April 14. 1999 LBNL-41813

Comment on "Nonlocality, counterfactuals and quantum mechanics." *

Henry P. Stapp
Lawrence Berkeley National Laboratory
University of California
Berkeley, California 94720

Abstract

A recent proof, formulated in the symbolic language of modal logic, claims to show that contemporary quantum theory, viewed as a set of rules that allow us to calculate statistical predictions among certain kinds of observations, cannot be imbedded in any rational framework that conforms to the principles that (1) the experimenters' choices of which experiments they will perform can be considered to be free choices, (2) outcomes of measurements are unique, and (3) the free choices just mentioned have no backward-in-time effects of any kind. This claim is similar to Bell's theorem, but much stronger, because no reality assumption alien to quantum philosophy is used. The paper being commented upon argues that some such reality assumption has been smuggled in. That argument is examined here and shown, I believe, to be defective.

*This work was supported by the Director, Office of Energy Research, Office of High Energy and Nuclear Physics, Division of High Energy Physics of the U.S. Department of Energy under Contract DE-AC03-76SF00098.

Disclaimer

One of the great lessons of quantum theory is that utmost caution must be exercised in reasoning about hypothetical outcomes of unperformed experiments. Yet Bohr [1] did not challenge the argument of Einstein, Podolsky, and Rosen [2] on the grounds that it was based on the simultaneous consideration of mutually exclusive possibilities. Rather he challenged the underlying EPR presumption that an experiment performed locally on one system would occur "without in any way disturbing " a faraway system. Bohr's own ideas rested heavily on the idea that experimenters could freely choose between alternative possible measurements, and the core of his answer to EPR was that although *"... there is in a case like that just considered no question of a mechanical influence of the system under investigation during the last critical stage of the measuring procedure."*..."there is essentially the question of *an influence of the very conditions which define the possible types of predictions about future behavior of the system."*

The adequacy of Bohr's answer and the nature of his intermediate position on the question of these influences have been much debated. The issue is of fundamental importance, because it concerns the nature of the causal structure of quantum theory, and its compatibility with an idea, drawn from the theory of relativity, that no influence of any kind can act backward in time in any frame.

The background is this. In relativistic classical physical theory the actual physical world is conceived to be one of a host of possible worlds that all obey the same laws of nature. With fixed initial conditions one can, by making a change in the Lagrangian in small space-time region, shift from the actual world to a neighboring possible world, and prove that the effects of this change are confined to times that lie later than the cause in every Lorentz frame. The change in the Lagrangian in the small region can be imagined to alter an experimenter's choice of which experiment he will soon perform in that region.

An analogous result holds in quantum field theory. However, in the quantum case that result is not the whole story: the eventual occurrence of the individual outcome must be described. In that connection, Bohr [3] mentions a discussion at the 1927 Solvay conference as to whether, as Dirac proposed, we should say that we are " concerned with a choice on the part of 'Nature' or, as suggested by Heisenberg, we should say that we have to do with a choice on the part of the 'observer' constructing the measuring instruments and reading their recording." It is just the possible effect of such a choice, made by an experimenter in one region, upon the outcome that appears to the observers located in another region that is the issue here.

The question, more precisely, is this: Is it possible to maintain in quantum mechanics, as one can in classical mechanics, the theoretical idea that the one real world that we experience can be imbedded in a set of possible worlds, each of which obeys the known laws of physics, if the following three conditions hold:

(1), The experimenters can be imagined to be able to freely choose between the different possible measurements that they might perform;

(2), If an experiment is performed then only one of the alternative possible outcomes will appear to observers who witness the outcome; and

(3), No free choice of the kind mentioned in (1) can have any effect on the truth-value of a statement whose truth-value is explicitly specified in terms of outcomes of possible observations that are localized earlier in time in some Lorentz frame.

