

What are the bearers of hidden states? On an important ambiguity in the formulation of Bell’s theorem

Joanna Luc

January 29, 2025

Abstract

Bell’s conclusion from his famous inequality was that any hidden variable theory that satisfies Local Causality is incompatible with the predictions of Quantum Mechanics (QM) for Bell’s Experiment. However, Local Causality does not appear in the derivation of Bell’s inequality. Instead, two other assumptions are used, namely Factorizability and Settings Independence. Therefore, in order to establish Bell’s conclusion, we need to relate these two assumptions to Local Causality. The prospects for doing so turn out to depend on the assumed location of the hidden states that appear in Bell’s inequality. In this paper, I consider the following two views on such states: (1) that they are states of the two-particle system at the moment of preparation, and (2) that they are states of thick slices of the past light cones of measurements. I argue that straightforward attempts to establish Bell’s conclusion fail in both approaches. Then, I consider three refined attempts, which I also criticise, and I propose a new way of establishing Bell’s conclusion that combines intuitions underlying several previous approaches.

1 Introduction

Einstein et al. (1935) famously argued that Quantum Mechanics (QM) cannot be regarded as a complete theory—there should exist some additional (“hidden” in the sense of not being captured by the quantum state) variables that explain in a local way the correlations that occur in quantum experiments. A major contribution to the investigation of theories postulating such variables (called “hidden variable theories”, in short “HVTs”) was made by Bell (1964), who proposed a general framework in which one can analyse all such theories at once. Within this framework, he derived an inequality (called “Bell’s inequality” or “Bell’s theorem”) that constrains the expectation values for the results of a special spin measurement experiment (which is now called “Bell’s Experiment”). Bell’s inequality is violated by QM predictions for Bell’s experiment, but it is satisfied by a class of HVTs, which Bell identifies with the class of local HVTs. Therefore, Bell’s own interpretation of his result was that it allows us to exclude all local HVTs by virtue of their being inconsistent with the predictions of QM for Bell’s Experiment. Since the results of actual realisations of Bell’s Experiment turned out to confirm QM predictions and violate Bell’s inequality, this interpretation—if correct—would lead to the conclusion that Nature itself is non-local, contrary to what Einstein expected (cf. Norsen 2009 and 2011, Maudlin 2014).

Locality is understood here in a way closely related to relativity: roughly, as the claim that no event can be causally influenced by anything from the outside of its past light cone. However, the exact formulation of this principle is a subtle issue. In his last paper, Bell’s preferred formulation of locality was the principle he called “Local Causality” (see section 7.1 for details), so his conclusion from his theorem can be stated as follows:

Bell’s Conclusion: All HVTs that satisfy Local Causality are inconsistent with the predictions

of QM for Bell’s Experiment.¹

However, Local Causality is not used in the proof of Bell’s inequality. Instead, two other assumptions are used: Factorizability and Settings Independence (see section 5 for details). Therefore, in order to retain Bell’s Conclusion, we need to relate Local Causality to these two assumptions. In this paper, I will investigate the prospects of doing this, but we will see that this task is far from trivial.

This paper is organised as follows. After introducing some basic concepts and notation (section 2), I will review two examples of HVTs (section 3) and provide a typology of such theories (section 4), in which of special interest will be locally deterministic theories, a distinguished subclass of locally causal theories. Next, I will recall the standard presentation of the assumptions of Bell’s theorem (section 5). The central topic of this paper will be the question of what exactly the hidden states invoked in Bell’s theorem encompass. The two main options considered in the literature are the hidden states of the two-particle system at the moment of preparation (section 6) and the hidden states of thick slices of the past light cones of measurements (section 7). In both cases, there are certain obstacles to establishing Bell’s Conclusion. However, there are some upgraded versions of both approaches that aim to overcome these obstacles, which will be the subject of section 8. My own proposal, presented in section 9, will build upon the insights of both approaches. The results of the current paper will be summarised in section 10.

One terminological remark is in place here. The terms “hidden variables” and “hidden states” can be misleading in at least three ways. First, the word “hidden” suggests that the postulation of these variables/states does not lead to any observable predictions. Although this might be the case for some HVTs, this is by no means a defining feature of hidden variables/states. Second, one might suppose that HVTs do not exhaust all theories that are alternatives to the standard QM. The argument for this claim would be that some such theories may not postulate hidden variables/states. However, the word “hidden” here merely indicates that the postulated variables/states involve something that is not captured (or at least not fully captured) by the quantum state and either supplements or replaces it. Clearly, every theory that attempts to account for the same phenomena as the standard QM but differs from it at the level of ontology must postulate some hidden variables/states *in this sense*. Third, the formulation of Bell’s theorem in terms of HVTs may misleadingly suggest that we can avoid the conclusion of non-locality by rejecting the assumption of the existence of hidden variables/states. However, this reasoning ignores the fact that the standard QM itself is non-local in some sense (although this issue is subtle; see section 4.1), so if there is any way to account for quantum phenomena in a local way, it requires some hidden variables/states in our sense. Despite these confusing connotations of the word “hidden”, I use the terms “hidden variables” and “hidden states” because this terminology has become widespread, and it makes the connection of this paper with the literature on Bell’s theorem more conspicuous.

2 Some basic concepts and notation

Since the concern with locality is motivated by relativity, our stage will be a spacetime with the Minkowski metric, which is used to define the light cone structure in the standard way. Additionally, we will refer to some fixed slicing of this spacetime (given by simultaneity hypersurfaces in some chosen inertial reference frame) and associated time parameter t . If an HVT is relativistic, in the end all its physical results must be independent of the choice of slicing.

Let R be a bounded spatiotemporal region, and let “ $t < R$ ” mean that time t is below R (i.e., the spatial slice at t lies below any spatial slice that intersects R). We define $\Sigma_{R,t}$ as the intersection of the spatial slice at t with the past (if $t < R$) or future (if $R < t$) light cone of R . If $t < R$, then for any $t' < t$ we also define $C(\Sigma_{R,t}, \Sigma_{R,t'})$ to be the part of the past light cone of R that lies between the surfaces $\Sigma_{R,t}$ and $\Sigma_{R,t'}$, including those surfaces themselves (such regions will informally be called “thick slices”).

We will be interested in accounting for Bell’s Experiment, in which a pair of particles is prepared in an

¹In this paper, I will use the expression “Bell’s Conclusion” as a label for this statement. This is not meant to suggest that this is the only conclusion that Bell derived from his theorem.

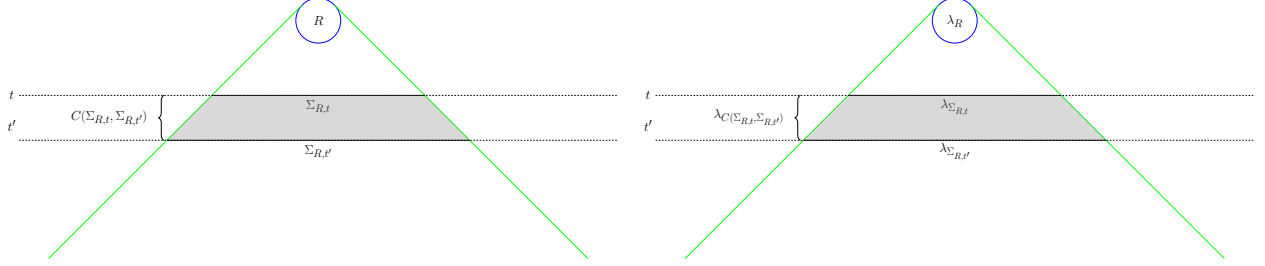


Figure 1: My notation for regions (left) and states (right); $C_{\Sigma_{R,t}, \Sigma_{R,t'}}$ covers only the shaded region.

entangled spin state

$$\psi_{12} = \frac{1}{\sqrt{2}} (\psi_1^+ \psi_2^- - \psi_1^- \psi_2^+). \quad (1)$$

Then, each of these particles moves from the source to one of two remote detectors, where spin in a chosen direction is measured on each of these particles. We will assume that one of two choices of direction can be made at each station (but the allowed choices do not need to be the same for both stations). The variable a will range over such possible choices in the left station, whereas b will range over the possible choices in the right station. Analogously, the variable A will range over possible outcomes of the measurement of spin in direction a , and B will range over possible outcomes of the measurement of spin in direction b (the values of the outcomes are, by convention, ± 1). We will denote the different choices of measurement settings with primes: $a' \neq a$ and $b' \neq b$; relatedly, A'/B' will range over the possible outcomes of spin measurement in directions a'/b' . By R_a and R_b , we will denote the spatiotemporal regions in which the respective choices of settings are made, whereas by R_A and R_B we will denote the spatiotemporal regions in which the measurements are performed and the outcomes are obtained (in a single run of the experiment). An additional assumption is that R_a and R_A are spacelike related to R_b and R_B (and, of course, R_a/R_b is in the causal past of R_A/R_B).

We will be interested in the following question posed by Bell: what constraints on possible HVTs can we derive from QM predictions for Bell's Experiment? For this purpose, we will introduce hidden states λ (the space of such states will be denoted by Λ), which might be different for different HVTs. Such λ encompasses all hidden variables considered in a given theory. For λ , Bell also used the term “the beables of the theory”, which he defines as “those entities in [the theory] which are, at least tentatively, to be taken seriously, as corresponding to something real” (Bell, 1990, p. 234). In other words, beables are “what some candidate theory posits as being physically real” (Norsen, 2009, p. 279), without the assumption that this theory is correct.

We will later need to ascribe such states to various regions of spacetime and space. For any spatiotemporal region R , by λ_R we denote a possible complete specification of hidden variables (of a given HVT) in R . Different such possible specifications will be denoted with primes (i.e., our convention is that $\lambda'_R \neq \lambda_R$). Analogously, we can talk about complete specifications of hidden variables on spatial slices (denoted by $\lambda_{\Sigma_{R,t}}$) or on thick slices (denoted by $\lambda_{C(\Sigma_{R,t}, \Sigma_{R,t'})}$). Finally, λ_t means the hidden state of the entire space at time t . My notation for regions and states is illustrated in Fig. 1.

3 Two examples of HVTs: True Spin Theory and Bohmian mechanics

The simplest and most natural candidate for hidden variables in the case of spin measurements is just the “true” values of spin in every direction that are possessed by each particle. As far as I know, the theory with such hidden variables has not received a name in the literature; I will call it “True Spin Theory”. According to this theory, when QM describes the spin in direction a of a pair of particles by means of an entangled state

(such as (1)), in which case no definite value of spin in direction a is ascribed to any of these particles, then the description provided by QM is incomplete. As such, it should be completed by ascribing to each of these particles a hidden state λ that consists of this particle’s spin in direction a (and in any other direction). In particular, if we consider measurements of spin in directions a, a', b and b' , then the part of λ that is relevant to determining the outcomes of these measurements will consist of four numbers corresponding to the values of spin in these four directions.

True Spin Theory can be shown to be inconsistent with the predictions of QM by using a reasoning that is generalised in the proof of Bell’s theorem (see, e.g., Norsen 2017, pp. 215–227). It should be stressed that True Spin Theory, as presented here, is not really a fully developed theory. For example, we have not said anything about how definite values of spin possessed by particles might change in interactions (if at all)—that is, we have not introduced any dynamical laws. For this reason, True Spin Theory should be regarded not as a single theory but as a family of possible theories that account for Bell’s Experiment in the way described above. The mentioned reasoning shows that this entire class of HVTs can be excluded as inconsistent with the predictions of QM.

However, True Spin Theory is not the only possible HVT. Another example of an HVT is Bohmian mechanics (see, e.g., Goldstein 2024), which, in contrast to True Spin Theory, is a fully developed theory with specified ontology and dynamical laws. In Bohmian mechanics, hidden variables are the true positions of particles. In the standard QM, the wave function only provides probabilities for finding a particle in a given region of space; so, if QM is a complete theory, then we need to abandon the idea that all particles at all times have definite positions, as was presupposed by classical physics. However, this idea of classical physics is restored by Bohmian mechanics, according to which the wave function guides how particles move, but any particle at any time has a definite position and not only a probability of being found in that position. The dynamics of this theory consists of the Schrödinger equation of the standard QM and an additional guiding equation which prescribes how the wave function influences the positions of particles.

These are only two examples of HVTs, but we can use them to illustrate the crucial point of Bell’s enterprise. Bohmian mechanics is consistent with the predictions of QM, but it is known to be non-local: if two particles are entangled, then what happens with one of them can have an immediate effect on what happens with the second one. (Of course, this is not the precise formulation of what locality means, on which more later.) On the other hand, True Spin Theory is supposed² to be local: the true spins are introduced in order to account for the outcome of the measurement on one of the entangled particles without referring to the other one. However, True Spin Theory is inconsistent with the predictions of QM. The interesting question is whether there exists a theory that has both virtues—that is, it is both local and consistent with the predictions of QM. Bell’s approach promises to give an answer (a negative one) to precisely this question. Before going into the details, let us say more about possible types of HVTs.

4 A typology of HVTs

In this section, I will distinguish some important types of HVTs: locally determinate vs. locally indeterminate, deterministic vs. indeterministic, and locally deterministic.

4.1 Locally determinate and locally indeterminate HVTs

Our notation for hidden states presupposes that they are locally determinate: if we can write λ_R for any spatiotemporal region R , this means that for any such region a hidden state in that region is well-defined. However, not all theories are like this. In particular, QM states are not locally determinate: if a pair of particles is in a state (1), then we cannot ascribe any spin state to one of these particles considered in isolation. Therefore, if an HVT includes ψ in an ineliminable way, then its ontology as a whole is locally indeterminate, even if the λ -part of that ontology is locally determinate.

This means that we can distinguish two types of HVTs. (1) Locally determinate HVTs: for any spatiotemporal region R and any possible state of that region, the restriction of that state to any subregion of R

²I say this in a cautious way because this is not a fully-developed theory.

is well-defined. In such theories, ψ must be eliminable (or at least excluded from the variables of the theory) and all the physical content of the theory must be captured by (locally well-defined) λ 's, which might then be called “complete states”. (2) Locally indeterminate HVTs: ψ is not eliminable and the physical content of the theory has two parts: λ (locally determinate) and ψ (locally indeterminate). For example, Bohmian mechanics is locally determinate if considered as a theory in which the only real physical quantity is the positions of particles (cf. Esfeld et al. 2014). However, Bohmian Mechanics might also be conceived of as a theory in which particles are described by both their true position and the QM state, in which case it would be an example of a locally indeterminate HVT.

Where should we locate the standard QM in this picture? I think that two variants of the standard QM need to be distinguished that differ in the ontological status they assign to the quantum state. If the quantum state is regarded as a beable of the standard QM, then this theory is locally indeterminate. This route is taken, among others, by Norsen (2017, p. 127), who emphasises that being locally indeterminate is a stronger kind of non-locality than causal non-locality:

[...] it would probably be most accurate to summarize the situation by describing quantum mechanics here as “not even non-local”. Remember what “locality” means: the causal influences that objects, moving and interacting in three-dimensional space, exert on one another, always propagate at the speed of light or slower. A theory which fails to provide a clear ontology of objects moving and interacting in three-dimensional space [...] doesn't even rise to the level of making the question, of whether causal influences always propagate at the speed of light or slower, or not, meaningful.

