The logic of EPR

ABNER SHIMONY

Professor Emeritus of Philosophy and Physics,
Boston University, Boston, MA 02215, U.S.A

.

Dedicated to my friend Georges Lochak for his passionate and courageous devotion to the foundations of physics.

ABSTRACT: A new version of EPR's argument for the incompleteness of the quantum mechanical description of certain physical systes is presented. The space-time structure is explicitly assumed to be that of special relativity theory, and the concept of locality is understood in the sense of that theory. The argumentation uses the logic of quantum conditional propositions and some applications of inductive logic. Contrary to the opinions of Ghirardi and Grassi and also of Redhead and La Rivière, there is no need to use the logic of counterfactual conditionals in order to reach EPR's conclusion relativistically.

1 Introduction

The argument of A.Einstein, B.Podolsky, and N.Rosen[1] (EPR, an acronym that I shall use both for the names of the authors and for the argument itself) concluded that the quantum mechanical description of certain physical systems is incomplete. It has three premisses: (a) *A sufficient condition for the existence of an element of physical reality* --"If, without in any way disturbing a system, we can predict with certainty (i.e., with probability equal to unity) the value of a physical quantity, then there exists an element of physical reality corresponding to this physical quantity; " (b) *perfect correlation* -- a pair of systems is prepared in a quantum state which ensures that there is perfect correlation between the outcomes of measurements of observables R and L and also between the outcomes of measurements of observables R′ and

L′, where R and R′ are incompatible observables of system I (i.e., are represented by non-commuting operators), and L and L′ are incompatible observables of system II, the two systems being well separated when the measurements are performed; (c)*locality* -- “Since at the time of measurement the two systems no longer interact, no real change can take place in the second system in consequence of anything that may be done to the first system.”

EPR is a quite informal argument, abstaining from a precise definition of “if ... then” and “can predict” in premiss (a) and from a specification of the conditions in which the antecedent of (c) is satisfied. It is not clear, for example, whether premiss (c), locality, is formulated in a relativistic or a non-relativistic space-time. J.C.Ghirardi and R.Grassi[2] and M.Redhead and P.La Rivière[3] all questioned the validity of a relativistic version of EPR unless certain inferences in the logic of counterfactual conditionals are legitimate, though the two pairs of collaborators disagreed with each other regarding the assumptions underlying these steps. I contend that there is a relativistic version of EPR which is valid without any recourse to counterfactual conditionals.

Sect. 2 presents a self-contained relativistic formulation of EPR, making explicit the meaning of crucial terms and the logic of the argumentation. This formulation, which uses the logic of conditional propositions and inductive logic but not the logic of counterfactuals, is essentially a sympathetic amplification of the original informal argument of EPR.

Sect. 3 comments briefly on the logic of counterfactuals. It maintains that counterfactual reasoning is not needed for a valid amplification of EPR, specifically not for a relativistic formulation. Relativistic locality does ensure the truth of certain counterfactual propositions, but it is an open question whether these suffice, without recourse to inductive logic, to derive the entirety of EPR’s conclusion concerning the existence of elements of physical reality. Once one has derived EPR’s elements of physical reality by another route, however, it is possible to use them to enlarge the class of counterfactual propositions to which truth values can be assigned.

2 A relativistic formulation of EPR

We shall consider an enormous ensemble ε of experimental arrangements, in each of which a measurement of either observable R or observable R′ (physical incompatibility preventing the measurement of both) is made in a space-time region **R**, and likewise a measurement is made of L or L′ (but not both) in space-time region **L**. The letters designating observables and regions are tacitly assumed to bear indices corresponding to enumeration in the ensemble, and the regions **R** and **L** (with the same tacit index) have space-like separation. Each observable is bivalent, with the two values labeled + and -. The statistics of outcomes in this ensemble is governed by a single quantum state Ψ, which is not endowed with an index and hence, according to one group of quantum mechanicians, is ascribable to each individual member of the ensemble, although there are others who ascribe it only to the ensemble as a whole -- a dispute which will not be considered here. It is assumed that the quantum state Ψ ensures perfect correlations between R and L -- specifically, with certainty either both outcomes are + or both are - if R and L are measured -- and also ensures perfect correlations between R′ and L′ if these are measured. (Note that EPR explicitly identify "certainty" and "with probability unity," disregarding the distinction commonly made by probability theorists. EPR's usage will be followed in the present paper.)

