CDD: 530.12

BELL'S THEOREM AND THE COUNTERFACTUAL DEFINITION OF LOCALITY

OSVALDO PESSOA JR.

Department of Philosophy
University of São Paulo
Av. Prof. Luciano Gualberto, 315
05508-900 SÃO PAULO, SP
BRAZIL

opessoa@usp.br

Received: 15.09.2009; Accepted: 31.11.2009

Abstract: This paper proposes a solution to the problem of non-locality associated with Bell's theorem, within the counterfactual approach to the problem. Our proposal is that a counterfactual definition of locality can be maintained, if a subsidiary hypothesis be rejected, "locality involving two counterfactuals". This amounts to the acceptance of locality in the actual world, and a denial that locality is always valid in counterfactual worlds. This also introduces a metaphysical asymmetry between the factual and counterfactual worlds. This distinction is analogous to what occurs in the derivations of Bell's theorem which assume hidden-variables, where macroscopic locality can be maintained at the price of rejecting outcome independence. This can be interpreted as non-locality at the level of potentialities, which might be identified with the non-locality of counterfactual worlds. Our solution, presented for the CHSH inequality, is falsifiable, and we test it with two other setups, Bell's original inequality and the EPR thought-experiment.

Keywords: Bell's theorem. Counterfactuals. Non-locality. Metaphysical asymmetry. Potentialities. EPR thought-experiment.

O TEOREMA DE BELL E A DEFINIÇÃO CONTRAFACTUAL DE LOCALIDADE

Resumo: Propõe-se uma solução ao problema da não-localidade associada ao teorema de Bell, dentro da abordagem contrafactual à questão. A proposta é que uma definição contrafactual de localidade pode ser mantida, se uma hipótese subsidiária for rejeitada, a "localidade envolvendo dois contrafactuais". Isso equivale à aceitação da localidade no mundo factual, mas à rejeição de que a localidade é sempre válida em mundos contrafactuais. Isso também introduz uma assimetria metafísica entre o mundo factual e os mundos contrafactuais. Tal distinção é análoga ao que ocorre nas derivações do teorema de Bell que supõem variáveis ocultas, onde a localidade macroscópica pode ser mantida às custas da rejeição da hipótese da "independência de resultados". Esta hipótese pode ser interpretada como uma não-localidade ao nível das potencialidades, o que poderia ser identificada com a não-localidade dos mundos contrafactuais. Nossa solução, apresentada para a desigualdade de CHSH, é falseável, e ela é testada com duas outras montagens, a desigualdade original de Bell e o experimento mental de EPR.

Palavras chave: Teorema de Bell. Contrafactuais. Não-localidade. Assimetria metafísica. Potencialidades. Experimento mental de EPR.

1. BELL'S THEOREM WITH HIDDEN VARIABLES

The most important result in the philosophy of physics, at least since 1952, has been Bell's theorem. John Stuart Bell (1964) derived an inequality, involving pairs of correlated quanta, which assumes certain hypotheses, such as the existence of hidden variables λ and the restriction that they act locally. The interest in this result is that quantum theory predicts that this inequality may be violated, and subsequent experiments confirmed this violation (Clauser & Shimony, 1978; Aspect & Grangier, 1986). In other words, Bell's theorem imposes severe limits on the class of realist interpretations of quantum theory known as *local hidden-variable theories*.

As a consequence, at least one of the assumptions used in building such local hidden-variable theories and deriving the inequalities must be rejected. Basically three assumptions are used (d'Espagnat, 1979).

1) *Realism.* A hidden-variables theory stipulates that the outcomes of every measurement have well-defined (albeit hidden) causes, existing at the time of measurement, which determine the outcomes in a unique way or only stochastically.

2) *Locality.* The influence of a hidden parameter on the outcome of a measurement cannot propagate faster than the speed of light. Thus, a measurement performed on the Earth cannot be influenced instantaneously by hidden variables in the star Sirius.

