Classical counterexamples to Bell's inequalities

Yuri F. Orlov
Cornell University, Ithaca, New York 14853
(Received 7 June 2001; published 18 March 2002)

This paper shows that a classical system containing a conventional yes/no decision-making component can behave like a quantum system of spin measurements in many ways (although it lacks a wave function) when, *in principle*, there are no deterministic decision procedures to govern the decision making, and when probabilistic decision procedures consistent with the system are introduced. Most notably, the system violates Bell's inequalities. Moreover, since the system is simple and macroscopic, its similarities to quantum systems arguably provide an insight into quantum mechanics and, in particular, EPR experiments. Thus, from the qualitative correspondences, decisions↔quantum measurements and the impossibility of deterministic decision procedures↔quantum noncommutativity, we conclude that the violation of Bell's inequalities in quantum mechanics does not require the existence of an unknown nonclassical nonlocality. It can merely be a result of local noncommutativity combined with nonlocalities of the classical type. The proposed classical decision-making system is a nonquantum theoretical construct possessing complementarity features in Bohr's sense.

DOI: 10.1103/PhysRevA.65.042106 PACS number(s): 03.65.Ud, 03.65.Ta

Here we describe and analyze a type of system, let us call it *S*, containing only classical components, including a conventional computer program generating yes/no decisions about the correctness of the orientations of given axes in accordance with probability distributions consistent with the properties of the system. (These decisions can be used to guide other systems.) We show that correlations between decisions made in two such systems, $S(1)$ and $S(2)$, having a common preparation history but separated in space as in the quantum Einstein-Podolski-Rosen (EPR) experiments [1], violate Bell's inequalities [2]. Such quantumlike behavior is not really surprising given the present author's description of a classical decision-making machine whose program cannot be deterministic [3]. And it bears out Kanter's conclusion that ''undecidability limits our knowledge on the spatial or temporal correlation functions even of classical systems'' [4], although the classical systems he analyzes contain neither decision-making components nor probabilistic decision procedures and, therefore, cannot violate Bell's inequalities.

The key to understanding the quantumlike properties of *S* is a theorem on the impossibility of deterministic decision procedures (IDP), the proof of which is given below. This theorem basically says that no logically consistent theory uniquely connecting the results of two consecutive decisions made in *S* is possible, in principle. Qualitatively, this theorem corresponds to the quantum-mechanical thesis that the results of two consecutive measurements of noncommuting observables cannot be uniquely connected. (None of the known theoretical ''deterministic'' models of quantum mechanics tries to *uniquely* connect results of two consecutive measurements of noncommuting observables, because they are models of *quantum* not classical mechanics.) The clear qualitative and, in part, quantitative similarity—which we will go on to demonstrate—between decisions made in *S* and measurements of spins placed on the same plane, and between IDP and quantum noncommutativity (as formulated above), means that *S* is the first nonquantum theoretical construct possessing complementarity features in Bohr's sense [5].

The main components of *S* are the following.

(a) An oriented axis $\vec{A}$ that lies and can be rotated in a homogeneous, isotropic plane. Let θ be the angle between $\vec{A}$ and a fixed axis. At time $t=t_k$, the angle of axis $\vec{A}$ is θ_k and the orientation of the axis is denoted by $\vec{A}_k$, $k=0,1,2,3,\ldots$.

(b) A generator *G* that is a computer program producing a sequence of random values of rotation angles $\Delta\theta_{k,k-1}$, $0 \leqslant |\Delta\theta_{k,k-1}| \leqslant 2\pi$, $k=2,3,\ldots$. When, at some time $t=t_k$, a value $\Delta\theta_{k+1,k}$ is produced, axis $\vec{A}$ changes its orientation, $\theta_k \to \theta_{k+1} \equiv \theta_k + \Delta\theta_{k+1,k}$, $\vec{A}_k \to \vec{A}_{k+1}$. This new orientation is fixed until the next rotation angle $\Delta\theta_{k+2,k+1}$ is produced at some time t_{k+1}. The initial orientation $\vec{A}_0$ is arbitrary. The initial condition for generator *G* is that at time $t=t_0$, *G* produces $\Delta\theta_{1,0}=0$, so the first random angle $\Delta\theta_{2,1}$ appears only at some time $t=t_1>t_0$.

(c) A decision-making component *C* that is a computer program written in accordance with rules I–VIII below. The initial decision c_0 is defined either as yes, $c_0=1$, or as no, $c_0=0$, by some factor(s) external to *S*, that is, it is *imposed* on *S*. Then *C* acts as follows. When some rotation angle, $\Delta\theta_{k+1,k}$, is produced by *G* and the new orientation $\vec{A}_{k+1}$ is established during the time interval $t_k<t<t_{k+1}$, *C* must decide the correctness of the new orientation, using a procedure based on rules I–VIII. *C* decides between yes (decision $c_{k+1}=1$) and no (decision $c_{k+1}=0$) on the basis of the previous decision c_k and the current rotation angle $\Delta\theta_{k+1,k}$. Once it has made decision c_{k+1}, *C* is ready to decide on the next rotation angle $\Delta\theta_{k+2,k+1}$.

