

Counterfactual Reasoning, Realism and Quantum Mechanics: Much Ado About Nothing?

Federico Laudisa¹ 

Received: 28 August 2017 / Accepted: 14 March 2018
© Springer Science+Business Media B.V., part of Springer Nature 2018

Abstract I show why old and new claims on the role of counterfactual reasoning for the EPR argument and Bell’s theorem are unjustified: once the logical relation between locality and counterfactual reasoning is clarified, the use of the latter does no harm and the nonlocality result can well follow from the EPR premises. To show why, after emphasizing the role of incompleteness arguments that Einstein developed before the EPR paper, I critically review more recent claims that equate the use of counterfactual reasoning with the assumption of a strong form of realism and argue that such claims are untenable.

1 Introduction

The debates on the EPR argument and the status of nonlocality, which still prove to be among the most hotly discussed in the foundations of physics today, continue to reserve a special place for the special form of reasoning that has been called ‘counterfactual’. At first sight, this appears hardly surprising given the role of counterfactual reasoning for scientific thought in general. If for instance we focus on the role that natural laws play in scientific theories, we realize that—whatever view of laws we might have—an essential task of laws themselves is to set boundaries for what is ‘possible’ and ‘impossible’ in nomological terms, namely for what might or might not happen compatibly with the laws and, consequently, for what *might have happened* (even if, in fact, it did not). In more detail, however, an option frequently adopted since the original formulation of the EPR argument has been to claim that

✉ Federico Laudisa
federico.laudisa@unimib.it

¹ Department of Human Sciences, University of Milan-Bicocca, Piazza dell’Ateneo Nuovo 1, 20126 Milan, Italy

counterfactual reasoning was an essential ingredient of Bell's theorem and to reject as a consequence the generality of Bell's theorem, due to the role that such form of reasoning is supposed to play when the very premises of the EPR argument itself are at work. Since the use of counterfactual reasoning was always taken in itself as an assumption at the outset of a form of realism that quantum mechanics cannot *in principle* support, then—so the argument goes—we can legitimately conclude that Bell's theorem need not imply that any theory preserving the predictions of quantum mechanics must be nonlocal. The idea that Bell's theorem, far from showing the microworld to be nonlocal, would be nothing but a further proof that “unperformed experiments have no results” (Peres 1978) is widespread, to such an extent that even many of those who take seriously nonlocality as a property of natural processes at the microlevel take for granted that in order to prove Bell's theorem we need to assume that unperformed experiments *do* have results. In a fairly recent paper, for instance, Stefan Wolf proposed an algorithmic-complexity view of nonlocality exactly with the motivation that it need not rely on the so-called counterfactual definiteness, taken by the author to be a standard assumption of Bell's theorem in any of its variants. According to Wolf “nonlocal correlations are usually understood through the outcomes of alternative measurements (on two or more parts of a system) that cannot altogether actually be carried about in an experiment. Indeed, a joint input–output—e.g., measurement–setting–outcome—behavior is nonlocal if and only if the outputs for *all* possible inputs cannot coexist consistently. It has been argued that this counterfactual view is how Bell's inequalities and their violations are to be seen. I propose an alternative perspective which refrains from setting into relation the results of mutually exclusive measurements, but that is based solely on data actually available” (Wolf 2015, 052102-1, emphasis in the original) and in the sequel the author further clarifies the point: “Nonlocal correlations are a fascinating feature of quantum theory. [...] More specifically, the difficulty manifests itself when alternatives are taken into account: the argument leading up to a Bell inequality relates outcomes of different measurements, only *one* of which can actually be realized. [...] Nonlocality is the impossibility of the outputs to *alternative inputs* to consistently coexist. Such reasoning assumes and concludes that certain pieces of *classical information exist or do not exist.*” (Wolf 2015, 052102-1, emphasis in the original).¹

In what follows, my aim is to show why these claims on the role of counterfactual reasoning for the EPR argument and Bell's theorem are unjustified: once the logical relation between locality and counterfactual reasoning is clarified, the use of the latter does no harm and the nonlocality result can well follow from the EPR premises. Section 2 is supposed, in a slightly historical vein, to emphasize some Einsteinian anticipations on the logical relation between locality and what we might call a “pre-determination” thesis: the subsequent discussion concerning the role of counterfactual reasoning in EPR-like contexts turns out to be largely consistent with the picture that these anticipations in fact suggest. The reference to these Einsteinian

¹ A side remark. Given a thesis X, it is quite peculiar to say that a certain argument *both* ‘assumes’ and ‘concludes’ X: the logical status of X within the argument—hence, the argument itself—changes dramatically if X is an assumption or the outcome of a derivation! As we will see later, this is exactly the heart of the matter with counterfactual definiteness in EPR-Bell kind of arguments.

reflections will allow us to briefly address the issue of the Einstein–Bell relation with respect to the link between locality and pre-determination. In Sect. 3, the central section, I critically discuss a fairly recent attempt to equate the use of counterfactual reasoning with the assumption of a strong form of realism: here I explore the reasons why this attempt—along with the many that preceded it—fails to be convincing. The final section draws some general conclusions.

2 Locality and Pre-existence: Einstein Versus Bell?

Although the EPR paper occupied the center of stage for many years, it is now well known that the formulation of the argument therein is only one of several formulations that Einstein provided for his claim concerning the incompleteness of quantum mechanics. In general terms, from the thought experiment proposed at the Solvay Conference in 1927 onward, his focus on the peculiar features of the state representation in quantum mechanics highly contributed *in retrospect* to realize how peculiar is this representation in many respects.² But also for our purposes the alternative formulations of an incompleteness argument turn out to be especially useful, since they show that it is locality rather than any pre-existence claim that is bound to play the fundamental role in highlighting a disconcerting feature of quantum mechanics.