A comment on inferences from the quantum state is important. An infinite number of conditional propositions can be based upon a single quantum state, each of the form :

If observable A is measured, then the probability of outcome a_i is p_i. (1)

In terms of a normalized vector ϕ which represents the specified quantum state and the projection E_i onto the eigenspace associated with the value a_i of observable A, the mathematical expression for p_i is :

$$p_i = |\langle\phi|E_i|\phi\rangle|^2 \ . \quad (2)$$

(In writing Eq. (2), I restrict attention to observables with discrete spectra, thereby avoiding the examples chosen by EPR, position and momentum, but allowing the spin observables considered by D.Bohm[4]; and I also restrict attention to pure quantum states. The following argumentation is not vitiated by these restrictions.)

Propositions of the form "If x then y", where x and y are themselves propositions, are conditional. But there are various kinds of conditionals -- material implication, strict implication, counterfactual conditional (to be considered in Sect. 3), and the conditional that appears in Proposition (1), which I shall call "the quantum conditional", since a standard name for it seems not to have been adopted in the literature. The quantum conditional may be usefully characterized by a comparison with material implication. When "If x then y" is construed as material implication, usually written as "$x \supset y$" or "$x \rightarrow y$", the conditional proposition has a definite truth value in all four rows of a truth table: it is true when x and y are both true, or when x is false and y is either true or false, and it is false when x is true and y is false. When "If x than y" is construed as a quantum conditional, as in Proposition (1), where "observable A is measured" is substituted for "x" and "the probability of outcome a_i is p_i" is substituted for "y", then the conditional proposition is definitely asserted to have a truth value in two rows of the truth table: it is true if x and y are both true and false if x is true and y is false; but when x is false there is no definite assignment of truth values to the conditional proposition. There are different versions of the quantum conditional, according to the interpretation of "no definite asignment of truth values." In a strong version it is meaningless to assign truth values when the antecedent, "observable A is measured" is false; this version may be Bohr's intention, since he regards the meaningfulness of locutions about truth and reality to be relative to the experimental arrangement[5]. In another version, which seems to be the minimal understanding of Proposition (1), there is no ascription of meaninglessness but simply an abstention from assigning a truth value when the antecedent is false, since there is no need for such assignments when actual phenomena are analyzed quantum mechanically. It should be noted that when the quantum state is given but the observable A mentioned in the antecedent of Proposition (1) is not given, either because objectively its choice is indeterminate or because it is unknown, then an infinite family of propositions of the form of (1), referring to different observables, can all be used contingently, with the contingent probabilities evaluated by Eq. (2). The availability of these contingent propositions shows how much information is contained in the quantum state. It should also be noted that there are interpretations of quantum mechanics according to which, at least in certain circumstances, Proposition (1) can be assigned a definite truth value when the antecedent is false, and accordingly the conditional in (1) is to be understood as a counterfactual conditional, rather than as a quantum conditional in either of the senses distinguished above. Such interpretations, however, will be disregarded in this paper until after the relativistic version of EPR is completed in Sect. 2. Once one reaches EPR's conclusion is that the quantum mechanical description of physical systems is in certain situations incomplete, new light is thrown upon the quantum conditional, which will discussed in Sect. 3. But in the process of presenting the relativistic EPR argument in Sect. 2 quantum mechanical implications will be drawn austerely, a definite truth value being assigned to Proposition (1) only when its antecedent is true.

