanations of Bell-inequality violations require fine-tuning. *New Journal of Physics 17* (2015), 033002.

[140] Wütrich, C. Can the world be shown to be indeterministic after all? In *Probabilities in Physics*, C. Beisbart and S. Hartmann, Eds. Oxford: Oxford University Press, 2011, pp. 365–389.

[141] Yearsley, J. M. The Leggett-Garg Inequalities and Non-Invasive Measurability. *Eprint arXiv:1310.2149v1* (2013).