

Classicality and Bell's Theorem

Márton Gömöri

*Department of Logic, Eötvös Loránd University, Budapest
Institute of Philosophy, Research Centre for the Humanities, Budapest*

Email: <gomorim@gmail.com>

Carl Hoefer

ICREA

Department of Philosophy, University of Barcelona

Email: <carl.hoefer@gmail.com>

Abstract

A widespread view among physicists is that Bell's theorem rests on an implicit assumption of "classicality," in addition to locality. According to this understanding, the violation of Bell's inequalities poses no challenge to locality, but simply reinforces the fact that quantum mechanics is not classical. The paper provides a critical analysis of this view. First we characterize the notion of classicality in probabilistic terms. We argue that classicality thus construed is not a mark of the validity of classical physics, nor of classical probability theory, contrary to what many believe. At the same time, we show that the probabilistic notion of classicality is not an additional premise of Bell's theorem, but a mathematical corollary of locality in conjunction with the standard auxiliary assumptions of Bell. Accordingly, any theory that claims to get around the derivation of Bell's inequalities by giving up classicality, in fact has to give up one of those standard assumptions. As an illustration of this, we look at two recent interpretations of quantum mechanics, Reinhard Werner's operational quantum mechanics and Robert Griffiths' consistent histories approach, that are claimed to be local and non-classical, and identify which of the standard assumptions of Bell's theorem each of them is forced to give up. We claim that while in operational quantum mechanics the Common Cause Principle is violated, the consistent histories approach is conspiratorial. Finally, we relate these two options to the idea of realism, a notion that is also often identified as an implicit assumption of Bell's theorem.

Keywords: classicality, Bell's theorem, Common Cause Principle, locality, conditional probability, interpretations of quantum mechanics

1 Introduction

Many physicists are unimpressed by Bell's theorem. A widespread view is that Bell's reasoning rests upon an implicit assumption of "classicality" that directly go against the fundamental principles of quantum mechanics (QM). According to such an understanding, the violation of Bell's inequalities poses no challenge to our causal picture of the world (locality, in particular), but simply reinforces the fact that QM is not classical. One proponent of such a view is Reinhard Werner, who concisely puts it like this (Werner 2014a, p. 4):

Bell showed (maybe against his own intentions ...) that classicality and locality together lead to false empirical conclusions. Of course, all the talk about the non-locality of quantum mechanics really says [is] that any classical extension violates locality ...

In line with this picture, physicist have developed various interpretations of quantum theory that are claimed to be local and non-classical. Among recent variants are Werner's operational quantum mechanics (Werner 2014a,b) and Robert Griffiths's consistent histories approach (Griffiths 2020).

Others object to this view. Criticizing Werner's position about the EPR argument and Bell's theorem, Tim Maudlin (2014b, pp. 1-2) writes:

Werner thinks that Bell and Einstein and I have all tacitly made an assumption of which we are unaware, an assumption he labels *C* for 'classicality'. ... Werner concedes that Bell proved that any classical theory that violates his inequalities must be non-local. But deny classicality and the arguments no longer go through. ... The condition *C* is easily stated: it is that the state space of a theory forms a simplex. Good. The space of density matrices in quantum theory does not form a simplex, so if one takes the possible physical states of a system to be given by the density matrices, then one's theory is not classical in this sense. That much is clear. But what is not at all clear is *where the assumption that the state space is a simplex is presupposed in either Einstein's or Bell's reasoning.*

It is interesting to compare this debate with an idea developed in the '80s by scholars like Itamar Pitowsky, Arthur Fine, and others. On their interpretation, Bell's inequalities have nothing to do with locality and causality. The violation of the inequalities, on their account, is an indication that the classical laws of probability are no longer applicable in the quantum domain. That is, there is a sense of classicality that directly entails Bell's inequalities, without any further assumption about locality. In Pitowsky's (1989, p. 8) words:

The set of axioms for classical probability entail that frequencies should obey an a-priori set of constraints [Bell-type inequalities] that are often violated by quantum frequencies. The violation itself has a-priori nothing to do with the principle of locality for it often occurs in cases where spatio-temporal aspects play no role whatever.

On Fine's (1982, p. 294) reading:

... hidden variables and the Bell inequalities are all about ... imposing requirements to make well defined precisely those probability

distributions for noncommuting observables whose rejection is the very essence of quantum mechanics.

Notice that Fine says not only that Bell's theorem is about probability rather than locality and causality, but also that Bell's inequalities follow from a condition (classicality) that contradicts QM, suggesting that the violation of the inequalities should not be surprising but should actually be expected in the first place.

In sum, the debate over the role of classicality in Bell's theorem leaves us with a confusing disagreement of the following positions:

- Bell's argument presupposes classicality in addition to locality (Werner)
- Classicality is nowhere referred to in Bell's argument; the only substantial assumption is locality (Maudlin)
- Bell's inequalities follow only from classicality; no further locality assumption is needed (Pitowsky and Fine)

This situation poses two straightforward questions: 1) Is there a common notion of classicality shared by all parties? 2) If yes, what role exactly does classicality thus construed play in Bell's theorem? The aim of this paper is to clarify these questions.

We will proceed as follows. Section 2 answers question 1 positively: we show that Werner's notion of classicality (condition C above) is equivalent with Pitowsky's and Fine's probabilistic conditions of the existence of a Kolmogorovian representation of quantum probabilities and the existence of joint distributions, respectively. Then, with an unambiguous notion of classicality at hand, Section 3 answers question 2: we demonstrate that classicality is *not* a presupposition of Bell's theorem but a *consequence* of the standard causal-statistical assumptions. Next, in Section 4, we investigate how the approaches of Werner and Griffiths can claim to get around Bell's theorem. In light of what we will have shown about classicality, it is clear that in getting around the derivation of Bell's inequalities each of the two approaches in question must violate one of the standard causal-statistical assumptions of Bell's theorem. We claim that while in Werner's operational quantum mechanics the Common Cause Principle is violated, in the consistent histories approach of Griffiths, the formulation of quantum theory turns out to be conspiratorial. Finally, in Section 5, we relate these two options to the idea of realism, a notion that is also often identified as an implicit assumption of Bell's theorem. The Appendix contains the proofs of the two central mathematical propositions that we formulate in the main text.

2 Classicality as a Probabilistic Notion

Recall Pitowsky's (1989) formalism. Let n be a natural number and S be a subset of $\{(i, j) | i < j; i, j = 1, 2, \dots, n\}$. Suppose we are given $n + |S|$ numbers ($|S|$ denotes the number of elements of S):

$$\begin{array}{ll} p_i & i = 1, 2, \dots, n \\ p_{ij} & (i, j) \in S \end{array} \quad (1)$$

with $0 \leq p_i, p_{ij} \leq 1$. We arrange these numbers in a so-called a *correlation vector*

$$\vec{p} = (p_1, p_2, \dots, p_n, \dots, p_{ij}, \dots) \in \mathbb{R}^{n+|S|} \quad (2)$$

where the index pairs $(i, j) \in S$ are ordered lexicographically. \vec{p} will be thought of as an array of experimentally ascertained probabilities of n outcomes of some measurements performed on a given system, and some of the correlations of these outcomes (the probabilities of their conjunctions). \vec{p} can be seen as a (partial) description of the system's "state," characterizing how the system is disposed to react to certain measurements performed on it.

As an example consider a 2×2 -type EPR–Bohm (EPRB) scenario. In each wing of a 2×2 EPRB experiment one selects from two given measurement settings (directions). Label by a_1, a_2 the settings on the left, by b_3, b_4 the settings on the right. Let A_1, A_2, B_3, B_4 denote the corresponding spin up outcomes. The probabilities of spin outcomes yielded by the experiment and predicted by QM can be arranged in a correlation vector of type $n = 4$, $S_{\text{EPR}} = \{(1, 3), (1, 4), (2, 3), (2, 4)\}$:

$$\vec{p}_{\text{EPR}} = (p_1, p_2, p_3, p_4, p_{13}, p_{14}, p_{23}, p_{24}) \in \mathbb{R}^{4+4} \quad (3)$$

with¹

$$\begin{aligned} p_i &= p(A_i|a_i) & i &= 1, 2 \\ p_j &= p(B_j|b_j) & j &= 3, 4 \\ p_{ij} &= p(A_i \cap B_j|a_i \cap b_j) & (i, j) &\in S_{\text{EPR}} \end{aligned} \quad (5)$$

\vec{p}_{EPR} provides a (partial) description of the spin state of the two-particle system (or an ensemble of such systems) prepared and measured in an EPRB experiment.

We now recall and precisely formulate notions of when such a description—and hence the system in question and its state—is regarded as "classical."

The first notion (Pitowsky 1989) requires that a correlation vector be composed of numbers that satisfy Kolmogorov's axioms, so they be classical probabilities.