These three assumptions can be called, "free choice", "unique outcomes", and "locality", respectively. It was shown in reference [4] that if, in a certain Hardy-type experiment, the 100%-certain predictions of quantum theory are assumed to be true in the class of "possible worlds", then imposing the three conditions listed obove leads to a logical contradiction with another prediction of quantum theory.

To obtain rigorous results in this domain it is necessary to formulate arguments within a formal logic, where each separate statement can be stated precisely, and the rules of inference connecting them are spelled out exactly.

A framework has been developed by philosophers and logicians for dealing, in a logically consistent way, with relationships between the real and possible worlds. It is called modal logic. It is designed to formalize in logically coherent rules what we normally mean by statements pertaining to these hypothetical worlds and their connections to the unique actual world. Although there are several versions of modal logic, which differ on fine points [5], they all adhere to certain general rules.

The proof given in [4] follows the general rules of modal logic. However, that does not guarantee that the proof is satisfactory. For modal logic was created by philosophers and logicians within a context in which the actual world and the physical laws that governed it were believed to be basically similar to what was imagined to exist in classical physics. But the quantum world is profoundly different from this classical idealization. Hence the entire question of the appropriate logic must be re-examined in a quantum context, where the very idea of the truth of statements about hypothetical worlds is greatly curtailed relative to classical physics. Utmost care must be taken not to introduce any notion of reality that is contrary to the philosophical principles of quantum theory.

The philosophy of Niels Bohr, as normally understood, allows one to imagine that a free choice made by an experimenter about which experiment he will soon perform would leave undisturbed an outcome that has actually already appeared in some earlier spacetime region to the observers of some other experiment. That notion is one of the ideas that is under scutiny here.

The first locality condition used in the proof expresses this condition. It is called LOC1. It states that if an experiment L2 is actually performed in a spacetime region L, and an experiment R2 is actually performed in a faraway region R that lies later in time than L (in some frame), and if an outcome L2+ actually appears to the observers stationed in the earlier region L, then that same result L2+ would appear to the observers in that earlier region L also in the alternative *possible world* in which everything is left unchanged except for (1), the free choice made later in time by the experimenter in R, and (2), the consequences of that later-in-time change: LOC1 asserts that the later free choice in R has no effect on the outcome that has already appeared earlier to the observers located in region L.

The basic idea of modal logic is that in a possible world W one can make true statements about possible worlds W' that are "hypothetical" relative to W. Some condition C is asserted to hold in these relatively hypothetical worlds W'. Normally, this condition is counterfactual, relative to W: normally, condition C is false in world W. The central problem of modal logic is to specify the necessary and sufficient conditions for the truth in world W of a statement SW of the form: "If C were true then statement S would be true"

Philosphers have identified the intuitive idea expressed by the strong claim that SW, issued from world W, is true. The idea is this: SW is true in W if and only if S is true in every world W' that differs from W only by the minimal changes *needed* to allow condition C to be true.

In our case there is a rigorous way to define this necessary and sufficent condition for the truth of SW that exactly fits both this idea and our logical need.

The condition C, in our case, will be of the form "If the experimenter in spacetime region X chooses to perform experiment E". The correct necessary and sufficient for the statement SW, issued in world W, to be true is that statement S be true in every possible world W' that is identical to W outside the forward lightcone of the spacetime region X, and in which C is satisfied.

This formulation imposes the strong condition that we must not assert that the statement S "would be true" unless its truth follows from our postulates, and that what our postulates entail is only that there be no effect of the change demanded by C outside the forward lightcone of the change in X that changes W to a possible hypothetical world W'.

The argument in reference [4], stated here in words, rather than the symbols of modal logic, begins as follows:

Suppose the actual world W is one in which L2 and R2 are performed and the outcome $g = R2+$ appears to the observers in R. Then a prediction of quantum theory, in the Hardy case under consideration in [4], entails that in the actual world W the outcome actually appearing to the observers in L must be $c = L2+$.

A second property of the Hardy state is that if L2 is performed in L and the outcome appearing in L is L2+ then if R1 is performed in R the outcome appearing in R will be R1-.