As is again well known, the main sources of misunderstanding surrounding the EPR argument in its original form are two: the first is the controversial reality condition and the second is the choice of two conjugate observables like position and momentum as the relevant observables. The latter proved especially misleading as far as the Bohr response to EPR was concerned: Bohr in fact questioned the argument on the basis of a somewhat obscure kind of disturbance theory of measurement,³ while the ingenuity of the thought experiment was exactly that of avoiding altogether any assumption that might imply any disturbance of a quantum sort.⁴ In any case, it was 8 years before EPR that Einstein first proposed a thought experiment in which the dilemma at the heart of the EPR argument—namely completeness versus locality—is spelled out with no reference to unclear ‘elements of reality’, ‘counterfactual definiteness’ and so on.

² This is one of the episodes in the history of contemporary physics where the received view of Einstein as a reactionary scientist, unable to follow the more promising lines of development of the new physics in the second half of his life, proves so dramatically inadequate. As an instance of this inadequacy, let me recall what Werner Heisenberg said in his introduction to *The Born–Einstein Letters*, commenting on the Einstein’s attitude toward quantum mechanics: “Most scientists are willing to accept new empirical data and to recognize new results, provided they fit into their philosophical framework. But in the course of scientific progress it can happen that a new range of empirical data can be completely understood only when the enormous effort is made to enlarge this framework and to change the very structure of the thought processes. In the case of quantum mechanics, Einstein was apparently no longer willing to take this step, or perhaps no longer able to do so” (Born 2005, xxxvii).

³ This theory is downgraded by Arthur Fine to the status of a ‘semantic’ theory of disturbance (Fine 1986, 35).

⁴ See the deep analysis of this point in the chapter 3 of Fine (1986); see also Maudlin (2014, 424010–12, 13).

The thought experiment contemplates a diaphragm S, with an opening O, and a hemispherical detection screen P (Fig. 1): incident particles indicated by the arrow typically are expected to diffract and produce a wave propagating toward P. If a single electron is sent toward the opening, at the end of the path it will be detected at some specific point on the screen, with the probability prescribed by the wave function ψ associated with the pre-detection state of the electron. According to Einstein, we might account for the situation in two possible, mutually exclusive ways:

Conception I – The de Broglie–Schrödinger waves do not correspond to a single electron, but to a cloud of electrons extended in space. The theory gives no information about individual processes, but only about the ensembles of an infinity of elementary processes.

Conception II – The theory claims to be a complete theory of individual processes.

[...] According to the first, purely statistical point of view $|\psi|^2$ expresses the probability that there exists at the point considered a *particular* particle of the cloud, for example at a given point on the screen. According to the second, $|\psi|^2$ expresses the probability that at a given instant *the same* particle is present at a given point (for example on the screen). Here, the theory refers to an individual process and claims to describe everything that is governed by laws.

After the presentation of the two options, between which Einstein clearly favours the first, he focuses on an especially awkward consequence of the second:

If $|\psi|^2$ were simply regarded as the probability that at a certain point a given particle is found at a given time, it could happen that *the same* elementary process produces an action *in two or several* places on the screen. But the interpretation according to which $|\psi|^2$ expresses the probability that *this* particle is found at a given point, assumes an entirely peculiar mechanism of action at a distance, which prevents the wave continuously distributed in space from producing an action *in two* places on the screen. [...] If one works solely with the Schrödinger waves, interpretation II of $|\psi|^2$ implies to my mind a contradiction with the postulate of relativity. (Bacciagaluppi and Valentini 2009, 441).

The negation of “an entirely peculiar mechanism of action at a distance”, which in the Einstein view follows straight from relativity, is exactly what we call “locality”: *if we assume it, then we decide to accept the good old intuition according to which there is a matter of fact concerning the pre-detection properties of particle. Since the ψ cannot in principle account for such properties, it follows that the quantum–mechanical description of the pre-detection state of the particle is incomplete.*⁵

⁵ A further, simplified version of the Einstein 1927 thought experiment is due to Hardy (1995). A slightly different formulation, inspired by the argument that Einstein proposed in 1927, is known as the *Einstein boxes* experiment and supposed to emphasize even more dramatically the dilemma between locality and completeness. This kind of thought experiment is reformulated by Einstein in the letter to Schrödinger of June 19th, 1935, in which he complains about the published version of the EPR argument: see the discussion of it in Fine (1986, 35–38), and Norsen (2005) for a thorough and more recent analysis.

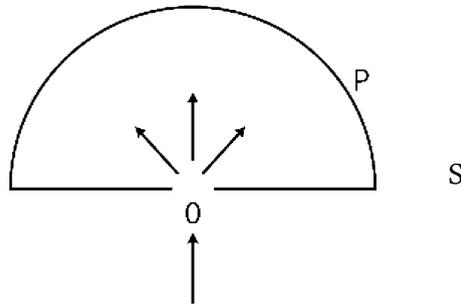


Fig. 1 The Einstein 1927 thought experiments

A further, standard source of misunderstanding—which resonates with the misunderstandings we have hinted at above—concerns the complex relationship between the presuppositions of the Einsteinian view of quantum mechanics and the Bell results. In what sense is this relationship complex and why can a more thorough investigation of it work effectively in both directions? For at least two, essential motivations. On the one hand, the Bell results are among the post-Einsteinian scientific achievements that most call into question some crucial features of the Einstein own image of the physical world. On the other hand, Bell’s theorem itself turned out to be a sort of lens from which to look at some of the original Einstein’s foundational views, a lens that more often than not produced serious distortions: not by chance, these distortions are often also distortions of the significance of the Bell results themselves.

