January 16, 1998 LBNL-41285

Meaning in Quantum Physics of Counterfactual Statements. *

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. The ambiguity he cites arise from his imposition of a certain criterion for the meaningfulness of 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 meaning within my proof of the counterfactual statement 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

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. But he raises an objection to the passage from Line 5 to Line 6 of my proof. That objection arises from his application of a criterion that he imposes on the meaningfulness, and the meaning, of counterfactual statements. I shall argue here that Mermin's criterion constitutes a needless and irrational constraint on logical argumentation.

Line 5 reads as follows:

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

Here S is the statement:

Statement S: 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 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, and 3), the requirement that if any one of the experiments being discussed is performed then nature will produce some particular outcome that conforms to the predictions of quantum theory.

The region L in which the possible experiments L2 and L1 are confined lies, in some frame, at times later than the region R in which of all the (macroscopic) events and hypothetical events referred to in statement S are confined.

LOC2 is 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 S is true.

I explained the meaning of LOC2:

"If a statement S refers only to choices between possible experiments and their possible observable phenomena that are all located in a space-time region lying earlier than some time T, then the truth of this statement S cannot depend upon which experiment is freely chosen and performed in a spatially separated region at a time later than T."

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 existence of the property specified by S does not depend on the later free choice.

Mermin does not appear to object to the logic of my proof. He objects rather to the assumption LOC2, which asserts that Line 6 follows from line 5. Since the conjunction of the various assumptions leads to a logical contradiction, one of these assumptions must fail: that is the point of the proof. I agree that the assumption most likely to fail is LOC2.

Speaking of the presumed failure of LOC2 Mermin says (at the end of his section III):

"This does not mean, however, that a choice made in the future (whether to measure L1 or L2 on the left) can influence events in the the present. It only means that *a choice made in the future can influence [the truth of] statements about events in the present that might have happened but did not.*" [His italics]

That characterization of the meaning of the failure of LOC2 is correct. There is no disagreement on that point. The failure of LOC2 asserts only that the validity of the property specified by S depends of the future free choice. And S is a statement whose truth depends on the truth of a counterfactual assertion. LOC2 does not claim that *what actually happens* is independent of the future free choice. [LOC1 that makes that sort of assumption.] This character of LOC2 is why I formulated my main conclusions in terms of "truths of statements", or of the "present facts" specified by such statements, rather than in terms of conditions on "events", or "what happens".

So where is the disagreement? We agree that LOC2 probably fails, and we agree that the failure of LOC2 means that the truth of the conclusion in statement S is not independent of the later free choice. And we agree that this conclusion is a contrary-to-fact statement.

The disagreement is, I believe, about the significance of the failure of LOC2.

I claim that a failure of LOC2 means 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 S 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 S 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 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 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 possibilties 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 S 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 S 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 S be the statement "x is positive". This statement S does not refer to y. But the proof that S is true under the conditions "y is positive" and "x = y" does refer to y, even though S does not.

Similarly here, the *meaning* of S does not depend on whether or not it is validity, or 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 m eaning, the meaning of statement S in Line 5 involves the fact that L2 was performed, because the proof of the validity of S in line 5 depends upon the fact that L2 was performed. Within this context-dependent construal of what S means there is no rationale for the idea that S should remain true if L2 is replaced by L1. For the fact that L2 is performed is imbedded in the very meaning of S.

That the meanings of counterfactuals should become context dependent was certainly no part of anything that I had in mind. I have explained what S 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 S 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.

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).