Under these conditions one can deduce the truth in world W of the statement SW: "If R1 had been performed in R then outcome observed by the observers in R would be R1-."

The truth of this statement SW in world W (i.e., when issued in world W) follows from the fact that the outcome R1- must appear to the observers in R in every possible world W' in which R1 is performed in R, and which is identical to W outside the forward lightcone from region R: in each such possible world W' the outcome in L is the same as it is in W, namely L2+, and hence a cited property of the Hardy state entails that if R1 were to be performed in R then the outcome appearing there would be R1-.

This argument, stated above in the general modal setting, boils down to the fact that the truth of the statement SW follows from the 100%-certain predictions of quantum theory in the Hardy State, coupled to the locality *assumption*, which asserts that the change from the actual world W in which R2 is measured in the later region R —and outcome R2+ appears in R—to any hypothetical world in which the experimenter in R chooses at the last minute to do something else, namely to measure R1 instead of R2, cannot affect what has already happened earlier in region L. Outcome L2+ must therefore appear in L in W', and hence R1- must appear in R in W'.

[I note, parenthetically, that in the modal context there are statements, like SW, that are true in the base world W, yet contain conditions C that are not true in W: indeed, the whole point of modal logic is precisely to accommodate such situations. This is mentioned because Unruh has claimed, in a report on this Comment on his paper, that my arguments here are flawed because SW cannot be true in W because SW contains a condition that is false in W. That argument fails to take account of the basic idea of modal logic.]

Ordinary logical principles entail that if (A and B) implies C then [(A and B and C) implies D] implies [(A and B) implies D)]. Also [(A and B) implies D] implies [A implies (B implies D)]. Take A to be The Hardy State and LOC1 and L2. Take B to be R2 and R2+. Take C to be L2+. Take D to be SW: "If condition R1 were to hold then R1-"

With these abbreviations, line 5 of my proof reads [A implies (B implies D)].

Abbreviating (B implies D) by SR one obtains line 5 of my proof:

LINE 5: (Hardy and LOC1 and L2) implies SR,

where SR is the statement:

SR: "If R2 is performed in R and outcome R2+ appears to the observers in R, then if R1, instead of R2, had been performed in R the outcome appearing to the observers in R would be R1-."

The form of this claim in line 5 is the same as a typical claim in classical mechanics: if the result of a certain measurement $R2$ is, say, $R2+$, then the deterministic laws of physics may allow one to deduce that if some alternative possible measurement $R1$ had been performed, instead of $R2$, then the result of that measurement $R1$ would necessarily have been $R1-$: knowledge of what happens in an actual experimental situation $R2$ may, with the help of deterministic laws, allow one to infer what would have happened if one had performed, instead, a different experiment $R1$.

Note that no outcome of any unperformed measured is asserted to exist unless that specific outcome is uniquely fixed by the explicitly stated assumptions, which include the assertion that some specific outcome appears to the observers of *some* actually performed experiment: every theoretically specified outcome is tracable to an actual outcome of the actual experiment via the deterministic laws and locality.

Unruh [6] says "...within quantum mechanics attributes do not have values unless they are actually measured."..."in the quantum case one must be extremely careful in carrying out such [counterfactual] arguments, and must ensure that one is not assuming a form of realism—that quantum attributes have values even if they have not been measured—together with the counterfactual discussion." But he immediately continues:

"I assume that such counterfactual statements may legitimately be made in certain circumstances. Given that one has established a correlation of a system A [say L]...with a system B [say R]... then one can make measurements on system B[R], and on the basis of the known correlations, make inferences about system A[L], even if system A[L] has not been directly measured. After all, if such reasoning were disallowed, the whole of von Neumann's argument about measurement would be invalid. In von Neumann's analysis it is precisely the use of correlations of measuring apparatuses with systems that allows us to deduce properties of the systems from measurements made on the measuring apparatus, even though no direct 'measurement' has been made."