If we focus in particular on the role that an implicit assumption of classically inspired form of realism—according to which physical systems are at any time endowed with a given set of measurement-independent properties—is supposed to play since the first reservations of Einstein about quantum mechanics, the current representation of the Einstein–Bell connecting line is still not free of ambiguities. Some instances of this representation prove to be simply wrong:

While a very tiny loophole in principle remains for local realism, it is a very safe position to assume that QM has definitely shown to be the right theory. Thus, a very deep question, namely whether or not events observed in the quantum world can be described by an underlying deterministic theory, has been answered by experiment, thanks to the momentous achievements of John Bell.” (Bertlmann and Zeilinger 2002, viii).

The existence and mathematical consistency of a pilot-wave-theoretical formulation of quantum mechanics proves this statement to be just false: any de Broglie–Bohm type of theory, being both deterministic and non-local by construction, does not run afoul of Bell’s theorem, which rules out only *local* completions—be they deterministic or indeterministic. Other statements, although closer to truth, are somewhat unclear. In his introduction to the second edition (2004) of the celebrated Bell collection *Speakable and Unspeakable in Quantum Mechanics*, Alain Aspect—a leading figure in the experimental work on the foundations of quantum mechanics and the Bell inequalities—for instance writes:

John Bell demonstrated that there is no way to understand entanglement in the framework of the usual ideas of a physical reality localized in space-time and obeying to causality. This result was the opposite to the expectations of Einstein, who had first pointed out, with his collaborators Podolsky and Rosen, the strong correlations between entangled particles, and analyzed these correlations in the framework of ideas of a local physical reality. The most remarkable feature of Bell's work was undoubtedly the possibility it offered to determine *experimentally* whether or not Einstein's ideas could be kept (Bell 2004, p. xvii, emphasis in the original).

The conjunction of 'physical reality localized in space-time' and 'causality' on one side and the expression 'local physical reality' on the other lead us to think that on the background the following presupposition is lurking: namely, that Bell's theorem rules out completions of quantum mechanics that are both local *and* 'realistic' and that the latter is the Einstein's heritage that the theorem itself forces us to abandon. The point is crucial, since according to some the assumption of realism—being quantum-mechanically untenable—undermines Bell's theorem as a *non-locality* theorem.

As far as Einstein is concerned, I think it can be safely said that Einstein would have longed for a formulation of quantum mechanics able to include what might be called a *Pre-Existence* condition: a standard physical system has at all times a whole bunch of pre-existing properties—encoded in its state—no matter whether there is a measurement interaction or not. As we attempted to show in the previous section, however, personal Einstein's preference for such a hypothetical formulation does *not* imply per se that Einstein's incompleteness arguments—the pre-EPR ones, the EPR itself and the post-EPR ones—*do need* to require independently such pre-existence condition. Pre-Existence and locality interact in an interesting way in the following statement, taken from the *Einstein Autobiographical Notes*:

There is to be a system which at the time t of our observation consists of two partial systems S_1 , and S_2 , which at this time are spatially separated [...] The total system is to be completely described through a known ψ -function ψ_{12} in the sense of quantum mechanics. All quantum theoreticians now agree upon the following: If I make a complete measurement of S_1 , I get from the results of the measurements and from ψ_{12} an entirely definite ψ -function ψ_2 of the system S_2 . The character of ψ_2 then depends upon what kind of measurement I undertake on S_1 . Now it appears to me that one may speak of the real factual situation of the partial system S_2 . Of this real factual situation, we know to begin with, before the measurement of S_1 , even less than we know of a system described by the ψ -function. But on one supposition we should, in my opinion, absolutely hold fast: the real, factual situation of the system S_2 is independent of what is done with the system S_1 which is spatially separated from the former. (Einstein 1949, 85).

What Einstein calls 'the real, factual situation' is clearly supposed to be a pre-existing feature of the system, but the assumption that 'we should hold fast'—i.e. the independence of S_2 from what is done with the system S_1 and viceversa—is

what *logically* guarantees that the system can be conceived as being in a well-defined real, factual situation at all times. Although the foundational views of Einstein show some ambivalence on whether pre-existence and locality are both fundamental assumptions or not, all his incompleteness arguments can be formulated with a clear logical relation between them, according to which the latter is a pre-requisite for the former.⁶

It is ironic to note that many misunderstandings of the Bell work derive *exactly* from attributing to Bell himself the endorsement of some sort of the pre-existence claim. Here the misunderstanding could not more complete. It was already in his paper «On the problem of hidden variables in QM» (published in 1966 but written in 1963, *before* the paper that derives for the first time the Bell inequality) that Bell remarked how unreasonable it was to require pre-existence. With reference to the ‘no-hidden-variable’ proofs provided in the preceding years by von Neumann, Gleason and Jauch-Piron, Bell wrote:

[...] these demonstrations require from the hypothetical dispersion free states not only that appropriate ensembles thereof should have all measurable properties of quantum mechanical states, but certain other properties as well. These additional demands appear reasonable when results of measurements are loosely identified with properties of isolated systems. They are seen to be quite unreasonable when one remembers with Bohr ‘the impossibility of any sharp distinction between the behaviour of atomic objects and the interaction with measuring instruments which serve to define the conditions under which the phenomena appear’. (Bell 2004, 1–2).

It is (the first instance of) what Abner Shimony called a «judo-like manoeuvre» (Shimony 1984, 121): to use the Bohr *dictum* itself in order to show how limited the results by von Neumann and the others turned out to be, exactly because they did assume something like the pre-existence claim.