A preliminary note on the rules. Since all the physical components of *S* are classical, *S* must obey the rules of classical physics and logic. Rules I–VIII are such rules. Some of them refer to decision procedures. There, as elsewhere in this discussion, a decision procedure is regarded as deterministic if the probability of the resulting decision equals 1 or 0. However, when *S* obeys rules I–V it turns out that, except for rotation angles divisible by π, *C*'s program must generate random decisions with some probabilities for them. In such cases, the decision procedure is regarded as probabilistic.

Rule I. For a decision c_{k+1}, only information about two values can be used: c_k, the previous decision, and $\Delta\theta_{k+1,k}$, the current rotation angle. (This means a first-order Markov process in the case of probabilistic decisions.) Thus, $c_{k+1}=f(c_k,\Delta\theta_{k+1,k})$.

Rule II. For a given rotation angle $\Delta\theta$, when a deterministic decision procedure is possible, only one of two procedures must be used: either the f_+ procedure, $1=f_+(1,\Delta\theta)$, $0=f_+(0,\Delta\theta)$, or the f_- procedure, $0=f_-(1,\Delta\theta)$, $1=f_-(0,\Delta\theta)$. (Sometimes we will refer to them as f_+ transitions, $(1,\Delta\theta\rightarrow1)$, $(0,\Delta\theta\rightarrow0)$, and f_- transitions, $(1,\Delta\theta\rightarrow0)$, $(0,\Delta\theta\rightarrow1)$. Rule II implies that deterministic transitions must be described by one-to-one functions uniquely defined by $\Delta\theta$.

Rule III. If $|\Delta\theta_{k+1,k}|=0$ or 2π, then $c_{k+1}=c_k$ (f_+ transitions).

Rule IV. If $|\Delta\theta_{k+1,k}|=\pi$, then $c_{k+1}=1-c_k\equiv\overline{c_k}$ (f_- transitions). (Later we will see what happens if rule IV is excluded from the list of rules.)

Rule V. Let us call pair $(c_k,\Delta\theta_{k+1,k})$—the previous decision and the current angle—a decision situation. The procedure for the next decision c_{k+1} is defined by this decision situation. If c_k,c_{k+1},c_{k+2} are three consecutive *certain* decisions, i.e., decisions made in accordance with deterministic procedures defined by the corresponding decision situations, then they must obey the following classical rule:

$$[(c_k,\Delta\theta_{k+1,k}\rightarrow c_{k+1})\wedge(c_{k+1},\Delta\theta_{k+2,k+1}\rightarrow c_{k+2})]\rightarrow(c_k,\Delta\theta_{k+1,k}+\Delta\theta_{k+2,k+1}\rightarrow c_{k+2}),$$

where $\rightarrow$ denotes logical implication. If it happens that $|\Delta\theta_{k+1,k}+\Delta\theta_{k+2,k+1}|=2\pi+\Delta\theta>2\pi$, then angle $\Delta\theta$ is used for the rotation.

Rules I–V lead to the following theorem.

Theorem on the impossibility of deterministic decision-procedures. In system *S*, deterministic decision procedures are impossible for an infinite countable set of rotation angles.

Proof. Consider the infinite countable set of rotation angles $\Delta\theta_j=\pi/2j$, $j=1,2,3,\ldots$. Take any *j*. Whichever deterministic function, f_+ or f_-, of rule II is used as the decision procedure for the corresponding angle $\Delta\theta_j$ it follows from rules I, II, and V that the function for the angle $2\Delta\theta_j=\pi/j$—which is the result of two consecutive transitions with the same angle $\Delta\theta_j$ and, therefore, with the same function, either f_+ or f_-—must be f_+. But then, as follows from applying rule V repeatedly, f_+ must also be the function for $j(2\Delta\theta_j)$, which is the result of the integer number *j* of the same f_+ transition. However, $j(2\Delta\theta_j)=\pi$ and according to rule IV, the function for $\Delta\theta=\pi$ must be f_-. Therefore, the use of any deterministic function for any rotation angle $\pi/2j$, $j=1,2,3,\ldots$ leads to contradiction. ■

Note that in our proof we could use a set of finite intervals $\delta\Delta\theta_l$ of the rotation angles, instead of a pointset of $\Delta\theta_j$'s. For example, $k\pi/j\leqslant(\Delta\theta)_{kj}<(k+1)\pi/j$, $k=0,1,2,3,\ldots$, $j=1,2,3,\ldots$, with f_+ or f_- assigned for every interval, i.e., the same function for every $\Delta\theta$ inside it. Indeed, according to rule III, when $k=0$, the function must be f_+ for any *j* since one of the rotation angles inside such intervals equals zero, $(\Delta\theta)_{0j}\in[0,\pi/j)$. Then the function must also be f_+ for any interval of rotation angles, which is the sum of such intervals, in particular, for the interval $j(\pi/j)\leqslant(\Delta\theta)_{jj}<(j+1)(\pi/j)$. But this violates rule IV, since rotation angle π lies inside this interval, $(\Delta\theta)_{jj}\in[\pi,(j+1)\pi/j)$. (A similar theorem with respect to spin measurements is presented in [6], but with no relation to Bell's inequalities.)