He goes on:

"However, great care is required in such counterfactual statements that one does not import into the statements a notion of reality. In particular, the truth of the statement made about system A[L] which relies on measurement made on system B[R] and on correlations which have been established between A[L] and B[R] in the state of the joint system is entirely dependent on the truth of the actual measurement which has been made on system B[R]. To divorce them is to effectively claim that the statement made about A[L] can have a value in and of itself, and independent of measurements which have been made on A[L]. This notion is equivalent to asserting the reality of the statement about A[L] independent of measurements, a position contradicted by quantum mechanics."

These quoted passages are the essence of Unruh's argument: everthing else hinges on them.

My proof is based squarely on the premise that R2 is actually performed and that the outcome R2+ actually appears to the observers stationed in

R. These two conditions, together with the third condition, namely that L2 is actually performed, entail, by virtue of a prediction of quantum theory—accepted as valid in the actual world—that the outcome L2+ actually appears to the observers stationed in L. This entailment is a consequence of exactly the von Neumann type of argument that Unruh has given as an example of reasoning that is valid in the quantum context. So my proof conforms exactly the conditions that Unruh demands, namely that R2 be actually measured and that the outcome R2+ appears to the actual observers of R2. No extra reality assumption is needed at this point, by Unruh's own criteria.

It is precisely by following the rules of modal logic that the reality structure became resolved in just this way that fits perfectly with the quantum requirements specified by Unruh.

Unruh's objections described above pertain to LOC1. But he raises an objection also to LOC2.

He says: "If it were true that one could deduce solely from the fact that a measurement had been made at L that some relation on the right must hold, then I would agree that this requirement [LOC2] would be reasonable."

The other assumptions, including LOC1, are of course needed. But, given those other assumptions, which I have specified, line 5 [L2 implies SR] asserts that SR can be "deduced solely from the fact that" L2 is performed. In particular, one does not need to *assume* that the outcome appearing to the observers stationed in L is L2+. For this fact is deduced from the given fact that R2 is performed and outcome R2+ appears to the observers of R2.

Unruh says that he "would agree that this requirement would be reasonable" if the stated premise were true, i.e., "If it were true that...[L2 implies SR]". As discussed above, he had previously given arguments that led him to believe that premise to be false. But the analysis just concluded shows, I believe, that those criticisms were linked to a misunderstanding of what is taken to be real in my proof and what is purely hypothetical/theoretical. Given the validity of line 5 [L2 implies SR], Unruh's statement acknowledges that LOC2 is all right.

Thus Unruh's objections hinge on his claim that my derivation of line 5 is incorrect.

Immediately after this qualified endorsement of LOC2 Unruh says: "However, if the truth of the relation on the right-hand side depended not only on which measurement had been made on the left, but also on the actual value obtained on the left, then no such locality condition would obtain."

He elaborates: "If it is the value [L2+] obtained on the left... which allows one to deduce [the truth of] the relation [SR] on the right, then [the truth of] that relation [SR] on the right cannot be independent of what is measured on the left, but rather is tied to that measured value. To assume otherwise, to assume that the [truth of the] relations between possible measurements on the right are independent of the values on the left that were used to derive [the truth of] those relations, is, in my opinion, simply another form of realism." [I have inserted the contents of the square brackets to make more precise what I believe Unruh to be saying.]

The logical steps in my proof allow me to replace this condition that outcome L2+ appears to the actual observers of L2 by the condition that outcome R2+ appears to the actual observers of R2, provided L2 is performed. This replacement of the assumption of one actual outcome by another one that entails it, by virtue of the predictions of quantum theory, is exactly the sort of replacement that is involved in von Neumann's analysis of the measurement process. *This* is the only part of the proof that pertains to reality: i.e., to what really happens in the actual world.