To sum up: the Einstein pre-1935 arguments do not directly address the two-particle situation that became familiar after the EPR argument. They are extremely useful, however, to point out the logical relation between locality and the sort of “pre-determinateness” thesis we referred to above. As a matter of fact, these arguments show that the idea of pre-determinateness—an idea Einstein was fond of anyway—turns out to be justified *on the basis of locality*: if the latter is assumed, pre-determinateness becomes rather natural. This logical relation, in which locality is prior to pre-determinateness and allows the latter to make sense, preserves its relevance when the debate moves forward in the EPR direction.

3 Counterfactual Realism: It Ain’t Necessarily So...

In spite of the arguments above, the controversy on an alleged realism assumption at the source of the Einstein–Bell arguments is still alive. I will now concentrate the discussion on a more recent proposal about the interpretation of Bell’s theorem, due

⁶ On the nature of Einstein’s realism, see e.g. Howard (1993) and Home and Whitaker (2007, ch. 8).

to Marek Zukovski and Caslav Brukner (ZB, from now on). Focusing on their formulation is useful, since they touch two recurrent points in the debate on the role of counterfactual reasoning and they do it systematically, in a way that is especially suitable to a clean discussion. ZB concentrate their critical analysis on what they take to be two distinct views, simply called view (A) and view (B):

- (A) The derivation of a Bell inequality relies only on the premise of locality, and nothing more.
- (B) The derivation of a Bell inequality does rely on the additional premise of determinism, or any of the notions listed in the abstract. However, these notions can be derived from the premises of locality, the freedom of an experimenter to choose the setting of his/her local apparatus and quantum predictions and nothing more (Zukovski and Brukner 2014, 424009-2).

The alleged refutations by ZB of both views (A) and (B) share the claim that, in both cases, what Bell called *local causality* is in fact “a *compound condition*” (Zukovski and Brukner 2014, 424009-2, emphasis in the original), namely the conjunction of locality *plus* an additional condition:

as to (A), the additional condition would be “the assumption of the existence of a (positive and normalized) joint probability distribution for the values of all possible measurements that could be performed on an individual system, no matter whether any measurement – and which measurement – is actually performed” (as we will see later, the reference to actual-or-possible measurement already hints at the role of counterfactual reasoning);

as to (B), the additional condition would be *counterfactual definiteness*, i.e. a condition that “allows one to assume the definiteness of the results of measurements, which were actually not performed on a given individual system. They are treated as unknown, but in principle *defined* values.” (Zukovski and Brukner 2014, 424009-2, emphasis in the original).

After all, both the assumptions that for ZB are implicit in view (A) and (B) respectively and that, conjoined with locality, make up the crucial, compound condition of Bell’s theorem seem to point at some sort of Classicality that the theorem would be about: therefore, the argument goes, the joint outcome of both refutations is the conclusion that Bell’s theorem, far from demonstrating nonlocality for the quantum world, would be just a sophisticated way of stating that the quantum world is not ‘classical’ (or, more modestly, that quantum mechanics is not a ‘classical’ theory). Let me review, then, the arguments about (A) and (B) in turn in the next two subsections.

3.1 View (A): Locality and Joint Probability Distributions

In the standard EPR-Bell context we suppose two parties, Alice and Bob, that we may associate with two different categories of events: the choice of a measurement setting and the production of an outcome. If we denote by x, y two possible settings and by A, B two possible outcomes respectively, we are generally interested to

correlations described by distributions $p(A, B | x, y)$. Therefore a statistical framework is introduced, in which expressions like $p(A, B | x, y)$ represent distributions of more general states λ , such that

$$p(A, B | x, y) = \int d\lambda \rho(\lambda) p(A, B | x, y, \lambda) \quad (1)$$

with $\rho(\lambda)$ being a suitable distribution over the ‘general’ state space Λ . Now, from the formal theory of probability we know that $p(A, B | x, y) = p(A | B, x, y) p(B | x, y)$, so that

$$p(A, B | x, y) = \int d\lambda \rho(\lambda) p(A | B, x, y, \lambda) p(B | x, y, \lambda); \quad (2)$$

moreover, locality requires that the outcomes A and B depend only on respective near-by causal factors, therefore (2) becomes

$$p(A, B | x, y) = \int d\lambda \rho(\lambda) p(A | x, \lambda) p(B | y, \lambda). \quad (3)$$

According to ZB, (3) allows to introduce a joint probability distribution along the following lines. If A_x [B_y] denote the outcome of an A -[B]-measurement when the chosen setting is x [y], and 1, 2 denote two possible choices for each setting, the desired joint distribution can be written as

$$p(A_{x=1}, A_{x=2}, B_{y=1}, B_{y=2} | \lambda) = p(A_{x=1} | \lambda) p(A_{x=2} | \lambda) p(B_{y=1} | \lambda) p(B_{y=2} | \lambda) \quad (4)$$

If this is the formal account, the ZB foundational criticisms are two. The first point is that the very introduction of the general state structure Λ amounts

to the introduction of additional hidden parameters, λ , which are *not* present in quantum mechanics. The λ [...] can pop up under many guises such as, e.g., ‘the physical state of the systems as described by any possible future theory’, ‘local beables’, ‘the real state of affairs’, ‘complete description of the state’, etc. Since λ do not appear in quantum mechanics, they are (good old) *hidden variables*. Anything on which one conditions probabilities which gives a different structure to formulas for probabilities than quantum mechanical formalism is a hidden variable per se. (Zukovski and Brukner 2014, 424009-3, emphasis in the original)

This remark is meant to suggest that the mere use of a Λ -statistical theory is nearly equivalent (or leads necessarily) to the independent assumption of pre-determined values for any relevant observable. The second point reinforces the first: the introduction of a joint probability “for all possible outcomes under all possible pairs of settings” presupposes that “[T]hese outcomes include, for a single run of a Bell experiment, the actual measured outcomes and on an equal footing those which *could have been potentially measured*.” (Zukovski and Brukner 2014, 424009-4, my emphasis). ZB argue, then, that these two points jointly undermine view (A): in their argument both the above points allegedly show that Bell ‘local causality’ brings with itself a hidden assumption of ‘pre-determination’, a circumstance that

makes ‘local causality’ a compound condition and that, as a consequence, is bound to change radically the interpretation of Bell’s theorem as a nonlocality result.