An alternative route, which excludes the quantum state from the beables of the standard QM, is taken by Bell himself. For him, the beables in ordinary QM are only “experimental procedures and experimental results” and the wave function is “a convenient but inessential mathematical device for formulating correlations” between such beables (Bell, 1976, p. 53). Since the mentioned beables are local, the standard QM, understood in this way, is locally determinate but at the price of excluding the quantum state from the ontology while retaining it in the formalism.

Thus, it seems that we have two very different readings of the standard QM: either

- (1) the quantum state *is a beable*, in which case the standard QM is not locally determinate and, as suggested by Norsen, it is “not even [causally] non-local” or
- (2) the quantum state *is not a beable*, in which case the only beables are the experimental settings and preparations, and the standard QM is not locally causal (in the sense of late Bell, see Definition 5), as argued by Bell (1976, p. 55 and 1990, pp. 240–241).

Since “the standard QM” is a rather vague term, both of these readings are acceptable, and we should consider two variants of this theory separately.

In what follows, I will restrict my considerations to locally determinate HVTs only. Perhaps my reasonings could also be applied, with some modifications, to locally determinate parts of locally indeterminate theories. However, the philosophical relevance of such an application would be unclear, given that such theories are already non-local in the sense that is stronger than causal non-locality. An immediate worry might be that the scope of my considerations will be, due to the mentioned restriction, too narrow. However, as we have seen, even the standard QM can be understood as a locally determinate theory if we reinterpret the role of the wave function in a way consistent with its not being a beable. (Importantly, this does not exclude its playing a role in determining the probabilities within that theory.) Similarly, any variant of QM that does not remove the wave function from the formalism can be understood as either locally indeterminate (if the wave function is included in the beables) or locally determinate (if the wave function is excluded from the beables), provided that all other beables are local. For example, some Bohmians have proposed understanding the wave function as “a component of physical law rather than of the reality described by the law” (Dürr et al., 1997, p. 33), which makes this variant of Bohmian mechanics locally determinate.

4.2 Deterministic, indeterministic and locally deterministic HVTs

Another important division of HVTs is between deterministic and indeterministic ones. For the purpose of defining them, we will use the concept of a solution of a theory T , which is any specification of hidden variables of T in the entire spacetime that is compatible with the laws of T , where “compatible” means simply “logically consistent”. Then, we can define:

Definition 1. If T is a locally determinate HVT, then T is deterministic iff for any t_0 and any λ_{t_0} , there is exactly one solution that is compatible with λ_{t_0} . Otherwise T is indeterministic.

In our terminology, “exactly one solution that is compatible with λ_{t_0} ” means “exactly one specification of hidden variables of T in the entire spacetime that is logically consistent with the laws of T and λ_{t_0} ”.

The above definition is (modulo differences in formulation and pace certain subtleties, which will not be discussed here³) standard in the contemporary literature about determinism and indeterminism (see, e.g., Earman 1986 and Butterfield 1998). However, it should be noted that in the literature about Bell’s theorem, deterministic HVTs are often defined in a different way, namely as theories that predict the measurement outcomes with probability 1, given the state of the system and the measurement settings.⁴ These two senses of determinism should not be conflated because, as we will see later (in the third paragraph of section 6), they are not equivalent.

An especially interesting subclass of deterministic theories is locally deterministic theories (cf. Bell 1976, p. 53). From the definition of determinism, it follows that for any region R , the physical contents of this region are uniquely determined by the full specification of the state of the world at some moment $t < R$. However, this does not exclude the possibility that what happens in R might depend on what happens at t arbitrarily far away spatially from R , which is in tension with the relativistic idea that what happens in R should not depend on what happens outside its past light cone. This leads to the concept of locally deterministic theories, in which to determine the physical contents of R , the information about the entire spatial slice at some $t < R$ is not needed and the part of that information that is confined to the intersection of the past light cone of R with the spatial slice at t suffices. More formally,

Definition 2. If T is a locally determinate HVT, then T is locally deterministic iff for any bounded spatiotemporal region R , for any t_0 such that $t_0 < R$ or $t_0 > R$ and for any $\lambda_{\Sigma_{R,t_0}}$, there exists λ_R such that all solutions compatible with $\lambda_{\Sigma_{R,t_0}}$ are also compatible with λ_R .

Since we have universal quantification over solutions, it follows from this definition that for any other $\lambda'_R \neq \lambda_R$, all solutions compatible with $\lambda_{\Sigma_{R,t_0}}$ are not compatible with λ'_R (a solution on the entire spacetime cannot be compatible with different states of some region of this spacetime). One can show that local determinism defined in this way entails determinism, so locally deterministic theories indeed form a subclass of deterministic theories (for the proof see Appendix A.1):

Theorem 1. Any locally determinate HVT that is locally deterministic is also deterministic.

This theorem is not as trivial as its verbal formulation might suggest because “local determinism” has not been defined as a conjunction of “determinism” and “locality” but by a condition that does not explicitly mention determinism. Later, we will see that any locally deterministic theory is also local in the sense of satisfying Local Causality $_{C(\Sigma,\Sigma)}$ (see Theorem 2 in section 7.2).

³One such subtlety is the physical equivalence of different solutions that are related by the symmetries of the theory. If a theory involves local symmetries (i.e., symmetries that are not the same at every spacetime point), such as diffeomorphisms in General Relativity and gauge symmetries in Electrodynamics, then there will be spurious witnesses of indeterminism—that is, solutions of the theory that are symmetry-related and identical on some initial segment but not identical overall. In such cases, the definition of determinism should be modified by taking into account the physical equivalence of such solutions (see, e.g., Butterfield 1989, pp. 7–9).

⁴This terminology was used by Bell in later commentaries on his 1964 paper (e.g., 1971 and 1981). However, in the 1964 paper itself, he talks only about measurement outcomes being predetermined or determined by settings together with λ ; the words “determinism” or “deterministic” do not appear there. Some other authors use terminology that indicates the difference between this concept of determinism and the usual one, such as “quasi-determinism” Malament (ated) or “outcome determinism” (Myrvold et al., 2024).

There exist both deterministic and indeterministic alternatives to the standard QM. For example, GRW (see, e.g., Ghirardi and Bassi 2024) is an indeterministic theory. True Spin Theories might be or might not be deterministic, depending on the details of their dynamics. What we know is only that they predict unique results of spin measurement. Bohmian mechanics is deterministic but not locally deterministic. This fact made this theory unattractive to Einstein, even though it was a proposal for how to complement QM, which is what he sought. This fact also led Bell to ask whether any possible completion of QM must be non-local. However, in order to pose this question properly, he needed to formulate a general definition of locality that can be applied to both deterministic and indeterministic theories, which he did using probability calculus (cf. section 7).

In order to apply any condition expressed in terms of probabilities to a deterministic theory, we would need a bridge principle that relates the two. Intuitively, deterministic theories are extreme cases of probabilistic theories, in the sense that in the former all probabilities are 0 or 1. However, this relation is not so straightforward because a deterministic theory might not mention probabilities at all. For example, in classical mechanics of pointlike particles there are no probabilities; instead, there are dynamical laws in the form of differential equations, which for many initial conditions have a unique solution. Moreover, it does not make any sense to ask what the probability of such-and-such behaviour of a particle is *simpliciter*. A particle is determined to either behave in this way or not, depending on the *initial conditions*; however, without the full initial conditions, there is not enough information to determine how it will behave. The moral is as follows: that a theory is deterministic does not mean that any probability we can think of has value 0 or 1 (cf. Norsen 2017, p. 17). Only conditional probabilities that are conditioned on sufficient information about initial conditions can be said to have values 0 or 1 (although this will usually be a proper extension of the language of a deterministic theory). This leads to the following Bridge Principle:⁵

Definition 3. Bridge Principle Between Determinism and 0-1 Probability: For any theory T and for any regions R_i and R_f , if all solutions that are compatible with λ_{R_i} are also compatible with λ_{R_f} , then $P(\lambda_{R_f}|\lambda_{R_i}) = 1$ according to T ; analogously, if all solutions that are compatible with λ_{R_i} are not compatible with λ_{R_f} , then $P(\lambda_{R_f}|\lambda_{R_i}) = 0$ according to T .

Conditional probabilities of other events might be introduced in a different way (e.g., by specifying the probability density over the initial states, as in Bohmian mechanics), but they might be not defined at all.

5 Assumptions of Bell's theorem

There are several versions of Bell's theorem, each of which involves a different inequality or set of inequalities. The most famous of them is the Bell-CHSH inequality, named after Bell, Clauser, Horne, Shimony and Holt (Clauser et al., 1969). The proofs of these various inequalities use two assumptions, which are called Factorizability and Settings Independence:⁶

Factorizability: $P(A, B|a, b, \lambda) = P(A|a, \lambda) \cdot P(B|b, \lambda)$.

Settings Independence: $P(\lambda|a, b) = P(\lambda)$.

⁵What would a probability space look like in such cases? Presumably, we should regard entire solutions of T as elementary events; an event corresponding to λ_R would then be the set of all solutions that ascribe state λ_R to region R . In what follows, we will write states (complete or coarse-grained) as arguments of a probability function, with the risk that such expressions will not always be well-defined. Another technical problem is that λ is continuous in some HVTs (e.g., in Bohmian mechanics, since particle positions are given by real numbers). This means that in many places we should use probability density instead of probability. For example, fine-grained formulation of Local Causality $_{C(\Sigma, \Sigma)}$ (i.e., Eq. (2)) should be in terms of probability densities, whereas coarse-grained formulation (i.e., Eq. (3)) should be in terms of probabilities obtained by integrating the respective probability densities. For reasons of space, I will not fully do justice to this problem here.

⁶Since we are considering only situations in which the experiment takes place, $P(\lambda)$ should be non-zero only for those λ that are consistent with Bell's Experiment being performed, which entails that every λ with non-zero probability (density) should be consistent with at least one choice of measurement settings. For this reason, we can think of $P(\lambda)$ as being a conditional probability $P(\lambda|\text{Bell's Experiment})$.

Starting with the latter, Settings Independence expresses the independence of the hidden state λ from the measurement settings a and b .⁷ It is also called “statistical independence”, “measurement independence”, “no conspiracy” and even “free will” or “free choice”. The latter two names are motivated as follows: measurement settings are chosen by humans, so if these choices are free, then they should not depend on the hidden state that is measured. The hidden state being such-and-such cannot prevent us from choosing these particular settings rather than others; moreover, it should not even influence the probability of our choice of these particular settings.

Factorizability is usually regarded as derived from some more basic principles. Jarrett (1984) suggested that Factorizability should be viewed as derived from two principles, which Shimony (1986) called “Parameter Independence” and “Outcome Independence” (Jarrett called them “locality” and “completeness”, but Shimony’s terminology has become more popular). As far as I know, this view on Factorizability as deriving from Parameter Independence and Outcome Independence was never endorsed by Bell. Instead, in his last paper (Bell, 1990), he regards Factorizability as derived from another principle, which he calls Local Causality. Its initial formulation is as follows (Bell, 1990, p. 239):⁸

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.

Subsequently in that paper, Bell provides a more precise formulation of this principle, which will be analysed in detail in section 7.1. The Bell-CHSH inequality, which follows from Factorizability and Settings Independence, turns out to be violated by the predictions of QM. Moreover, experimental tests have confirmed its violation in a way predicted by QM (starting with Aspect et al. 1982; for an overview see Myrvold et al. 2024, sections 4 and 5). Therefore, we need to conclude that at least one of our assumptions (which are, to recall, Factorizability and Settings Independence) is false. Most commentators, including Bell and Clauser et al. (1969), regard Settings Independence as well justified on independent grounds, so their conclusion is that Factorizability is false. However, according to Bell (1990), the proper conclusion should be that Local Causality is false, which indeed follows if Factorizability can be derived from Local Causality. The aim of the rest of this paper will be to investigate how one can retain Bell’s Conclusion.

In the above considerations, we did not specify to what exactly the hidden state λ is ascribed (i.e., what the bearer of this state is). It turns out that, in the literature, there are two main ways of thinking about λ : it is regarded either as a state of the two-particle system at the moment of preparation, or as a state of the thick slices of the past light cones of measurements. The next two sections, 6 and 7, will analyse these two options in more detail.

6 Hidden state as a state of the two-particle system at the moment of preparation

The first option for the choice of λ is to take it to be the state of the measured system *only*. For example, Myrvold et al. (2024) write:

This conceptual framework first of all postulates an ensemble of pairs of systems, the individual systems in each pair being labeled as 1 and 2. Each pair of systems is characterized by a “complete state” λ which contains the entirety of the properties of the pair at the moment of generation.

We will denote this state by $\lambda_{t_P,1,2}$, to be read as “the hidden state of the system consisting of particles 1 and 2 at the time of preparation of this system for Bell’s Experiment, t_P ”. This expression should not

⁷ Technically, Settings Independence is a statement about the probability (density) of λ conditioned on settings, not the other way around. So, literally it says that the probability of λ does not depend on settings, not that the probability of settings does not depend on λ . However, this does not mean that the violation of this condition would be a causal influence of settings on the hidden state (such an influence, if it occurred, would be retrocausal because the hidden state is provided at a time *before* the choice of settings). This is just a probabilistic independence condition, which is, as such, symmetric, so the direction of conditionalization in the formula does not need to be aligned with the direction of causation in the case of the violation of this condition (cf. footnote 12).

⁸All page references to Bell’s papers are from the 2004 reprint.

be understood as presupposing the existence of localised, classical-like particles; “the system consisting of particles 1 and 2” might be a physical system of any kind that corresponds, in a given HVT, to what we ordinarily recognise as two particles on which the measurement is performed. What matters for our purposes is that $\lambda_{t_P,1,2}$ is the state of the measured system at the moment of preparation, which is located in the intersection of the light cones of measurements (see Fig. 4). I prefer the term “preparation” over “generation” because the latter suggests that the particles did not exist before t_P , which might or might not be the case. This way of thinking about the hidden state agrees with “early” Bell (1964, p. 15 and 1971, p. 36) and with many of his followers and commentators, including Clauser et al. (1969, p. 881), Shimony (1993, p. 130), van Fraassen (1982, p. 32) and many others. It seems that identifying λ in Bell’s theorem with $\lambda_{t_P,1,2}$ is predominant in the literature, but this is difficult to assess due to the large size of that literature. The idea behind this choice of λ is that the correlations between the outcomes of measurements on remote particles can be explained by the particles’ interaction in the past, the results of which are encoded in the subsequent states of each of them.

The problem with this choice of λ is that the explanation of correlations between measurement outcomes by the properties of the measured system at the time of preparation does not exhaust all possibilities for local explanation of such correlations. Only in some special cases can we be sure that there are no other possible local explanations. In True Spin Theory, according to which hidden variables are the true values of spin in every direction possessed by each particle, it is arguably sufficient to consider $\lambda_{t_P,1,2}$. This is the case at least if we assume that (i) the measurement of spin reveals its true value and (ii) this value does not change between the time of the preparation of the system and the time of measurement. Assumption (i) seems to be justified just by the definition of spin measurement (at least under True Spin Theory’s assumption that every particle has a definite spin all the time), whereas assumption (ii) follows from assuming that the system is appropriately isolated between its preparation and measurement. The latter does not mean that the state of the system does not change at all during this interval; in particular, particles are supposed to move (from the source to the detector), so their positions *do* change during that interval (or rather, what changes is the wave function that prescribes the probabilities of particles being found in various positions if measured, unless we consider a variant of True Spin Theory that postulates that particles have definite positions in addition to definite values of spin in all directions). What is required to not change between the time of preparation and the time of measurement is only the particles’ true values of spin. If we accept these assumptions, then probabilities of the form $P(A|a, \lambda_{t_P,1,2})$ and $P(B|b, \lambda_{t_P,1,2})$ all have values 0 or 1. This is because the outcome of spin measurement should only depend on the direction in which we measure spin and the particle’s true value of spin in that direction; however, this true value is the same at the time of measurement as it was at the time of preparation.