Definition 1. Correlation vector \vec{p} admits a classical probability space representation iff there exists a classical probability space (X, \mathcal{A}, μ) and $E_1, E_2, \dots, E_n \in \mathcal{A}$ such that

$$\begin{aligned} p_i &= \mu(E_i) & i &= 1, 2, \dots, n \\ p_{ij} &= \mu(E_i \cap E_j) & (i, j) &\in S \end{aligned} \quad (6)$$

The second notion (Fine 1982) requires that the probability values in a correlation vector arise from a joint distribution as marginal probabilities.

¹Note that the non-signaling character of the EPRB scenario means that these probabilities obey

$$\begin{aligned} p(A_i|a_i) &= p(A_i|a_i \cap b_j) \\ p(B_j|b_j) &= p(B_j|a_i \cap b_j) \\ & (i, j) \in S_{\text{EPR}} \end{aligned} \quad (4)$$

Definition 2. Correlation vector \vec{p} is *extractable from a joint distribution* iff there exist 2^n numbers² $p_{\alpha_1 \dots \alpha_n}, \alpha_1, \dots, \alpha_n \in \{+, -\}$ such that

$$\sum_{\alpha_1, \dots, \alpha_n \in \{+, -\}} \begin{array}{l} 0 \leq p_{\alpha_1 \dots \alpha_n} \leq 1 \\ p_{\alpha_1 \dots \alpha_n} = 1 \end{array} \quad (7)$$

and

$$\begin{aligned} p_i &= \sum_{\substack{\alpha_1, \dots, \alpha_n \in \{+, -\} \\ \alpha_i = +}} p_{\alpha_1 \dots \alpha_n} & i = 1, 2, \dots, n \\ p_{ij} &= \sum_{\substack{\alpha_1, \dots, \alpha_n \in \{+, -\} \\ \alpha_i, \alpha_j = +}} p_{\alpha_1 \dots \alpha_n} & (i, j) \in S \end{aligned} \quad (8)$$

Consider the set Ω of all possible correlation vectors that can be experimentally realized—with fixed type of measurements, but varying ways in which the system is prepared before the measurements are carried out. One assumes that Ω is a convex set in $\mathbb{R}^{n+|S|}$, so that the statistical mixture of realizable probabilities is also realizable. Ω can be associated with the system’s “state space.” The third notion of classicality (Barrett 2007; Werner 2014) characterizes the state space of a system as a convex set: it requires that the state space of a classical system be a *simplex*, that is, every state has a unique decomposition as a convex combination of extreme points of state space. Since correlation vectors don’t necessarily provide a *complete* description of the system’s “state” in the sense of specifying the probabilities of atomic events, here we give a slightly modified formulation of this idea, one where Ω itself is not required to be a simplex, but be obtainable as a *projection* (that is, partial description) of one.

Definition 3. Let $\Omega \subset \mathbb{R}^{n+|S|}$ be a set of correlation vectors. Ω is *projectable from a probability simplex* iff there exists a probability simplex $\Delta_d \subset \mathbb{R}^d$ with d number of vertices for some positive integer d , a linear map $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}^{n+|S|}$, and sets of indices $R_i \subseteq \{1, 2, \dots, d\}, i = 1, 2, \dots, n$ such that

$$\Omega \subseteq \varphi(\Delta_d) \quad (9)$$

and for all $\vec{p} \in \Omega, \boldsymbol{\pi} \in \Delta_d$, if $\varphi(\boldsymbol{\pi}) = \vec{p}$ then

$$\begin{aligned} p_i &= \sum_{r \in R_i} \pi_r & i = 1, 2, \dots, n \\ p_{ij} &= \sum_{r \in R_i \cap R_j} \pi_r & (i, j) \in S \end{aligned} \quad (10)$$

where π_r is the r th component of $\boldsymbol{\pi}$.³

In the foundations of QM literature the above notions of classicality are often used interchangeably. For special, EPRB-type correlation vectors, the equivalence of the first two notions is an immediate consequence of results by Fine (1982) and Pitowsky (1989). However, all three notions are in fact equivalent, for generic correlation vectors (for proof see Appendix):

²Numbers $p_{\alpha_1 \dots \alpha_n}$ encode the probabilities of the 2^n number of atomic events constructed from the n events in question.

³Numbers π_r ($r = 1, 2, \dots, d$) are probabilities of some atomic events. Index set R_i corresponds to the set of those atomic events where the i th event of the n events we talk about occurs.

Proposition 4. Consider a set of correlation vectors $\Omega \subset \mathbb{R}^{n+|S|}$. The following conditions are equivalent:

- (i) For all $\vec{p} \in \Omega$, \vec{p} admits a classical probability space representation.
- (ii) For all $\vec{p} \in \Omega$, \vec{p} is extractable from a joint distribution.
- (iii) Ω is projectable from a probability simplex.

The consequence of Proposition 4 is that classicality as a probabilistic notion has an unambiguous meaning. Many hold that classicality thus construed is an implicit assumption of Bell’s theorem, and giving up this assumption then provides a way to get around the violation of Bell’s inequalities—a particularly natural way, it is held, given that classicality is already in contradiction with the fundamental principles of QM. In the next section we will argue that this picture is mistaken: in fact, classicality is *not* a presupposition of Bell’s theorem, in addition to the standard causal-statistical assumptions, rather it is a *corollary* of those.

3 Classicality and Bell’s Theorem

Bell’s theorem can be and has been formulated in various different ways. Here we consider a commonly accepted derivation that is more general than Bell’s original 1964 reasoning in that it doesn’t presuppose perfect correlations and is based on the notion of a common cause.

The probabilities measured in an EPRB experiment and encoded in correlation vector \vec{p}_{EPR} ((3)–(5)) in general display statistical correlations between outcomes in the two wings. In general, we have

$$p(A_i \cap B_j | a_i \cap b_j) \neq p(A_i | a_i \cap b_j) p(B_j | a_i \cap b_j) \quad (i, j) \in S_{\text{EPR}} \quad (11)$$

Since the two wings are space-like separated, the only way these correlations can be explained is by assuming the existence of some correlated properties, commonly described by a “hidden variable,” that the particles carry with themselves right from their emission and that are responsible for the outcomes (even if in a probabilistic sense).⁴ As many have rightly emphasized (e.g. Bell 2004, p. 143–144; Norsen 2007, pp. 318–319; Maudlin 2014a, p. 5), these pre-existing properties *are not presupposed but inferred* in Bell’s reasoning. Given the experimentally verified statistics, the fundamental presuppositions from which the existence of these properties derives are in most general terms captured by the following two principles.

- 1) *Locality*: There can exist no direct causal connection between space-like separated events.

⁴As noted by Fine (1989), one could also simply take it as a brute fact that the laws of physics enforce certain correlations on measurement outcomes in EPRB-type systems, independently of whether the measurements are space-like separated or not. Note that if the laws enforce correlations between space-like separated events in this way, then the laws are, in an important sense, non-local in themselves: what the laws dictate for the future behavior of system B is not affected by just what has happened in the recent causal past of B, but instead may depend on what is happening in distant, space-like separated regions. A second thing to note is that this brute law-explanation claim—if the non-locality that it entails is not interpreted in a causal way—is in contradiction with the Common Cause Principle which we introduce below.

- 2) *Common Cause Principle*: Robust probabilistic correlations do not occur in nature as a matter of pure accident, or *mere hap*. Any such correlation must be brought about either by direct or by common causal connection.⁵

1–2 entail the existence of a common cause—which we may think of as the physical event that determines the physical properties of the particles after emission—in terms of which the EPRB correlations (11) can be explained.⁶ What it means to explain these correlations is characterized by a second pair of assumptions.

Let C_k ($k \in K$) denote a partition of events, describing the common cause.⁷

- 3) *Factorization*:

$$p(A_i \cap B_j | a_i \cap b_j \cap C_k) = p(A_i | a_i \cap C_k) p(B_j | b_j \cap C_k) \quad (12)$$

$(i, j) \in S_{\text{EPR}}, k \in K$

- 4) *No-conspiracy*:

$$p(C_k | a_i \cap b_j) = p(C_k) \quad (13)$$

$(i, j) \in S_{\text{EPR}}, k \in K$

Both factorization and no-conspiracy are statistical independence conditions. Factorization expresses the requirement that conditionalizing on the common cause C_k leaves no residual correlation between A_i and B_j , given the chosen measurement angles *on their respective sides*. No-conspiracy expresses the assumption that the choice about which angles to measure, which is something that can be done at the last moment and by any selection-procedure one likes, can not influence, nor be influenced by, the common cause, and hence the two must be uncorrelated.