As to the first point, I wish to stress that equating the mere use of a Λ -statistical theory with an assumption of definiteness of values for all observables—even those for which quantum mechanics prevents such definiteness—looks like a *petitio principii*. On the contrary, the introduction of a general framework like a Λ -statistical theory is relatively innocent and perfectly reasonable by a foundational point of view: provided that we already know that quantum mechanics is nonlocal (via entanglement), we wonder whether it might be seen as a ‘fragment’ of a more general theory which might recover the condition of locality at a ‘higher’ level (Bell 1975, reprinted in Bell 2004, 55). The λ states just need be general enough to encompass quantum states,⁷ but nowhere it is required that this sort of ‘inclusive’ feature is in itself equivalent to the assumption that the Λ -statistical theory must be a ‘classical’ theory in the sense of pre-determined values for any meaningful physical quantity.

As to the second point, I remark that the mere definition of what a joint probability distribution is in the appropriate context need not imply by itself that the definition is applicable. But if I assume locality, the outcome probabilities factorization in (4) obtains *as a consequence* of locality: it is exactly locality that imposes on the form of distribution the very constraint we need in order to be sure that the desired joint probability distribution actually exists. This is hardly surprising and is exactly what we would expect: if an outcome is sensitive to causal factors that may affect it only locally, the probability of each outcome is independent from the probability of any other outcome.

By this point of view, in order for the Bell argument to go through we do not need to assume the existence of a total joint probability distribution for all possible values: it is just locality that justifies the factorization of the specific probabilities we are interested in. If, on the other hand, we follow ZB in assuming that the factorization in (4) is really the restriction of a total joint probability distribution, we are in fact *unnecessarily imposing from the outside* the existence of values for all observables. Should we push the above argument a bit further, we might then *ex absurdo* add any condition we like to locality—let me call it the Whatever-Condition—and claim that Bell’s theorem does not show the quantum world to be nonlocal but rather that it does not satisfy the Whatever-Condition! On the other hand, it is equally unsurprising that, if I assume an equivalence between locality in the sense of (3) and the existence of a joint probability distribution for all values as if these values were already defined (no matter whether they are measured or not), I can go *from* the existence of a joint probability distribution *to* locality: if values are fixed, they are insensitive to whatever else takes place remotely and therefore locality obtains.⁸

⁷ It should be pointed out that λ can just be ψ .

⁸ The standard reference for all arguments relying on the equivalence between factorizability and existence of a joint probability distribution is Fine (1982): see also Mermin (1992), where the author writes: “Bell’s theorem, although it does rely on the properties of a particular state, proves the non-existence of a joint distribution for a set of observables *required to have one by an apparently common-sense notion of physical locality*” (27, my emphasis). In addition to the argument defended above in the

3.2 View (B): Counterfactual Reasoning, the EPR Argument and All That

As recalled above, the ZB refutation of View (B) amounts to showing that

any attempts to derive determinism, or hidden-variables, via an EPR argument are futile. The argument of EPR was based on an assumption, which is often not noticed or is forgotten: *counterfactual definiteness*. This assumption allows one to assume the definiteness of the results of measurements, which were actually not performed on a given individual system. They are treated as unknown, but in principle *defined* values. This is in striking disagreement with quantum mechanics and the complementarity principle. (Zukovski and Brukner 2014, 424009-2, emphasis in the original).

The lesson that we learn from the joint refutations of views (A) and (B), then, “is that Bell’s theorem, in all its forms, tells us *not* what quantum mechanics *is*, but what quantum mechanics *is not*.” (Zukovski and Brukner 2014, 424009-8, emphasis in the original). With respect to the refutation of view (B), the ZB stance is widely shared. In an extensive review on Bell’s theorem appeared on the *American Journal of Physics* in 2010, Guy Blaylock claimed “that the minimal assumptions behind Bell’s predictions are locality and ‘counterfactual definiteness’ (that we can postulate a single definite result from a measurement even when the measurement is not performed)” (Blaylock 2010, 111). In the same journal in 2013 Lorenzo Maccone—while providing “an extremely simple proof of Bell’s inequality”—claims that “Bell’s theorem can be phrased as ‘quantum mechanics cannot be both local and counterfactual-definite’” (Maccone 2013, 854),⁹ whereas Stephen Boughn most recently states: “Bell’s theorem clearly makes use of counterfactual definiteness; his inequality involves the correlations of the spins of the two particles in each of two different directions that correspond to non-commuting spin components. *This use of counterfactuals is entirely appropriate because it is used to investigate a test for classical hidden variable theories.*” (Boughn 2017, 648, my emphasis).¹⁰

I recall here that the view (B) is the claim that determinism (or any cognate notion) can be derived from the premises of locality, the freedom of an experimenter to choose the setting of his/her local apparatus and quantum predictions and nothing

Footnote 8 continued

main text, it can be shown that the Fine deterministic hidden variable model implicitly assumes locality anyway (Shimony 1984; Laudisa 1996).