This is why nondeterministic (i.e., probabilistic) decision procedures are needed for $\Delta\theta\neq k\pi$, $k=0,1,2,\ldots$.

[Our IDP theorem may not hold if the angle space is not continuous but consists, for example, only of angles $\Delta\theta_k=\pi k/(2N+1)$ with a fixed integer *N*. In such a case, a deterministic decision procedure is possible. See formulas (36)–(38) in [6].]

Rule VI. In a case not covered by rules II–IV, a conditional probability of random decisions, $p(c_{k+1})=p(c_{k+1}|c_k,\Delta\theta_{k+1,k})$, must be introduced.

Rule VII. The probability of a decision c_k, $p(c_k)$, is defined as in the classical axiomatic theory of probability. The probabilities of decisions, $p(c_{k+1})=p(c_{k+1}|c_k,\Delta\theta_{k+1,k})$, in program *C*, and the probability of the appearance of a rotation angle $w=w(\Delta\theta)$ in program *G*, are mutually independent.

Before turning to rule VIII, we introduce formulas (1)–(3) below, which follow from rules VI and VII.

Since there are only two decisions, $c_k=1$ or 0 and $\overline{c_k}\equiv(1-c_k)=0$ or 1, we have, independently of the values of c_{k-1} and $\Delta\theta_{k,k-1}$,

$$p(c_k)\geqslant0,\quad p(c_k\vee\overline{c_k})=p(c_k)+p(\overline{c_k})=1,\quad p(c_k\wedge\overline{c_k})=0.$$

From rule III,

$$p(c_{k+1}|c_k0)=\delta_{c_k,c_{k+1}}=p(c_{k+1}|c_k,\pm2\pi),$$

and from rule IV,

$$p(c_{k+1}|c_k,\pm\pi)=\delta_{(1-c_k),c_{k+1}}. \tag{1}$$

For the same reason, namely, there being only two possible decisions for c_{k+1} in any decision situation $c_k,\Delta\theta_{k+1,k}$, we can always express the relative probabilities of these decisions as $\cos^2\varphi$ and $\sin^2\varphi$, where $\varphi=\varphi(\Delta\theta_{k+1,k})$. Assuming that transitions $(1,\Delta\theta\rightarrow0)$ and $(0,\Delta\theta\rightarrow1)$ have equal probabilities (which is consistent with rule II when $p=1$ or 0), the following probabilities $p_{c_k,c_{k+1}}(\Delta\theta)$ for transitions $(c_k,\Delta\theta\rightarrow c_{k+1})$, $\Delta\theta\equiv\Delta\theta_{k+1,k}$, must be built into *C*'s program:

$$(1,\Delta\theta\rightarrow1),\quad p_{11}(\Delta\theta)=\cos^2\varphi(\Delta\theta),$$

$$(1,\Delta\theta\rightarrow0),\quad p_{10}(\Delta\theta)=\sin^2\varphi(\Delta\theta), \tag{2a}$$

$$(0,\Delta\theta\rightarrow1),\quad p_{01}(\Delta\theta)=\sin^2\varphi(\Delta\theta),$$

$$(0,\Delta\theta\rightarrow0),\quad p_{00}(\Delta\theta)=\cos^2\varphi(\Delta\theta), \tag{2b}$$

$$\varphi(\Delta\theta)=\frac{\Delta\theta}{2}h(\Delta\theta),\quad h(0)=h(\pm\pi)=h(\pm 2\pi)=2k+1. \tag{3}$$

The boundary conditions (3) follow from rules III and IV. In Eq. (3), the signs on different sides of the equations are not interdependent.

Rule VIII. Function $h(\Delta\theta)$, obeying the boundary conditions (3), is chosen for S by some person or system outside S.

Note that $h(\Delta\theta)\equiv 1$ is one of the possible functions. For simplicity we assume that the probabilistic decision procedure for the yes/no decisions is defined by Eqs. (2a) and (2b), with $h\equiv 1$.