It is worth noting that the four conditions are in fact inextricably intertwined. Both factorization and no-conspiracy incorporate the Common Cause Principle in a trivial sense: if there could be *robust* correlations in the world occurring as a matter of pure accident, then requiring these independence conditions would have no ground, for any such accidental correlation could spoil these independencies. Further, factorization is often formulated as a joint result of two conditions: 1) outcome independence

$$p(A_i \cap B_j | a_i \cap b_j \cap C_k) = p(A_i | a_i \cap b_j \cap C_k) p(B_j | a_i \cap b_j \cap C_k) \quad (14)$$

⁵The first formulation of the argument from locality to pre-existing properties is that of EPR of course. The EPR argument famously employs the Reality Criterion. As is argued by Gömöri and Hofer-Szabó (2021), the Reality Criterion is just a special case of the Common Cause Principle for perfect correlations.

⁶In Bell’s argument this is where the so-called “hidden variable” terms λ come in. At times physicists such as Werner and Griffiths present this introduction of a factor accounting for measurement results as an unacceptable move amounting to the presupposition of realism, or “classical realism.” We feel it is better to keep the mathematical notion of classicality (Werner 2014a, p. 3) separate from the more general idea that particles have properties that play a role in explaining measurement outcomes and the correlations observable in them, which Werner fails to do (as does Griffiths (2020, p. 15)). In Section 5 we briefly return to the question of how realism in the latter sense is related to the assumptions of Bell’s argument.

⁷The notion of common cause we employ here is what Hofer-Szabó et al. (2013, Sec. 7) terms a *common cause system*.

which is a characterization of the common cause as a screener-off; and 2) parameter independence

$$p(A_i|a_i \cap b_j \cap C_k) = p(A_i|a_i \cap C_k) \quad (15)$$

$$p(B_j|a_i \cap b_j \cap C_k) = p(B_j|b_j \cap C_k) \quad (16)$$

which is taken to be required by locality. Finally, note that no-conspiracy, as a statement about *statistical* independence, is also a compound condition: it not only incorporates the idea of the autonomy of measurement choice (sometimes referred to as no-conspiracy in a narrower sense), according to which the settings of measurements can neither be directly influenced by the C_k -s, nor can there be common causal connection between them. But it also incorporates the idea of no retrocausation. This is because statistical independence (13) could also break down in a way that, reversely, the measurement choices have an effect on the C_k -s, and since the measurement choice can be made at the last moment before the measurement, while the C_k -s, characterizing the common cause, are localized at the emission of particles, this would involve retrocausal connection.

Conditions 1–4 above will be referred to as the *standard causal-statistical assumptions of Bell’s theorem*. We do not claim that these conditions cover every detail that the derivation of Bell’s inequalities rests upon⁸—though we believe they condense the substantial assumptions—, nor are these conditions independent or non-redundant, as we have just seen. None of this will be relevant to our argument; the main ingredient of which is the mathematical fact that given these four assumptions correlation vector \vec{p}_{EPR} must be classical. To formulate and prove this we will use Pitowsky’s characterization of classicality as is encapsulated in Definition 1. The following statement is a consequence of results by Fine (1982) in conjunction with Proposition 4. In the Appendix we give a more explicit proof of it based on Hofer-Szabó (2020).

Proposition 5. *Suppose that there is a partition of events $C_k (k \in K)$ for which (12)–(13) hold. Then \vec{p}_{EPR} admits a classical probability space representation.*

Conditions 1-2 plus the EPRB statistics imply the existence of a common cause (hidden variable) assumed to be characterized by conditions 3-4. Conditions 3-4 imply that \vec{p}_{EPR} must be classical. Thus, in sum, the standard causal-statistical assumptions of Bell’s theorem imply that \vec{p}_{EPR} is a classical correlation vector.

In light of this result, the following remarks about the conceptual terrain are in order. Firstly, as Pitowsky (1989) and Fine (1982) proved, classicality (the mathematical condition) *alone* implies Bell’s inequalities. Therefore it is strictly speaking incorrect to say, as Werner (2014a) does, that classicality is an *additional* premise of Bell’s theorem, *on top of the standard causal-statistical assumptions*. Werner here seems to have in mind the introduction of the common cause C_k (“a hidden state λ ”), which he takes to either imply, or be equivalent to, classicality. But as we have seen, the necessity of introducing C_k follows from the standard causal-statistical assumptions alone, so it is not an *additional* assumption.

⁸For example, the argument obviously employs the laws of classical logic. Some suggest (e.g. Santos 1986; Weingartner 2009) that these too qualify as premises that might be given up in light of the violations of Bell’s inequalities.

Secondly, the Pitowsky–Fine derivation of Bell’s inequalities is often interpreted as a demonstration that Bell’s theorem has nothing to do with locality, causality, etc., but is instead essentially about probability. We have two remarks on this view. The first one is simply logic: Since Bell’s inequalities can be derived from two alternative sets of premises (the standard causal-statistical assumptions on the one hand, and classicality on the other), the violation of the inequalities implies that *both* sets of premises must contain a false one. That is, *both* classicality and the standard causal-statistical assumptions have to be violated in the world.

But this picture is still, on its own, potentially misleading. For while the standard causal-statistical assumptions (locality, the Common Cause Principle, no-conspiracy, etc.) are all robust physical/metaphysical principles which we have strong reasons to assume, the mathematical condition of classicality *in itself* is completely unreasonable and unmotivated. This is because the components of correlation vector \vec{p}_{EPR} , (5), are *conditional* probabilities:

$$\vec{p}_{\text{EPR}} \left\{ \begin{array}{l} p_1 = p(A_1|a_1) \\ p_2 = p(A_2|a_2) \\ p_3 = p(B_3|b_3) \\ p_4 = p(B_4|b_4) \\ p_{13} = p(A_1 \cap B_3|a_1 \cap b_3) \\ p_{14} = p(A_1 \cap B_4|a_1 \cap b_4) \\ p_{23} = p(A_2 \cap B_3|a_2 \cap b_3) \\ p_{24} = p(A_2 \cap B_4|a_2 \cap b_4) \end{array} \right. \quad (17)$$

Values of conditional probabilities pertaining to different conditions do not form a probability measure in general, and so it makes no sense in general to require that these values obey Kolmogorov’s axioms, that is, that they be representable in a classical probability space in accord with Definition 1.⁹

The only way classicality is motivated in EPRB is that, as stated by Proposition 5, it is a mathematical corollary of the standard causal-statistical assumptions of Bell’s theorem. So what the Pitowsky–Fine derivation of Bell’s inequalities provides is not a new understanding of Bell’s theorem, but yet another way of deriving Bell’s inequalities from the standard assumptions:

standard causal-statistical assumptions

↓

classicality

↓

Bell’s inequalities¹⁰

Therefore, the mathematical condition of classicality makes sense in cases where the standard causal-statistical assumptions apply: where we have space-like separated subsystems that are assumed to behave locally, etc. Importantly,

⁹This observation was first made by Szabó (1995).

¹⁰The second arrow here could be a biconditional (\Downarrow), as proved by Pitowsky (1989) and Fine (1982), but it is the downward direction that matters for our discussion.

when one or more of those assumptions does not hold, classicality (again, the *mathematical condition*) may fail to hold, even in *classical* physical systems.

One way that this can happen is if the system in question is composed of *time-like* separated subsystems that are allowed to interact. As an example, imagine Laurel and Hardy on a teeter-totter. Assume that Hardy weights twice as much as Laurel. We want to see if Laurel and Hardy go up or down, under the following conditions:

- a_1 : Laurel sits 1.5 meter away from the center of the teeter-totter
- a_2 : Laurel sits 1 meter away from the center
- b_3 : Hardy sits 1.5 meter away from the center
- b_4 : Hardy sits 0.5 meter away from the center

Introduce the following outcome events:

- A_1 : Laurel goes up when he sits 1.5 meter away from the center
- A_2 : Laurel goes up when he sits 1 meter away from the center
- B_3 : Hardy goes down when he sits 1.5 meter away from the center
- B_4 : Hardy goes down when he sits 0.5 meter away from the center

Suppose that the teeter-totter experiment is performed repeatedly. Elementary physics entails the following probabilities:¹¹

$$\begin{aligned}
 p_{13} &= p(A_1 \cap B_3 | a_1 \cap b_3) = 1 \\
 p_{14} &= p(A_1 \cap B_4 | a_1 \cap b_4) = 0 \\
 p_{23} &= p(A_2 \cap B_3 | a_2 \cap b_3) = 1 \\
 p_{24} &= p(A_2 \cap B_4 | a_2 \cap b_4) = \frac{1}{2}
 \end{aligned}
 \tag{18}$$

Further, assuming that both Laurel and Hardy sit far from the center and close to the center half of the times, independently of each other, we have

$$\begin{aligned}
 p_1 &= p(A_1 | a_1) = \frac{1}{2} \\
 p_2 &= p(A_2 | a_2) = \frac{3}{4} \\
 p_3 &= p(B_3 | b_3) = 1 \\
 p_4 &= p(B_4 | b_4) = \frac{1}{4}
 \end{aligned}
 \tag{19}$$

Now, correlation vector $\vec{p}_{\text{LH}} = (p_1, p_2, p_3, p_4, p_{13}, p_{14}, p_{23}, p_{24})$ is not classical since, for example, $p_1 < p_{13}$. There certainly cannot exist a classical probability space with events E_1 and E_3 in it such that

$$\begin{aligned}
 p_1 &= \mu(E_1) \\
 p_{13} &= \mu(E_1 \cap E_3)
 \end{aligned}
 \tag{20}$$

as that would entail $\mu(E_1) < \mu(E_1 \cap E_3)$, in contradiction with Kolmogorov's laws of probability. But this fact by no means implies the break down of classical probability theory, let alone classical physics, in any sense. It is simply that the values of classical *conditional* probabilities pertaining to *different* conditions— $p_1 = p(A_1 | a_1)$ and $p_{13} = p(A_1 \cap B_3 | a_1 \cap b_3)$ —do not form a probability measure.