3) *Induction.* When experimentally testing Bell's inequality, a large set of pairs of particles must be simultaneously measured with three or four different settings of the macroscopic apparatus. One must assume that the measurement performed on each pair of particles is independent of what happens in any other measurement, in order to guarantee a fair sampling.

Which of the assumptions should be rejected? Realism could be maintained by abandoning locality, as was done by David Bohm (1952) in his "causal" interpretation. Or alternatively, the hypothesis that one may ascribe "elements of reality" to something other than observable events may be seen as the culprit, as was done by the orthodox interpretation as a response to Einstein, Podolsky & Rosen (1935) (for a brief response, see Pauli, 1949). Induction is often neglected, but it could play an essential role (Leggett, 1987, p. 880), as could other hypotheses related to loopholes in experimental tests (see review by Lalöe, 2001, pp. 672-4).

One additional assumption not spelled out above may be called *determinism in measurements*: the set of hidden variables at a certain instant determines *in a unique way* the measurements of *any* observable. It is reminiscent of Leibniz's principle of sufficient reason: something can only exist (in our case, a measurement outcome) if there is a cause that makes it exist. A denial of this assumption leads to *stochastic hidden variable theories*, for which the hidden parameters furnish only probabilities for different outcomes. Around 1974, it was proved that a version of Bell's theorem also applies to such stochastic theories.

This is remarkable because standard quantum mechanics itself may be considered a stochastic hidden variable theory. Quantum mechanics furnishes probabilities based on the *state* or wavefunction $\psi(\mathbf{r})$ of a physical system, and the parameters defining this state may be taken to be the hidden variables of the stochastic realist interpretation. The realist assumption mentioned above is considerably weakened, amounting, in this case, merely to the thesis that the parameters defining a wavefunction correspond to something in reality. If one accepts this *weak realism* and the assumption of induction, the trilemma leads to the conclusion that quantum mechanics must be *non-local*. What does this amount to?

Around 1984, an important distinction was made between two types of non-locality, "parameter independence" and "outcome independence" (see discussion in Shimony, 1993). The first is what we will call *macroscopic locality*: for two separated but correlated particles, the probability of obtaining a result for a measured observable in the first particle *is independent of what observable* is being simultaneously measured at the other particle. The second assumption, outcome independence, asserts that the probability for the first particle *is independent of what result* is obtained in the simultaneous measurement of the second particle.

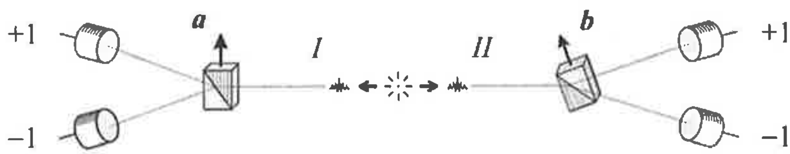
Granting this analysis, the assumptions involved in Bell's theorem are now: weak realism, induction, macroscopic locality and outcome independence. The consensus between those who accept weak realism is that *outcome independence* should be abandoned. In other words, accepting some kind of reality associated to the quantum mechanical state, we must accept that the outcome associated to one particle (say on a planet of Sirius) is instantaneously correlated with outcomes of experiments in a far away location (such as on the Earth), even if in an uncontrollable way.

Certain authors, especially Henry Stapp (1985), have derived Bell's theorem without assuming hidden-variables, but only making assumptions about *counterfactual* situations. With this, he concluded that non-locality could be proved without any assumption about realism, but this claim of generality has been convincingly criticized, on the grounds that he implicitly assumes determinism in measurements (Clifton, Butterfield & Redhead, 1990; Dickson, 1993). But it is still instructive to consider his derivation, taking into account all assumptions involved.

2. BELL'S THEOREM WITH COUNTERFACTUALS

There are different versions of the inequality associated with Bell's theorem, and different ways of deriving it. The version which will be now presented is known as the "CHSH inequality", due to Clauser, Horne, Shimony & Holt (see Clauser & Shomony, 1978, pp. 1889-90, or Redhead, 1987, pp. 82-6).