⁹ A remark in sociology of science: since the *American Journal of Physics* is the intended journal of US physics teachers, it is an revealing sign that in two issues of the journal in two different years readers are pedagogically taught that a ghostly assumption like counterfactual definiteness does play an essential role in the EPR argument and in the ensuing Bell theorem!

¹⁰ For older, notable statements in the same direction see e.g. Unruh (1999, 2001). It should be noted that, historically, there is an interesting antecedent to the discussion in the text, namely the debate surrounding the counterfactual formulation of the Bell theorem proposed especially by Eberhard (1977, 1978), Stapp (1985a, b) and Kraus (1989). In 1990, Butterfield et al. by applying the very conceptual resources of the Lewisian theory of counterfactual conditionals to the Eberhard–Stapp–Kraus formulation, could show that such formulation in fact embodied an implicit assumption of determinism: in that context such assumption is de facto equivalent to counterfactual definiteness, which in turn leads us back to the issue that I discuss in the text.

more. If ‘QF’ stands for the quantum formalism, ZB depict view (B) in a compact form as

$$\text{Freedom \& QF \& Locality} \Rightarrow \text{Determinism}$$

Usually this implication is traced back to the very EPR argument but ZB claim with no further justification that this reconstruction

is wrong. The logical structure of the EPR argumentation about elements of reality is

$$\text{Freedom \& QF \& Locality \& Counterfactual Definiteness}$$

$$\Rightarrow \text{for specific observables elements of reality exist for the EPR state}$$

The thesis of the implication is true, provided one does not make unfounded generalization of it to *arbitrary* observables and *arbitrary* states. Indeed, for the EPR state and momentum and position observables (P and Q), elements of reality seem to be a consistent notion. However this is arrived at by considering two situations of which only one can be the case in the given run of the experiment (measuring either P or Q). This is counterfactual definiteness at work (Zukovski and Brukner 2014, 424009-6, emphasis in the original).

The flaw in this argument is similar to the one discussed in the previous section, namely the supposition that the existence of the elements of physical reality is *added* on top with no other justification than to show that this ‘classical-like’ supposition cannot hold in a quantum world. If, on the other hand, we take seriously (1) the fact that it is the locality assumption that is prior, and (2) that it is exactly this assumption that legitimates to talk of elements of physical reality corresponding to spin measurements along different axes, we realize that the above appeal to counterfactual reasoning does no harm. Take the EPR argument in the Bohm formulation. *If* the theory by which we try to account for EPR correlation is assumed to be local, *then* the choice of a certain axis for spin measurement *here* is totally unaffected by what is the chosen axis for spin measurement *there*, and this holds perfectly the same also in a counterfactual sense. Suppose in the actual world the experimenter has chosen the axis z : if her measuring operations cannot affect by definition what axis is chosen—and what outcome has been obtained—on the other side, this holds naturally for *whatever choice* and this makes it invalid to claim that in an ordinary EPR argument we have adopted a counterfactual definiteness assumption *in addition* to locality.

Let us consider the point in a specific formulation of the EPR argument in the Bohm version, in which we have a composite quantum system $S_1 + S_2$ of a pair of spin-1/2 particles S_1 and S_2 . The composite system is prepared at a time t_0 in the singlet state Ψ

$$\Psi = 1/\sqrt{2}(|1, +>_{\mathbf{n}}|2, ->_{\mathbf{n}} - |1, ->_{\mathbf{n}}|2, +>_{\mathbf{n}}),$$

where \mathbf{n} denotes a generic spatial direction. We take into account the measurements concerning the spin components along given directions, whose possible outcomes

are only two (conventionally denoted by ‘+ 1’ and ‘− 1’). We assume also that the spin measurements on S_1 and S_2 are performed when S_1 and S_2 occupy two mutually isolated spacetime regions R_1 and R_2 . According to QM, we know that if the state of $S_1 + S_2$ at time t_0 is Ψ , then the (reduced) states of the subsystems S_1 and S_2 at time t_0 are respectively

$$\begin{aligned}\rho(1, \Psi) &= 1/2(\mathbf{P}_{|1,+>\mathbf{n}} + \mathbf{P}_{|1,->\mathbf{n}}), \\ \rho(2, \Psi) &= 1/2(\mathbf{P}_{|2,+>\mathbf{n}} + \mathbf{P}_{|2,->\mathbf{n}}),\end{aligned}\tag{RS}$$

(‘RS’ stands for ‘reduced states’) so that, for any \mathbf{n} ,

$$\begin{aligned}\text{Prob}_{\rho(1,\Psi)}(\text{spin}_{\mathbf{n}} \text{ of } S_1 = +1) &= \text{Prob}_{\rho(1,\Psi)}(\text{spin}_{\mathbf{n}} \text{ of } S_1 = -1) = 1/2 \\ \text{Prob}_{\rho(2,\Psi)}(\text{spin}_{\mathbf{n}} \text{ of } S_2 = +1) &= \text{Prob}_{\rho(2,\Psi)}(\text{spin}_{\mathbf{n}} \text{ of } S_2 = -1) = 1/2\end{aligned}$$

Moreover, if we perform at a time t a spin measurement on S_1 along \mathbf{n} with outcome + 1 (− 1), a spin measurement on S_2 along \mathbf{n} at a time $t' > t$ will give with certainty the outcome − 1 (+ 1), namely for any \mathbf{n} (‘AC’ stands for ‘AntiCorrelation’)

$$\begin{aligned}\text{Prob}_{\Psi}[(\text{spin}_{\mathbf{n}} \text{ of } S_1 = +1) \ \&\ (\text{spin}_{\mathbf{n}} \text{ of } S_2 = -1)] &= \\ \text{Prob}_{\Psi}[(\text{spin}_{\mathbf{n}} \text{ of } S_1 = -1) \ \&\ (\text{spin}_{\mathbf{n}} \text{ of } S_2 = +1)] &= 1.\end{aligned}\tag{AC}$$