¹¹When Hardy sits twice as close to the center as Laurel, they in principle balance. In this case assume small perturbations to determine how they move, so that half of the times Laurel, half of the times Hardy ends up going up. That's why we have 1/2 in the last row of (18).

At the same time, this example witnesses an obvious violation of the standard assumptions of Bell’s theorem.¹² The outcomes are correlated, for we have:

$$\frac{1}{2} = p(A_2 \cap B_4 | a_2 \cap b_4) \neq p(A_2 | a_2 \cap b_4) p(B_4 | a_2 \cap b_4) = \frac{1}{2} \cdot \frac{1}{2} \quad (21)$$

The obvious explanation of this correlation is the *direct* physical influence between the two ends of the teeter-totter, which ensures that when one end goes down, the other end goes up. Since there is direct causal connection, the Common Cause Principle no longer demands the existence of a common cause satisfying factorization and no-conspiracy. Indeed, there is just no such an event in the example. For instance, B_4 as a potential direct cause of A_2 does screen off correlation (21), meaning that partition $\{C_k\}_{k=1,2} = \{B_4, \neg B_4\}$ satisfies outcome independence (14), as well as factorization (12). But B_4 fails to satisfy no-conspiracy (13), for the obvious reason that B_4 can only occur when the corresponding “measurement” b_4 is performed, so there is strong correlation between B_4 and b_4 . Again, the fact that there is no common cause C_k satisfying the standard Bell assumptions comes as no surprise since, unlike in the EPRB scenario, events on the two sides of the teeter-totter can, and in fact do, causally influence each other, with no contradiction to locality.

With these remarks in mind, it is worth mentioning a strand of approaches to get around Bell’s theorem. The experimental violation of Bell’s inequalities in conjunction with the Pitowsky–Fine derivation of the inequalities implies that \vec{p}_{EPR} does not admit a classical probability space representation. A reaction shared by many scholars is that the way to evade this problem is generalizing the notion of probability space, relaxing some of Kolmogorov’s axioms, so that under the generalized notion of probability space \vec{p}_{EPR} does admit a probability space representation. There are many proposals in this direction, for an overview see Feintzeig and Fletcher (2017). In light of what we said about classicality, here we briefly mention two possible concerns with these approaches. First of all, it must be emphasized that there is a clear sense in which the probability values in \vec{p}_{EPR} can be represented in a *classical* probability space: not as absolute probabilities of events, as Definition 1 requires, but as *conditional* probabilities conditioned on different conditioning events, in line with (17).¹³ Similarly in our example: \vec{p}_{LH} is not a classical correlation vector, but no one would take this as evidence of Kolmogorov’s probability rules being violated by the teeter-totter system. If one wants to write down the numbers in \vec{p}_{LH} as values of probabilities in a probability space, they will be *conditional* probabilities in a *classical* probability space. Accordingly, to accommodate the Bell inequality violating correlation vectors in a probability space it is not necessary to generalize the notion of probability. Secondly, even if one did that, it must be clear that this move doesn’t save locality. This is because classicality (in the sense of Definition 1), as we have argued, is *not* among the premises of the *standard* derivation of Bell’s

¹²It is easy to verify that probabilities (18)–(19) violate Bell’s inequalities, more precisely, the following Clauser–Horne inequality:

$$-1 \leq p_{24} + p_{23} + p_{13} - p_{14} - p_2 - p_3 \leq 0$$

¹³The thesis that quantum probabilities can always be interpreted as classical conditional probabilities is called the Kolmogorovian Censorship Hypothesis by Szabó (1995). That fact that this is indeed always possible mathematically, has been proved in various forms (Bana and Durt 1997; Szabó 2001; Rédei 2010).

inequalities—it is just not a condition that one could deny in order to evade the derivation of the inequalities from locality, no-conspiracy, etc. (without also having to deny one of these standard assumptions).¹⁴

Note that what we said about classicality is equally true for the existence of non-contextual hidden variables: 1) Bell’s inequalities can be derived from the existence of non-contextual hidden variables (Shimony 1984, pp. 30-31). 2) Many believe that this makes the derivation from the standard causal-statistical assumptions irrelevant. 3) But the derivation from non-contextual hidden variables does not invalidate the derivation from the standard causal-statistical assumptions: the violation of Bell’s inequalities implies that *both* non-contextual hidden variables and locality (or another one of the standard causal-statistical assumptions) must go. 4) Furthermore, as with classicality, the existence of non-contextual hidden variables *in itself* is not well-motivated. Our teeter-totter system is again a good example, since it displays an obvious violation of non-contextuality in the following sense. Let C_k now describe not a common cause but the system’s “ontic state,” that is, for the teeter-totter system, all the physical factors together, including the weights of Laurel and Hardy and the small perturbations in play in the balance case (but excluding conditions a_1, a_2, b_3, b_4), that go into determining which one of the two goes up and down. Non-contextuality is the condition that the ontic state determines the probability of the outcomes of each measurement independently of what other measurements are simultaneously performed;¹⁵ which is, in our case, is nothing but the condition of parameter independence (15)–(16). The violation of this condition is due to the fact that, even when C_k is given,¹⁶ the outcome on one side (whether Laurel/Hardy goes up or down) depends on the measurement choice on the other side (where Hardy/Laurel sits, respectively). Again, contextuality in this sense comes as no surprise since the two ends can and do physically interact. 5) That said, the existence of non-contextual hidden variables, just as classicality, does follow when we have space-like separated subsystems that are assumed to behave locally, etc., that is, where the standard causal-statistical assumptions apply. For in that case C_k in assumptions 3–4 will just be a non-contextual hidden variable. The most well-known derivation of non-contextual hidden variables from locality, etc. is the EPR argument.

Indeed, the tight connection between classicality, non-contextual hidden variables, and the standard causal-statistical assumptions is especially transpar-

¹⁴Note that the function $p(\cdot)$ itself (but not $p(\cdot|X)$ for varying $X!$) is indeed assumed to obey Kolmogorov’s axioms. That is, in the derivation of Bell’s inequalities we assume that there is a classical probability space (X, \mathcal{A}, p) where $a_1, a_2, b_3, b_4, A_1, A_2, B_3, B_4$ and $C_k (k \in K)$ are represented as events, and probability measure p , in terms of which we formulate conditions (12)–(13), obeys Kolmogorov’s axioms. This is explicitly assumed in the proof of Proposition 5 (see Appendix). It must be emphasized however that the classicality of $p(\cdot)$ (but not of $p(\cdot|X)$ for varying $X!$) is an analytic consequence of p being interpretable as relative frequency of events occurring in the runs of an EPRB experiment; and without such an interpretation the violation of Bell’s inequalities cannot be said to be empirically confirmed (cf. Szabó 2001, Sec. 2.2).

¹⁵This notion of non-contextuality is what Hofer-Szabó (2022) calls *simultaneous* non-contextuality.