Instead of following the usual derivation, which assumes the existence of hidden-variables, we will consider the derivation which is based on hypotheses involving *counterfactuals*, as done by Stapp (1985). As our experimental setup, consider the generation of entangled pairs of photons, which are detected in opposite sides of the source. On each side, the polarization can be measured by means of prisms oriented at any desired angle (in the plane orthogonal to the paths of the photons), followed by two detectors, an upper one and a lower one (see Figure below).



Consider a certain pair *k* of detected photons. Suppose that the prism on the left-hand side of the apparatus is oriented at an angle ***a***. In this case, the observable being measured on this side is denoted by $I(a)_k$. If a photon is detected in the upper photocell, we assign the value +1 to this observable; if it is detected in the lower detector, the value attributed to the outcome of the measurment of $I(a)_k$ is –1. Contrary to the case of in which hidden-variables are assumed, here $I(a)_k$ does not denote a preexisting possessed value of the system, but simply the macroscopic outcome of the measurement. In this sense, the counterfactual approach has a more operational flavor, although it does work with the non-operational notion of possible measurement outcome.

Now, instead of orienting the prism at angle ***a***, we could have put it at angle ***a'***. If this setting had been fixed for pair *k*, then we would have obtained a value for observable $I(a')_k$, which could be either +1 or –1.

The same thing would happen for the observable measured on the right-hand side of the apparatus: for angles $\boldsymbol{b}$ or $\boldsymbol{b'}$, we would measure observables $II(b)_k$ or $II(b')_k$, respectively.

To derive the CHSH inequality, we write out the following expression involving products of measurement results obtained on each pair k of photons: $I(a)_k \cdot II(b)_k + I(a)_k \cdot II(b')_k + I(a')_k \cdot II(b)_k - I(a')_k \cdot II(b')_k$. The difficulty is that only a single experiment can be performed on pair k, so that if $I(a)_k$ and $II(b)_k$ are measured, the other observables are not. Let us then spell out the following assumptions.

i) Suppose that the actual experiment for pair k involved prisms at angles $\boldsymbol{a}$ and $\boldsymbol{b}$, furnishing values $I(a)_k$ and $II(b)_k$.

ii) Assume that if the prism at the right had been set at angle $\boldsymbol{b'}$, then the value obtained at the left would still be $I(a)_k$ (the same value as in *i*), while that at the right the outcome would be some $II(b')_k$.

iii) Symmetrically to this previous supposition, assume that if the prism at the left had been set at angle $\boldsymbol{a'}$, then the value obtained at the right would still be $II(b)_k$ (the same value as in *i*), while that at the left the outcome would be some $I(a')_k$.

iv) Assume that if the prisms were set at angles $\boldsymbol{a'}$ and $\boldsymbol{b'}$, then the outcomes obtained would have been $I(a')_k$ (the same as in *iii*) and $II(b')_k$ (the same as in *ii*), respectively.

Accepting these assumptions, it follows that the following relation holds for pair k:

$$I(a)_k \cdot II(b)_k + I(a)_k \cdot II(b')_k + I(a')_k \cdot II(b)_k - I(a')_k \cdot II(b')_k = \pm 2. \qquad (1)$$

To see this, just rewrite the left-hand side of the equation as $I(a)_k \cdot (II(b)_k + II(b')_k) + I(a')_k \cdot (II(b)_k - II(b')_k)$, and notice that either $(II(b)_k + II(b')_k)$ or $(II(b)_k - II(b')_k)$ is zero, while the other has value $+2$ or -2 (Redhead, 1987, p. 84).