However, assumptions (i) and (ii) cease to be valid if hidden variables are different from the true values of spin. An example here is Bohmian Mechanics. In this theory, it is not the case that the hidden state of the two-particle system at the time of preparation together with the choice of the direction of spin measurement uniquely determine the outcome of that measurement; moreover, the hidden state of the two-particle system will change between the time of preparation and the time of measurement. That hidden state, to recall, is the true positions of the two particles; since these particles need to move from the source to the detector, $\lambda_{t_P,1,2}$ will be different from $\lambda_{t_M,1,2}$, where t_M is the time of measurement.⁹ Moreover, even though $\lambda_{t_M,1,2}$ is uniquely determined by λ_{t_P} (because Bohmian mechanics is deterministic), it is not uniquely determined by $\lambda_{t_P,1,2}$ or even by $\lambda_{\Sigma_{R_A,t_P} \cup \Sigma_{R_B,t_P}}$ (because Bohmian mechanics is not locally deterministic).¹⁰ Therefore, by our Bridge Principle Between Determinism and 0-1 Probability (Definition 3), in Bohmian mechanics it is *not* the case that probabilities $P(A|a, \lambda_{t_P,1,2})$ and $P(B|b, \lambda_{t_P,1,2})$ have values 0 or 1. This means that

⁹ There are several subtleties related to t_M that will not matter for our discussion in the main text. First, it will typically not be the case that both measurements happen at exactly the same time (in a given reference frame). To avoid ambiguity, we can define t_M to be the time of the first of the two measurements (or of the last of them). Second, in different reference frames the temporal relation between the two measurements will be different. However, now we are working with a fixed reference frame.

¹⁰ I consider here the version of Bohmian mechanics according to which the wave function is not a beable but has a lawlike status because the version including the wave function as a beable is not locally determinate (see section 4.1). The failure of local indeterminism is due to the fact that in Bohmian mechanics the motion of a given particle can be influenced by the positions of arbitrarily remote particles.

Bohmian mechanics is not outcome deterministic *despite being deterministic*.

In general, $P(A|a, \lambda_{t_P,1,2})$ and $P(B|b, \lambda_{t_P,1,2})$ having values 0 or 1 in accordance with T is not guaranteed by T being deterministic. If T is locally deterministic, then $\lambda_{t_M,1,2}$ is uniquely determined by $\lambda_{\Sigma_{R_A,t_P} \cup \Sigma_{R_B,t_P}}$ but still not by $\lambda_{t_P,1,2}$.¹¹

Let us use “Factorizability $_{t_P,1,2}$ ” and “Settings Independence $_{t_P,1,2}$ ” to denote the conditions of Factorizability and Settings Independence introduced in section 5, in which “ λ ” is substituted with “ $\lambda_{t_P,1,2}$ ”. Settings Independence $_{t_P,1,2}$ seems very plausible: its violation would mean that the state of the measured system is correlated with the choices of measurement settings, so for some states of the system we cannot make or at least are less likely to make some types of measurements on it. If $\lambda_{t_P,1,2}$ lies in the causal past of R_a and R_b , it *can* influence the events that happen in these regions; however, the point is that there are so many *other* events in the causal past of the choices of settings that the influence of $\lambda_{t_P,1,2}$ becomes insignificant (unless there exists some specific causal linkage between the preparation of the system and the choice of settings, which seems unlikely; cf. Shimony et al. 1976). It is also possible (at least in principle) to delay the preparation of the system so that it is spacelike related to the choices of measurement settings (see, e.g., Müller and Placek 2023, Fig. 6), in which case nothing that happens in regions R_a and R_b can be influenced by $\lambda_{t_P,1,2}$ in a locally causal way. Some authors regard the violation of *this* variant of Settings Independence (i.e., Settings Independence $_{t_P,1,2}$) as an option worth considering,¹² but for now let us assume that it holds. Then, from Bell’s theorem it follows that there is no HVT that reproduces the predictions of QM and satisfies Factorizability $_{t_P,1,2}$. However, this does not answer the question of whether there exists an HVT that reproduces the predictions of QM and satisfies Factorizability with a different choice of λ (i.e., encompassing more than the measured system but not confined to the past light cones of R_A and R_B , so that it remains relevant for local explanation of quantum correlations). Moreover, it is not clear how Factorizability $_{t_P,1,2}$ could be derived from Local Causality and why Factorizability with this particular choice of λ (and not some other) should be derivable from Local Causality. However, such a derivation (for some λ) is indispensable for establishing Bell’s Conclusion.

At this point, one could respond that the proof of Bell’s theorem works in the same way, no matter what λ exactly is, so the verdict for *any* variant of Factorizability will be the same as for Factorizability $_{t_P,1,2}$. However, we will see that for some choices of λ , Settings Independence becomes less plausible. In particular, it will be shown that for λ that encompasses full thick slices of the past light cones of measurements, any locally deterministic theory violates Settings Independence.

7 Hidden state as a state of thick slices of the past light cones of measurements

The second option for the choice of λ is to take it to be the state of thick slices of the past light cones of measurements. This option was favoured by late Bell, and it is analysed in this section.

¹¹In the superdeterministic models LPW_f, C-LPW_f, LPW_p and C-LPW_p developed by Ciepielewski et al. (2021), which are locally deterministic, $\lambda_{t_M,1,2}$ is uniquely determined by $\lambda_{t_P,1,2}$. However, this is not a typical feature of locally deterministic theories. The mentioned models possess this feature because, in these models, any point (\vec{x}, t) is associated with a copy of the entire state of the universe at t , which evolves in accordance with the dynamics of Bohmian mechanics. Since the latter is deterministic, the laws of these models together with the state at any single spacetime point determine the evolution of the entire universe (under an additional assumption that the copies of the entire state of the universe are identical at every point).

¹²There are two variants of this view, depending on the direction of causal influence. According to superdeterminism, causation is directed towards the future (i.e., either $\lambda_{t_P,1,2}$ influences the settings, or something in their common past is responsible for the correlation between them). According to retrocausal approaches, causal influence travels into the past from settings to $\lambda_{t_P,1,2}$. Superdeterminism is considered, for example, by Hossenfelder and Palmer (2020) and Ciepielewski et al. (2021); retrocausality is considered, for example, by Price (1994) and Adlam (2022) and reviewed by Wharton and Argaman (2020) and Friederich and Evans (2023).

7.1 Late Bell’s concept of Local Causality

Bell (1976, p. 53) introduces the concept of Local Causality as the generalisation of the concept of local determinism that is also applicable to indeterministic theories. In his last paper, Bell (1990, p. 239) formulates this principle as follows:

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

He then provides a more precise formulation of it (1990, pp. 239–240):

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 local beables in a space-time region 3 (Fig. 4).



Fig. 4. Full specification of what happens in 3 makes events in 2 irrelevant for predictions about 1 in a locally causal theory.

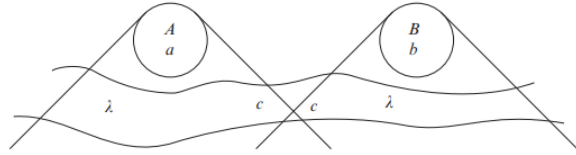


Fig. 6. Diagram for CHSH inequality derivation.

Figure 2: From Bell (1990, pp. 240 and 242).

To this formulation, Bell adds the following commentary (Bell, 1990, p. 240):

It is important that region 3 completely shields off from 1 the overlap of the backward light cones of 1 and 2. And it is important that events in 3 be specified completely. Otherwise the traces in region 2 of causes of events in 1 could well supplement whatever else was being used for calculating probabilities about 1. The hypothesis is that any such information about 2 becomes redundant when 3 is specified completely.

This is in striking disagreement with the choice of λ as $\lambda_{t_P,1,2}$, which lies in the overlap of the past light cones of R_A and R_B , that is, Bell’s “overlap of the backward light cones of 1 and 2” (cf. Fig. 4).¹³

Here is a proposal for how to formulate this condition mathematically (cf. a similar formulation in Norsen 2009, p. 277):

Definition 4. Local Causality $_{C(\Sigma,\Sigma)}$, fine-grained version: For any bounded region R and any bounded region R' that is space-like related to it, for any $t, t' < R$ that lie above the intersection of the past light cones of R and R' and such that $t' < t$, and for any $\lambda_R, \lambda_{R'}$ and $\lambda_{C(\Sigma_{R,t}, \Sigma_{R,t'})}$,¹⁴

$$P(\lambda_R | \lambda_{C(\Sigma_{R,t}, \Sigma_{R,t'})}, \lambda_{R'}) = P(\lambda_R | \lambda_{C(\Sigma_{R,t}, \Sigma_{R,t'})}). \quad (2)$$

¹³That Bell’s formulation of his theorem and its assumptions changed over time has been recognised in the literature (see, e.g., Wiseman 2014 and Brown and Timpson 2016). However, the difference that is most often pointed out is that early Bell was concerned only with local HVTs that are “deterministic” in the sense of predicting the outcomes of spin measurements uniquely (and, relatedly, he considered only the quantum states that lead to the prediction of perfect (anti-)correlations, since these are the states relevant for the EPR argument), whereas later he wanted to also encompass local HVTs that give only probabilistic predictions. Here, I have assumed from the start that Bell’s theorem should be applicable to theories of both types; the differences between early and late Bell that I discuss are only those regarding the localisation of λ .

¹⁴The quantification over λ_R is the quantification over all hidden states of a given HVT that can be ascribed to region R (other quantifiers are understood analogously). To make this fully rigorous, we should proceed as follows: define function λ from the set of regions to all hidden states of a given HVT and replace “for all λ_R ” with “for all $\lambda_1 \in \lambda(R)$ ”.

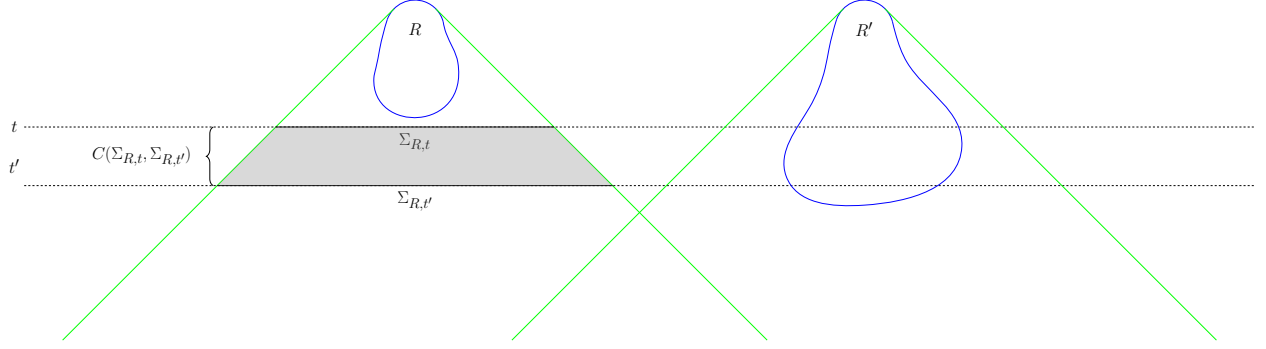


Figure 3: Relations between the regions mentioned in the formulation of Local Causality $C(\Sigma, \Sigma)$. In the proof of Factorizability $C(\Sigma, \Sigma)$, we substitute for R either region of the choice of settings (i.e., R_a/R_b) or the region of measurement together with the region of the choice of settings (i.e., $R_A \cup R_a/R_B \cup R_b$). For details see Appendix A.2.

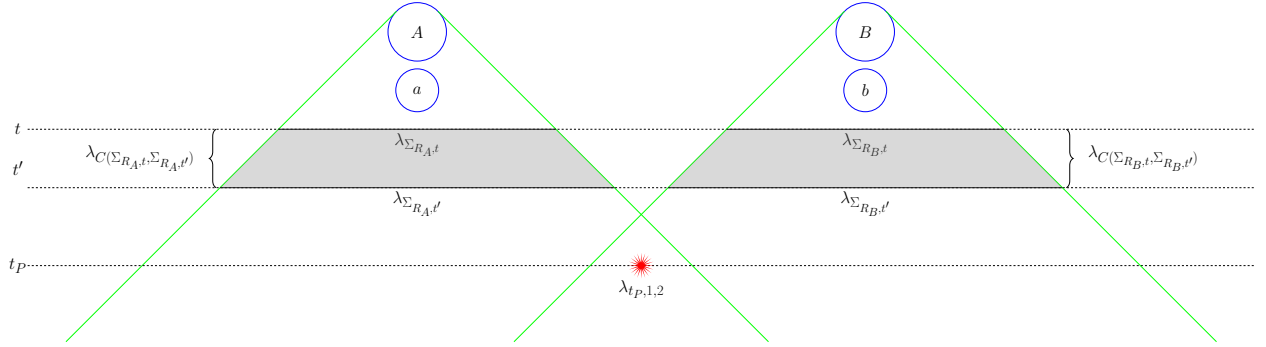


Figure 4: Comparison of two different choices of λ in Bell's theorem, $\lambda_{t_p,1,2}$ vs. $\lambda_{C(\Sigma, \Sigma)}$.

The respective regions are depicted in Fig. 3. Regions R and R' are assumed to be bounded but not necessarily connected, which will be exploited later (see Appendix A.2). Notice that in concrete calculations this principle can be used in two ways: either “from left to right” (i.e., we can remove conditionalization on what happens in R') or “from right to left” (i.e., we can harmlessly add conditionalization on what happens in R').

The main difference with Bell's (1990) formulation is that the above formulation uses a fixed slicing in order to define the region that is supposed to screen off the region R_A from the region R_B . This might seem to be in tension with relativity, which rejects the existence of a distinguished slicing; however, if we add the requirement that Local Causality $C(\Sigma, \Sigma)$ should be satisfied in any slicing, then the tension disappears. We could also consider regions of more “wavy” shapes, as depicted in Fig. 2, but I prefer using $\Sigma_{R,t}$ to define these regions because this will make it easier to relate Local Causality $C(\Sigma, \Sigma)$ to locally deterministic theories.

