Counterfactuals in Quantum Mechanics

Lev Vaidman

Counterfactuals in quantum mechanics appear in discussions of (a) ▶ nonlocality, (b) pre- and post-selected systems, and (c) ▶ interaction-free measurement; Quantum interrogation. Only the first two issues are related to counterfactuals as they are considered in the general philosophical literature:

> *If it were that $\mathcal{A}$, then it would be that $\mathcal{B}$.*

The truth value of a counterfactual is decided by the analysis of similarities between the actual and possible counterfactual worlds [1].

The difference between a counterfactual (or counterfactual conditional) and a simple conditional: *If $\mathcal{A}$, then $\mathcal{B}$,* is that in the actual world $\mathcal{A}$ is not true and we need some "miracle" in the counterfactual world to make it true. In the analysis of counterfactuals out of the scope of physics, this miracle is crucial for deciding whether $\mathcal{B}$ is true. In physics, however, miracles are not involved. Typically:

> $\mathcal{A}$: *A measurement $\mathcal{M}$ is performed*

> $\mathcal{B}$: *The outcome of $\mathcal{M}$ has property $\mathcal{P}$.*

Physical theory does not deal with the questions of which measurement and whether a particular measurement is performed? Physics yields conditionals: "If $\mathcal{A}_i$, then $\mathcal{B}_i$". The reason why in some cases these conditionals are considered to be counterfactual is that several conditionals with incompatible premises $\mathcal{A}_i$ are considered with regard to a single system.

The most celebrated example is the Einstein–Podolsky–Rosen (▶ EPR problem) argument in which incompatible measurements of the position or, instead, the momentum of a particle are considered. Stapp has applied a formal calculus of counterfactuals to various EPR-type proofs [2,3] and in spite of extensive criticism [4–9], continues to claim that the nonlocality of quantum mechanics can be proved without the assumption "reality" [10].

Let me give here just the main point of this controversy. Stapp provides elaborate arguments in which an *a priori uncertain* outcome of a measurement of O in one location might depend on the measurements performed on an entangled quantum particle in another location. But if *anything* is different in a counterfactual world, the outcome of the measurement of O need not be the same as in the actual world. The core of the difficulty is this randomness of the outcomes of quantum measurements. The formal philosophical analysis of counterfactuals which uses similarity criteria, presupposes that in a counterfactual world which is identical to the actual world in all relevant aspects up until the measurement of O, the outcome has to be the same. Thus, Stapp's analysis tacitly adopts the *counterfactual definiteness* [4,5] which is essentially equivalent to "reality" or ▶ hidden variables and which is absent in the conventional quantum theory.

Important examples of quantum counterfactuals are *elements of reality*. Consider the following *definition* [11]:

If we can *infer* with certainty that the result of measuring at time t of an observable O is o, then, at time t, there exists an element of reality $O = o$.

C

If we consider several elements of reality which cannot be verified together, we obtain counterfactuals. A celebrated example is the Greenberger–Horne–Zeilinger (▶ GHZ) entangled state of three spin-$\frac{1}{2}$ particles [4, 13]:

$$|\Psi\rangle = \frac{1}{\sqrt{2}}(|\uparrow\rangle_A|\uparrow\rangle_B|\uparrow\rangle_C - |\downarrow\rangle_A|\downarrow\rangle_B|\downarrow\rangle_C). \tag{1}$$

We consider spin component measurements of these three particles in the x and y directions. The counterfactuals (the elements of reality) have a more general form than merely "the value of O is o", they are properties of a set of three measurements:

$$\begin{aligned}
\{\sigma_{Ax}\}\{\sigma_{Bx}\}\{\sigma_{Cx}\} &= -1, \\
\{\sigma_{Ax}\}\{\sigma_{By}\}\{\sigma_{Cy}\} &= 1, \\
\{\sigma_{Ay}\}\{\sigma_{Bx}\}\{\sigma_{Cy}\} &= 1, \\
\{\sigma_{Ay}\}\{\sigma_{By}\}\{\sigma_{Cx}\} &= 1.
\end{aligned} \tag{2}$$

Here $\{\sigma_{Ax}\}$ signifies the outcome of a measurement of σ_x of particle A, etc. Since one cannot measure for the same particle both σ_x and σ_y at the same time, this is a set of counterfactuals. It is a very important set because no local hidden variable theory can ensure such outcomes with certainty; there is no solution for the set of equations (2).

Lewis's theory of counterfactuals is asymmetric in time [14]. The counterfactual worlds have to be identical to the actual world during the whole time before $\mathcal{A}$, but not after. This creates difficulty in applications of counterfactuals to physics and especially to quantum mechanics because "before" and "after" are not absolute concepts. Different Lorentz observers might see different time ordering of measurements performed at different places. Finkelstein [15] and Bigaj [16] have attempted to define time asymmetric counterfactuals to overcome this difficulty. But in my view, the time asymmetry of quantum counterfactuals is an unnecessary burden [17]. We *can* consider a time symmetric (or time neutral) definition of quantum counterfactuals.