We can now see that the classical system S possesses some features of quantumlike complementarity. According to [5], complementarity means "the existence of different aspects of the description of a physical system, seemingly incompatible but both needed for a complete description of the system." Consider two arbitrary nonparallel axes $\vec{A}$ and $\vec{B}$, $\theta_B-\theta_A\equiv\Delta\theta_{BA}\neq k\pi$. Let an external observer of decisions in S select cases where a specific decision, say $c_A=1$, about axis $\vec{A}$ (more precisely, about any axis inside a small interval of angles around θ_A) is followed by some decision about axis $\vec{B}$. He can observe, in principle, all events in the system without disturbing it. What he will see is the random distribution of the decisions c_B with the probabilities $p(c_B=1)=\cos^2(\Delta\theta_{BA}/2)$, $p(c_B=0)=\sin^2(\Delta\theta_{BA}/2)$. If, instead, he selects cases where a specific decision about axis $\vec{B}$, say $c_B=1$, is preceded by some decision about axis $\vec{A}$, then he will see the random distribution of the c_A decisions with the same probability distribution. If he tries to find some hidden deterministic chains of events connecting these different decisions, he will realize that there are no such chains. (This point will be discussed later.) If he asks the system designer to make the decision-making program deterministic, the answer will be that such a change is impossible without changing the fundamentals of the system. And finally, if he decides to observe (and that means to select) a chain of step-by-step decisions leading from c_A to c_B, for example, $c_A, c_C, c_B, \theta_{BA}=\theta_{BC}+\theta_{CA}$, he will realize that any such intermediate observation, made without any physical influence on the system, will change the final probability distribution. These are fundamental features of complementary observables, let us call them here P_A and P_B, which are not compatible in any classical logical sense but, at the same time, are inseparable parts of the whole system.

The exact characteristics of an observable P_X in S are the full set of its possible numerical values, {0,1}, here the same for all axes; the direction of the axis to which its values are assigned; its relations to other observables, in this case, to P_Y's for different Y's, $\vec{Y}\neq\vec{X}$; and the methods of observing it, which means in this case methods of observing results of decisions c_X's. The difference between P_X and c_X is the same as the difference between, say, momentum and one of its numerical values obtained from measurements.

Up to this point, there is no observable quantitative difference between the system S and a quantum system of consecutive measurements of $\frac{1}{2}$-spin projections on different axes, when spins and axes are placed in the same plane. In such a quantum system, we can consider projectors (operators) $\hat{P}_A$, $\hat{P}_A\equiv(1+2\hat{s}_A)/2$, instead of spin-projection operators $\hat{s}_A$, where $\vec{A}$'s are different axes in the plane. It follows from the formula for P_A that this observable is measured simultaneously with s_A, with the possible outcomes (eigenvalues) 1 ($s_A=+\frac{1}{2}$) or 0 ($s_A=-\frac{1}{2}$). $\hat{P}_A$ and $\hat{P}_B$ do not commute if $\theta_{BA}\neq k\pi$ and are complementary for this reason, while in system S the complementarity appears as a result of the IDP theorem. All the qualitative and quantitative phenomena described in the above-thought experiments made on system S can be observed in this quantum system, if we consider consecutive measurements of spin projections on randomly chosen axes in the quantum system as corresponding to consecutive decisions in S.

(The analogy between the two systems is, of course, limited. There is no place for operators and wave functions in system S. More important, the quantization of spin in quantum mechanics is a result of rotation symmetry in *three-dimensional* space, whereas in system S, whose axes are placed on a plane, the "quantization"—discreteness of the yes/no decisions—is merely a feature of classical logic and exists independently of the existence of the third dimension.)

But even though S lacks a wave function, the probability phase (3) has some properties of quantum phases. Let us change the structure of S a little and link the probability phase $\varphi(\Delta\theta)$ to some external factors governing the phase development between the system's decisions; the only task of G, then, is to introduce the times t_k's when decisions must be made in accordance with the previous decisions and the rotation angles developed under the influence of these external factors. Now, it follows from the rules of the system that the phase advance $\varphi_{k+1}-\varphi_k\equiv\varphi_{k+1,k}$ developed before the c_{k+1}-decision is made may be forgotten after this decision is made, since the next decision c_{k+2} does not depend on it. This is a feature of every first-order Markov chain. But in S there is much more than that. In the classical Markov chain of 1, 0 events, the probabilities of three consecutive outcomes must obey the classical relation

$$\begin{aligned} p(c_k,\Delta\theta_{k+2k}\to c_{k+2}) &= p(c_k,\Delta\theta_{k+1,k}\to c_{k+1})p(c_{k+1},\Delta\theta_{k+2,k+1}\to c_{k+2}) \\ &\quad + p(c_k,\Delta\theta_{k+1,k}\to\overline{c_{k+1}})p(\overline{c_{k+1}},\Delta\theta_{k+2,k+1}\to c_{k+2}), \end{aligned} \tag{4}$$