Let us suppose now to perform at time $t_1 > t_0$ a spin measurement on S_1 with outcome + 1. Therefore, according to (AC), a spin measurement on S_2 along \mathbf{n} at a time $t_2 > t_1$ will give with certainty the outcome − 1. Let us suppose now to assume the following criterion:

REALITY – If, without interacting with a physical system S , we can predict with certainty – or with probability 1 - the value \mathbf{q} of a quantity \mathbf{Q} pertaining to S , then \mathbf{q} represents an objective property of S (denoted by $[\mathbf{q}]$).

Then, for $t_2 > t_1$ $[\text{spin}_{\mathbf{n}} = - 1]$ represents an objective property of S_2 . But might the objective property $[\text{spin}_{\mathbf{n}} = - 1]$ of S_2 have been somehow “created” by the spin measurement on the distant system S_1 ? The answer is negative if we assume the following condition:

LOCALITY – No objective property of a physical system S can be influenced by operations performed on physical systems that are isolated from S .

At this point, **LOCALITY** allows us to state the existence of the property $[\text{spin}_{\mathbf{n}} = - 1]$ for the system S_2 also at a time t' such that $t_0 > t' > t_1$. Namely, if we assume that the measurement could not influence the validity of that property at that time, it follows that the property *was holding already at time t'* , a time that *precedes* the measurement performed on the other subsystem. But at time t' the state of $S_1 + S_2$ is the singlet state Ψ , therefore according to (RS) the state of S_2 is the reduced state $\rho(2, \Psi) = 1/2(\mathbf{P}_{|2,+>\mathbf{n}} + \mathbf{P}_{|2,->\mathbf{n}})$, that prescribes for the property $[\text{spin}_{\mathbf{n}} = - 1]$ of S_2 only a probability 1/2. Let us consider finally the following condition:

COMPLETENESS – Any objective property of a physical system S must be represented within the physical theory that is supposed to describe S .

It follows that there exist properties of physical systems that, according to the **REALITY** condition are objective—like $[\text{spin}_n = -1]$ for S_2 —but that QM does not represent as such: therefore QM is not complete.

If we focus on the question “Might the objective $[\text{spin}_n = -1]$ of S_2 have been somehow “created” by the spin measurement on the distant system S_1 ?”, we realize that the only acceptable possibility for justifying the existence of the property $[\text{spin}_n = -1]$ for the system S_2 also at a time t' is its pre-measurement definiteness *under the hypothesis that LOCALITY holds*: such definiteness, therefore, is *justified* by **LOCALITY** and nowhere the argument needs to *assume independently* that spin properties are definite. This is equivalent to stress that assuming **REALITY** is not equivalent to assuming at the outset the existence of definite, pre-measurement properties: **REALITY** is simply a sufficient criterion for a property of physical system to be objective, namely holding in a measurement-independent way.¹¹ If the existence of definite, pre-measurement properties is not required as an *independent* assumption, the same holds for the assumption of counterfactual definiteness, that “allows one to assume the definiteness of the results of measurements, which were actually not performed on a given individual system. They are treated as unknown, but in principle *defined* values.” (Zukovski and Brukner 2014, 424009-2).

4 Conclusions

Alongside with a wide consensus on the relevance of the EPR-like arguments and Bell’s theorem to deep foundational issues concerning fundamental physics, it is interest to note that a different, ‘deflationary’ attitude toward these issues is alive, shared in the past also by physicists of such high profile as Richard Feynman, Murray Gell-Mann, Jim Hartle and Robert B. Griffiths. In a lecture held in 1983 Feynman claimed that Bell’s theorem “is not a theorem that anybody thinks is of any particular importance. We who use quantum mechanics have been using it all the time. It is not an important theorem. It is simply a statement of something that we know is true—a mathematical proof of it.” (quoted in Whitaker 2016, 493), while Murray Gell-Mann and Jim Hartle wrote:

The EPR or EPRB situation is no more mysterious. There, a choice of measurements, say σ_x or σ_y for a given electron, is correlated with the behavior of σ_x or σ_y for another electron because the two together are in a singlet state even though widely separated. Again, the two measurement situations (for σ_x and σ_y) decohere from each other. But here, in each, there is also a correlation between the information obtained about one spin and the information that can be obtained about the other. This behavior, although

¹¹ Maudlin goes as far as claiming that the EPR criterion—what I call here **REALITY**—is an *analytic* criterion, namely a criterion that simply follows from the very meanings of the words that compose it (Maudlin 2014, 424010-6).

unfortunately called “non-local” by some authors, involves no non-locality in the ordinary sense of quantum field theory and no possibility of signaling outside the light cone. The problem with “local realism” that Einstein would have liked is not the locality but the *realism*. (Gell-Mann and Hartle 1989, 340).¹²

A key strategy of particular expressions of this deflationary attitude lies in pointing out counterfactuality as the style of reasoning that carries with itself the allegedly ‘realistic’ burden of the Einstein–Bell arguments. In the above pages I have attempted to develop some arguments to counteract this attitude: if agreed upon, these arguments lead us to see that the real challenge is not the search for imaginary flaws in the Einstein–Bell work, but rather the exercise of scientific (and metaphysical!) imagination in order to figure out what a non-local natural world might look like.