Taking the mean value of eq.(1) for all the particles in the experiment, it is easy to see that it must lie between –2 and +2 (since for some pairs we get –2 and for others +2):

$$\langle I(a)_k \cdot II(b)_k \rangle + \langle I(a)_k \cdot II(b')_k \rangle + \langle I(a')_k \cdot II(b)_k \rangle - \langle I(a')_k \cdot II(b')_k \rangle \leq 2. \quad (2)$$

Quantum mechanics predicts that there are cases that violate ineq.(2), such as for the "singlet state", expressed in the following way:

$$\tfrac{1}{\sqrt{2}} \cdot \left|\varphi_0\right\rangle_1 \otimes \left|\varphi_{90}\right\rangle_2 \; - \; \tfrac{1}{\sqrt{2}} \cdot \left|\varphi_{90}\right\rangle_1 \otimes \left|\varphi_0\right\rangle_2 , \quad (3)$$

where $|\varphi_0\rangle$ and $|\varphi_{90}\rangle$ are eigenstates for linear polarization. In this optical case, the mean value is given by $\langle I(a)_k \cdot II(b)_k \rangle = -\cos 2(\boldsymbol{b}-\boldsymbol{a})$, so for certain angles $\boldsymbol{a}$, $\boldsymbol{a'}$, $\boldsymbol{b}$, $\boldsymbol{b'}$ the left-hand side of the above inequality is greater than 2. Experiments confirm the predictions of quantum mechanics, so one or more of the assumptions used in the derivation of ineq.(2) must be abandoned!

What assumptions have we used?

a) Counterfactual definiteness. If an experiment has not been performed, we can still be sure that if it had been made, definite values would have been obtained.

b) Locality (in counterfactual language). For a measurement performed at a certain location at time *t*, if a remote piece of apparatus were (counterfactually) modified, that modification would not alter the value obtained for the measurement at time *t*. This hypothesis (together with hypothesis *a*) justifies assumptions *ii* and *iii* above.

c) Locality involving two counterfactuals. For a counterfactual measurement, that is, for an experiment not performed at a specific location (but which could have been performed), the value that would have been obtained, with a certain setting of a remote piece of apparatus, would not be altered by a further counterfactual modification in the remote apparatus. This justifies assumption *iv* above.

d) Induction. In comparing ineq.(2) with the predictions of quantum mechanics, one must assume that the experimental samples represent in a fair way the ensembles considered by the theory.

The novel proposal of this paper is the separation of hypotheses *b* and *c*, which are usually lumped together under the name "locality" (Stapp, 1985; Leggett, 1986). By doing so, we can reject hypothesis *c* without abandoning locality (hypothesis *b*), counterfactual definiteness, or induction. Once again, we stress that no claim is being made that the counterfactual approach is valid for any interpretation of quantum theory, as suggested by Stapp. Our point is that hypotheses *a*, *b*, and *d* may be sustained, if only hypothesis *c* is rejected. We will now examine the metaphysical consequences of our proposal and establish a criterion for its refutation.

3. METAPHYSICS OF POTENTIALITIES AND COUNTERFACTUALS

What we have proposed is more or less analogous to what was described in section 1, where macroscopic locality was maintained at the cost of rejecting outcome independence. One way of picturing outcome *dependence* is to interpret the quantum-mechanical state in a realistic way, as a *potentiality* (Bohm, 1951, pp. 132-3; Margenau, 1954; Heisenberg, 1958, p. 54; Redhead, 1987, pp. 48-9). When a measurement is performed on one of the particles of the correlated pair, an instantaneous collapse of the entangled state takes place, modifying the state of the other particle, and ensuring perfect anticorrelation (assuming eq. 3). The non-local state collapse is an example of violation of outcome independence. One may then say that locality is maintained at the level of actualities (macroscopic locality), but *not at the level of potentialities.*

In the counterfactual approach, the analysis is analogous. One can maintain locality (hypothesis *b*) if one rejects "locality involving two counterfactuals" (hypothesis *c*), which amounts to the assertion that, in a counterfactual world, one cannot assume locality. This introduces an *asymmetry* between the factual world, where locality (defined by recourse to a counterfactual modification at a distance) is valid, and counterfactual worlds, where locality cannot be always valid. This asymmetry between reality and possibility makes intuitive sense (there must be *some* difference between both), but it is generally ignored in metaphysical discussions of counterfactual worlds (see, for example, Divers, 2002). The factual world is not just one among the possible worlds: its materialization (becoming) endows it with certain properties (such as locality) which are absent (or at least might be absent) from counterfactual worlds.