where $\overline{c_{k+1}}=1-c_{k+1}$, $\Delta\theta_{k+2,k}=\Delta\theta_{k+2,k+1}+\Delta\theta_{k+1,k}$. But probabilities (2a) and (2b) in S violate this relation. For example, when the summed rotation angle on the left side of Eq. (4) equals π, the left side equals zero for $c_{k+2}=c_k$, in accordance with rule IV and formula (1), while the right side can be positive if none of the two intermediate angles equals 0 or π. The deep reason for this violation of classical rules in Eq. (4) is that, according to IDP, c_{k+1} is forbidden to be either 1 or 0 in the infinite number of cases when rotation angles $\Delta\theta_{k+1,k} \neq k\pi$. So if c_{k+1} nevertheless appears to be 1 or 0 at some time t'_k, we must assume that system S has made an unpredictable jump (has "collapsed") at this time, because any predictable transition to a certain c_{k+1} value would violate IDP. But the only parameter that can jump here is phase φ. We conclude that the development of phase φ in S not only may begin anew after the intervention of any decision, but *must* begin anew, regardless of what the system's physical decision-making mechanism is. Thus, the collapse of the probability phase as a result of a decision (measurement, in quantum mechanics) is a consequence of the IDP theorem (noncommutativity in quantum mechanics).

Note that the conditional probabilities (2a) and (2b) cannot appear in classical-mechanics systems that lack components to which the IDP theorem applies.

What physically distinguishes our probabilities (2a) and (2b) from the usual classical ones? Classical probabilities of observations of events having numerical values are introduced when we lack knowledge about those values, although those values are present before we begin our observations. In such cases, the observations reveal the objectively existing numbers. Our analysis of S has shown that quantumlike probabilities appear when there can be no algorithm to help us get knowledge about certain numerical values in any decision situation belonging to an infinite set of decision situations. But such values may result from decisions (measurements, in quantum mechanics), since procedures for decisions (measurements) are not algorithms. However, in such cases, questions about whether the observed numerical values were present before the observations cannot be answered, in principle, because the answers presuppose the existence of an algorithm that we have proved does not exist. Nevertheless, it is clear that those values could not have been present with certainty (i.e., with probability equal to 1), since such a presence would violate IDP. The following examples show what can be present in an ordinary classical case, and cannot be present in a quantumlike one.

Take a statistical ensemble of S's prepared as follows. In line with (c), some external factor defines the initial decision c_0 (and only c_0, not the following decisions) imposed on each S. Let c_0 be imposed randomly with probabilities $w(c_0)$,

$$w(1) = w(0) = \tfrac{1}{2}. \tag{5}$$

If this external factor were, say, some external classical generator of random numbers, then by analyzing the mechanism of this generator we could, in principle, have full information about the physical conditions preceding every c_0 decision and, therefore, could precisely predict every c_0. Then the theory of those predictions could be translated into the needed algorithm. The mere possibility of such a theory makes this ensemble of S's classical.

Let us now turn to a single S. The IDP theorem implies that it is impossible to have a physical theory that helps an observer of S to predict the decisions—c_{k+1} values—made in any decision situation $(c_k, \Delta\theta_{k+1,k})$, $k=1,2,\ldots$. Bearing this in mind, we have already introduced the probabilities (2a) and (2b) of c_{k+1}'s. Analyzing, again, the physical and mathematical structure of all parts of S involved in generating random events (decisions 1 and 0) can give us a physical theory connecting the concrete conditions inside S with its concrete decisions. Nevertheless, the ensemble of such decisions is not a classical one because the concrete conditions inside S, being uniquely connected with c_{k+1}, are not (and cannot be) uniquely connected with the values c_k, $\Delta\theta_{k+1,k}$. Only the probability distribution, with a given $h(\Delta\theta)$, is uniquely connected with them. The only physical "theory" possible in such a case is an infinite set of *a posteriori* conclusions that at times t_k, $k=0,1,2,\ldots$, there were such-and-such connections between the values c_k, $\Delta\theta_{k+1,k}$ and states of S. The necessity of this infinite description manifests the lack of an algorithm—which, by definition, is a finite text.