¹⁶Note that in this simple example not only non-contextuality but also no-signaling (4) is violated. One can give examples of more finely-tuned classical interacting systems where the measurement statistics obey no-signaling, but the underlying ontic state fails to satisfy non-contextuality (and Bell’s inequalities and classicality are also violated). For such an example see e.g. Hofer-Szabó (2021), Sec. 4.

ent in the *deterministic* case. Suppose we have parallel measurement directions in the two wings of EPRB, with *perfect* correlation between outcomes of measurements in the same direction. Perfect correlation can only be explained by a deterministic common cause (Hofer-Szabó et al. 2013, p. 15, Proposition 2.7). This fact, together with no-conspiracy (13), imply that $p(A_i|a_i \cap b_j \cap C_k), p(B_j|a_i \cap b_j \cap C_k) \in \{0, 1\}, (i, j) \in S_{\text{EPR}}$. Parameter independence (15)–(16) further entails

$$p(A_i|a_i \cap C_k), p(B_j|b_j \cap C_k) \in \{0, 1\} \quad (i, j) \in S_{\text{EPR}} \quad (22)$$

Now introduce the following events:

$$\begin{aligned} C_{A_i} &= \bigcup_{\substack{k \in K \\ p(A_i|a_i \cap C_k)=1}} C_k \\ C_{B_j} &= \bigcup_{\substack{k \in K \\ p(B_j|b_j \cap C_k)=1}} C_k \\ &(i, j) \in S_{\text{EPR}} \end{aligned} \quad (23)$$

In conjunction with (22), the standard causal-statistical assumptions imply (see formula (32) in Appendix):

$$\begin{aligned} p(A_i|a_i) &= p(C_{A_i}) \\ p(B_j|b_j) &= p(C_{B_j}) \\ p(A_i \cap B_j|a_i \cap b_j) &= p(C_{A_i} \cap C_{B_j}) \\ &(i, j) \in S_{\text{EPR}} \end{aligned} \quad (24)$$

That is, the *conditional* probabilities of outcome events figuring in \vec{p}_{EPR} must be equal with the *absolute* probabilities of the events that predetermine these outcomes. This gives Proposition 5 a straightforward interpretation: since on the right hand side we have absolute probabilities of events that are representable in a classical probability space, the values of conditional probabilities on the left hand side must also be so representable: \vec{p}_{EPR} must be classical. On the other hand, what (24) says is that the measurement outcomes A_i, B_j simply reveal the pre-existing properties C_{A_i}, C_{B_j} —which is exactly the idea behind non-contextual hidden variables. It must be stressed that equalities (24), and thus both classicality and the existence of non-contextual hidden variables, is a consequence of the standard causal-statistical assumptions of Bell’s theorem, including, importantly, the causal separation of the subsystems we consider.

All this implies that when the system in question is *not* composed of space-like separated subsystems then we have no automatic reason to expect that classicality, the existence of non-contextual hidden variables and Bell’s inequalities will be satisfied in the first place. A *quantum* example of such a system is precisely that of the spin-3/2 Neon atom, as discussed in Griffiths (2020): Looking at expected values of spins along different axes, one can derive predicted values that violate a Bell inequality (specifically the CHSH inequality). Griffiths (2020, p. 3) urges us, on the basis of this example, to view things in this way (similar to Pitowsky and Fine quoted in the introduction):

Thus the violation of the CHSH inequality in this case has nothing to do with nonlocality. Instead it has everything to do with the fact that in quantum mechanics, unlike classical mechanics, physical properties and variables are represented by *noncommuting operators*.

Let us make a few remarks here. Firstly, one way to characterize the significance of noncommuting operators is what Fine (1982) suggests in the passage quoted in the introduction: according to the accepted view physical variables represented by noncommuting operators cannot be measured simultaneously and so in general they need not, and in fact they do not, have a joint distribution, in contradiction with what's required by classicality in the sense of Definition 2. Notice however that the failure of existence of joint distributions in that sense is not specific to QM. Indeed, the same holds in our Laurel and Hardy example: since all three definitions of classicality are equivalent, and classicality in the sense of Definition 1 is violated in the Laurel and Hardy case, this means that classicality in the sense of Definition 2 also fails to hold for this simple classical physical system. Therefore, the absence of joint distributions that comes with noncommuting operators in the formalism doesn't seem to mark off quantum *physics* from classical *physics*.

Secondly, there indeed is an *intuition* behind non-contextual hidden variables that might come from (simple cases in) classical physics. In the Neon atom case, one might say “the value of its spin along any given direction ought to be well-defined at all times; after all, spin might be roughly analogous to a classical angular momentum vector, the projection of which along any direction always has a well-defined value.” But the idea of non-contextual hidden variables not only incorporates the assumption that there are real, well-defined properties existing at all times; but, crucially, it brings with it a specific, and rather simplified, picture of how these properties are revealed in measurements: that the outcome of a measurement only depends on the corresponding property at present, irrespective of what other measurements, that is physical interactions, take place. However, this latter picture of the measurement process is something that is generally, perhaps even typically, *not* true in classical physics. The Laurel and Hardy example is again a case in point. The physics of it incorporates real properties that are well-defined at all times: the weights of Laurel and Hardy, their distances from the center, the density distribution of the board, etc. All these real properties altogether will determine the measurement outcomes, that is which end will go up and down. But this determination is contextual: whether Laurel goes up or down will depend not only on a property of Laurel but also on where Hardy sits. Since non-contextual hidden variables do not exist in general even for classical physical systems, it doesn't seem terribly surprising, in itself, that their existence is not provided for generic physical systems, including the Neon atom.¹⁷

Finally, we agree with Griffiths that the violation of a Bell inequality doesn't necessarily have to do with non-locality, as exemplified by the Neon case. Indeed, the same is true of the violation in the Laurel and Hardy example. While Laurel and Hardy are somewhat spatially separated, that is inessential; what matters is that the events A_1, B_3 etc. are (a) not *space-like separated* and indeed are (b) causally connected. So the violation of a Bell inequality in general is neither surprising nor relevant to whether there is locality in the world. By contrast, when a Bell inequality has been derived for a setup like EPRB experiments, with locality as a fundamental physical assumption among the premises, and the inequality is found to be violated in actual experiments, this does bear on the

¹⁷To see how spin is understood as a contextual property in the Bohmian version of QM see e.g. Daumer et al. (1996).

principle of locality!¹⁸ By analogy: to say that the violation of Bell’s inequalities in EPRB has nothing to do with non-locality because the inequalities can be violated in cases where locality perfectly holds, is no better than saying that the violation of energy conservation in a closed system has no bearing on the laws of thermodynamics because energy conservation can be violated in an open system in complete harmony with the those laws.

4 Two Ways to Get Around Bell’s Theorem

Classicality is thus not a condition that one could deny without also having to deny one of the standard causal-statistical assumptions of Bell’s theorem. If any theory of quantum phenomena is to avoid the derivation of Bell’s inequalities, it has to give up one of those standard premises. We will now look at two versions of standard QM, Werner’s operational quantum mechanics and Griffiths’ consistent histories approach, both of which are claimed to evade Bell’s theorem by giving up classicality, that is, claimed to be local non-classical quantum theories. We describe what the two versions are claimed to say about the example of the EPRB scenario, and identify which one of the standard causal-statistical assumptions of Bell’s theorem each theory is in fact forced to give up.

Operational quantum mechanics (OQM) is basically a variant of standard QM in which quantum states ψ are treated as purely epistemic, i.e., as tools for calculating what results to expect from measurements made in various scenarios. In the EPRB case this means that the only role of ψ is to recover, through Born’s rule, the measurement statistics encapsulated in \vec{p}_{EPR} :

$$\begin{aligned} \left\langle \psi_s, \hat{A}_i \otimes \hat{I} \psi_s \right\rangle &= p(A_i | a_i) & i = 1, 2 \\ \left\langle \psi_s, \hat{I} \otimes \hat{B}_j \psi_s \right\rangle &= p(B_j | b_j) & j = 3, 4 \\ \left\langle \psi_s, \hat{A}_i \otimes \hat{B}_j \psi_s \right\rangle &= p(A_i \cap B_j | a_i \cap b_j) & (i, j) \in S_{\text{EPR}} \end{aligned} \quad (25)$$

where \hat{A}_i, \hat{B}_j are projection operators pertaining the outcomes A_i, B_j respectively, and ψ_s is the singlet state.

It is not clear in what sense OQM is meant to be local. On the one hand, Werner seems to suggest that locality simply consists in no-signaling (4), that is in a *statistical* independence condition:

... in the operational approach no prediction about B changes when or if a measurement or other procedure is carried out on A. This independence is built into the structure of quantum theory. This is also the same as the no-signalling condition and the possibility of tracing out system A, getting a reduced state for B, which does not change (and so is undisturbed) whatever happens just to A. (Werner 2014b, p. 4)

On the other hand, later he explicitly condemns the conflation of mere statistical dependence and physical disturbance:

¹⁸As we noted in footnote 4, one could take the results as showing that we should give up the Common Cause Principle, as Fine urges, but this leaves us with a sort of non-locality built into the laws of nature themselves. The only other escape routes run through rejecting no-conspiracy, which is a price so high that only a handful of physicists have been willing to consider paying it (see e.g. Hossenfelder and Palmer 2020).

