January 19, 1998 LBNL-41285

Meaning of Counterfactual Statements in Quantum Physics. *

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

Abstract

David Mermin suggests that my recent proof pertaining to quantum nonlocality is undermined by an essential ambiguity pertaining to the meaning of counterfactual statements in quantum physics. The ambiguity he cites arise from his imposition of a certain criterion for the meaningfulness of such counterfactual statements. That criterion conflates the meaning of a counterfactual statement with the details of a proof of its validity in such a way as to make the meaning of such a statement dependent upon the context in which it occurs. That dependence violates the normal demand in logic that the meaning of a statement be defined by the words in the statement itself, not by the context in which the statement occurs. My proof conforms to that normal requirement. I describe the context-independent meaning within my proof of the counterfactual statements in question.

*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

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor The Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial products process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or The Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or The Regents of the University of California and shall not be used for advertising or product endorsement purposes.

Lawrence Berkeley Laboratory is an equal opportunity employer.

David Mermin[1] has nicely summarized and explained the main ideas of my recent proof[2] of the incompatibility of some predictions of quantum theory with a certain formulation of the notion that no causal influence of any kind acts backward in time in any Lorentz frame. However, he avers that I have overlooked an "implicit reference to the future in a statement that explicitly talks only of the past." This claimed implicit reference to the future arises from his imposition of a criterion that he proposes on the meaningfulness of counterfactual statements. I shall argue here that it is not reasonable to impose this constraint on logical argumentation in a quantum context.

The statement in question refers to an experiment L2 that is performed in a region L that lies later in time than a region R in which one of two alternative possible experiments, R1 or R2, might be performed. This statement is Line 5 of my proof, and it reads as follows:

Line 5: If experiment L2 is performed then SR is true.

Here SR is the statement:

Statement SR: If experiment R2 was performed and the outcome was +, then if, instead, experiment R1 had been performed the outcome would have been –.

Line 5 was proved without making any assumption of determinism, counterfactual definiteness, or hidden variables. It was proved using only 1), the requirement that the choices made by the two experimenters can be treated as free variables, 2) a locality condition LOC1 that expresses the requirement that *what happens* at an earlier time cannot be disturbed by what an experimenter freely chooses to do at a later time, 3), the requirement that if any one of the experiments being discussed is actually performed then the pair of outcomes that will appear to observers in the pair of regions L and R will conform to the predictions of quantum theory, and 4), if on the basis of the assumptions just listed one can predict with certainty (i.e., with probability unity) which outcome of some alternative possible m easurement would have appeared if it had been performed then one can say that if that measurement had been performed then that predicted outcome would have appeared.

Assumption 3 asserts that some particular pair of outcomes will appear to the observers in the pair of regions L and R. Thus it excludes many-worlds interpretations. The locality assumption LOC1 assumes that some particular outcome appears in region L, for the actually performed measurement L2, for it is this outcome that is asserted by LOC1 to be independent of which experiment will be freely chosen and performed in region R. This independence condition is supposed to hold in every Lorentz frame, and in particular in frames where region R is later in time than region L.

Assumption 4 has a certain resemblance to the Einstein-Podolsky-Rosen[3] criterion of physical reality. However, unlike that criterion, it does not give a sufficient condition for some hidden element of physical reality to exists even if it is not measured. Rather it affirms that the concept of what "would have appeared" includes the demand that the usual laws of nature "would hold".

An assumption LOC2 was defined as follows:

"If SR is proved true under the condition that L2 is freely chosen in L then SR must be true also under the condition that L1 is freely chosen in L."

Thus LOC2 is simply the assumption that if Line 5 is true then Line 6 is true, where Line 6 is obtained from Line 5 by replacing L2 by L1:

Line 6: If L1 is performed then SR is true.

Notice that LOC2 [unlike LOC1] does not claim that "what happens" in the earlier region is independent of the later choice: it claims rather that the truth of SR (i.e., the validity of the property specified by SR) does not depend on the later free choice made in L.

The structure of my proof is that a set of assumptions leads to a logical contradiction. Thus at least one of the assumptions must fail, and LOC2 seems to be the most likely candidate.

I claim that a failure of LOC2 means that there is some sort of influence backward in time: Mermin, I believe, is attempting to deny this. However, his argument is not simply that, because SR refers to a contrary-to-fact situation, its truth or failure has no physical significance.

I would argue against that simple immediate dismissal by examining the

meanings of the statements under consideration.

Line 5 says that statement SR is true if L2 is performed. This means that, if L2 is performed, then if under the condition that R2 is performed nature produces the result +, then if, instead, R1 had been freely chosen nature would have produced the outcome –.

This result says that if L2 is performed then there is a constraint on nature. This constraint does not specify which outcome nature will produce in R in any single situation. Rather it relates what nature produces in R in an actual situation defined there to what nature would have produced in R in an alternative situation that the experimenter in R could have set up there but did not.

Because we are given, by virtue of the given (for the present discussion) Line 5, the validity of this constraint within nature that pertains to a contrary-to-fact situation in the earlier (in some frame) region R, we cannot just summarily dismiss *all* such connections as meaningless within nature. We must face the question of whether this already established constraint pertaining to nature's possible productions in region R will continue to hold if at the later time the experimenter in L makes the other choice.