Thus, even though all parts of S are classical and our probability axioms are classical, we can expect probabilities (2a) and (2b) to violate Bell's inequalities since the functional forms of these probabilities are consistent with the structure and rules of S, to which the IDP theorem is integrally bound. To confirm this, let us first check a Bell inequality not for an ensemble of pairs of S's but for an ensemble (5) of single systems, using arguments analogous to those developed in [7]. Consider three oriented axes $\vec{A}$, $\vec{B}$, $\vec{C}$ with angles between them $\Delta\theta_{AB}$, $\Delta\theta_{BC}$, $\Delta\theta_{AC} \equiv \Delta\theta_{AB} + \Delta\theta_{BC}$. A researcher investigating Bell's inequalities selects cases when S's in the ensemble have made decisions about the correctness of these three orientations (each inside some small interval $\pm d\Delta\theta/2$ equal for all axes). Let us denote here eight triplets of such possible decisions as $\vec{A}^+, \vec{B}^+, \vec{C}^+$; $\vec{A}^+, \vec{B}^+, \vec{C}^-$;...; $\vec{A}^-, \vec{B}^-, \vec{C}^-$, instead of our c_A, c_B, c_C (which would be the triplets 1, 1, 1; 1, 1, 0;...; 0, 0, 0). In S, $\vec{X}^\pm$ yes/no decisions for any axis $\vec{X}$ are random consequences of the previous decisions and the corresponding rotation angles. We will see later that if our ensemble is prepared as ensemble (5) is, then there exist the normalized unconditional probabilities $p(\vec{X}^+) = w = \frac{1}{2}$ (of decision "yes") and $p(\vec{X}^-) = 1 - w = \frac{1}{2}$ (of decision "no") about the orientation of axis $\vec{X}$. Assuming $p(\vec{X}^\pm) = \frac{1}{2}$, we can calculate the joint probabilities for $\vec{A}$, $\vec{B}$, and $\vec{C}$ needed for any Bell inequality, in the three cases of transitions $(\vec{A}^+, \Delta\theta_{AB} \to \vec{B}^-)$, $(\vec{A}^+, \Delta\theta_{AC} \to C^-)$, and $(\vec{B}^-, \Delta\theta_{BC} \to C^+)$. From Eqs. (2a) and (2b) with $h \equiv 1$,

$$p(\vec{A}^+\vec{B}^-) = p(\vec{B}^-|\vec{A}^+)p(\vec{A}^+) = w\sin^2\frac{\Delta\theta_{AB}}{2} = \frac{1}{2}\sin^2\frac{\Delta\theta_{AB}}{2}, \tag{6}$$

$$p(\vec{A}^+\vec{C}^-) = p(\vec{C}^-|A^+)p(\vec{A}^+) = w\sin^2\frac{\Delta\theta_{AC}}{2} = \frac{1}{2}\sin^2\frac{\Delta\theta_{AC}}{2}, \tag{7}$$

$$p(\vec{A}^-\vec{B}^-) = p(\vec{B}^-|\vec{A}^-)p(\vec{A}^-) = (1-w)\cos^2\frac{\Delta\theta_{AB}}{2} = \frac{1}{2}\cos^2\frac{\Delta\theta_{AB}}{2}, \tag{8}$$

Using Eqs. (6) and (8),

$$p(\vec{B}^-)=w\sin^2\frac{\Delta\theta_{AB}}{2}+(1-w)\cos^2\frac{\Delta\theta_{AB}}{2}=\frac{1}{2}$$
$$=p(\vec{B}^+)\quad\text{when}\quad w=\frac{1}{2}. \tag{9}$$

The result (9) shows that, indeed, if $p(\vec{A}_k^\pm)=\frac{1}{2}$, then $p(\vec{A}_{k+1}^\pm)=\frac{1}{2}$. And since $p(\vec{A}_0^\pm)=\frac{1}{2}$ at the beginning of our Markov chain, we conclude (by induction) that Eq. (9) is correct for any axis at any time,

$$p(\vec{A}^\pm)=p(\vec{B}^\pm)=p(\vec{C}^\pm)=\tfrac{1}{2},\quad w=\tfrac{1}{2}. \tag{10}$$

Further,

$$p(\vec{C}^+\vec{B}^-)=p(\vec{C}^+|\vec{B}^-)p(\vec{B}^-)=\frac{1}{2}\sin^2\frac{\Delta\theta_{BC}}{2},\quad w=\frac{1}{2}. \tag{11}$$

Now assume that these probabilities behave classically. Then they must obey the corresponding Bell inequality [7], in particular,

$$p(\vec{A}^+\vec{B}^-)\leqslant p(\vec{B}^-\vec{C}^+)+p(\vec{A}^+\vec{C}^-). \tag{12}$$

However, the corresponding inequality following from Eqs. (6), (7), and (11),

$$\sin^2\frac{\Delta\theta_{AB}}{2}\leqslant\sin^2\frac{\Delta\theta_{BC}}{2}+\sin^2\frac{\Delta\theta_{AC}}{2}, \tag{13}$$

is violated, for example, when $\Delta\theta_{AB}=9\pi/8$, $\Delta\theta_{BC}=2\pi/8$, and $\Delta\theta_{AC}=11\pi/8$, with the left side of Eq. (13) equal to 0.962 and the right equal to 0.838. Thus, our probabilities (2a) and (2b) indeed do not behave classically.