... if you condition on the outcome of a measurement on A, you get a modified state for B. That is just another way to look at correlation, but never, not even in classical probability, can this be confused with a physical disturbance. The state change only becomes effective when the results from the two labs are brought together and are jointly analyzed, which can happen centuries later.¹⁹ (Werner 2014b, p. 4)

One can nonetheless accept that OQM is local in the sense that the theory doesn't state—in fact, since it only talks about measurement statistics, it has no words to state—the existence of physical disturbance between the two wings. The problem is that in the very same way, the theory doesn't contain any kind of causal mechanism that could serve to correlate the outcomes in the two wings. Indeed, since the ontology of OQM only consists of measurement events and perhaps information states of agents, there are no common causes in it either that could explain the EPRB correlations. Thus, OQM violates the Common Cause Principle, the demand that there be no robust regularities without some sort of causal explanation—which was one of the standard assumptions of Bell's theorem. It is by giving up the Common Cause Principle's requirement that OQM is able to evade the derivation of Bell's inequalities. In effect, the defender of OQM says: “Two guys are flipping coins on opposite sides of the Earth, whenever they get a “Flip!” command from their cousin Alice. There is no causal connection—neither direct, nor common causal—between the outcomes. But the two coins always land oppositely. Deal with it.”

Griffiths' consistent histories approach (CH) to QM is an expression of Bohr's complementarity idea. In CH, every maximal set of compatible measurements is associated with a so-called “framework.” In the EPRB case we have four frameworks pertaining to the four measurement pairs in S_{EPR} . Relative to a given framework, in CH, quantum systems possess real properties that measurements simply reveal. Mathematically, one can model this by assigning to each framework a “small” probability space $(X_{ij}, \mathcal{A}_{ij}, \mu_{ij})$, $(i, j) \in S_{\text{EPR}}$.²⁰ In each of these spaces we have a partition of events $C_{ij}^{++}, C_{ij}^{+-}, C_{ij}^{-+}, C_{ij}^{--} \in \mathcal{A}_{ij}$ corresponding to the particles having spin properties + or – in the chosen pair of directions $(i, j) \in S_{\text{EPR}}$. Then Born's rule is said to recover the probabilities of these properties:

$$\begin{aligned} \left\langle \psi_s, \hat{A}_i \otimes \hat{I} \psi_s \right\rangle &= \mu_{ij} (C_{ij}^{++} \cup C_{ij}^{+-}) \\ \left\langle \psi_s, \hat{I} \otimes \hat{B}_j \psi_s \right\rangle &= \mu_{ij} (C_{ij}^{++} \cup C_{ij}^{-+}) \quad (i, j) \in S_{\text{EPR}} \quad (26) \\ \left\langle \psi_s, \hat{A}_i \otimes \hat{B}_j \psi_s \right\rangle &= \mu_{ij} (C_{ij}^{++}) \end{aligned}$$

These properties are thought to be revealed in measurements and hence the

¹⁹If quantum states are epistemic, one may wonder that the measurement at A does not change the state for B, since it clearly does change the epistemic state of the observer at A who does the measurement! Werner does not say what it means to say that a state change has “become effective,” nor what happens to the state for B if no measurement is made in that wing (so that the results cannot be jointly analyzed).

²⁰For more on employing “large” versus “small” probability spaces for describing quantum experiments, see e.g. Butterfield (1992); Gömöri and Placek (2017).

probabilities of these properties return the components of \vec{p}_{EPR} :²¹

$$\begin{aligned} \mu_{ij} (C_{ij}^{++} \cup C_{ij}^{+-}) &= \mu_{ij} (A_i) = p_i \\ \mu_{ij} (C_{ij}^{++} \cup C_{ij}^{-+}) &= \mu_{ij} (B_j) = p_j \\ \mu_{ij} (C_{ij}^{++}) &= \mu_{ij} (A_i \cap B_j) = p_{ij} \end{aligned} \quad (i, j) \in S_{\text{EPR}} \quad (27)$$

The essential tenet of CH is the “single framework rule”: properties associated with different frameworks, and represented in different probability spaces, do not coexist and cannot be talked about at the same time. In particular, properties $C_{ij}^{\alpha\beta}$ do not coexist for different $(i, j) \in S_{\text{EPR}}$, and hence the particles possess spin only in one direction at a time. This notion is what encapsulates Bohr’s idea of complementary descriptions.

In CH, properties $C_{ij}^{\alpha\beta}$ operate as (deterministic) common causes in the sense that they are assumed to obey outcome independence in each probability space separately:

$$\mu_{ij} (A_i \cap B_j | C_{ij}^{\alpha\beta}) = \mu_{ij} (A_i | C_{ij}^{\alpha\beta}) \mu_{ij} (B_j | C_{ij}^{\alpha\beta}) \quad (i, j) \in S_{\text{EPR}}, \alpha, \beta = +, - \quad (28)$$

Notice however that neither parameter independence, nor no-conspiracy can be written down in the “frameworks” formalism. This is because that would require an identification of the $C_{ij}^{\alpha\beta}$ -s across different frameworks (probability spaces), and that’s exactly what the single framework rule forbids doing. Nonetheless, there is a clear sense in which no-conspiracy is violated in CH. Consider the ensemble of runs of an EPRB experiment. The presence of a spin property of the system, in a given run, depends on the framework in which we choose to describe the system, which in turn depends on the measurements we choose to perform in the given run. Indeed, spin property $C_{ij}^{\alpha\beta}$ will only be assumed to be present in a given run if we choose to perform measurements a_i and b_j . This means that there is strong correlation, over the runs of the EPRB experiment, between the properties we ascribe to the system and the measurements we choose to perform—which is a violation of no-conspiracy. Again, here we use ‘correlation’ in a relative frequency sense rather than a probabilistic sense expressible in terms of probability spaces $(X_{ij}, \mathcal{A}_{ij}, \mu_{ij})$. One way to phrase the point is that in CH, the very existence of the $C_{ij}^{\alpha\beta}$ properties depends on human choices and thus cannot have an antecedent probability. Therefore, the right hand side of equation (13) cannot exist, per the single framework rule.

Thus, in the consistent histories approach, it is the violation of another standard premise of Bell’s theorem, that of no-conspiracy, which blocks the derivation of Bell’s inequalities.

5 Conclusion

As we mentioned in the introduction, many physicists adamantly insist that EPRB-type phenomena do not show that there is genuine non-locality in natural phenomena. Werner and Griffiths are just two prominent voices making such claims in recent years. While they advocate different interpretations of

²¹Here we assume that the measurement outcomes (but not the revealed properties!) can be identified across different frameworks so that we have $A_i, B_j \in \mathcal{A}_{ij}$, $(i, j) \in S_{\text{EPR}}$. Note also that the second equation in each row of (17) makes use of no-signaling (4).

the standard QM formalism, both options can be related to a form of anti-realism. The instrumentalist attitude implicit in OQM is certainly a source of not asking for explanation of correlations, and of not being bothered by the violation of the Common Cause Principle (cf. Lewis 2019). The violation of no-conspiracy we found in the CH approach can also be a result of some form of anti-realism, in which the measurement process is thought to have a role in *constituting* the property of the system measured. Such a constitution relation brings about correlation between performing the measurement and ascribing the property—a correlation which is not due to a causal link but due to a logical/analytic connection. The idea is reminiscent of the Bohrian position that Einstein lamented: the moon is only there when you look at it.

In our view, adopting an anti-realist stance of the above sorts does not save locality. The reason is that the very formulation of locality requires a sort of realism (cf. Norsen 2007). One piece of evidence for this is that in neither of the two “anti-realist” versions of QM in question can one meaningfully formulate the condition of parameter independence (15)–(16), a condition that is usually taken to be a requirement of locality. In OQM there are no C_k -s, so parameter independence cannot be written down. Similarly, as we have seen above, to write down parameter independence in the histories formalism, we need to have an identification of the $C_{ij}^{\alpha\beta}$ -s across different “frameworks”; but that’s exactly what we are forbidden to have in the CH approach.

If this dialectical situation is acknowledged, then the philosophers impressed by Bell’s theorem and the EPRB experiments can come to a peaceful agreement with physicists who are not so impressed. Those physicists prefer a strong dose of anti-realism in their physics, rather than a realistic physics that incorporates non-locality explicitly. The CH advocate embraces a form of conspiracy between the pre-existing properties that are revealed by measurement and the choices we make of what to measure. The OQM advocate must admit that nature somehow displays (enforces?) the inequality-violating correlations, and that nothing in the properties of the measured particles pre-determine (or at least causally explain) what the results would be (cf. Alice’s two friends flipping coins on opposite sides of the world). In both approaches, it must remain a mystery how nature can display these correlations between chancy events at space-like separation.²² As philosophers, we would only ask that the physicists refrain from making two sorts of statements (i) Saying that the QM treatment of EPRB is perfectly local (though they can perfectly well say that the QM treatment is not overtly non-local!). (ii) Saying that Bell did not prove what many philosophers think he proved, because he made a tacit and inappropriate presupposition of “classicality” in his argument. As we have seen (Section 3), “classicality” is a consequence of the standard causal and statistical assumptions made in Bell’s argument, not a separate, tacit assumption.

Appendix

Here we give the proofs of Proposition 4 and 5.