If this constraint upon what nature is allowed to produce in R holds if L2 is chosen later, but does not hold if L1 is chosen later, then the free choice to be made later in L is influencing or disturbing nature in R *in some sort of way.*

What argument does Mermin mount against this position?

The essence of Mermin's argument seems to be contained in three consecutive sentences near the end of his section III;

1) You cannot make the inference asserted by the counterfactual conditional if the subsequent choice of experiment on the left turns out to be L1.

2) *A statement about the specific outcome of an experiment at a given time that was not actually performed, may derive its truth or falsity or its very meaning from events in the future that had not yet happened at that*

given time [His italics]

3) The statement (S_2) [the conclusion part of SR (:=S)] and hence the full statement (S) implicitly does refer to a choice taking place on the left at the later time.

Statement 1) is certainly true if the clause"you cannot make the inference" means "you cannot make the inference by using only LOC1 and the predictions of quantum theory". If one adds LOC2 then one can make such inferences. If the intent is to say that any such inference is false, and hence that LOC2 is false, then I would agree that this is probably correct: I agree that LOC2 is probably false.

Statement 2) is true in the sense that the *truth* of statement S in the case that L2 is performed is derived, in our proof, from the fact that L2 is performed. However, the *meaning* of statement S, and its parts, pertains only to possibilities associated with region R, including nature's possible productions of outcomes there. So there is an essential distinction here between the *validity* of a statement and its *meaning*: they should not be conflated in the way done in statement 2).

Statement 3) is a point of clear disagreement. I have specified, above, the *meaning* of SR as a certain constraint on nature's possible productions of outcomes on the region R. This *meaning* does not depend upon which choice some experimenter will make later in L. On the other hand, the *proof* that this property SR holds under the condition that L2 is performed does depend, of course, upon the fact that L2 is performed.

As an example of the usual situation in logical analysis consider two real numbers x and y. Let SR be the statement "x is positive". This statement SR does not refer to y. But the proof that SR is true under the conditions "y is positive" and "x = y" does refer to y, even though SR does not.

Similarly here, the *meaning* of SR does not depend on whether or not it is valid, or on the details of some special context in which it happens to be valid.

Mermin suggest that some sort of problem with the step from Line 5 to

Line 6 arises if one demands conformity to a criterion based on my assertion:

"Theoretical assumptions often allows one say with certainty, on the basis of the outcome of a certain experiment. what would have happened if an alternative possible apparatus had been used."

In conformity with this format, Line 5 says that the theoretical assumptions LOC1 plus QT plus L2 allows one to say with certainty, on the basis of the outcome + of experiment R2, that outcome – would have occurred if apparatus R1 had been used.

To get Line 6 one changes L2 to L1 and adds LOC2.

So my arguments do in this way conform to the format that I specified.

Mermin claims in his abstract that my "conclusion contains an essential ambiguity as regards the meaning of the expression 'statement referring only to phenomena confined to an earlier time.' " But in his text he makes it clear that this ambiguity arises "If all counterfactual statements are to be interpreted according to this criterion...".

The ambiguity that Mermin cites arises, then, directly from his effort to impose a certain criterion for the meaningfulness of contrary-to-fact statements. This criterion was supposed to capture the condition of comformity to the format that I discussed above. But I have already explained how my argument conforms to that format as I conceived it. Mermin proposes another meaning.

Mermin's criterion is essentially that counterfactual statements are meaninful only insofar as they are provable, and that therefore the meaning is tied to the proof, and hence to the context. Given that notion of meaning, the meaning of statement SR in Line 5 involves the fact that L2 was performed, because the proof of the validity of SR in line 5 depends upon the fact that L2 was performed. Within this context-dependent construal of what SR means there is no rationale for the idea that SR should remain true if L2 is replaced by L1. For the fact that L2 is performed is imbedded in the very meaning of SR.

That the meanings of counterfactuals should become context dependent

was certainly no part of anything that I had in mind. I have explained what SR means in the context of my proof. That meaning conforms to the normal demand in logic that the *meaning* of a statement should be specified by the words appearing in the statement itself, not by some particular context in which the statement happens to be provable.

The meaning that I ascribe to SR is not arbitrary or capricious. After all, the introduction of the counterfactual condition was done in the first line of the proof, by the assertion of LOC1. But what did the counterfactual condition mean there? It meant that if nature produced an outcome in a region in a certain actual situation there, then it is logically possible to assume that nature would produce that same outcome in that region under those very same conditions there if the later faraway choice had gone the other way. Line 5 is merely that same assumption transformed by means of two one-hundred-percent certain predictions of quantum theory. These two predictions constitute two other constraints on nature's productions of outcomes. Line 5 is just the result of combining these three constraints on nature's production of outcomes. So its meaning is naturally in terms of such constraints. There is no logical need to introduce a contrary idea of meaning that makes the meaning of statements depend on the context in which they appear.

References

1. N.D. Mermin, "Nonlocal character of quantum theory?", American Journal of Physics, preceding article.

2. H.P. Stapp, "Nonlocal character of quantum theory", American Journal of Physics, **65**, 300-304 (1997).

3. A. Einstein, B. Podolsky, and N. Rosen, "Can quantum mechanical description of physical reality be considered complete?", The Physical Review, **47**, 777-780 (1935).