It is not difficult to find the concrete cause of this phenomenon. Three axes $\vec{A}$, $\vec{B}$, $\vec{C}$ are involved in Bell's inequalities. When there are only two axes, then according to the IDP theorem, it is impossible to assign certain numbers, c_k and c_{k+1}, to both; however, it is still possible to assign the classical conditional and joint probabilities (2a), (2b), (7), (8),... without inconsistency. It is when there are three axes that the joint probabilities become inconsistent, because in such cases some decisions for intermediate angles are involved. Take, for example, one of the consistency relations of classical probability theory used to deduce Bell's inequalities,

$$p(A_{k+2}{}^+\vec{A}_k{}^+)=p(A_{k+2}{}^+A_{k+1}{}^+\vec{A}_k{}^+)$$
$$+p(A_{k+2}{}^+A_{k+1}{}^-\vec{A}_k{}^+), \tag{14}$$

in which we have ordered time, $t_k<t_{k+1}<t_{k+2}$. When $p(A_k{}^+)=1$, Eq. (14) can be rewritten as

$$p(A_{k+2}{}^+|A_k{}^+)=p(A_{k+2}{}^+A_{k+1}{}^+|A_k{}^+)$$
$$+p(A_{k+2}{}^+A_{k+1}{}^-|A_k{}^+). \tag{15}$$

But as we already saw in Eq. (4), which is equivalent to Eq. (15), if $\Delta\theta_{k+2,k}=\pi$ and $\Delta\theta_{k+2k+1}\neq 0,\ \pi$, then the left side of Eq. (15) equals zero [see rule (IV) and formula (1)], while the right side is positive.

Since rule IV of our system is, as we can see, directly involved in the violation of Bell's inequalities, it makes sense to look at what happens if we exclude it (and only it) from the set of system rules. If rule IV is excluded, the proof of the IDP theorem is destroyed because a *deterministic* procedure for all decisions $(c_k,\Delta\theta\rightarrow c_{k+1})$ is now possible: $c_{k+1}\equiv c_k$. This means that only f_+ is permitted. So for any $\Delta\theta_{k+1,k}$ and any pair of axes, $\vec{X}$, $\vec{Y}$, we now have $p(\vec{X}^\pm|\vec{Y}^\pm)=1$, $p(\vec{X}^-|\vec{Y}^+)=p(\vec{X}^+|\vec{Y}^-)=0$. (A similar analysis is made in [3].) These probabilities radically differ from Eqs. (2a) and (2b), which obey rule IV. With such a deterministic decision procedure, inequality (12) is trivially satisfied as $0\leqslant 0$.

Turning finally to the usual form of Bell's inequalities, let us describe a classical EPR type of decision-making system comprising an ensemble of pairs—$(S(1),S(2))$—of the systems that we analyzed earlier. There is an infinite set of axes oriented isotropically in all directions in a plane. Every pair of S's is prepared at some initial time $t_{A1}=t_{A2}=t_A$ in the following way. The same arbitrary axis $\vec{A}$ is assigned to each system of a pair. The initial yes/no decision about the correctness of its orientation $(c_0(1),c_0(2))$ is imposed randomly, so $p(\vec{A}^\pm)=\frac{1}{2}$ for both systems; these decisions are mutually opposite, i.e., either $(1,0)\equiv(\vec{A}^+,\vec{A}^-)$ or $(0,1)\equiv(\vec{A}^-,\vec{A}^+)$. After an initial decision is imposed on it, each system of a pair makes its own first decision—$c_1(1)$, $c_1(2)$—as follows. One system, either $S(1)$ or $S(2)$ (at random), makes this decision *at the time of the preparation*, t_A, and the other system *postpones* it. After this, $S(1)$ and $S(2)$ are transported to two unconnected places, 1 and 2; and later, at times $t_1>t_A$, $t_2>t_A$, their respective generators—$G(1)$, $G(2)$—will produce their own *local* rotation angles. Each system must remember the decision imposed on/made by it during preparation.

Recall that according to the design of these systems, the very first rotation angle equals zero, $\Delta\theta_{1,0}(1)=\Delta\theta_{1,0}(2)=0$, see (b), above. Therefore, the first decisions that the systems make themselves, whether postponed or not, are $c_1(1)=c_0(1)$, $c_1(2)=c_0(2)$, see rule III. Let system $S(1)$, for example, postpone its first decision. Then $G(1)$, generator of $S(1)$, must remember that its first rotation angle has yet to be generated. Given the preparation, the first rotation angle of $S(1)$ can equal only $\Delta\theta_{1,0}=0$, whereas the (second) rotation angle of $S(2)$, $\Delta\theta_{2,1}$, will be random. The times t_1, t_2 of the local decisions of $S(1)$, $S(2)$ about the correctness of the orientation of their local axes need not be correlated.