²²As Lewis (2019) points out, the dialectical situation may be different in an Everettian framework, although there is no consensus among philosophers about whether EPRB phenomena involve non-locality in that framework, or not.

Proof of Proposition 4. We show that (a) (i) implies (ii), (b) (ii) implies (iii), and (c) (iii) implies (i).

(a) Let \vec{p} be an arbitrary element of Ω and suppose that \vec{p} admits a classical probability space representation. We demonstrate that \vec{p} is extractable from a joint distribution. Since \vec{p} admits a classical probability space representation, there is a classical probability space (X, \mathcal{A}, μ) and $E_1, E_2, \dots, E_n \in \mathcal{A}$ such that (6) holds. As (X, \mathcal{A}, μ) is a classical probability space, events $E_1, E_2, \dots, E_n \in \mathcal{A}$ in it has a joint distribution which marginalizes to (6). More precisely, with notation

$$\begin{aligned} E_i^+ &:= E_i \\ E_i^- &:= X \setminus E_i \\ & \quad i = 1, 2, \dots, n \end{aligned}$$

let

$$p_{\alpha_1 \dots \alpha_n} := \mu(E_1^{\alpha_1} \cap E_2^{\alpha_2} \dots \cap E_n^{\alpha_n}) \quad \alpha_1, \alpha_2, \dots, \alpha_n \in \{+, -\}$$

Since μ is a Kolmogorovian probability measure, (7) will obviously hold. Moreover, due to (6) and the law of total probability,

$$\begin{aligned} p_i &= \mu(E_i) = \sum_{\substack{\alpha_1, \dots, \alpha_n \in \{+, -\} \\ \alpha_i = +}} \mu(E_1^{\alpha_1} \cap E_2^{\alpha_2} \dots \cap E_n^{\alpha_n}) = \\ &= \sum_{\substack{\alpha_1, \dots, \alpha_n \in \{+, -\} \\ \alpha_i = +}} p_{\alpha_1 \dots \alpha_n} \quad i = 1, 2, \dots, n \end{aligned}$$

$$\begin{aligned} p_{ij} &= \mu(E_i \cap E_j) = \sum_{\substack{\alpha_1, \dots, \alpha_n \in \{+, -\} \\ \alpha_i, \alpha_j = +}} \mu(E_1^{\alpha_1} \cap E_2^{\alpha_2} \dots \cap E_n^{\alpha_n}) = \\ &= \sum_{\substack{\alpha_1, \dots, \alpha_n \in \{+, -\} \\ \alpha_i, \alpha_j = +}} p_{\alpha_1 \dots \alpha_n} \quad (i, j) \in S \end{aligned}$$

that is, (8) is also satisfied. Thus, \vec{p} is extractable from a joint distribution.

(b) Suppose the for all $\vec{p} \in \Omega$, \vec{p} is extractable from a joint distribution. We will demonstrate that Ω is projectable from the 2^n -vertex probability simplex $\Delta_{2^n} \subset \mathbb{R}^{2^n}$. Instead of $r \in \{1, 2, \dots, 2^n\}$, it will be convenient to label the components of vectors in \mathbb{R}^{2^n} by $(\alpha_1, \dots, \alpha_n) \in \{+, -\}^n$, where indices $(\alpha_1, \dots, \alpha_n)$ are ordered lexicographically, with + preceding -. With this notation, introduce the following sets of indices: $R_i := \{(\alpha_1, \dots, \alpha_n) \in \{+, -\}^n \mid \alpha_i = +\}$, $i = 1, 2, \dots, n$; and let map $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}^{n+|S|}$ be defined by

$$\begin{aligned} (\varphi(\mathbf{x}))_i &= \sum_{(\alpha_1, \dots, \alpha_n) \in R_i} x_{\alpha_1 \dots \alpha_n} \quad i = 1, 2, \dots, n \\ (\varphi(\mathbf{x}))_{ij} &= \sum_{(\alpha_1, \dots, \alpha_n) \in R_i \cap R_j} x_{\alpha_1 \dots \alpha_n} \quad (i, j) \in S \end{aligned} \quad (29)$$

for all $\mathbf{x} \in \mathbb{R}^{2^n}$. φ is obviously linear. Now, since for all $\vec{p} \in \Omega$, \vec{p} is extractable from a joint distribution, for all $\vec{p} \in \Omega$ there exist 2^n numbers

$p_{\alpha_1 \dots \alpha_n}, \alpha_1, \dots, \alpha_n \in \{+, -\}$ such that (7) and (8) hold. For each $\vec{p} \in \Omega$, consider the vector $\pi \in \mathbb{R}^{2^n}$ for which $\pi_{\alpha_1 \dots \alpha_n} = p_{\alpha_1 \dots \alpha_n}$. Due to (7), $\pi \in \Delta_{2^n}$ for all $\vec{p} \in \Omega$. Moreover, due to (8) and the definition of φ , (29), $\varphi(\pi) = \vec{p}$. Hence, $\Omega \subseteq \varphi(\Delta_{2^n})$. Furthermore, suppose that for a $\pi' \in \Delta_{2^n}$, $\varphi(\pi') = \vec{p}$ for some $\vec{p} \in \Omega$. Then, again due to (29),

$$\begin{aligned} p_i &= \sum_{(\alpha_1, \dots, \alpha_n) \in R_i} \pi'_{\alpha_1 \dots \alpha_n} & i = 1, 2, \dots, n \\ p_{ij} &= \sum_{(\alpha_1, \dots, \alpha_n) \in R_i \cap R_j} \pi'_{\alpha_1 \dots \alpha_n} & (i, j) \in S \end{aligned}$$

which is nothing but (10) in terms of indices $(\alpha_1, \dots, \alpha_n)$ instead of r . Thus, Ω is projectable from the probability simplex $\Delta_{2^n} \subset \mathbb{R}^{2^n}$.

(c) Suppose that Ω is projectable from a probability simplex $\Delta_d \subset \mathbb{R}^d$. Consider the set of vertices of Δ_d , $X = \{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_d\}$, and its subset algebra \mathcal{A} . We show that each $\vec{p} \in \Omega$ admits a classical probability space representation over measurable space (X, \mathcal{A}) . Since Ω is projectable from the probability simplex Δ_d , there is a linear map $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}^{n+|S|}$, and sets of indices $R_i \subseteq \{1, 2, \dots, d\}, i = 1, 2, \dots, n$ such that (9) and (10) hold. Let $E_i := \{\mathbf{e}_r | r \in R_i\} \in \mathcal{A}, i = 1, 2, \dots, n$. Consider an arbitrary $\vec{p} \in \Omega$. Since $\Omega \subseteq \varphi(\Delta_d)$, there is a $\pi \in \Delta_d$ such that $\varphi(\pi) = \vec{p}$. Now, define μ in the following way:

$$\mu : \mathcal{A} \rightarrow [0, 1], \mu(E) := \sum_{r: \mathbf{e}_r \in E} \pi_r$$

μ is obviously a probability measure on \mathcal{A} , and

$$\begin{aligned} \mu(E_i) &= \sum_{r: \mathbf{e}_r \in E_i} \pi_r = \sum_{r \in R_i} \pi_r = p_i & i = 1, 2, \dots, n \\ \mu(E_i \cap E_j) &= \sum_{r: \mathbf{e}_r \in E_i \cap E_j} \pi_r = \sum_{r \in R_i \cap R_j} \pi_r = p_{ij} & (i, j) \in S \end{aligned}$$

that is, (6) is satisfied. Thus, (X, \mathcal{A}, μ) with $E_1, E_2, \dots, E_n \in \mathcal{A}$ provides a classical probability space representation of \vec{p} . \square

Proof of Proposition 5. As Pitowsky (1989) showed, the fact that a correlation vector $\vec{p} \in \mathbb{R}^{n+|S|}$ admits a classical probability space representation is equivalent with the fact that \vec{p} lies within the so-called *classical correlation polytope*

$$c(n, S) = \left\{ \sum_{\varepsilon \in \{0,1\}^n} \lambda_\varepsilon \vec{u}^\varepsilon \mid \lambda_\varepsilon \geq 0, \sum_{\varepsilon \in \{0,1\}^n} \lambda_\varepsilon = 1 \right\} \subset \mathbb{R}^{n+|S|}$$

the convex hull of the *classical vertex vectors* $\vec{u}^\varepsilon \in \mathbb{R}^{n+|S|}$ ($\varepsilon \in \{0, 1\}^n$) whose components are defined as

$$\begin{aligned} u_i^\varepsilon &= \varepsilon_i & i = 1, \dots, n \\ u_{ij}^\varepsilon &= \varepsilon_i \varepsilon_j & (i, j) \in S \end{aligned}$$

We will show that $\vec{p}_{\text{EPR}} \in c(4, S_{\text{EPR}})$.