References

- Bacciagaluppi, G., & Valentini, A. (Eds.). (2009). *Quantum theory at the crossroad. Reconsidering the 1927 Solvay conference*. Cambridge: Cambridge University Press.
- Bell, J. S. (1975). *The theory of local beables*, CERN publication TH-2053-CERN (reprinted in Bell 2004, pp. 52–62).
- Bell, J. S. (2004). *Speakable and unspeakable in quantum mechanics* (2nd ed.). Cambridge: Cambridge University Press.
- Bertlmann, R. A., & Zeilinger, A. (Eds.). (2002). *Quantum (un)speakables. From Bell to quantum information*. Berlin: Springer.
- Blaylock, G. (2010). The EPR paradox, Bell’s inequality, and the question of locality. *American Journal of Physics*, 78, 111–120.
- Born, M. (2005). *The Born–Einstein letters 1916–1955*. New York: McMillan.
- Boughn, S. (2017). Making sense of Bell’s theorem and quantum nonlocality. *Foundations of Physics*, 47, 640–657.
- Butterfield, J., Clifton, R., & Redhead, M. L. G. (1990). Nonlocal influences and possible worlds—A Stapp in the wrong direction. *British Journal for Philosophy of Science*, 41, 5–58.
- Eberhard, P. H. (1977). Bell’s theorem without hidden variables. *Il Nuovo Cimento*, 38B, 75–80.
- Eberhard, P. H. (1978). Bell’s theorem and different concepts of locality. *Il Nuovo Cimento*, 46B, 392–419.
- Einstein, A. (1949). Autobiographical notes. In P. A. Schilpp (Ed.), *Albert Einstein philosopher-scientist*. New York: Harper Torchbooks.
- Fine, A. (1982). Hidden variables, joint probability and the Bell inequalities. *Physical Review Letters*, 48, 291–295.
- Fine, A. (1986). *The Shaky Game. Einstein realism and the quantum theory*. Chicago: University of Chicago Press.
- Gell-Mann, M., & Hartle, J. (1989). Quantum mechanics in light of quantum cosmology. In *Proceedings of the third international symposium on the foundations of quantum mechanics, Tokyo* (pp. 321–343).
- Goldstein, S., Norsen, T., Tausk, D., & Zanghì, N. (2011). Bell’s theorem. *Scholarpedia*, 6(10), 8378.
- Griffiths, R. B. (2011). EPR, Bell and quantum locality. *American Journal of Physics*, 79, 954–965.

¹² In the so-called *consistent-histories* approach, that Gell-Mann, Hartle, Griffiths, Omnes (and others) have developed in detail, it is also claimed that locality can be preserved for quantum mechanics: see for instance Griffiths (2011). Arguments that show this claim to be far from convincing can be found in Goldstein et al. (2011) and Maudlin (2011).

- Hardy, L. (1995). The EPR argument and nonlocality without inequalities for a single photon. *Annals of the New York Academy of Sciences*, 755, 600–615.
- Home, D., & Whitaker, A. (2007). *Einstein's struggles with quantum theory*. Berlin: Springer.
- Howard, D. (1993). Was Einstein really a realist? *Perspectives on Science*, 1, 204–251.
- Kraus, K. (1989). Quantum theory does not require action at a distance. *Foundations of Physics Letters*, 2, 1–6.
- Laudisa, F. (1996). Nonlocality: A defense of widespread beliefs. *Studies in History and Philosophy of Modern Physics*, 27, 297–313.
- Maccione, L. (2013). Simplest proof of Bell's inequality. *American Journal of Physics*, 81, 854–859.
- Maudlin, T. (2011). How Bell reasoned: A reply to Griffiths. *American Journal of Physics*, 79, 966–970.
- Maudlin, T. (2014). What Bell did. *Journal of Physics A: Mathematical and Theoretical*, 47, 424010.
- Mermin, N. D. (1992). Not quite no simply no hidden variables. *American Journal of Physics*, 60, 25–27.
- Norsen, T. (2005). Einstein's boxes. *American Journal of Physics*, 73, 164–176.
- Peres, A. (1978). Unperformed measurements have no result. *American Journal of Physics*, 46, 745–747.
- Shimony, A. (1984). Contextual hidden variable theories and Bell's inequalities. *British Journal for the Philosophy of Science*, 35, 25–45 (reprinted in *Search for a Naturalistic World View. Vol. II Natural Science and Metaphysics*, Cambridge University Press, Cambridge, 1993, pp. 104–129, reference to the reprinted edition).
- Stapp, H. P. (1985a). Bell's theorem and the foundations of physics. *American Journal of Physics*, 53, 306–317.
- Stapp, H. P. (1985b). EPR: What has it taught us? In P. Lahti & P. Mittelstaedt (Eds.), *Symposium on the foundations of modern physics* (pp. 637–652). Singapore: World Scientific.
- Unruh, W. G. (1999). Nonlocality, counterfactuals, and quantum mechanics. *Physical Review A*, 59, 126–130.
- Unruh, W. G. (2001). Is quantum mechanics non-local? In T. Placek & J. Butterfield (Eds.), *Non-locality and modality* (pp. 125–136). Dordrecht: Kluwer.
- Whitaker, A. (2016). Richard Feynman and Bell's theorem. *American Journal of Physics*, 84, 493–494.
- Wolf, S. (2015). Nonlocality without counterfactual signalling. *Physical Review A*, 92, 052102.
- Zukovski, M., & Brukner, C. (2014). Quantum nonlocality—It ain't necessarily so.... *Journal of Physics A: Mathematical and Theoretical*, 47, 424009.