Now the original argument of EPR will be amplified as follows. It is assumed that free choice between measuring R or R′ is exercised by a decision made in region **R** and free choice between mesuring L or L′ is independently exercised by a decision in region **L**, with non-negligible probabilities of all four choices; the difficult philosophical problem of understanding freedom is left in limbo. The ensemble ε is thus partitioned into four non-empty, mutually exclusive, and exhaustive subensembles,

$$\varepsilon = \varepsilon(RL) \cup \varepsilon(RL') \cup \varepsilon(R'L) \cup \varepsilon(R'L'), \quad (3)$$

where the letters inside the parentheses denote which observables are measured. Since the ensemble ε is enormous and the probability of membership in each subensemble is non-negligible, each of the four subensembles is also enormous.

Attention is now focused upon the subensemble ε(RL). The quantum state Ψ was characterized as predicting perfect correlation between the outcomes of mesuring R and L : either both outcomes are + or both are -. Perfect correlation can be expressed in terms of the product observable RL, which has outcome +1 if measurements of both R and L have the same outcome and -1 if the outcomes of R and L are different: making the appropriate substitutions into Proposition (1) yields :

"If observable RL is measured,
the probability of outcome +1 is 1". (1')

If the system under consideration belongs to ε(RL), then Proposition (1') is true in view of the characterization of the state Ψ and the interpretation of "if...then" as the quantum conditional. Hence, if the outcome of the measurement of R is +, an observer in **R** can infer with certainty (i.e., probability unity) that the outcome of the measurement of L is also +, and correspondingly if the outcome of the measurement of R is -. Note that the word "infer" is used in place of EPR's word "predict," because **R** and **L** have space-like separation, and therefore the outcome of the measurement of L cannot be invariantly later than the outcome of the measurement of R. Checking the correctness of the prediction, of course, requires the registration of the measurement of L and propagation of the registered information to a region in the overlap of the forward light-cones of **R** and **L**; but if the registration and propagation processes are accurate, then the information thus obtained can be safely ascribed to the outcome of measuring L in region **L**.

Contact can now be established with EPR's original argumentation. Because of the space-like separation of **R** and **L** the locality condition -- labeled (c) in Sect. 1 -- is satisfied: no change can take place in **L** as a consequence of anything done in **R**. (Relativistic space-time structure is evidently superior to non-relativistic for legitimating this step.) Because of the perfect correlation between the outcomes of measuring R and L, one can infer with certainty the outcome of the latter, and because of locality this is done without in any way disturbing **L**.The antecedent of EPR's condition for the existence of an element of physical reality, labeled (a) in Sect. 1, is (after the replacement of "predict" by "infer") satisfied regarding the observable L in region **L**.

Consequently, *the existence of an element of physical reality corresponding to L is demonstrated for a system in ε(RL)*. It must be recognized, of course, that from the standpoint of an observer in **R**, this conclusion is contingent, because that observer cannot know whether the observable L was measured and hence does not know whether the system under consideration belong to the subensemble ε(LR). But since the system under consideration does as a matter of fact belong to ε(RL), the contingent inference is objectively valid, and furthermore its validity can be checked by an observer located far enough forward along the world-line of the observer in **R** to receive information about **L.** Obviously, the foregoing reasoning can be repeated after switching **R** and **L** to conclude that there is an element of physical reality in **R** corresponding to the observable R. It is reasonable to amplify EPR's conclusion slightly by saying that the elements of physical reality demonstrated in **R** and **L** are located within these regions in the backward light-cones of the decisions to measure R and L respectively, since relativity theory does not otherwise allow them to be causally efficacious in the measurement processes.