The above version is not quite suitable for our purposes because all arguments of the probability function are complete specifications of hidden variables (beables) in respective regions, whereas we are interested in representing more coarse-grained specifications. For example, we are not interested in all differences between possible specifications of beables in the region R_a (i.e., possible λ_{R_a} 's), so we should consider their equivalence classes (each of which corresponds to one choice of measurement setting in region R_a) instead of single such specifications. In order to capture this, we will introduce a further piece of notation. Let d_a be a (coarse-grained) description of an event that can happen in a region R_a that can be formulated within a given HVT. Then, by $\overline{\lambda_{R_a; d_a}}$ we denote the set of all and only those λ_{R_a} 's in which d_a is true. Notice that since we use d_a to define $\overline{\lambda_{R_a; d_a}}$, not all sets of states count as coarse-grained states but only those definable within the theory. This leads to another version of Local Causality:

Definition 5. Local Causality $_{C(\Sigma, \Sigma)}$, coarse-grained version: For any bounded region R and any bounded region R' that is space-like related to it, for any $t, t' < R$ that lie above the intersection of the past light cones of R and R' and such that $t' < t$, for any $\lambda_R, \lambda_{R'}$ and $\lambda_{C(\Sigma_{R,t}, \Sigma_{R,t'})}$, and for any coarse-grainings $\overline{\lambda_{R,d_R}}$ and $\overline{\lambda_{R',d_{R'}}}$,

$$P(\overline{\lambda_{R,d_R}} | \lambda_{C(\Sigma_{R,t}, \Sigma_{R,t'})}, \overline{\lambda_{R',d_{R'}}}) = P(\overline{\lambda_{R,d_R}} | \lambda_{C(\Sigma_{R,t}, \Sigma_{R,t'})}). \quad (3)$$

Notice that all complete specifications of beables are now replaced by coarse-grained ones, except for the region $C(\Sigma_{R,t}, \Sigma_{R,t'})$, where we want to retain a complete specification of beables, in agreement with Bell. For any coarse-grained state, what matters are not the probabilities associated with complete states that fall under it, but only the probabilities associated with the entire equivalence class. In our description of experiments, instead of the above complex notation for coarse-grained states, we will use the values of chosen settings and the values of obtained outcomes, which will be denoted, as previously, by a, b, A, B etc.

It can be shown (see Appendix A.2) that Local Causality $_{C(\Sigma, \Sigma)}$ (in the coarse-grained version) entails Factorizability $_{C(\Sigma, \Sigma)}$, that is,

$$P(A, B | a, b, \lambda_{C(\Sigma, \Sigma)}) = P(A | a, \lambda_{C(\Sigma, \Sigma)}) \cdot P(B | b, \lambda_{C(\Sigma, \Sigma)}), \quad (4)$$

where $\lambda_{C(\Sigma, \Sigma)}$ denotes the pair consisting of $\lambda_{C(\Sigma_{R_A,t}, \Sigma_{R_A,t'})}$ and $\lambda_{C(\Sigma_{R_B,t}, \Sigma_{R_B,t'})}$.

We can also define a variant of Settings Independence for our choice of hidden variables:

Definition 6. Settings Independence $_{C(\Sigma, \Sigma)}$:

$$P(\lambda_{C(\Sigma, \Sigma)} | a, b) = P(\lambda_{C(\Sigma, \Sigma)}). \quad (5)$$

This completes our formulation of the assumptions of Bell's theorem for hidden states encompassing entire thick slices in the past light cones of measurement regions R_A and R_B .

7.2 Locally deterministic theories vs. Local Causality $_{C(\Sigma, \Sigma)}$ and Settings Independence $_{C(\Sigma, \Sigma)}$

One can show that any locally deterministic HVT satisfies Local Causality $_{C(\Sigma, \Sigma)}$ but violates Settings Independence $_{C(\Sigma, \Sigma)}$. This is the content of the following two theorems (for proofs see Appendix A.3):

Theorem 2. Any locally deterministic HVT satisfies Local Causality $_{C(\Sigma, \Sigma)}$.

Theorem 3. Any locally deterministic HVT violates Settings Independence $_{C(\Sigma, \Sigma)}$.

In fact, all locally deterministic theories violate even a much weaker condition, which might be called Settings Compatibility $_{C(\Sigma, \Sigma)}$ and which states that any $\lambda_{C(\Sigma, \Sigma)}$ is compatible with any choice of settings in $R_a \cup R_b$. In locally deterministic theories, any $\lambda_{C(\Sigma, \Sigma)}$ is compatible with *exactly one* choice of such settings.

One can object to the above reasoning by pointing out that it relies on the following two additional implicit assumptions:

Physical Nature of Human Choices: Human choices are physical processes.

Universality of Laws: The laws of T apply to all physical processes.

If the Universality of Laws holds, then all physical processes must conform to the laws of T . If, in addition, the Physical Nature of Human Choices holds, then, in particular, the choice of measurement settings by human experimenters must conform to the laws of T . This is precisely what we assumed implicitly in the proof of Theorem 3 when we applied the laws of T to the physical contents of regions R_a and R_b . Therefore, our conclusion that any locally deterministic theory violates Settings Independence $_{C(\Sigma, \Sigma)}$ can be avoided by rejecting one of these assumptions.

If we reject the Universality of Laws (while retaining the Physical Nature of Human Choices), then a locally deterministic theory can be made consistent with Settings Independence $_{C(\Sigma, \Sigma)}$, provided that the

domain of application of the laws of T is appropriately restricted.¹⁵ More specifically, we need to assume that T does not apply to some types of physical processes, which include choices of measurement settings. Importantly, rejecting the Universality of Laws does not need to mean that we just create a gap in the natural order to make room for humans’ freedom as there might be other types of physical processes that are not subject to the laws of T , some of which might not be in any way related to humans.

The drawback of this strategy is that by rejecting the Universality of Laws we also block our proof that any locally deterministic theory satisfies Local Causality $_{C(\Sigma, \Sigma)}$. This is because this proof assumes that the laws of such a theory apply to all physical processes that happen in the region consisting of $R_A \cup R_B$ and the part of the past light cones of R_A and R_B lying above $\Sigma_{R_A, t} \cup \Sigma_{R_B, t}$ (including $\Sigma_{R_A, t} \cup \Sigma_{R_B, t}$ itself). Therefore, this strategy is not helpful for establishing Bell’s Conclusion.

We can also reject the Physical Nature of Human Choices while retaining the Universality of Laws. The problem with this option is that in most actual realisations of Bell’s Experiment, the measurement settings are not chosen by humans but by means of some physical processes (an exception is an experiment done by The BIG Bell Test Collaboration, see Abellán et al. 2018). This would mean that only experiments involving human choices are actual tests of Local Causality and all others are not, which might seem surprising given that both types of experiments yield the same results (namely, violation of the Bell-CHSH inequality in agreement with QM predictions).

Some statements made by Bell himself might be interpreted as the rejection of either the Physical Nature of Human Choices or the Universality of Laws (or both). For example, Bell (1976, p. 61) writes:

It has been assumed that the settings of instruments are in some sense free variables—say at the whim of experimenters—or in any case not determined in the overlap of the backward light cones.

In a subsequent paper, he elaborates on this fragment as follows (Bell, 1977, p. 101):

A respectable class of theories, including contemporary quantum theory as it is practised, have ‘free’ ‘external’ variables in addition to those internal to and conditioned by the theory. These variables are typically external fields or sources. They are invoked to represent experimental conditions. They also provide a point of leverage for ‘free willed experimenters’, if reference to such hypothetical metaphysical entities is permitted. I am inclined to pay particular attention to theories of this kind [...]

If the mentioned external parameters are irreducibly external in the sense that they cannot be explained by applying the theory to a larger system, then this indeed looks like a violation of the Universality of Laws.

However, soon after, Bell weakens the assumption that the experimental settings are free to the assumption that they are effectively or sufficiently free “for the purpose at hand”. He invokes the example of a deterministic number generator (Bell 1977, pp. 102–103; this is also the line that he takes in his last paper—see Bell 1990, p. 244):

A particular output is the result of combining so many factors, of such a lengthy and complicated dynamical chain, that it is quite extraordinarily sensitive to minute variations of any one of many initial conditions. It is the familiar paradox of classical statistical mechanics that such exquisite sensitivity to initial conditions is practically equivalent to complete forgetfulness of them. To illustrate the point, suppose that the choice between two possible outputs, corresponding to a and a' depended on the oddness or evenness of the digit in the millionth decimal place of some input variable. Then fixing a or a' indeed fixes something about the input—i.e., whether the millionth digit is odd or even. But this peculiar piece of information is unlikely to be the vital piece for any distinctively different purpose [...]

¹⁵This seems to be somehow similar to the approach of Müller and Placek (2023), with their distinction between nature-induced indeterminism and agent-induced indeterminism. However, this is not quite the same because they do not talk about laws.

Bell does not specify which particular “distinctively different purpose” is relevant for establishing Bell’s Conclusion. If we take it to be “rendering the far-away outcome B and setting b redundant for the task of determining the probabilities of obtaining A ” (Seevinck and Uffink, 2011, p. 445), then we obtain the approach that will be discussed in section 8.2.

7.3 Further discussion of the violation of Settings Independence $_{C(\Sigma, \Sigma)}$ in locally deterministic theories

The fact that all locally deterministic theories violate Settings Independence $_{C(\Sigma, \Sigma)}$ (and even a much weaker condition of Settings Compatibility $_{C(\Sigma, \Sigma)}$) might be worrying because Settings Independence is sometimes argued to be indispensable for experimental practice. If these arguments were able to establish that Settings Independence $_{C(\Sigma, \Sigma)}$ holds, then we could exclude all locally deterministic theories without using Bell’s theorem, merely on the basis of Theorem 3. However, under closer scrutiny such typical arguments against Settings Independence turn out to be much less convincing in the case of Settings Independence $_{C(\Sigma, \Sigma)}$ than in the case of Settings Independence $_{t_P, 1, 2}$.

Let us briefly look at four such arguments. The first argument is that Settings Independence is indispensable for the procedure of randomization in empirical sciences such as medicine. Goldstein et al. (2011) give the following example:

[...] if you are performing a drug versus placebo clinical trial, then you have to select some group of patients to get the drug and some group of patients to get the placebo. The conclusions drawn from the study will necessarily depend on the assumption that the method of selection is independent of whatever characteristics those patients might have that might influence how they react to the drug.

However, they clearly have in mind Settings Independence $_{t_P, 1, 2}$ rather than Settings Independence $_{C(\Sigma, \Sigma)}$. This is visible in another fragment of Goldstein et al. (2011) (emphasis mine):

in practice, one assesses the applicability of [Settings Independence] to a given experiment by examining the care with which the experimental design precludes any non-conspiratorial dependencies *between the preparation of the systems and the settings of instruments*.

Moreover, medicine, biology and social sciences are usually or even always interested in coarse-grained variables, not in full specifications of the physical state of the systems they are investigating, whereas Settings Independence $_{C(\Sigma, \Sigma)}$ asserts the dependence of the choice of settings on the *full specification* of the state in $C(\Sigma, \Sigma)$. Even if there is a correlation of the form $P(\text{placebo}|\lambda_{\text{patient}_1, t}) \neq P(\text{placebo}|\lambda_{\text{patient}_2, t})$, where $\lambda_{\text{patient}, t}$ is the full specification of the physical state of the patient at the time when the assignment to drug/placebo groups is made, this does not undermine the results of such studies because $\lambda_{\text{patient}, t}$ contains a lot of information that is causally irrelevant for the process of patients’ recovery.

The second argument is that “it is wrong on methodological grounds to worry seriously about [violation of Settings Independence] if no specific causal linkage is proposed” (Shimony et al. 1976; cited from Shimony 1993, p. 168). Again, these authors clearly have in mind Settings Independence $_{t_P, 1, 2}$ and their argument cannot be extended to Settings Independence $_{C(\Sigma, \Sigma)}$. In the latter case, we do not have a “causal linkage”, but the dependence of the choice of settings on $\lambda_{C(\Sigma, \Sigma)}$ is guaranteed by the laws of the locally deterministic theory.

The third argument, which also has its roots in the paper by Shimony et al. (1976), is that rejecting Settings Independence amounts to abandoning “in advance the whole enterprise of discovering the laws of nature by experimentation” (cited from Shimony 1993, p. 168). The issue of discovering physical laws is more relevant for the discussion of Settings Independence $_{C(\Sigma, \Sigma)}$ because we test hypotheses about such laws by preparing a system in some initial state and checking whether its evolution agrees with that prescribed by these hypothetical laws, so the full specification of a physical state (and not merely a coarse-grained one) becomes important. In order to learn about the laws of nature from such experiments, it has to be possible to measure any aspect of the final state of the system, no matter what its initial state is (and perhaps also

our likelihood of measuring this particular aspect should not depend on the initial state, although this is less clear). If this condition were violated, then it could happen that for any initial state we cannot measure exactly those aspects of the final state which disagree with these hypothetical laws. However, as before, this requires only Settings Independence $_{t_P,1,2}$ and not Settings Independence $_{C(\Sigma,\Sigma)}$.¹⁶

The fourth argument is that violation of Settings Independence “requires an atypical fine-tuning of the initial state of the universe” (see, e.g., Baas and Bihan 2023, p. 10). Again, this might be true for Settings Independence $_{t_P,1,2}$ (although one can construct a theory in which the correlations between settings and $\lambda_{t_P,1,2}$ are lawlike and, as such, independent of the initial state; see, e.g., Ciepielewski et al. 2021) but not for Settings Independence $_{C(\Sigma,\Sigma)}$. This is reflected in the fact that our proof of Theorem 3 does not make any assumptions about the initial state of the universe.

8 Attempts at improvement

The results of sections 6 and 7 pose a serious threat to Bell’s Conclusion. We have considered two possible choices of λ , neither of which leads straightforwardly to any conclusion about Local Causality $_{C(\Sigma,\Sigma)}$:

- If we consider $\lambda = \lambda_{t_P,1,2}$, then we do not take into account other possible local explanations of correlations between measurement outcomes. Moreover, the relationship between Factorizability $_{t_P,1,2}$ and Local Causality $_{C(\Sigma,\Sigma)}$ is not clear.
- The choice of $\lambda = \lambda_{C(\Sigma,\Sigma)}$ might seem better because Factorizability $_{C(\Sigma,\Sigma)}$ follows from Local Causality $_{C(\Sigma,\Sigma)}$, but then we cannot run the proof of Bell’s theorem for locally deterministic theories (which form an important subclass of locally causal theories) because they violate Settings Independence $_{C(\Sigma,\Sigma)}$ (at least if we assume the Universality of Laws and the Physical Nature of Human Choices).

As far as I can see, there are four possible ways to try to recover Bell’s Conclusion in the face of these difficulties:

- just regard Factorizability $_{t_P,1,2}$ rather than Local Causality $_{C(\Sigma,\Sigma)}$ as the proper formulation of Local Causality;
- instead of complete states $\lambda_{C(\Sigma,\Sigma)}$, use states $\bar{\lambda}$ specified incompletely (so that Settings Independence $_{\bar{\lambda}}$ is satisfied) but sufficiently (so that Factorizability $_{\bar{\lambda}}$ is satisfied);
- derive Factorizability $_{t_P,1,2}$ from Local Causality $_{C(\Sigma,\Sigma)}$, perhaps using some additional but plausible assumptions;
- consider some different localisations of hidden states.

The first three options will be analysed in the following subsections. Notice that the first and third options regard $\lambda_{t_P,1,2}$ as the proper choice of λ for Bell’s theorem; the second option uses $\lambda_{C(\Sigma,\Sigma)}$ but takes a coarse-graining of it, whereas the last option looks for yet another choice of λ .

8.1 Local Causality is Factorizability $_{t_P,1,2}$

Why can we not be content with showing that Factorizability $_{t_P,1,2}$ is false? Can this principle (instead of Local Causality $_{C(\Sigma,\Sigma)}$) be regarded as the proper formulation of Local Causality? Bell (1990, p. 243) was clearly unwilling to accept this, since he writes:

Very often such factorizability is taken as the starting point of the analysis. Here we have preferred to see it not as the formulation of ‘local causality’, but as a consequence thereof.