The general strategy of counterfactual theory is to find counterfactual worlds closest to the actual world. In the standard approach, the worlds must be close only before the measurement. In the time-symmetric approach, the counterfactual worlds should be close to the actual world both before and after the measurement at time t. Quantum theory allows for a natural and non-trivial definition of "close" worlds as follows: *all outcomes of all measurements performed before and after the measurement of O at time t are the same in the actual and counterfactual worlds.*

A peculiar example of time symmetric counterfactuals is the *three box paradox* [18]. Consider a single particle prepared at time t_1 in a ▶ superposition of being in three separate boxes:

$$|\Psi_1\rangle = \frac{1}{\sqrt{3}}(|A\rangle + |B\rangle + |C\rangle). \qquad (3)$$

At a later time t_2 the particle is found in another superposition:

$$|\Psi_2\rangle = \frac{1}{\sqrt{3}}(|A\rangle + |B\rangle - |C\rangle). \qquad (4)$$

For this pre- and post-selected particle, a set of counterfactual statements, which are *elements of reality* according to the above definition, is:

$$\mathbf{P}_A = 1,$$
$$\mathbf{P}_B = 1. \qquad (5)$$

Or, in words: if we open box A, we find the particle there for sure; if we open box B (instead), we also find the particle there for sure. Indeed, not finding the particle in box A (or B) collapses the pre-selected state (3) to a state which is orthogonal to the post-selected state (4).

Beyond these counterfactual statements, there are numerous manifestations of the claim that in some sense, this single particle is indeed in two boxes simultaneously. A single photon which interacts with this particle scatters as if there are two particles: one in A and one in B, but two or more photons (▶ light quantum) do not "see" two particles. Many photons see this single particle as two particles if the photons interact weakly with the particle. Indeed, there is a useful theorem which says that if a strong measurement of an observable O yields a particular outcome with probability 1, (i.e. there is an element of reality) then a weak measurement yields the same outcome. Sometime this is called a *weak-measurement element of reality* [19]. The outcomes of weak measurements are *weak values* (▶ weak value and weak measurements):

$$(\mathbf{P}_A)_w = 1,$$
$$(\mathbf{P}_B)_w = 1. \qquad (6)$$

Contrary to the set of counterfactuals above, the weak measurements can be performed simultaneously both in box A and box B. Thus, the existence of counterfactuals helps us to know the outcome of real (weak) measurement.

The three-box paradox and other time-symmetric quantum counterfactuals have raised a significant controversy [11, 20, 21, 21–28]. It seems that the core of the controversy is that quantum counterfactuals about the results of measurements of ▶ observables, and especially "elements of reality" are understood as attributing values to observables which are not observed. But this is completely foreign to quantum mechanics. Unperformed experiments have no results! "Element of reality" is

just a shorthand for describing a situation in which we know with certainty the outcome of a measurement *if* it is to be performed, which in turn helps us to know how weakly coupled particles are influenced by the system. Having "elements of reality" does not mean having values for observables. The semantics are misleading since "elements of reality" are not "real" in the ontological sense.

An attempt to give counterfactuals some ontological sense, at the cost of placing artificial constraints on the context in which counterfactuals are considered, was made by Griffiths [29]. He showed that counterfactuals have no paradoxical features when only ▶ *consistent histories* are considered. Another recent step in this direction are quantum counterfactuals in very restrictive "measurement-ready" situations [30].

Penrose [31] used the term "counterfactuals" in a very different sense:

> *Counterfactuals* are things that might have happened, although they did not in fact happen.

In interaction-free measurements [32], an object is found because it might have absorbed a photon, although actually it did not. This idea has been applied to "counterfactual computation" [33], a setup in which the outcome of a computation becomes known in spite of the fact that the computer did not run the algorithm (in case of one particular outcome [34]).

In the framework of the ▶ Many-Worlds Interpretation, Penrose's "counterfactuals" are counterfactual only in one world. The physical Universe incorporates all worlds, and, in particular, the world in which Penrose's "counterfactual" is actual, the world in which the "counterfactual" computer actually performed the computation.