The foregoing argument is based upon the assumption that the system of interest belongs to the subensemble ε(RL). However, the decisions to measure R rather than R′ and L rather than L′ are by assumption made freely, and therefore membership in the subensemble ε(RL) is the result of two independent free choices, one made in **R** and the other in **L**. The elements of physical reality which determine the outcomes of the measurement of R and the measurement of L must be present in every member of the entire ensemble ε, since any member of the entire ensemble could be selected to be in ε(RL), and it would be an unbelievable coincidence if all selected to belong to the very large subensemble ε(RL) were endowed with the elements of physical reality that guaranteed perfect correlations between the outcomes of measuring R and L while a non-negligible fraction of the entire ensemble ε lacked this property. Accepting such a coincidence in preference to attributing the property to all members of ε is contrary both to informal inductive logic and to probabilistic refinements of inductive logic. This conclusion cannot be eluded by hypothesizing that it is the choice of R that endows the part of the system in **R** with an element of physical reality determining the outcome of measuring R, and likewise that the choice of L endows the part of the system in **L** with the element of physical reality determining the outcome of measuring L -- because such a hypothesis could not account for the agreement with certainty of the two outcomes. Hence, *the elements of physical reality sufficient to determine the outcomes of measurements of R and L, if these observables were indeed measured, are present in every member of the ensemble ε.* To my knowledge, the first to use inductive logic explicitly in a discussion of EPR was B. d'Espagnat[6].

The foregoing reasoning can be paralleled, substituting R′ for R and L′ for L. The first stage is the demonstration that every system in the subensemble ε(R′ L′) is endowed with elements of physical reality determining concordant outcomes of the measurements of R′ and L′. The second stage is to extend this conclusion by induction to the whole of the ensemble ε. Combining the conclusion regarding the observables R and L with that concerning R′ and L′ yields the result that *for each member of the ensemble ε both observables R and R′ in region **R** and both observables L and L′ in region **L** are assigned definite values, even though only one of each pair is measured.* This result shows that the quantum mechanical description, according to which the two members of each pair are incompatible, is incomplete.

To summarize, EPR's conclusion that the quantum mechanical description of some physical systems is incomplete has been reached by an amplification of their original argument. The assumptions made here were slight variants of the three assumptions of EPR, labeled (a), (b), and (c) in Sect. 1, and the reasoning was a combination of the logic of quantum conditionals and a standard procedure of inductive logic.

3 Counterfactual Conditionals

A proposition of the form "If x then y" is a counterfactual conditional if the proposition denoted by "x" is false, and nevertheless the proposition is assigned a definite truth value.

In ordinary prose, the conditional is usually written with a subjunctive, "If x were true, then y." Sometimes the subjunctive is used without the protocol that the antecedent is false, even though the most interesting case is that in which x is false.

Whatever convention is adopted on the matter of restricting attention to a false antecedent, it is clearly understood in the literature that there is not a truth-table for the counterfactual conditional. In the locution of logical theory, the counterfactual conditional is not truth-functional. The truth value of "If x were true, then y" depends upon the content of the propositions designated by "x" and "y" and upon contextual information. It is generally recognized that the problem of formulating criteria for assigning truth values to counterfactual conditionals is subtle and difficult. The closest to a standard work on these criteria is the book of D. Lewis[7], but that is not canonical. In the context of quantum mechanics and special relativity theory a contribution to the logic of counterfactuals, to be summarized below, has been made by H. Stapp[8] and by A.Shimony and H.Stein[9].

At this point, what I wish to emphasize is that the demonstration of EPR's conclusions in Sect. 2 never made use of counterfactual conditionals. In proving for a member of ε(RL) that there is an element of physical reality in region **L** determining the outcome of the measurement of observable L, use was made of a quantum conditional proposition in a situation where the antecedent is true. The extension to other members of the ensemble in which L′ is measured, and therefore the antecedent is not true, was achieved by an inductive extrapolation from ε(RL) to the whole of ε, not by an illegitimate assignment of a truth value to the quantum conditional proposition nor by a switch to counterfactual logic.