¹⁶There are other problems with testing HVTs in this way; for example, for some such theories we might not be able to know the complete initial and final state of the system sufficiently well to test the predictions derived from the hypothetical laws. However, this problem has nothing to do with Settings Independence.

He does not provide an in-depth justification for this conceptual choice, but I believe that he had good reasons for it.

Bell’s focus on Local Causality $_{C(\Sigma, \Sigma)}$ seems to be in line with his broader views about physics. Factorizability $_{t_P, 1, 2}$ is a condition that is applicable only to situations of a very special type—namely, the instances of Bell’s Experiment (and other appropriately similar experiments). This is because, in its formulation, Factorizability $_{t_P, 1, 2}$ involves concepts such as measurement settings and measurement outcomes, and presupposes a certain structure of the experiment. In contrast, Local Causality $_{C(\Sigma, \Sigma)}$ is a very general physical principle—in fact, it is applicable to any physical situation. Moreover, Local Causality $_{C(\Sigma, \Sigma)}$ does not involve in its formulation the concept of measurement, and one needs to remember that for Bell, the use of the concept of measurement in the formulation of the Copenhagen interpretation of QM was its main disadvantage (cf. Bell 1989). For this reason, Local Causality $_{C(\Sigma, \Sigma)}$ seems more relevant for the foundations of physics than Factorizability $_{t_P, 1, 2}$.

Another possible objection to Factorizability $_{t_P, 1, 2}$ is that it relies not only on the idea of locality but also on the common cause principle (cf. Barandes 2024, pp. 8–9). This principle, which was formulated by Reichenbach, asserts that whenever events A and B are correlated, that is,

$$P(A, B) \neq P(A) \cdot P(B), \quad (6)$$

then there exists another event C such that

$$P(A, B|C) = P(A|C) \cdot P(B|C) \quad (7)$$

(for the full statement, see Hitchcock and Rédei 2021).

In contradistinction to Factorizability $_{t_P, 1, 2}$, Local Causality $_{C(\Sigma, \Sigma)}$ does not involve the concept of common cause in any way. In Definitions 4 and 5, $\lambda_{C(\Sigma_{R,t}, \Sigma_{R,t'})}$ is not supposed to be a common cause of some two events but a cause of events in region R . The exact justification that formulas (2) and (3) correctly capture the intuitive idea of local causality is another matter. One possible justification relies on analogical reasoning. In the case of locally deterministic theories, state $\lambda_{C(\Sigma_{R,t}, \Sigma_{R,t'})}$ uniquely determines state λ_R , so information about state in any other region is redundant for λ_R given $\lambda_{C(\Sigma_{R,t}, \Sigma_{R,t'})}$. Now, Local Causality $_{C(\Sigma, \Sigma)}$ is supposed to be the generalisation of this idea to theories that are locally causal (in an intuitive sense) but not necessarily locally deterministic. If we replace “unique determination of state” with “unique determination of the *probability of state*”, we obtain Definitions 4 and 5 (but now we must remember that information about events lying within the light cone of R after t might *not* be redundant).

8.2 The idea of an incomplete but sufficient specification of hidden states

The idea of considering coarse-grained versions of $\lambda_{C(\Sigma, \Sigma)}$ arises from Bell’s expression “what happens in the backward light cone of 1 is already sufficiently specified, for example by a full specification of local beables in a space-time region 3” (cf. the full quote in section 7.1). We obtained the result that any locally deterministic theory violates Settings Independence $_{C(\Sigma, \Sigma)}$ because we have taken λ to be the *complete* specification of the physical contents of the region $C(\Sigma, \Sigma)$. However, Bell might be understood as saying only that λ should be a *sufficient* specification; it might be, for example, a complete specification, but perhaps some incomplete specifications are sufficient as well. This idea appears in Norsen’s (2009, p. 283 and 2011, p. 1266) analysis of Bell and was taken up by Seevinck and Uffink (2011). The latter authors propose a precisification of the concept of sufficiency. The crucial question is, of course, for what purpose should the specification of λ be sufficient? Their answer is that it should be sufficient “for the purpose of rendering the far-away outcome B and setting b redundant for the task of determining the probabilities of obtaining A ” (Seevinck and Uffink, 2011, p. 445). This means that we should consider λ ’s that satisfy the following condition (cf. Seevinck and Uffink 2011, p. 445):

$$P(A|a, b, B, \bar{\lambda}) = P(A|a, \bar{\lambda}) \text{ and } P(B|a, b, A, \bar{\lambda}) = P(B|b, \bar{\lambda}). \quad (8)$$

Clearly, something less than the full $\lambda_{C(\Sigma, \Sigma)}$ might satisfy this condition. For example, in a locally deterministic theory,

$$\lambda = \left\{ \left\{ \lambda_{C(\Sigma_{R_A, t}, \Sigma_{R_A, t'})}, \lambda_{C(\Sigma_{R_B, t}, \Sigma_{R_B, t'})} \right\}, \left\{ \lambda'_{C(\Sigma_{R_A, t}, \Sigma_{R_A, t'})}, \lambda'_{C(\Sigma_{R_B, t}, \Sigma_{R_B, t'})} \right\} \right\} \quad (9)$$

where $\lambda_{C(\Sigma_{R_A,t},\Sigma_{R_A,t'})}$ is consistent with a , $\lambda_{C(\Sigma_{R_B,t},\Sigma_{R_B,t'})}$ is consistent with b , $\lambda'_{C(\Sigma_{R_A,t},\Sigma_{R_A,t'})}$ is consistent with $a' \neq a$ and $\lambda'_{C(\Sigma_{R_B,t},\Sigma_{R_B,t'})}$ is consistent with $b \neq b'$, is sufficient in this sense. This example also shows that incomplete specifications of the physical contents of some region R can be constructed by taking classes of λ_R 's instead of separate such states.¹⁷ The hope is that we can impoverish our complete states (by replacing them with classes of such states) so that they will still be sufficient, but by omitting information we will lose the correlation between these states and the choice of settings.

There are two problems with the above concept of sufficient specification. The first is that, in general, we do not even know whether the complete specification is sufficient! In a non-local theory, the complete specification of beables in the region $C(\Sigma,\Sigma)$ is not sufficient in the above sense—this is the whole point of the theory being non-local. Therefore, *a fortiori*, we cannot exclude that no incomplete specification is sufficient. We need to phrase our question differently: Is it the case that for any HVT that satisfies Local Causality $_{C(\Sigma,\Sigma)}$, there exists a coarse-graining of Λ (call it $\bar{\Lambda}$), such that the elements of $\bar{\Lambda}$ satisfy sufficiency condition (8) and Settings Independence $_{\bar{\lambda}}$? If so, then we can run Bell's theorem with this $\bar{\lambda}$ and show that any such HVT is inconsistent with the predictions of QM.

The second problem with the proposal of Seevinck and Uffink (2011) is that their sufficiency condition is unnecessarily strong. This is because what we really want is to make Bell's theorem applicable to any HVT that satisfies Local Causality $_{C(\Sigma,\Sigma)}$, and for this purpose, it is sufficient that at least one version of Factorizability and Settings Independence (but with the same $\bar{\lambda}$!) is satisfied by this HVT. Therefore, I suggest that instead of (8) we should just use Factorizability $_{\bar{\lambda}}$.

With these amendments, the reasoning one can use to exclude all HVTs that satisfy Local Causality $_{C(\Sigma,\Sigma)}$ as inconsistent with empirically confirmed predictions of QM is as follows. Assume that for any theory satisfying Local Causality $_{C(\Sigma,\Sigma)}$, there exists $\bar{\Lambda} \subseteq \mathcal{P}(\Lambda)$ such that $\bigcup \bar{\Lambda} = \Lambda$, the elements of $\bar{\Lambda}$ are pairwise disjoint, and both Factorizability $_{\bar{\lambda}}$ and Settings Independence $_{\bar{\lambda}}$ hold. Let T be any theory satisfying Local Causality $_{C(\Sigma,\Sigma)}$. Then, there exists $\bar{\Lambda}$ such that both Factorizability $_{\bar{\lambda}}$ and Settings Independence $_{\bar{\lambda}}$ are satisfied (by our assumption). However, Factorizability $_{\bar{\lambda}}$ and Settings Independence $_{\bar{\lambda}}$ entail Bell's inequality, which is violated. Therefore, T cannot be true. Since T was an arbitrary HVT that satisfies Local Causality $_{C(\Sigma,\Sigma)}$, we have excluded in this way all HVTs that satisfy Local Causality $_{C(\Sigma,\Sigma)}$.

The problem with this reasoning is that it is not obvious that for every theory satisfying Local Causality $_{C(\Sigma,\Sigma)}$ at least one such $\bar{\Lambda}$ exists. Could it be the case that whenever we coarse-grain our hidden states so that they satisfy Settings Independence $_{\bar{\lambda}}$, they also cease to satisfy Factorizability $_{\bar{\lambda}}$? It seems that one might as well run the reasoning that is opposite to that in the previous paragraph: if our theory reproduces the predictions of QM, then its coarse-grained states cannot satisfy both Settings Independence $_{\bar{\lambda}}$ and Factorizability $_{\bar{\lambda}}$. How can one break the impasse between these two lines of reasoning?¹⁸

One possible response has been proposed by Goldstein et al. (2011). Although they do not use the terminology of incompleteness and sufficiency, their proposal is effectively the same as the one described above. Instead of talking about coarse-grained versions of $\lambda_{C(\Sigma,\Sigma)}$, they consider $\lambda_{C(\Sigma,\Sigma)}$ as divided into several beables, some of which are irrelevant for the measurement outcome, while the others either influence the outcome only via their influence on measurement settings or in some other way (the latter correspond to our $\bar{\lambda}$). Their solution to our problem is that theories in which $\lambda_{C(\Sigma,\Sigma)}$ cannot be divided in this way are called “conspiratory” and are excluded from the scope of Bell's theorem. However, one might worry that in this approach, Bell's theorem would be inapplicable to an excessively large class of theories, some of which cannot be excluded on independent grounds. The crucial question is whether we have equally strong reasons for Settings Independence $_{\bar{\lambda}}$ as we had for Settings Independence $_{t_P,1,2}$. This is unclear because the arguments

¹⁷This distinguishes my approach from that of Seevinck and Uffink (2011), who prefer to talk in terms of variables instead of in terms of classes of complete states. However, any statements made within the former approach can be translated into statements within the latter approach because any specification of some variables of a theory T in a region R corresponds to a class of possible complete states of region R . Another difference is that Seevinck and Uffink regard a and b as labels of probability functions, while I keep the convention that they are arguments of a single probability function.

¹⁸Hofer-Szabó (2015) argues that it is logically possible that the required coarse-graining exists. However, our question is different: does the required coarse-graining exist for all HVTs that satisfy Local Causality $_{C(\Sigma,\Sigma)}$ (or at least for a large class of them, which leaves outside only those HVTs satisfying Local Causality $_{C(\Sigma,\Sigma)}$ that are implausible for some independent reason)?

for the latter (see section 7.3) concern $\lambda_{t_P,1,2}$, which is the state of the measured system at the moment of preparation, whereas $\bar{\lambda}$ is a much more encompassing state with a different localisation in spacetime, so these arguments cannot be adapted straightforwardly to support Settings Independence $_{\bar{\lambda}}$. However, other equally strong arguments may exist; one argument sketched by Bell (see the last quotation in section 7.2) can be read as supporting Settings Independence $_{\bar{\lambda}}$.

8.3 Attempts to derive Factorizability $_{t_P,1,2}$ from Local Causality $_{C(\Sigma,\Sigma)}$

The motivation for the next approach can be formulated as follows. Settings Independence $_{t_P,1,2}$ is plausible, so if we could derive Factorizability $_{t_P,1,2}$ from Local Causality $_{C(\Sigma,\Sigma)}$, then we could show that all locally causal theories are excluded as inconsistent with QM by the following reasoning:

$$\begin{array}{l} \text{Local Causality}_{C(\Sigma,\Sigma)} \Rightarrow \text{Factorizability}_{t_P,1,2} \\ \text{Settings Independence}_{t_P,1,2} \\ (\text{Factorizability}_{t_P,1,2} \wedge \text{Settings Independence}_{t_P,1,2}) \Rightarrow \text{Bell's inequality} \\ \neg \text{Bell's inequality} \\ \hline \neg \text{Local Causality}_{C(\Sigma,\Sigma)} \end{array}$$

The problematic step in this reasoning is, of course, the first one. It is clear that Factorizability $_{t_P,1,2}$ does not follow logically from Local Causality $_{C(\Sigma,\Sigma)}$ because they involve different states (i.e., $\lambda_{t_P,1,2}$ and $\lambda_{C(\Sigma,\Sigma)}$, respectively) that are ascribed to different regions and are not definitionally related to each other. Therefore, if there is any chance of saving the first step, this would require some additional assumptions. Of course these additional assumptions should be independently plausible and have some physical meaning; for example, just adding “Local Causality $_{C(\Sigma,\Sigma)} \Rightarrow \text{Factorizability}_{t_P,1,2}$ ” as an additional assumption would not do the job. On the other hand, the additional assumptions should not entail Factorizability $_{t_P,1,2}$ on their own, independently of Local Causality $_{C(\Sigma,\Sigma)}$, since it is the latter principle that is our proper subject.

An attempt to derive Factorizability $_{t_P,1,2}$ from Local Causality $_{C(\Sigma,\Sigma)}$ has been made by Ciepielewski et al. (2021). They propose decomposing $\lambda_{C(\Sigma,\Sigma)}$ into four parts: $\lambda_{C(\Sigma,\Sigma),1,2}$ (which pertains to the measured system), $\lambda_{C(\Sigma,\Sigma),a}$ and $\lambda_{C(\Sigma,\Sigma),b}$ (which influence or determine the settings a and b , respectively), and $\lambda_{C(\Sigma,\Sigma),E}$ (i.e., everything else). Then, Factorizability $_{C(\Sigma,\Sigma)}$,

$$P(A, B|a, b, \lambda_{C(\Sigma,\Sigma)}) = P(A|a, \lambda_{C(\Sigma,\Sigma)}) \cdot P(B|b, \lambda_{C(\Sigma,\Sigma)}), \quad (10)$$

becomes

$$\begin{aligned} &P(A, B|a, b, \lambda_{C(\Sigma,\Sigma),E}, \lambda_{C(\Sigma,\Sigma),a}, \lambda_{C(\Sigma,\Sigma),b}, \lambda_{C(\Sigma,\Sigma),1,2}) = \\ &P(A|a, \lambda_{C(\Sigma,\Sigma),E}, \lambda_{C(\Sigma,\Sigma),a}, \lambda_{C(\Sigma,\Sigma),b}, \lambda_{C(\Sigma,\Sigma),1,2}) \cdot P(B|b, \lambda_{C(\Sigma,\Sigma),E}, \lambda_{C(\Sigma,\Sigma),a}, \lambda_{C(\Sigma,\Sigma),b}, \lambda_{C(\Sigma,\Sigma),1,2}). \end{aligned} \quad (11)$$

Now, assuming that (i) $\lambda_{C(\Sigma,\Sigma),E}$ is irrelevant, whereas (ii) $\lambda_{C(\Sigma,\Sigma),a}$ and $\lambda_{C(\Sigma,\Sigma),b}$ influence the outcome only via a and b , which were already taken into account, they conclude that only $\lambda_{C(\Sigma,\Sigma),1,2}$ remains relevant, so from (11) they obtain

$$P(A, B|a, b, \lambda_{C(\Sigma,\Sigma),1,2}) = P(A|a, \lambda_{C(\Sigma,\Sigma),1,2}) \cdot P(B|b, \lambda_{C(\Sigma,\Sigma),1,2}). \quad (12)$$

However, this reasoning is contestable. Ciepielewski et al. (2021, p. 11) justify (i) by saying that λ_E “is irrelevant by definition”. This is true only if the measurement outcomes depend only on the measurement settings and the state of the system at the time of preparation. However, they might also be influenced by other factors, such as the microstates of the measurement devices or some details of the environment. Moreover, the state of the system might change between the time of preparation and the time of measurement, being influenced by $\lambda_{C(\Sigma,\Sigma),E}$ (among other things). Also (ii) might be false because the same factor can influence both the measurement settings and something else that influences the measurement outcomes (not via the settings).