Returning to our analogy between the approaches that assume hidden variables and counterfactuals, there seems to be a connection between a realistic assumption of *potentialities* and a more operational approach involving *counterfactuals*. We have cornered non-locality either at the microscopic level of potentialities (wavefunctions, quantum potentials) or at the macroscopic level of counterfactuals (possible observable events). Might we try to *identify* microscopic unobservable potentialities and macroscopic counterfactually observable events?

4. A CRITERION FOR REFUTATION

The conclusions of section 2 were based on a specific form of Bell's inequality (ineq. 2), derived by him in 1971. This form may be derived either by the hidden-variable approach or by the counterfactual approach. There are many other different inequalities, some of which may only be derived by the hidden-variable approach, not by the counterfactual one. For those derivable counterfactually, it could turn out that hypotheses *b* and *c* cannot be separated (as was done above, in section 2). If this is the case, then the "solution" of Bell's paradox presented at the end of section 2 is not general, and the asymmetry described in section 3 is also falsified.

Our proposal, therefore, satisfies the criterion of falsifiability or refutability, which Karl Popper (1959) considered the most important characteristic of a scientific hypothesis. We show below that our proposal passes two tests, Bell's original inequality and the EPR thought-experiment. Other situations, however, could falsify our proposal, such as a counterfactual version of Bell's theorem without inequalities (Greenberger *et al.*, 1990) or of the Leggett (2003) inequality, which remain as open problems.

(a) Bell's original inequality

Consider Bell's original derivation in 1964, based on local hidden variables, which furnished an inequality which may be rewritten as:

$$| \langle I(a)_k \cdot II(b)_k \rangle - \langle I(a)_k \cdot II(b')_k \rangle | \leq 1 + \langle I(b')_k \cdot II(b)_k \rangle . \qquad (4)$$

Instead of four pairs of measurements, this inequality uses only three, so apparently hypothesis *c* (locality involving two counterfactuals) does not have to be used, assuming that the actual experimental setting has been $I(a)_k$ and $II(b)_k$. However, the derivation of ineq.(4) makes implicit use of the property of "strict anti-correlation", which may be expressed as: $\langle I(a')_k \cdot II(b')_k \rangle = -1$, where $\boldsymbol{a'}=\boldsymbol{b'}$. This implies that the values for $I(b')_k$ and $II(b')_k$ in ineq.(4) must have opposite signs, but this is only warranted if assumption (*iv*) of section 2 is valid. The rejection of this assumption by means of the rejection of hypothesis *c* blocks the derivation of the inequality.

(b) EPR thought-experiment

The thought-experiment devised by Einstein, Podolsky & Rosen (EPR) (1935) may be briefly stated as follows. For the entangled state of eq.(3), a scientist *could measure*, with the equipment at the left hand side, the state of linear polarization associated to the prism setting $\boldsymbol{a}$=0°, and according to the outcome (+1 or –1) he could be sure of the "element of reality" associated to the measurement at the far-away right hand side, set at $\boldsymbol{b}$=0°, which would give the opposite value (strict anti-correlation). But instead of doing this measurement, he *could also choose to measure* the incompatible observable associated to ***a'***=45°, and with this he could be certain of the outcome obtained at the far-away piece of apparatus, set at ***b'***=45°. However, argue EPR, the choice of the scientist cannot affect in any way the situation at the far-away region, which is their assumption of locality. Therefore, well-defined values for both incompatible observables must exist simultaneously at the far-away region, and thus quantum theory is incomplete (since it does not associate well-defined values for incompatible observables).

The EPR thought-experiment does not involve any actual measurements, but only measurements in two different counterfactual worlds. However, by rejecting hypothesis *c* we deny that locality must be valid in a counterfactual world. Therefore, EPR's argument does not follow.

The counterfactual definition of locality (hypothesis *b*) only applies if a measurement is actually performed. One could modify EPR'a argument, and consider that one measurement (at the left-hand side) is performed, and that the other one (at the same side) is counterfactual. While locality (hypothesis *b*) could be applied for the actual measurement, it could not be applied for the counterfactual scenario (since hypothesis *c* has been rejected), so the incompleteness argument would not follow.