Both Ghirardi/Grassi[2] and Redhead/La Rivière[3] maintained that without counterfactual reasoning there is no way to reach the conclusion that an element of physical reality exists in **L** which determines what would be the outcome of measuring observable L if actually L′ were measured. This thesis is plausible, but it turns out not to be true, as shown by the argumentation of Sect. 2. An advantage of dispensing with counterfactuals in reaching EPR's conclusions is the avoidance of the controversy between these two groups of authors on the exact criteria for assigning truth values to counterfactual conditionals.

It is nevertheless interesting to inquire whether the logic of counterfactuals could provide a procedure essentially different from that of Sect. 2 to EPR's conclusion. I am not sure of the answer to this question. I can see part of a procedure which substitutes the logic of counterfactual conditionals for inductive logic, but I do not see how to complete the argument. One first considers the subensemble ε(RL) and argues exactly as in Sect. 2 that there is an element of physical reality in **L** determining the outcome of the measurement of L, which is indeed measured. Likewise there is an element of physical reality in **R** determining the outcome of the measurement of R, which is indeed measured. One can then reach the same conclusion concerning the subensemble ε(RL′) by applying the following sufficient condition for the truth of a counterfactual conditional that is implicit in Ref. 8 and explicit in Ref. 9:

> If x asserts a matter of fact about a space-time region **x**, then 'if x were true then y' is true in a possible world W if and only if y is true in every possible world W′ in which x is true and which does not differ from W outside the extended future light-cone of **x**.
>
> (Condition)

(By "extended future light-cone of **x**" is meant the union of **x** itself with the future light-cone of **x**). The Condition is reasonably a consequence of relativistic locality. The Condition can now be applied by substituting "observable L is measured" for "x" and "the outcome of measuring L has a value agreeing with the outcome of the measurement of R in **R"** for "y". Suppose W is a possible world in which the system of interest belongs to the subsensemble ε(RL′). Then, in view of the substitutions just made, x is false in W, but y is true in every possible world W′ in which x is true and which agrees with W outside the extended future light-cone of **x**; the reason is that R and L are both measured in such a W′, thus making the system of interest belong to the subensemble ε(RL) in possible world W′. Accordingly, by the Condition, the counterfactual proposition "if x were true then y" is true. Thus there is valid reasoning to the conclusion that there is an element of physical reality in **L** determining what the outcome of a measurement of L would have been if L had been measured, proceeds just as when the system belongs to ε(RL).

If the system belongs to ε(R′L) one can parallel the foregoing reasoning to infer that an element of physical reality exists in **R** which would determine the outcome of the measurement of R had it been measured. In order to ensure agreement with the outcome of the measurement of L there must also be an element of reality in **L** which determined the outcome of the measurement of L prior to the decision to make that measurement.

Continuing in this way we have the following results: there are elements of physical reality determining the outcomes of actual or possible measurements of both R and L and both R′ and L′ when the system is in either of the subensembles ε(RL′) or ε(R′L); there are elements of physical reality determining the outcomes of measuring R and L when the system is in ε(RL), but it has not been established that there are elements of physical reality determining what the outcomes of R′ and L′ would have been in this case; and there are elements of physical reality determining the outcomes of measurements of R′ and L′ when the system is in ε(R′L′), but it has not been established that there are elements of physical reality determining what the outcomes of measurements of R and L would have been in this case.

There are thus two gaps in the procedure aiming at establishing, by use of counterfactual reasoning and dispensing with inductive reasoning, that there are elements of physical reality that would determine the outcomes of measurements of all four of R, R′, L, and L′ in the entire ensemble ε. Can this gap be filled? I am not sure. Perhaps a reasonable broadening of the foregoing Condition for the assignment of truth values to counterfactual propositons could be proposed which would suffice to fill this gap. Lewis's discussion in Ref. 7 may be helpful for this purpose. I shall leave this question open, since my primary purpose -- to establish EPR's conclusion without invoking counterfactual logic -- has been fulfilled.