Importantly, what they ultimately want to derive is not (12) but Factorizability $_{t_P,1,2}$, which is

$$P(A, B|a, b, \lambda_{t_P,1,2}) = P(A|a, \lambda_{t_P,1,2}) \cdot P(B|b, \lambda_{t_P,1,2}). \quad (13)$$

This is because for the proof of Bell’s theorem to work, we need the same λ in both Factorizability and Settings Independence. However, (12) and (13) are not equivalent. As already observed, the state of the measured system might change between the time of preparation and the time of measurement. Therefore, $\lambda_{C(\Sigma,\Sigma),1,2}$ might not be the same as $\lambda_{t_P,1,2}$ (and if position is one of the hidden variables, they surely will be different). Denoting the state of the system just by λ , without any indices, obscures this issue.

The above-mentioned problem with (i) was recognised by Ciepielewski and Okon (2022), which motivated them to make a different attempt to derive Factorizability $_{t_P,1,2}$ from Local Causality $_{C(\Sigma,\Sigma)}$ (cf. also their criticism of some previous derivations by other authors, pp. 6–14). This time, their starting point is not Factorizability $_{C(\Sigma,\Sigma)}$ but the following mathematical fact:

$$P(A, B|a, b, \lambda_{t_P,1,2}) = P(A|B, a, b, \lambda_{t_P,1,2}) \cdot P(B|a, b, \lambda_{t_P,1,2}). \quad (14)$$

The main idea of this new derivation is as follows (Ciepielewski and Okon, 2022, p. 15):

[...] assuming the principle of local causality, under what circumstances would it be illegitimate to remove b or B from the first term on the right-hand side of [(14)]? For that to be the case, two conditions are required. First, there must be something, besides a and λ , which is relevant for the prediction. Second, that something must be (at least partially) encoded in b or B .

Then they claim that the first assumption is plausible, whereas the second amounts to the violation of what they call “Microstate Independence”:

$$P(\lambda_A|b, B) = P(\lambda_A) \text{ and } P(\lambda_B|a, A) = P(\lambda_B), \quad (15)$$

where λ_A is everything that influences the outcome A besides a and $\lambda_{t_P,1,2}$ (and analogously for λ_B). But what exactly does it mean to say that λ_A is “everything that influences the outcome A besides a and $\lambda_{t_P,1,2}$ ”? How to express this idea in terms of probabilities? Ciepielewski and Okon (2022) do not give the answer. Perhaps this can be captured by the following condition, which is similar to Seevinck and Uffink’s sufficiency condition:

$$P(A|a, b, B, \lambda_{C(\Sigma,\Sigma)}, \lambda_A, \lambda_{t_P,1,2}) = P(A|a, \lambda_A, \lambda_{t_P,1,2}) \text{ and } P(B|a, b, A, \lambda_{C(\Sigma,\Sigma)}, \lambda_B, \lambda_{t_P,1,2}) = P(B|b, \lambda_B, \lambda_{t_P,1,2}). \quad (16)$$

Therefore, instead of assuming the existence of $\bar{\lambda}$ that satisfies both Factorizability $_{\bar{\lambda}}$ and Settings Independence $_{\bar{\lambda}}$, we need to assume the existence of some λ_A and λ_B that satisfy (15) and (16). But are these assumptions more plausible?

Even more importantly, does Factorizability $_{t_P,1,2}$ follow from these assumptions? Ciepielewski and Okon (2022, p. 17) claim that it does:

[...] the principle of local causality implies that anything outside of the past light cone of 1, which does not enhance the prediction for A given a and λ , can be removed from the conditional in the probability of A (and similarly for region 2). Then, we notice that, because of microstate independence, this is the case for b and B in the first term on the right and for a in the second, so such terms can be removed, leading to factorizability.

However, Local Causality $_{C(\Sigma,\Sigma)}$ allows for removing factors from the conditional probability of A only if there is another appropriate factor that we condition upon, which needs to cover the entire thick slice of the past light cone of R_A . This is not satisfied here because Eq. (14) does not involve any factor that covers the entire thick slice of the past light cone of R_A . Notice also that λ_A does not appear anywhere in (14), so it is not clear how (15) and (16) could justify removing any terms from (14).

9 A new proposal

The proposal for establishing Bell’s Conclusion that I want to put forward here is in some sense a combination of the second, third and fourth strategies listed in the second paragraph of section 8. It is the closest to the

third strategy: we will derive something close to Factorizability $_{t_P,1,2}$ from Local Causality $_{C(\Sigma,\Sigma)}$. However, instead of $\lambda_{t_P,1,2}$, we will use a bit more encompassing state (the fourth strategy), and we will also appeal to a coarse-graining in order to bypass the problem of the possible changes of $\lambda_{t_P,1,2}$ in time (the second strategy).

Let us begin with the following observation. In principle, some factors different from the state of the two particles at t_P , the effects of which are encoded in their subsequent states, could explain the correlation between the outcomes of measurements made on them (cf. the fourth paragraph in section 8.3). But what else could these be? All other factors, except for the arrangement of the measurement devices and the state-preparing device, could be different in every run of Bell's Experiment, so they cannot be responsible for the statistical regularities in the outcomes. Butterfield (1992, pp. 39–40) formulates this idea in the following way:

Consider an experiment that gets repeated many times. Its repetitions will not share all their properties; any two are liable to differ in myriad ways (such as temperature and humidity). But it may be that no property divides the class of repetitions into those with the property and those without it so as to give different statistics in the two subclasses. In such a case, call the experiment's statics "homogeneous". [...] The more the repetitions vary among themselves while the statistics nevertheless remain homogeneous, the more we expect that the properties common to all the repetitions are exactly all the properties on which the chances of the phenomena at issue depend.

Therefore, when looking for the causes of the correlation between measurement outcomes, it is reasonable to restrict ourselves to the two-particle system at t_P , or at least to a close neighbourhood of it (in case it is the state-preparing device or something always present in its immediate vicinity rather than the two-particle system itself that causes the correlation). We will call this state $\lambda_{t_P,1,2,+}$; it includes $\lambda_{t_P,1,2}$ and presumably something more, but it is located entirely within the intersection of the past light cones of R_A and R_B at time t_P .

Let us turn to our argument. Assume for reductio that Local Causality $_{C(\Sigma,\Sigma)}$ holds (i.e., that some HVT satisfying Causality $_{C(\Sigma,\Sigma)}$ is true). From Local Causality $_{C(\Sigma,\Sigma)}$, we can derive Factorizability $_{C(\Sigma,\Sigma)}$ (see section 7.1),

$$P(A, B|a, b, \lambda_{C(\Sigma,\Sigma)}) = P(A|a, \lambda_{C(\Sigma,\Sigma)}) \cdot P(B|b, \lambda_{C(\Sigma,\Sigma)}). \quad (17)$$

Let $\Lambda_{1,2,+}$ be the set of possible states of our enlarged system (including two particles and "something more") at a single time. Assume that there exists¹⁹ a coarse-graining of $\Lambda_{1,2,+}$, denoted by $\tilde{\Lambda}_{1,2,+}$, such that in Bell's Experiment,

(A1) the aspects of the state of our enlarged system captured by this coarse-graining do not change in time, that is,

$$\forall_{t'' \in [t_P, t_M]} \tilde{\lambda}_{t'', 1, 2, +} = \tilde{\lambda}_{t_P, 1, 2, +}, \quad (18)$$

where t_M is the time of measurements (cf. footnote 9), and

(A2) those aspects of the state of our enlarged system are also the only aspects of $\lambda_{C(\Sigma,\Sigma)}$ that are relevant for probabilities of measurement outcomes conditioned on measurement setting and $\lambda_{C(\Sigma,\Sigma)}$, that is,

$$P(A, B|a, b, \lambda_{C(\Sigma,\Sigma)}) = P(A, B|a, b, \tilde{\lambda}_{t, 1, 2, +}) = P(A, B|a, b, \tilde{\lambda}_{t_P, 1, 2, +}), \quad (19)$$

$$P(A|a, \lambda_{C(\Sigma,\Sigma)}) = P(A|a, \tilde{\lambda}_{t, 1, 2, +}) = P(A|a, \tilde{\lambda}_{t_P, 1, 2, +}) \quad (20)$$

and

$$P(B|b, \lambda_{C(\Sigma,\Sigma)}) = P(B|b, \tilde{\lambda}_{t, 1, 2, +}) = P(B|b, \tilde{\lambda}_{t_P, 1, 2, +}). \quad (21)$$

¹⁹It should be stressed that the assumption here is that there *exists* at least one such coarse-graining, not that any coarse-graining satisfies (A1)–(A2). For this reason, we should not be worried about Simpson-like effects, which in our case would amount to possible violations of (A1)–(A2) by other coarse-grainings.

The second equality in each of (19), (20) and (21) is justified by (A1): since $t_P < t < t_M$, we have $\tilde{\lambda}_{t,1,2,+} = \tilde{\lambda}_{t_P,1,2,+}$.

Now, if we apply (19) to the LHS of (17), and (20) together with (21) to the RHS of (17), we obtain Factorizability $\widetilde{\text{Factorizability}}_{t_P,1,2,+}$, that is,

$$P(A, B|a, b, \tilde{\lambda}_{t_P,1,2,+}) = P(A|a, \tilde{\lambda}_{t_P,1,2,+}) \cdot P(B|b, \tilde{\lambda}_{t_P,1,2,+}). \quad (22)$$

If we additionally assume Settings Independence $\widetilde{\text{Settings Independence}}_{t_P,1,2,+}$,

$$P(\tilde{\lambda}_{t_P,1,2,+}|a, b) = P(\tilde{\lambda}_{t_P,1,2,+}), \quad (23)$$

then we will have the full package of assumptions that are needed to derive Bell's inequality with $\lambda = \tilde{\lambda}_{t_P,1,2,+}$ (i.e., Factorizability $\widetilde{\text{Factorizability}}_{t_P,1,2,+}$ and Settings Independence $\widetilde{\text{Settings Independence}}_{t_P,1,2,+}$). Since Bell's inequality is violated, either Local Causality $\widetilde{\text{Local Causality}}_{C(\Sigma, \Sigma)}$ is false or there is no coarse-graining $\tilde{\Lambda}_{1,2,+}$ that satisfies (A1), (A2) and Settings Independence $\widetilde{\text{Settings Independence}}_{t_P,1,2,+}$.

Why should we believe that such a coarse-graining exists? Let us begin with observing that Settings Independence $\widetilde{\text{Settings Independence}}_{\lambda_{t_P,1,2,+}}$, that is,

$$P(\lambda_{t_P,1,2,+}|a, b) = P(\lambda_{t_P,1,2,+}), \quad (24)$$

is already plausible (even without taking a coarse-graining of $\lambda_{t_P,1,2,+}$). The justification for Settings Independence $\widetilde{\text{Settings Independence}}_{t_P,1,2,+}$ is the same as the justification for Settings Independence $\widetilde{\text{Settings Independence}}_{\lambda_{t_P,1,2}}$. The entire state $\lambda_{t_P,1,2,+}$ is confined to the intersection of the light cones of R_A and R_B at t_P . Therefore, the past light cones of R_a and R_b include plenty of factors that belong to neither the region covered by $\lambda_{t_P,1,2,+}$ nor its causal future, so they are able to interfere with the influence of $\lambda_{t_P,1,2,+}$ on the measurement settings. After coarse-graining, this version of Settings Independence becomes even more plausible: if there were any factors in $\lambda_{t_P,1,2,+}$ correlated with the choices of measurement settings, they could be lost by taking an equivalence class of such states, but no new correlations can be introduced in this way.²⁰

What about (A1) and (A2)? An important role is played in their justification by common-cause-like intuitions and continuous-action intuitions.²¹ First, whatever is responsible for the *correlation* between measurement outcomes should lie in the common past of the two measurements, that is, in the intersection of the past light cones of R_A and R_B . Otherwise it could influence at most one of the measurement outcomes but not both; however, the correlation involves both outcomes. A factor that lies outside of the intersection of the past light cones of R_A and R_B could spoil the correlation between the two particles, but it cannot create any correlation between them because it can influence at most one of them. Second, we have observed that $\lambda_{t_P,1,2}$ can change between t_P and t_M ; *a fortiori* this is true about $\lambda_{t_P,1,2,+}$. However, this aspect of $\lambda_{t_P,1,2,+}$ that is responsible for the correlation should better last unchanged until the measurements on both wings are made, since otherwise its influence on measurement outcomes would require some kind of action at a temporal distance (cf. Adlam 2018).²² The idea that causation between temporally distant events needs to always be mediated by intermediate events is known in the literature about Bell's theorem as "action-by-contact" (Evans et al., 2013) or "continuous action" (Wharton and Argaman 2020 and Adlam

²⁰This part of my reasoning can be underpinned formally. Assume that X and X' are such that (i) $X \cap X' = \emptyset$, (ii) $P(X|Y) = P(X)$ and (iii) $P(X'|Y) = P(X')$. Let their coarse-grained counterpart be $\tilde{X} = X \cup X'$. We have $P(\tilde{X}|Y) = P(X \cup X'|Y) = P(X|Y) + P(X'|Y) = P(X) + P(X') = P(\tilde{X})$ (where the second and the last equality comes from (i), whereas the third equality comes from (ii) and (iii)). This can be generalised to any \tilde{X} that consists of more than two events.

²¹Importantly, however, we do not assume that $\lambda_{t_P,1,2,+}$ is a common cause in the sense of satisfying Factorizability $\widetilde{\text{Factorizability}}_{t_P,1,2,+}$ because this would render Local Causality $\widetilde{\text{Local Causality}}_{C(\Sigma, \Sigma)}$ redundant in the derivation of Bell's inequality, which would again block our attempt to establish Bell's Conclusion.

²²Adlam (2018, p. 2) defines the condition of Temporal Locality in the image of Factorizability, not Local Causality $\widetilde{\text{Local Causality}}_{C(\Sigma, \Sigma)}$. However, we can easily find a formulation of Temporal Locality that is closer to the latter:

Temporal Locality $\widetilde{\text{Temporal Locality}}_{C(\Sigma, \Sigma)}$, fine-grained version: For any bounded region R , for any $t < R$ and any $t' < t$, for any λ_R and $\lambda_{C(\Sigma_{R,t}, \Sigma_{R,t'})}$, for any bounded region $R'' < t'$ and any $\lambda_{R''}$,

$$P(\lambda_R | \lambda_{C(\Sigma_{R,t}, \Sigma_{R,t'})}, \lambda_{R''}) = P(\lambda_R | \lambda_{C(\Sigma_{R,t}, \Sigma_{R,t'})}). \quad (25)$$

2022). Notably, this idea is clearly present in Bell’s (1990, p. 239) initial formulation of Local Causality (“[t]he direct causes (and effects) of events are near by”), but it is not reflected in Local Causality $_{C(\Sigma, \Sigma)}$, which does not prohibit events in the remote past but lying in the past-light cone of R to influence the probability of a given state in R .