Bibliography

ASPECT, A., GRANGIER, P. "Experiments on Einstein-Podolsky-Rosen-type correlations with pairs of visible photons". In: R. Penrose and C. J. Isham (eds.). *Quantum Concepts in Space and Time.* Oxford: Clarendon, pp. 1-15, 1986.

BELL, J. S. "On the Einstein-Podolsky-Rosen Paradox", *Physics 1*, 195-200, 1964. Reprinted in Bell, *Speakable and Unspeakable in Quantum Mechanics.* Cambridge: Cambridge University Press, pp. 14-21, 1987.

BOHM, D. *Quantum theory*. New York: Prentice-Hall, 1951.

———. "A suggested interpretation of quantum theory in terms of 'hidden' variables, I. II.". *Physical Review* 85, 166-93, 1952.

CLAUSER, J. F., SHIMONY, A. "Bell's theorem: experimental tests and implications", *Reports of Progress in Physics* 41, 1881-1927, 1978.

CLIFTON, R., BUTTERFIELD, J., REDHEAD, M. "Nonlocal Influences and Possible Worlds". *British Journal for the Philosophy of Science* 41, 5-58, 1990.

D'ESPAGNAT, B. "The quantum theory and reality". Scientific American 241 (nov.), 128-40, 1979.

DICKSON, M. "Stapp's Theorem without Counterfactual Commitments: Why it Fails Nonetheless". *Studies in History and Philosophy of Modern Physics* 24, 791-814, 1993.

DIVERS, J. *Possible worlds*. London: Routledge, 2002.

EINSTEIN, A., PODOLSKY, B., ROSEN, N. "Can quantum-mechanical description of physical reality be considered complete?". *Physical Review* 47, 777-80, 1935.

GREENBERGER, D. M., HORNE, M. A., SHIMONY, A., ZEILINGER, A. "Bell's theorem without inequalities". *American Journal of Physics* 58, 1131-43, 1990.

HEISENBERG, W. *Physics and Philosophy*. London: Allen & Unwin, 1958.

LALÖE, F. "Do we really understand quantum mechanics? Starnge correlations, paradoxes and theorems". *American Journal of Physics* 69, 655-701, 2001.

LEGGETT, A. J. "Quantum mechanics at the macroscopic level". In: J. de Boer, E. Dal and O. Ulfbeck (eds.). *The Lessons of Quantum Theory*. Amsterdam: Elsevier, pp. 35-57, 1986.

———. "Book reviews". *Foundations of Physics* 17, 875-82, 1987.

———. "Nonlocal hidden-variables theories and quantum mechanics: an incompatibility theorem". *Foundations of Physics* 33, 1469-93, 2003.

MARGENAU, H. "Advantages and Disadvantages of Various Interpretations of the Quantum Theory". *Physics Today* 7 (Oct.), 6-13, 1954.

PAULI, W. "Editorial". *Dialectica* 2, 307-11, 1948.

POPPER, K. R. *The Logic of Scientific Discovery*. London: Hutchinson, 1959.

REDHEAD, M. *Incompleteness, Non-Locality, and Realism*. Oxford: Clarendon, 1987.

SHIMONY, A. "Controllable and Uncontrollable Non-Locality". In: *Search for a Naturalistic Wordlview*, vol. 2. Cambridge: Cambridge University Press, pp. 130-9, 1993.

STAPP, H.P. "Bell's Theorem and the Foundations of Quantum Physics". *American Journal of Physics* 53, 306-17, 1985.

ISSN 0100-6045

MANUSCRITO

Revista Internacional de Filosofia

Issues in the Philosophy of Physics

Edited by

Décio Krause
Otávio Bueno

Vol. 33 - n.1 - jan.-jun. 2010

CENTRO DE LÓGICA, EPISTEMOLOGIA
E HISTÓRIA DA CIÊNCIA - UNICAMP