The other part, the counterfactual part, pertains to a theoretical (hypothetical) world, that conforms to a counterfactual condition. It may, or may not, be possible to construct a theoretical-hypothetical world that conforms to the locality condition LOC1, but the theoretical conditions pertaining to this hypothetical world are not 'reality' conditions, for they are not about reality. They are about the possibility of imposing a certain kind of locality condition.

On that point I contrast my argument here with (1), the argument of

Einstein, Podolsky, and Rosen, which was designed to show, specifically, that "reality" had elements that were not described by quantum theory, and with (2), hidden-variable arguments that assume from the outset (or on the basis of the Einstein-Podolsky-Rosen argument) that reality contains some extra elements. I maintain strictly, and throughout, an extremely conservative stance with regard to reality: I allow as realities only actually performed experiments, and the outcomes appearing to observers of actually performed experiments. Thus Unruh's claim that I have made an unwarranted 'reality' assumptions does not, I believe, accord with the way that the concept of reality actually enters into my argument. The only realities that I recognize are what the actual experimenters actually do or actually experience: I use no EPR-Bell idea of some "property" lurking behind these realities. And the only theoretical values that I allow are those directly deducible from real outcomes via the *explicitly stated* assumptions.

Von Neumann's analysis is based on arranging different measurements in tandem, so that the observed fact about the outcome of the final measurment in the sequence fixes an earlier fact, without any need to specify explicitly the intermediate facts. My proof follows the same line, making use of LOC1 to shift over, at one stage, from facts about the actual world to theoretical assertions about a class of hypothetical world. The eventual contradiction arises from a conflict between values that follow from the explicitly stated theoretical assumptions.

I do use, in the arguments, the normal conventional relationships between statements involving logical words like "and" and "implies", but not in any way that contravenes quantum philosophy.

There is an ambiguity in the meaning of 'depend upon' in Unruh's assertion that the truth of SR "cannot be independent of... ." What the truth of SR depends upon might mean the basic condition that defines whether SR is true. Or it might mean some particular condition that is sufficient to ensure that this basic condition is satisfied. Or it might mean some third condition that enters into a proof that this basic condition is satisfied under that particular condition.

Unruh's statement uses to the third meaning of "depends upon", whereas the meaning that is needed in my proof is the first meaning.

A key function of logic is to organize the reasoning process so it does not have to carry along the entire proof of a statement in order to give that statement a well defined meaning. Indeed, one generally sets out to prove the truth of some statement without even knowing whether there is a proof. Thus the definition of the condition under which a statement is true needs to be separable from a proof that the statement is true. In proving that line 5 is true what has to be shown is that the *defining condition* for SR to be true holds under condition L2.

The defining condition for a statement to be true is supposed to be specified by the words in that statement: if one abandons that idea then one is, I believe, departing from the realm of rational analysis, as it is normally understood.

In the end, the significance of the proof lies in how it can be used. The purpose of this proof is to place a stringent condition on the possibilities of imbedding contemporary quantum theory—which is a set of practical rules that allow us to calculate statistical predictions about connections between certain kinds of observations—in an rational framework that conforms to the general requirements of "free choices", "unique outcomes", and "locality". I believe that my proof shows, by means of a rational argument completely concordant with quantum philosophy, that no such framework exists, and that the claim made in the paper being commented upon, namely that some hidden reality assumption has been smuggled in, is not supported there by any clear or valid argument.

References

[1] N. Bohr, Phys. Rev. **48**, 696 (1935).

[2] A. Einstein, B. Podolsky, and N. Rosen, Phys. Rev. **47**, 777 (1935).

[3] N. Bohr in Albert Einstein: Philosopher-Scientist. P.A. Schilpp, Tudor, New York, 1951. pg.223.

[4] H.P. Stapp, Am. J. Phys. **65**, 300 (1997).

[5] David Lewis, Philosophical Papers, Vol II, Oxford U.P. (1986); Counterfactuals, Blackwell Press, Oxford, (1973).

[6] W. Unruh, Phys. Rev. A59 126-130 (1999). (quant-ph/9710032).