I do not know whether the existence of a coarse-graining that satisfies (A1) and (A2) can be derived in a formal way from some more basic principles, so I will leave the justification of this assumption in the above form.

10 Summary

In this paper, an important ambiguity in the understanding of hidden states in the literature about Bell’s theorem has been analysed. Hidden states in the context of Bell’s theorem are usually considered either as the states of the two-particle system at the moment of preparation or as the states of thick slices of the past light cones of measurements. In both cases, there are problems with establishing Bell’s Conclusion, which states that all HVTs that satisfy Local Causality are inconsistent with the predictions of QM for Bell’s Experiment. In the first approach, the link between Factorizability $_{t_P, 1, 2}$ and Local Causality $_{C(\Sigma, \Sigma)}$ is missing. In the second approach, Settings Independence $_{C(\Sigma, \Sigma)}$ is violated by all locally deterministic theories (unless we reject either Physical Nature of Human Choices or the Universality of Laws, but this also undermines the proof that all locally deterministic theories satisfy Local Causality $_{C(\Sigma, \Sigma)}$). Since the condition of Local Causality $_{C(\Sigma, \Sigma)}$ was intended by Bell to be a generalisation of local determinism, locally deterministic theories form a very important (even a paradigmatic one) subclass of locally causal theories. Therefore, the failure to establish Bell’s Conclusion for this class of theories cannot be ignored.

Three possible ways of improving this situation have been identified. One can stay with $\lambda = \lambda_{t_P, 1, 2}$ and try to derive Factorizability $_{t_P, 1, 2}$ from Local Causality $_{C(\Sigma, \Sigma)}$. This, however, requires some additional assumptions, and it is not clear what they should be. One can also stay with $\lambda = \lambda_{C(\Sigma, \Sigma)}$ and consider coarse-grained versions of these states. However, to establish Bell’s Conclusion, we need to assume that there exists a coarse-graining $\bar{\lambda}$ for which both Factorizability $_{\bar{\lambda}}$ and Settings Independence $_{\bar{\lambda}}$ hold, and it is not clear how to justify this assumption. Finally, one can consider some other choices of λ . My own proposed approach combines the ideas of all three strategies.

Let me list what I think are the major results of this paper. First of all, our considerations show that it *does* matter where exactly the λ that occurs in Bell’s theorem is located. Second, we have pointed out that the relationship between deterministic theories and 0-1 probabilities is not straightforward, and we have proposed the Bridge Principle, which allows us to relate them (Definition 3). Third, we have provided a precise definition of locally deterministic theories (Definition 2) and shown that they are deterministic (Theorem 1), that they satisfy Local Causality $_{C(\Sigma, \Sigma)}$ (Theorem 2) and that they violate Settings Independence $_{C(\Sigma, \Sigma)}$ (Theorem 3); the latter two results are based on the aforementioned Bridge Principle. Fourth, this Bridge Principle has also allowed us to distinguish clearly between deterministic and outcome deterministic theories (see the third paragraph of section 6). Finally, it seems that there is at least one plausible way of establishing Bell’s Conclusion; the one that I have presented relies on some additional assumptions that allow one to derive the relevant variant of Factorizability from Local Causality $_{C(\Sigma, \Sigma)}$ (see section 9).

Assuming that the above considerations are correct, let us reflect on the scope of the conclusions that we can derive from Bell’s theorem. If “non-local” means “either locally indeterminate or locally determinate but violating Local Causality $_{C(\Sigma, \Sigma)}$ ”, then we can strengthen what we called “Bell’s Conclusion” as follows:

Strengthened Bell’s Conclusion: All HVTs consistent with the predictions of QM for Bell’s Experiment are non-local.

Temporal Locality does not play any official role in our reasoning that leads to Bell’s Conclusion, but the underlying idea is used in our justification of (A1) and (A2). Moreover, it is interesting for us because it allows one to reconcile the satisfaction of Factorizability $_{t_P, 1, 2, +}$ with the violation of Factorizability $_{C(\Sigma, \Sigma)}$: we can explain this by assuming that $\bar{\lambda}_{t_P, 1, 2, +}$ acts at a distance directly on measurement outcomes, without the mediation of anything in the region $C(\Sigma, \Sigma)$.

If a given HVT is locally indeterminate, this follows from the above definition of non-locality; and if a given HVT is locally determinate, this follows from Bell's Conclusion. Since Nature realises the predictions of QM for Bell's Experiment, if we assume that one of HVTs needs to be a true description of Nature, we can conclude that Nature itself is non-local in the above sense. However, I want to emphasise that this concept of non-locality is not just the violation of Local Causality $_{C(\Sigma, \Sigma)}$ but a disjunction whose second disjunct is an even deeper form of non-locality—namely, the failure of local determinateness.

References

- Abellán, C., Acín, A., Alarcón, A., Alibart, O., Andersen, C. K., Andreoli, F., Beckert, A., Beduini, F. A., Bendersky, A., Bentivegna, M., Bierhorst, P., Burchardt, D., Cabello, A., Cariñe, J., Carrasco, S., Carvacho, G., Cavalcanti, D., Chaves, R., Cortés-Vega, J., Cuevas, Á. C., Delgado, A., de Riedmatten, H., Eichler, C., Farrera, P., Fuenzalida, J., García-Matos, M., Garthoff, R., Gasparinetti, S., Gerrits, T., Jouneghani, F. G., Glancy, S., Gómez, E. S., González, P., Guan, J. Y., Handsteiner, J., Heinsoo, J., Heinze, G., Hirschmann, A., Jiménez, O., Kaiser, F., Knill, E., Knoll, L. T., Krinner, S., Kurpiers, P., Larotonda, M. A., Larsson, J., Lenhard, A., Li, H. B., Li, M., Lima, G., Liu, B., Liu, Y., Grande, I. H. L., Lunghi, T., Ma, X., Magaña-Loaiza, O. S., Magnard, P., Magnoni, A. S., Marti-Prieto, M., Martínez, D., Mataloni, P., Máttar, A., Mazzera, M., Mirin, R. P., Mitchell, M. W., Nam, S. W., Oppliger, M., Pan, J.-W., Patel, R., Pryde, G. J., Rauch, D., Redeker, K., Rieländer, D., Ringbauer, M., Roberson, T., Rosenfeld, W., Salathé, Y., Santodonato, L., Sauder, G., Scheidl, T., Schmiegelow, C. T., Sciarrino, F., Seri, A., Shalm, L. K., Shi, S., Slussarenko, S. S., Stevens, M. J., Tanzilli, S., Toledo, F., i Brugués, J. T., Ursin, R., Vergyris, P., Verma, V. B., Walter, T., Wallraff, A., Wang, Z., Weinfurter, H., Weston, M. M., White, A. G., Wu, C., Xavier, G. B., You, L., Yuan, X., Zeilinger, A., Zhang, Q., Zhang, W. H., and Zhong, J. (2018). Challenging local realism with human choices. *Nature*, 557:212 – 216.
- Adlam, E. (2018). Spooky Action at a Temporal Distance. *Entropy*, 20(1).
- Adlam, E. (2022). Two Roads to Retrocausality. *Synthese*, 200(5):1–36.
- Aspect, A., Grangier, P., and Roger, G. (1982). Experimental Realization of Einstein-Podolsky-Rosen-Bohm Gedankenexperiment: A New Violation of Bell's Inequalities. *Phys. Rev. Lett.*, 49:91–94.
- Baas, A. and Bihan, B. L. (2023). What Does the World Look Like According to Superdeterminism. *British Journal for the Philosophy of Science*, 74(3):555–572.
- Barandes, J. A. (2024). New Prospects for a Causally Local Formulation of Quantum Theory. <https://arxiv.org/abs/2402.16935v1>.
- Bell, J. S. (1964). On the Einstein-Podolsky-Rosen paradox. *Physics*, 1:195–200. Reprinted in ?, ch. 2.
- Bell, J. S. (1971). Introduction to the hidden-variable question. In d'Espagnat, B., editor, *Foundations of quantum mechanics (Proceedings of the International School of Physics 'Enrico Fermi', course II)*, pages 171–181. Academic, New York. Reprinted in ?, ch. 4.
- Bell, J. S. (1976). The theory of local beables. *Epistemological Letters*, 9:11–24. Reprinted in ?, ch. 7.
- Bell, J. S. (1977). Free variables and local causality. *Epistemological Letters*, 15:79–84. Reprinted in ?, ch. 12.
- Bell, J. S. (1981). Bertlmann's socks and the nature of reality. *Journal de Physique Colloque*, C2(suppl. au no 3, Tome 42):C2–41–46. Reprinted in ?, ch. 16.
- Bell, J. S. (1989). Against 'measurement'. In *62 Years of Uncertainty: Erice, 5-14 August 1989*. Plenum Publishers. Reprinted in ?, ch. 23.

- Bell, J. S. (1990). La nouvelle cuisine. In Sarlemijn, A. and Kroes, P., editors, *Between Science and Technology*, pages 97–115. Elsevier, Amsterdam. Reprinted in ?, ch. 24.
- Brown, H. R. and Timpson, C. G. (2016). Bell on Bell’s theorem: the changing face of nonlocality. In Bell, M. and Gao, S., editors, *Quantum Nonlocality and Reality: 50 years of Bell’s theorem*, pages 91–123. Cambridge University Press, Cambridge.
- Butterfield, J. (1989). The Hole Truth. *British Journal for the Philosophy of Science*, 40(1):1–28.
- Butterfield, J. (1992). David Lewis Meets John Bell. *Philosophy of Science*, 59(1):26–43.
- Butterfield, J. (1998). Determinism and indeterminism.
- Ciepielewski, G. S. and Okon, E. (2022). From locality to factorizability: a novel escape from Bell’s theorem. Preprint, <https://arxiv.org/abs/2204.05162>.
- Ciepielewski, G. S., Okon, E., and Sudarsky, D. (2021). On Superdeterministic Rejections of Settings Independence. *British Journal for the Philosophy of Science*, 74(2):435–467.
- Clauser, J. F., Horne, M. A., Shimony, A., and Holt, R. A. (1969). Proposed Experiment to Test Local Hidden-Variable Theories. *Phys. Rev. Lett.*, 23:880–884.
- Dürr, D., Goldstein, S., and Zanghì, N. (1997). Bohmian mechanics and the meaning of the wave function. In Cohen, R., Horne, M., and Stachel, J., editors, *Experimental Metaphysics—Quantum Mechanical Studies for Abner Shimony*, volume 1 of *Boston Studies in the Philosophy of Science*. Kluwer Academic Publishers, Boston.
- Earman, J. (1986). *A Primer on Determinism*. D. Reidel.
- Einstein, A., Podolsky, B., and Rosen, N. (1935). Can Quantum-Mechanical Description of Physical Reality Be Considered Complete? *Phys. Rev.*, 47:777–780.
- Esfeld, M., Lazarovici, D., Hubert, M., and Dürr, D. (2014). The Ontology of Bohmian Mechanics. *British Journal for the Philosophy of Science*, 65(4):773–796.
- Evans, P., Price, H., and Wharton, K. (2013). New Slant on the EPR-Bell Experiment. *British Journal for the Philosophy of Science*, 64(2):297–324.
- Friederich, S. and Evans, P. W. (2023). Retrocausality in Quantum Mechanics. In Zalta, E. N. and Nodelman, U., editors, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Winter 2023 edition.
- Ghirardi, G. and Bassi, A. (2024). Collapse Theories. In Zalta, E. N. and Nodelman, U., editors, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Fall 2024 edition.
- Goldstein, S. (2024). Bohmian Mechanics. In Zalta, E. N. and Nodelman, U., editors, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Summer 2024 edition.
- Goldstein, S., Norsen, T., Tausk, D. V., and Zanghì, N. (2011). Bell’s theorem. *Scholarpedia*, 6(10):8378. Revision #91049.
- Hitchcock, C. and Rédei, M. (2021). Reichenbach’s Common Cause Principle. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Summer 2021 edition.
- Hofer-Szabó, G. (2015). Local Causality and Complete Specification: A Reply to Seevinck and Uffink. In Mäki, U., Votsis, I., Ruphy, S., and Schurz, G., editors, *Recent Developments in the Philosophy of Science: EPSA13 Helsinki*, volume 1 of *European Studies in Philosophy of Science*. Springer, Cham.

- Hossenfelder, S. and Palmer, T. (2020). Rethinking Superdeterminism. *Frontiers in Physics*, 8:139.
- Jarrett, J. P. (1984). On the Physical Significance of the Locality Conditions in the Bell Arguments. *Noûs*, 18(4):569–589.
- Malament, D. B. (undated). Notes on Bell’s theorem. Available at <https://sites.socsci.uci.edu/~dmalamen/courses/prob-determ/PDnotesBell.pdf>, accessed 23 January 2025.
- Maudlin, T. (2014). What Bell Did. *Journal of Physics A: Mathematical and Theoretical*, 47(42):424010.
- Müller, T. and Placek, T. (2023). On the Condition of Setting Independence. *European Journal for Philosophy of Science*, 13(4):1–20.
- Myrvold, W., Genovese, M., and Shimony, A. (2024). Bell’s Theorem. In Zalta, E. N. and Nodelman, U., editors, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Spring 2024 edition.
- Norsen, T. (2009). Local Causality and Completeness: Bell Vs. Jarrett. *Foundations of Physics*, 39(3):273–294.
- Norsen, T. (2011). John S. Bell’s concept of local causality. *American Journal of Physics*, 79(12):1261–1275.
- Norsen, T. (2017). *Foundations of Quantum Mechanics: An Exploration of the Physical Meaning of Quantum Theory*. Springer, Cham.
- Price, H. (1994). A Neglected Route to Realism About Quantum Mechanics. *Mind*, 103(411):303–336.
- Seevinck, M. P. and Uffink, J. (2011). Not Throwing Out the Baby with the Bathwater: Bell’s Condition of Local Causality Mathematically ‘Sharp and Clean’. In Dieks, D., Gonzalo, W., Uebel, T., Hartmann, S., and Weber, M., editors, *Explanation, Prediction, and Confirmation*, pages 425–450. Springer.
- Shimony, A. (1986). Events and processes in the quantum world. In Penrose, R. and Isham, C. J., editors, *Quantum Concepts in Space and Time*, pages 182–203. Oxford University Press, Oxford. Reprinted in Shimony (1993), pp. 140–162.
- Shimony, A. (1993). *Search for a Naturalistic World View*, volume 2. Cambridge University Press, Cambridge.
- Shimony, A., Horne, M. A., and Clauser, J. (1976). Comment on ‘The theory of local beables’. *Epistemological Letters*, 13:1–8. Reprinted in Shimony (1993), pp. 163–167.
- van Fraassen, B. C. (1982). The Charybdis of Realism: Epistemological Implications of Bell’s Inequality. *Synthese*, 52(1):25–38.
- Wharton, K. B. and Argaman, N. (2020). Colloquium: Bell’s theorem and locally mediated reformulations of quantum mechanics. *Rev. Mod. Phys.*, 92:021002.
- Wiseman, H. M. (2014). The two Bell’s theorems of John Bell. *Journal of Physics A: Mathematical and Theoretical*, 47:424001.