Literature

1. D. Lewis: *Counterfactuals*. Oxford, Blackwell (1973).
2. H.P. Stapp: *S-Matrix interpretation of quantum theory*. Phys. Rev. D 3, 1303 (1971).
3. H.P. Stapp: *Nonlocal character of quantum theory*. Am. J. Phys. 65, 300 (1997).
4. B. Skyrms: *Counterfactual definiteness and local causation*. Phil. Sci. 49, 43 (1982).
5. M. Redhead: *Incompleteness, Nonlocality, and Realism: A Prolegomenon to the Philosophy of Quantum Mechanics*. New York, Oxford University Press (1987).
6. R.K. Clifton, J.N. Butterfield, M. Redhead: *Nonlocal influences and possible worlds – a Stapp in the wrong direction*. Br. J. Philos. Sci. 41, 5 (1990).
7. D. Mermin: *Nonlocal character of quantum theory?* Am. J. Phys. 66, 920 (1998).
8. W. Unruh: *Nonlocality, counterfactuals, and quantum mechanics*. Phys. Rev. A 59, 126 (1999).
9. A. Shimony, H. Stein: *Comment on Nonlocal character of quantum theory*, Am. J. Phys. 69, 848 (2001).
10. H.P. Stapp: *Comments on Shimony's an analysis of Stapp's 'a Bell-type theorem without hidden variables'*, Found. Phys. 36, 73 (2006).
11. L. Vaidman: *The meaning of elements of reality and quantum counterfactuals*: Reply to Kastner. Found. Phys. 29, 856 (1999).

12. D.M. Greenberger, M.A. Horne, A. Zeilinger: *Going beyond Bell's theorem*. In Bell Theorem, Quantum Theory and Conceptions of the Universe, M. Kafatos, ed., p. 69, Dordrecht, Kluwer, (1989).
13. N.D. Mermin: *Quantum mysteries revisited*. Am. J. Phys. 58, 731 (1990).
14. D. Lewis: *Counterfactual dependence and time's arrow*. Nous 13, 455 (1979).
15. J. Finkelstein: *Space-time counterfactuals*. Synthese 119, 287 (1999).
16. T. Bigaj: *Counterfactuals and spatiotemporal events*. Synthese 142, 1 (2004).
17. L. Vaidman: *Time-symmetrized counterfactuals in quantum theory*. Found. Phys. 29, 755 (1999).
18. Y. Aharonov, L. Vaidman: *Complete description of a quantum system at a given time*. J. Phys. A 24, 2315 (1991).
19. L. Vaidman: *Weak-measurement elements of reality*. Found. Phys. 26, 895 (1996).
20. W.D. Sharp, N. Shanks: *The rise and fall of time-symmetrized quantum mechanics*. Philos. Sci. 60, 488 (1993).
21. R.E. Kastner: *Time-symmetrised quantum theory, counterfactuals and 'advanced action'*. Stud. Hist. Philos. Mod. Phy. 30 B, 237 (1999).
22. L. Vaidman: *Defending time-symmetrised quantum counterfactuals*. Stud. Hist. Philos. Mod. Phy. 30 B, 337 (1999).
23. R.E. Kastner: *The three-box paradox and other reasons to reject the counterfactual usage of the ABL rule*. Found. Phys. 29, 851 (1999).
24. R.E. Kastner: *The nature of the controversy over time-symmetric quantum counterfactuals*. Phil. Sci. 70, 145 (2003).
25. L. Vaidman: (2003) *Discussion: Time-Symmetric Quantum Counterfactuals*. e-print: PITT-PHIL-SCI000001108 (2003).
26. U. Mohrhoff: *Objective probabilities, quantum counterfactuals, and the ABL rule* A response to R. E. Kastner. Am. J. Phys. 69, 864 (2001).
27. K.A. Kirkpatrick: *Classical three-box 'paradox'*. J. Phys. A 36, 4891 (2003).
28. T. Ravon, L. Vaidman: *The three-box paradox revisited*. J. Phys. A 40, 2882 (2007).
29. R.B. Griffiths: *Consistent quantum counterfactuals*. Phys. Rev. A 60, R5 (1999).
30. D.J. Miller: *Counterfactual reasoning in time-symmetric quantum mechanics*. Found. Phys. Lett. 19, 321 (2006).
31. R. Penrose: *Shadows of the Mind*. Oxford, Oxford University Press (1994).
32. A.C. Elitzur, L. Vaidman: *Quantum mechanical interaction-free measurements*. Found. Phys. 23, 987 (1993).
33. G. Mitchison, R. Jozsa: *Counterfactual Computation*. Proc. R. Soc. Lond. A 457, 1175 (2001).
34. L. Vaidman: *Impossibility of the counterfactual computation for all possible outcomes*. Phys. Rev. Lett. 98, 160403 (2007).

Covariance

K. Mainzer

Covariance means form invariance, i.e. the form of a physical law is unchanged (invariant) with respect to transformations of reference systems. Covariance can be distinguished from ▶ invariance which refers to quantities and objects [2]. The covariant formulation of laws implies that the form of laws is independent of the state of motion in a reference system that an observer takes. In that sense, all fundamental