It is clear that once EPR's conclusion has been reached by the procedure of Sect. 2, counterfactual reasoning concerning all four observables R, R′, L, and L′ is legitimated for all systems in the ensemble ε. In general, the existence of elements of physical reality supplementing the quantum mechanical description of physical systems greatly extends the range of legitimacy of counterfactual reasoning. For example, Shimony and Stein, in Ref. 9, have argued that Stapp, in Ref. 8, needed to assume the existence of non-quantum mechanical elements of physical reality as a justification for a crucial counterfactual inference, even though this assumption is contrary to his program.

A final remark is a correction of a remark about EPR's argument that I made in an earlier publication[10] concerning the phrase "can predict" in EPR's sufficient condition for the existence of an element of physical reality:

> 'can predict' may be used in the strong sense of having data at hand sufficient for the prediction, or in the weak sense of being able to make a measurement that would provide data sufficient for the pre-

diction. The EPR argument goes through only if 'can predict' is used in the weak sense.

The last sentence is incorrect. The first stage of the argument presented in Sect. 2 concerned only systems in ε(RL), in which R is measured, thereby actually providing data sufficient to infer with certainty the outcome of the measurement of L. (As noted previously, it is appropriate to substitute "infer" for "predict."). The second stage of the argument in Sect. 2, which extends the result of the first stage to the whole of the ensemble ε, is accomplished by induction, not by broadening EPR's sufficient condition through a weakening of "can predict (infer)." To be sure, the program of trying to reach EPR's conclusion by substituting counterfactual reasoning for inductive reasoning uses the weak sense of "can predict (infer)," but as noted above, there is at present a gap in the achievement of this program.

References

[1] Albert Einstein, Boris Podolsky, and Nathan Rosen, "Can quantum-mechanical description of physical reality be considered complete?", Physical Review **47**, 777-780 (1935).

[2] GianCarlo Ghirardi and Renata Grassi, "Outcome predictions and property attribution: the EPR argument reconsidered," Studies in the History and Philosophy of Science **25**, 397-423 (1994). Also GianCarlo Ghirardi, "Note added in proof," in R.S. Cohen, Michael Horne, and John Stachel (eds.), *Potentiality, Entanglement and Passion-at-a-Distance* (Kluwer, Dordrecht, 1997), pp. 103-104.

[3] Michael Redhead and Patrick La Rivière, "The relativistic EPR argument," in R.S. Cohen, Michael Horne, and John Stachel (eds.), *Potentiality, Entanglement and Passion-at-a-Distance* (Kluwer, Dordrecht, 1997), pp. 207-215.

[4] David Bohm, *Quantum Theory* (Prentice-Hall, Englewood Cliffs, NJ, 1951).

[5] Niels Bohr, Atomic Physics and Human Knowledge (Science Editions, New York, 1961), pp. 60-61.

[6] Bernard d'Espagnat, *Conceptual Foundations of Quantum Mechanics*, 1st edition (Benjamin, Menlo Park, 1971) 2nd edition (1976) 3rd edition (1989).

[7] David Lewis, *Counterfactuals* (Harvard University Press, Cambridge, MA, 1973).

[8] Henry P. Stapp, “Nonlocal character of quantum theory,” American Journal of Physics 65, 4257-4260 (1997).

[9] Abner Shimony and Howard Stein, “On quantum nonlocality, special relativity, and counterfactual reasoning,” in Abhay Ashtekar, Robert S. Cohen, Don Howard, Juergen Renn, Sahotra Sarkar, Abner Shimony (eds.), *Revisiting the Foundations of Relativistic Physics: Festschrift in Honor of John Stachel* (Kluwer, Dordrecht, 2001). Also, “On Stapp’s ‘Nonlocal character of quantum theory,” American Journal of Physics **69** (2001).

[10] Abner Shimony, *Search for a Naturalistic World View*, vol. 2 (Cambridge University Press, Cambridge, UK, 1973), p. 187.

(Manuscrit reçu le 17 Septembre 2000)