First, note that any correlation vector $\vec{p} \in \mathbb{R}^{n+|S|}$ for which

$$p_{ij} = p_i p_j \quad (i, j) \in S \quad (30)$$

is an element of $c(n, S)$. Indeed, it is easy to verify that such a vector can be expressed as a convex combination of classical vertex vectors \vec{u}^ε with coefficients

$$\lambda_\varepsilon = \prod_{i=1}^n p_i^*, \quad \text{where } p_i^* = \begin{cases} p_i, & \text{if } \varepsilon_i = 1 \\ 1 - p_i, & \text{if } \varepsilon_i = 0 \end{cases}$$

Now suppose that there is a classical probability space (X, \mathcal{A}, p) where $a_1, a_2, b_3, b_4, A_1, A_2, B_3, B_4$ and $C_k (k \in K)$ are represented as events $(a_1, a_2, b_3, b_4, A_1, A_2, B_3, B_4, C_k \in \mathcal{A}, k \in K)$, $\{C_k\}_{k \in K}$ is a partition of \mathcal{A} , and probability measure p satisfies (12)–(13). Since $C_k (k \in K)$ partition \mathcal{A} , one can apply the law of total probability with respect to conditional measures $p(\cdot | a_i \cap a_j)$ to receive

$$\begin{aligned} p(A_i | a_i \cap b_j) &= \sum_{k \in K} p(A_i | a_i \cap b_j \cap C_k) p(C_k | a_i \cap b_j) \\ p(B_j | a_i \cap b_j) &= \sum_{k \in K} p(B_j | a_i \cap b_j \cap C_k) p(C_k | a_i \cap b_j) \\ p(A_i \cap B_j | a_i \cap b_j) &= \sum_{k \in K} p(A_i \cap B_j | a_i \cap b_j \cap C_k) p(C_k | a_i \cap b_j) \end{aligned} \quad (31)$$

$(i, j) \in S_{\text{EPR}}$

Taking into account no-signaling (4) and assumptions (12)–(13), equations (31) simplify to

$$\begin{aligned} p(A_i | a_i) &= \sum_{k \in K} p(A_i | a_i \cap C_k) p(C_k) \\ p(B_j | b_j) &= \sum_{k \in K} p(B_j | b_j \cap C_k) p(C_k) \\ p(A_i \cap B_j | a_i \cap b_j) &= \sum_{k \in K} p(A_i | a_i \cap C_k) p(B_j | b_j \cap C_k) p(C_k) \end{aligned} \quad (32)$$

$(i, j) \in S_{\text{EPR}}$

Now introduce correlation vectors $\vec{p}^k \in \mathbb{R}^{4+4} (k \in K)$ with components

$$\begin{aligned} p_i^k &= p(A_i | a_i \cap C_k) & i = 1, 2 \\ p_j^k &= p(B_j | b_j \cap C_k) & j = 3, 4 \\ p_{ij}^k &= p(A_i | a_i \cap C_k) p(B_j | b_j \cap C_k) & (i, j) \in S_{\text{EPR}} \end{aligned} \quad (33)$$

Observe that correlation vectors \vec{p}^k are of the form (30), and thus they all lie in $c(4, S_{\text{EPR}})$. Furthermore, notice that what (32) says is

$$\vec{p}_{\text{EPR}} = \sum_{k \in K} p(C_k) \vec{p}^k$$

where $p(C_k) \geq 0$, $\sum_{k \in K} p(C_k) = 1$ since $\{C_k\}_{k \in K}$ is a partition. This means that \vec{p}_{EPR} is a convex combination of elements of $c(4, S_{\text{EPR}})$, which entails that $\vec{p}_{\text{EPR}} \in c(4, S_{\text{EPR}})$ as $c(4, S_{\text{EPR}})$ is a convex set. \square

Acknowledgment

Márton Gömöri acknowledges the support of the following Hungarian research grants: Hungarian Eötvös Scholarship and grants no. K-115593 and K-134275

of the National Research, Development and Innovation Office. Carl Hofer acknowledges the support of the following Catalan and Spanish research grants: 2021-SGR-00276, FFI2016-76799-P, PID2020-115114GB-I00, and CEX2021-001169-M funded by MCIN/AEI/10.13039/501100011033.

We wish to thank Győző Egri, Balázs Gyenis and Gábor Hofer-Szabó for valuable feedback.

References

- Bana, G. and T. Durt (1997): Proof of Kolmogorovian Censorship, *Foundations of Physics* 27, 1355–1373.
- Barrett J. (2007): Information processing in generalized probabilistic theories, *Phys. Rev. A* 75, 032304.
- Bell, J.S. (2004): *Speakable and Unsayable in Quantum Mechanics*, Cambridge University Press.
- Butterfield, J. (1992): Bell’s theorem: What it takes, *British Journal for the Philosophy of Science* 43, 41–83.
- Daumer, M. et al. (1996): Naive realism about operators, *Erkenntnis* 45, 379–397.
- Feintzeig, B.H., and S.C. Fletcher (2017): On Noncontextual, Non-Kolmogorovian Hidden Variable Theories, *Foundations of Physics* 47, 294–315.
- Fine, A. (1982): Hidden variables, joint probability, and the Bell inequalities, *Phys. Rev. Lett.* 48, 291–295.
- Fine, A. (1989): Do Correlations Need to be Explained?, in: J. Cushing and E. McMullin (eds.) *Philosophical Consequences of Quantum Theory* (pp. 175–194), University of Notre Dame Press.
- Gömöri, M., and T. Placek (2017): Small Probability Space Formulation of Bell’s Theorem, in: G. Hofer-Szabó and L. Wroński (eds.) *Making it Formally Explicit* (pp. 109–127), European Studies in Philosophy of Science, vol 6. Springer, Cham.
- Gömöri, M., and G. Hofer-Szabó (2021): On the meaning of EPR’s Reality Criterion, *Synthese* 199, 13441–13469.
- Griffiths, R.B. (2020): Nonlocality claims are inconsistent with Hilbert-space quantum mechanics, *Phys. Rev. A* 101, 022117.
- Hofer-Szabó, G., M. Rédei and L.E. Szabó (2013): *The Principle of the Common Cause*, Cambridge University Press.
- Hofer-Szabó, G. (2020): On the Three Types of Bell’s Inequalities, in: M. Hemmo and O. Shenker (eds.) *Quantum, Probability, Logic* (pp. 353–374), Jerusalem Studies in Philosophy and History of Science, Springer, Cham.
- Hofer-Szabó, G. (2021): Causal contextuality and contextuality-by-default are different concepts, *Journal of Mathematical Psychology* 104, 102590.
- Hofer-Szabó, G. (2022): Two concepts of noncontextuality in quantum mechanics, *Studies in History and Philosophy of Science* 93, 21–29.

- Hossenfelder, S. and T. Palmer (2020): Rethinking Superdeterminism, *Front. Phys.* 8, 139.
- Lewis P.J. (2019): Bell's Theorem, Realism, and Locality, in: Cordero A. (eds) *Philosophers Look at Quantum Mechanics* (pp. 33–43), Synthese Library (Studies in Epistemology, Logic, Methodology, and Philosophy of Science), vol. 406. Springer, Cham.
- Maudlin, T. (2014a): What Bell did, *J. Phys. A: Math. Theor.* 47, 424010.
- Maudlin, T. (2014b): Reply to Comment on 'What Bell did', *J. Phys. A: Math. Theor.* 47, 424012.
- Norsen, T. (2007): Against 'Realism', *Foundations of Physics* 37, 311–340.
- Pitowsky, I. (1989): *Quantum Probability – Quantum Logic*. Springer.
- Rédei, M. (2010): Kolmogorovian Censorship Hypothesis for general quantum probability theories, *Manuscrito – Revista Internacional de Filosofia* 33, 365–380.
- Santos, E. (1986): The Bell inequalities as tests of classical logic, *Phys. Lett. A* 115, 363–365.
- Shimony, A. (1984): Contextual Hidden Variables Theories and Bell's Inequalities, *The British Journal for the Philosophy of Science* 35, 25–45.
- Szabó, L.E. (1995): Is quantum mechanics compatible with a deterministic universe? Two interpretations of quantum probabilities. *Foundations of Physics Letters* 8, 421–440.
- Szabó, L.E. (2001): Critical reflections on quantum probability theory, in: M. Rédei and M. Stoeltzner (eds.) *John von Neumann and the Foundations of Quantum Physics* (pp. 201–219), Kluwer Academic Publishers, Dordrecht.
- Weingartner, P. (2009): Matrix-Based Logic for Application in Physics, *The Review of Symbolic Logic* 2(1), 132–163.
- Werner, R. (2014a): Comment on 'What Bell did,' *J. Phys. A: Math. Theor.* 47, 424011.
- Werner, R. (2014b): What Maudlin replied to, <<https://arxiv.org/pdf/1411.2120.pdf>>