April 13, 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

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.

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

SR: If experiment R2 was performed in region R and the outcome + appeared to the observers in R then if, instead, experiment R1 had been performed in region R the outcome – would have appeared to the observers in R.

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 appears to observers at an earlier time cannot depend upon what some 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 a unique pair of outcomes will appear to observers in the pair of regions L and R; 4) the demand that this pair of outcomes will conform to the predictions of quantum theory; and 5) the demand that 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 measurement would have appeared to the observers of the outcome if that measurement had been performed then one can assert that if that measurement had been performed then that predicted outcome would have appeared to those observers.

Assumption 3 asserts that if some particular pair of measurements is actually performed in the pair of regions (L,R) then some one outcome will appear to the observers in L and some one outcome will appear to the observers in R. This assumption excludes many-worlds interpretations, in which each possible outcome appears to some observer. The locality assumption LOC1 presupposes that if the measurement L2 is actually performed then some one particular outcome must appear to the observers in region L, for it is this particular (though perhaps non-predetermined) 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 at least in some Lorentz frame in which region R is definitely later in time than region L. Lorentz invariance is thus not explicitly assumed. Assumption 5 has a certain resemblance to the Einstein-Podolsky-Rosen[3] criterion of physical reality. However, unlike that criterion, it does not propose a sufficient condition for some hidden element of physical reality to exist even though 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, instead of L2, is freely chosen in L."

Thus LOC2 is the claim that the assertion that Line 5 is true can be replaced by the assertion that 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 appears" 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. LOC2 seems to be the candidate most likely to fail.

I claim that a failure of LOC2 would mean 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 conditions that R2 is performed in region R and the outcome appearing to the observers in region R is +, then if R1, instead of R2, had been freely chosen in region R the outcome appearing to the observers in region R would have been –."

This asserts that if L2 is performed then there is a constraint on what appears to observers. This constraint does not specify which outcome will appear to the observers in region R in any single situation. Rather it connects what might appear to the observers in R in the actual situation defined there to what, in that case, would have appeared to the observers in R in an alternative situation that the experimenter in region 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 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. We must face the question of whether this already established constraint pertaining to what can appear in region R at an earlier time would continue to hold if at the later time the experimenter in L were to make the other free choice.

If this constraint upon what can appears to observers 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 *in some way* the possibilities for what can appear to the observers in region R.

What argument does Mermin mount against this?

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 the (trial) assumption LOC2 is probably false.

Statement 2) is true in the sense that the *truth or falsity* of statement S in the case that L2 is performed is derived in my proof from the fact that L2 is performed. However, the *meaning* of statement S, and of its parts, pertains only to possibilities for outcomes that might appear to the observers in region R. Hence there is an essential distinction here between the *truth or falsity* of the 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 relating an outcome that could have appeared to the observers in region R in the actual experiment performed there to what could then have appeared to those same observers if, instead, a certain alternative possible experiment had been performed there. This *meaning* does not depend upon which choice some other experimenter will make later in L. On the other hand, the *proof* that this property SR is true under the condition that L2 is performed does, of course, depend 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 true, or on the details of some special context in which it happens to be true..

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 if if the alternative measurement R1 had been performed then the outcome (–) would have appeared.

To get Line 6 I invoked the further (trial) theoretical assumption LOC2 to say that this same relationship between possible outcomes in region R continues to hold if the apparutus corresponding to L2 is replaced by the apparatus corresponding to L1.

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, according to this construal, 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 does conform 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 reference to some special conditions under which the statement can be proved to be true.

In summary, I believe that the argument should not be judged not on the basis of an arbitrary criterion that makes the meaning of statements depend on the context in which they appear. It should be judged rather on its meaning as expressed directly by the words in the statements occurring in the proof: it is within this meaning that the import of the proof lies.

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).