A Proofs

This appendix collects proofs of the theorems from the main part of the paper.

A.1 Locally deterministic theories are deterministic

Theorem 1. Any locally determinate HVT that is locally deterministic is also deterministic.

Proof. Let T be a locally determinate HVT that is not deterministic. This means that there exists time t_0 and state λ_{t_0} such that two different solutions, $\lambda_{\text{spacetime}} \neq \lambda'_{\text{spacetime}}$, are compatible with λ_{t_0} and the laws of T . Since $\lambda_{\text{spacetime}}$ and $\lambda'_{\text{spacetime}}$ are different, there exists a spatiotemporal region R on which they differ. That is, $\lambda_{\text{spacetime}}$ is compatible with some λ_R and $\lambda'_{\text{spacetime}}$ is compatible with some λ'_R such that $\lambda_R \neq \lambda'_R$. If region R partially overlaps with Σ_{t_0} , then there exists a bounded and connected region R' such that $R' \subsetneq R$, R' does not overlap with Σ_{t_0} and the restrictions of $\lambda_{\text{spacetime}}$ and $\lambda'_{\text{spacetime}}$ to R' are different (i.e., $\lambda_{R'} \neq \lambda'_{R'}$). This is because $\lambda_{\text{spacetime}}$ and $\lambda'_{\text{spacetime}}$ by definition do not differ on Σ_{t_0} but do differ on R , so they must differ on some part of R that does not overlap with Σ_{t_0} . If R does not overlap with Σ_{t_0} , let R' be any bounded and connected subregion of R on which $\lambda_{\text{spacetime}}$ and $\lambda'_{\text{spacetime}}$ differ. Since R' does not overlap with Σ_{t_0} and is bounded and connected, it must be wholly located either below or above Σ_{t_0} —that is, $t_0 < R'$ or $t_0 > R'$. Assume without loss of generality that $t_0 < R'$ (the other case is analogous). Consider our initial conditions λ_{t_0} restricted to the intersection of the past light cone of R' with Σ_{t_0} —that is, $\lambda_{\Sigma_{R',t_0}}$. Since there exist two different specifications of hidden variables in region R' that are compatible with $\lambda_{\Sigma_{R',t_0}}$ (namely, $\lambda_{R'}$ and $\lambda'_{R'}$), our theory T is not locally deterministic. Therefore, T not being deterministic entails that it is also not locally deterministic. By contraposition, if T is locally deterministic, then it is also deterministic. \square

A.2 Derivation of Factorizability $_{C(\Sigma,\Sigma)}$ from Local Causality $_{C(\Sigma,\Sigma)}$

We will derive some consequences for Bell's Experiment from the coarse-grained version of Local Causality $_{C(\Sigma,\Sigma)}$, which will enable us to prove Factorizability $_{C(\Sigma,\Sigma)}$. Consider any times t and t' such that $t' < t$, R_a and R_b lie above t , and t' is above the intersection of the past light cones of R_A and R_B . Let a and b be possible experimental settings chosen in regions R_a and R_b , and let A and B be their possible outcomes obtained in regions R_A and R_B . By a, b, A and B we also denote the equivalence classes of $\lambda_{R_a}, \lambda_{R_b}, \lambda_{R_A}$ and λ_{R_B} , which correspond to the respective settings and outcomes. Additionally, $\lambda_{C(\Sigma,\Sigma)}$ denotes the pair consisting of $\lambda_{C(\Sigma_{R_A,t},\Sigma_{R_A,t'})}$ and $\lambda_{C(\Sigma_{R_B,t},\Sigma_{R_B,t'})}$, and $C(\Sigma,\Sigma)$ denotes the respective region. From Local Causality $_{C(\Sigma,\Sigma)}$ it follows that for any $\lambda_{C(\Sigma_{R_A,t},\Sigma_{R_A,t'})}$ and any $\lambda_{C(\Sigma_{R_B,t},\Sigma_{R_B,t'})}$,

$$P(A, a|B, b, \lambda_{C(\Sigma_{R_B,t},\Sigma_{R_B,t'})}, \lambda_{C(\Sigma_{R_A,t},\Sigma_{R_A,t'})}) = P(A, a|\lambda_{C(\Sigma_{R_A,t},\Sigma_{R_A,t'})}). \quad (26)$$

In more detail, in the transition from the LHS to the RHS, we have used Eq. (3) with the following substitutions: $R_A \cup R_a$ for R , A and a for $\overline{\lambda_{R,d_R}}$, $R_B \cup R_b \cup C(\Sigma_{R_B,t},\Sigma_{R_B,t'})$ for R' , and finally, B, b and $\lambda_{C(\Sigma_{R_B,t},\Sigma_{R_B,t'})}$ for $\overline{\lambda_{R',d_{R'}}}$. Notice that under these choices, regions R and R' are bounded (as they must be) but not connected.

It also follows from Local Causality $_{C(\Sigma,\Sigma)}$ that for any $\lambda_{C(\Sigma_{R_A,t},\Sigma_{R_A,t'})}$ and any $\lambda_{C(\Sigma_{R_B,t},\Sigma_{R_B,t'})}$,

$$P(a|b, \lambda_{C(\Sigma_{R_A,t},\Sigma_{R_A,t'})}, \lambda_{C(\Sigma_{R_B,t},\Sigma_{R_B,t'})}) = P(a|\lambda_{C(\Sigma_{R_A,t},\Sigma_{R_A,t'})}). \quad (27)$$

Finally, it follows from Local Causality $_{C(\Sigma,\Sigma)}$ that for any $\lambda_{C(\Sigma_{R_A,t},\Sigma_{R_A,t'})}$ and any $\lambda_{C(\Sigma_{R_B,t},\Sigma_{R_B,t'})}$,

$$P(A, a|\lambda_{C(\Sigma_{R_A,t},\Sigma_{R_A,t'})}) = P(A, a|\lambda_{C(\Sigma_{R_A,t},\Sigma_{R_A,t'})}, \lambda_{C(\Sigma_{R_B,t},\Sigma_{R_B,t'})}) \quad (28)$$

and

$$P(a|\lambda_{C(\Sigma_{R_A,t},\Sigma_{R_A,t'})}) = P(a|\lambda_{C(\Sigma_{R_A,t},\Sigma_{R_A,t'})}, \lambda_{C(\Sigma_{R_B,t},\Sigma_{R_B,t'})}), \quad (29)$$

where we harmlessly added the conditionalization on the second hidden state.

From these we can prove Factorizability $_{C(\Sigma, \Sigma)}$:

$$\begin{aligned}
P(A, B|a, b, \lambda_{C(\Sigma, \Sigma)}) &= P(A, B|a, b, \lambda_{C(\Sigma_{R_A, t}, \Sigma_{R_A, t'})}, \lambda_{C(\Sigma_{R_B, t}, \Sigma_{R_B, t'})}) = \\
&= \frac{P(A, B, a, b, \lambda_{C(\Sigma_{R_A, t}, \Sigma_{R_A, t'})}, \lambda_{C(\Sigma_{R_B, t}, \Sigma_{R_B, t'})})}{P(a, b, \lambda_{C(\Sigma_{R_A, t}, \Sigma_{R_A, t'})}, \lambda_{C(\Sigma_{R_B, t}, \Sigma_{R_B, t'})})} = \\
\frac{P(A, a|B, b, \lambda_{C(\Sigma_{R_A, t}, \Sigma_{R_A, t'})}, \lambda_{C(\Sigma_{R_B, t}, \Sigma_{R_B, t'})})P(B, b, \lambda_{C(\Sigma_{R_A, t}, \Sigma_{R_A, t'})}, \lambda_{C(\Sigma_{R_B, t}, \Sigma_{R_B, t'})})}{P(a|b, \lambda_{C(\Sigma_{R_A, t}, \Sigma_{R_A, t'})}, \lambda_{C(\Sigma_{R_B, t}, \Sigma_{R_B, t'})})P(b, \lambda_{C(\Sigma_{R_A, t}, \Sigma_{R_A, t'})}, \lambda_{C(\Sigma_{R_B, t}, \Sigma_{R_B, t'})})} &= \\
\frac{P(A, a|\lambda_{C(\Sigma_{R_A, t}, \Sigma_{R_A, t'})})P(B, b, \lambda_{C(\Sigma_{R_A, t}, \Sigma_{R_A, t'})}, \lambda_{C(\Sigma_{R_B, t}, \Sigma_{R_B, t'})})}{P(a|\lambda_{C(\Sigma_{R_A, t}, \Sigma_{R_A, t'})})P(b, \lambda_{C(\Sigma_{R_A, t}, \Sigma_{R_A, t'})}, \lambda_{C(\Sigma_{R_B, t}, \Sigma_{R_B, t'})})} &= \quad (30) \\
\frac{P(A, a|\lambda_{C(\Sigma_{R_A, t}, \Sigma_{R_A, t'})}, \lambda_{C(\Sigma_{R_B, t}, \Sigma_{R_B, t'})})P(B, b, \lambda_{C(\Sigma_{R_A, t}, \Sigma_{R_A, t'})}, \lambda_{C(\Sigma_{R_B, t}, \Sigma_{R_B, t'})})}{P(a|\lambda_{C(\Sigma_{R_A, t}, \Sigma_{R_A, t'})}, \lambda_{C(\Sigma_{R_B, t}, \Sigma_{R_B, t'})})P(b, \lambda_{C(\Sigma_{R_A, t}, \Sigma_{R_A, t'})}, \lambda_{C(\Sigma_{R_B, t}, \Sigma_{R_B, t'})})} &= \\
P(A|a, \lambda_{C(\Sigma_{R_A, t}, \Sigma_{R_A, t'})}, \lambda_{C(\Sigma_{R_B, t}, \Sigma_{R_B, t'})}) \cdot P(B|b, \lambda_{C(\Sigma_{R_A, t}, \Sigma_{R_A, t'})}, \lambda_{C(\Sigma_{R_B, t}, \Sigma_{R_B, t'})}) &= \\
P(A|a, \lambda_{C(\Sigma, \Sigma)}) \cdot P(B|b, \lambda_{C(\Sigma, \Sigma)}) &.
\end{aligned}$$

In the consecutive steps of this derivation we have used the following: definition of $\lambda_{C(\Sigma, \Sigma)}$; definition of conditional probability; theorem about conditional probability; (26) and (27); (28) and (29); theorem about conditional probability and definition of conditional probability; and finally, definition of $\lambda_{C(\Sigma, \Sigma)}$.

A.3 Locally deterministic HVTs satisfy Local Causality $_{C(\Sigma, \Sigma)}$ but violate Settings Independence $_{C(\Sigma, \Sigma)}$

Theorem 2. Any locally deterministic HVT satisfies Local Causality $_{C(\Sigma, \Sigma)}$.

Proof. In a locally deterministic theory, $P(\lambda_R|\lambda_{C(\Sigma_{R, t}, \Sigma_{R, t'})})$ is equal to 0 or 1 because $\lambda_{\Sigma_{R, t}}$ uniquely determines the complete state in R (we use our Bridge Principle Between Determinism and 0-1 Probability, Definition 3). Therefore, adding any further conditionalization either does not change the value of probability or makes it ill-defined (the latter will happen if we add conditionalization over some $\lambda_{R''}$ that is inconsistent with $\lambda_{C(\Sigma_{R, t}, \Sigma_{R, t'})}$). In particular, adding conditionalization in the same way as in the formulation of the fine-grained version of Local Causality $_{C(\Sigma, \Sigma)}$ will not change the value of conditional probability of λ_R . Similarly, $P(\overline{\lambda_{R; d_R}}|\lambda_{C(\Sigma_{R, t}, \Sigma_{R, t'})})$ equals 0 or 1 for any coarse-grained state $\overline{\lambda_{R; d_R}}$. This entails the coarse-grained version of Local Causality $_{C(\Sigma, \Sigma)}$. \square

Theorem 3. Any locally deterministic HVT violates Settings Independence $_{C(\Sigma, \Sigma)}$.

Proof. Assume that T is locally deterministic, and consider Bell's Experiment with the usual notation. Consider any states in regions $R_a, R_b, \Sigma_{R_A, t}$ and $\Sigma_{R_B, t}$, denoted by $\lambda_{R_a}, \lambda_{R_b}, \lambda_{\Sigma_{R_A, t}}$ and $\lambda_{\Sigma_{R_B, t}}$. Since T is locally deterministic, $\lambda_{\Sigma_{R_A, t}}$ is compatible with exactly one state in R_a , and analogously for $\lambda_{\Sigma_{R_B, t}}$ and R_b . Therefore, any pair of states $\lambda_{\Sigma_{R_A, t}}, \lambda_{\Sigma_{R_B, t}}$ is compatible with exactly one pair of states $\lambda_{R_a}, \lambda_{R_b}$. *A fortiori*, any pair of states $\lambda_{C(\Sigma_{R_A, t}, \Sigma_{R_A, t'})}, \lambda_{C(\Sigma_{R_B, t}, \Sigma_{R_B, t'})}$ is compatible with exactly one pair of states $\lambda_{R_a}, \lambda_{R_b}$.

Assume without loss of generality that $\lambda_{C(\Sigma_{R_A, t}, \Sigma_{R_A, t'})}$ is consistent with a but not with a' , and $\lambda_{C(\Sigma_{R_B, t}, \Sigma_{R_B, t'})}$ is consistent with b but not with b' . Then, from our Bridge Principle Between Determinism and 0-1 Probability (Definition 3) it follows that $P(\lambda_{C(\Sigma, \Sigma)}|a, b) \neq 0$ and $P(\lambda_{C(\Sigma, \Sigma)}|a', b') = 0$. Therefore, they cannot both be equal to $P(\lambda_{C(\Sigma, \Sigma)})$, which is a violation of Settings Independence $_{C(\Sigma, \Sigma)}$. \square

It is worth pointing out that Theorem 3 cannot be strengthened to “any deterministic HVT violates Settings Independence $_{C(\Sigma, \Sigma)}$ ”, where “locally deterministic” has been replaced with “deterministic”. In the

case of theories that are deterministic but not locally deterministic, it is the very first step of the proof that will not work. This is because in such a theory, for any t below $R_a \cup R_b$,

$$P(\lambda_{R_a}|\lambda_t) = 0/1 \text{ and } P(\lambda_{R_b}|\lambda_t) = 0/1 \quad (31)$$

but not necessarily

$$P(\lambda_{R_a}|\lambda_{\Sigma_{R_a,t}}) = 0/1 \text{ and } P(\lambda_{R_b}|\lambda_{\Sigma_{R_b,t}}) = 0/1. \quad (32)$$

In other words, in a theory that is deterministic but not locally deterministic, all physical events in $R_a \cup R_b$ are uniquely determined by the full initial conditions at t , but not necessarily by the part of these initial conditions that is contained within $\Sigma_{R_a,t} \cup \Sigma_{R_b,t}$. However, this does not mean that the choices of measurement settings are more “free” in theories that are deterministic but not locally deterministic. This only reveals the fact that Settings Independence $_{C(\Sigma,\Sigma)}$ has built a certain kind of locality into its formulation: it states that the choices of measurement settings are independent of the conditions in the past *restricted* to the past light cone of measurement, and not that the choices of measurement settings are independent of the conditions in the past, *full stop*.