In line with the conditions of any EPR type of experiment, a researcher investigating correlations between decisions made by the paired systems has access to both the local decisions about the axis orientations, and to the orientations of the local axes. Moreover, the researcher's "nonlocality" permits him or her to measure angles between axes belonging to paired systems. But he does not know whose first decision—$S(1)$'s or $S(2)$'s—was postponed. Let him first select three local orientations that are *identical* for $S(1)$ and $S(2)$: $\vec{A}$, $\vec{B}$, and $\vec{C}$ (each inside some small interval $\pm d\Delta\theta/2$), such that $\Delta\theta_{AB}+\Delta\theta_{BC}=\Delta\theta_{AC}$. Then let him select the following pairs of local decisions: $(\vec{A}^+,\vec{B}^+)$, $(\vec{B}^-,\vec{C}^-)$, and $(\vec{A}^+,\vec{C}^+)$. [The notation: the first pair, for example, means either that $S(1)$'s decision about the correctness of the $\vec{A}$ orientation is $c_A=1$ and $S(2)$'s decision about the correctness of the $\vec{B}$ orientation is $c_B=1$, or that $S(2)$'s decision about the correctness of the $\vec{A}$ orientation is $c_A=1$ and $S(1)$'s decision about the correctness of the $\vec{B}$ orientation is $c_B=1$.]

It is easy to see that the Bell inequality corresponding to this case,

$$p(\vec{A}^+,\vec{B}^+)\leqslant p(\vec{B}^-,\vec{C}^-)+p(\vec{A}^+,\vec{C}^+), \tag{16}$$

is violated for the same angles between the axes as in Eq. (12). Indeed, we know that at one of two locations, place 1 in our example, $\Delta\theta(1)=0$. Let the decision observed (by the researcher) at this place be "no." [According to Eq. (16), the decision at place 2 is also "no."] Given our researcher's selection of decisions, displayed in Eq. (16), the decision at place 1 can be either about axis $\vec{B}$ [decision $\vec{B}^-(1)$] or about axis $\vec{C}$ [decision $\vec{C}^-(1)$], that is, either transition $\vec{B}^-(1)$, $0\rightarrow\vec{B}^-(1)$ or transition $\vec{C}^-(1)$, $0\rightarrow\vec{C}^-(1)$. These are *local* transitions at place 1. Since $\Delta\theta_{BC}\neq 0$, π, these two cases at place 1 are directly connected to two mutually incompatible subensembles of our ensemble. Now, the preparation of every system pair is such that the decision at place 2 would certainly be "yes" if the rotation angle there were also zero. In such a case, there would be a deterministic connection between $S(1)$ and $S(2)$, namely, either $(\vec{B}^-(1),\vec{B}^+(2))$ or $(\vec{C}^-(1),\vec{C}^+(2))$. But since $\Delta\theta(1)=0$, $\Delta\theta(2)$ must be random and, given the researcher's selection of two "no" decisions in Eq. (16), the selected rotation angle at place 2 for the rotation from $\vec{B}$ to $\vec{C}$ must be $\Delta\theta_{BC}$. So the selected case at place 2 is either transition $\vec{B}^+(2)$, $\Delta\theta_{BC}\rightarrow\vec{C}^-(2)$ or transition $\vec{C}^+(2)$, $-\Delta\theta_{BC}\rightarrow\vec{B}^-(2)$. These are *local* transitions at place 2. Since $p(\vec{B}^\pm)=p(\vec{C}^\pm)=p(\vec{A}^\pm)=\frac{1}{2}$ in our ensemble, as in Eq. (10), $p(\vec{B}^{-1}(1),\vec{C}^-(2))=p(\vec{C}^{-1}(1),\vec{B}^-(2))=\frac{1}{2}\sin^2(\Delta\theta_{BC}/2)$, as in Eq. (11). There are two more mutually incompatible subensembles corresponding to the exchange $1\leftrightarrow 2$, that is, $\Delta\theta(2)=0$. So the total $p(\vec{B}^-,\vec{C}^-)=2\sin^2(\Delta\theta_{BC}/2)$. Similar arguments apply to the other correlations in Eq. (16).

In a slightly different procedure, the researcher could select decisions $\vec{B}^-$ only at place 1 and decisions $\vec{C}^-$ only at place 2. Then the following two chains of arguments—logical implications—would lead to the same results:

$$(1)\quad \vec{B}^-(1)\rightarrow\vec{B}^+(2)\rightarrow\left[p(\vec{C}^-(2))=\sin^2\frac{\Delta\theta_{BC}}{2}\right],$$

$$\text{so}\ \ p(\vec{B}^-(1),\vec{C}^-(2))=\frac{1}{2}\sin^2\frac{\Delta\theta_{BC}}{2},$$

$$(2)\quad \vec{C}^-(2)\rightarrow\vec{C}^+(1)\rightarrow\left[p(\vec{B}^-(1))=\sin^2\frac{\Delta\theta_{BC}}{2}\right],$$

$$\text{so}\ \ p(\vec{B}^-(1),\vec{C}^-(2))=\frac{1}{2}\sin^2\frac{\Delta\theta_{BC}}{2}.$$

It is impossible, in principle, to combine Eqs. (1) and (2) into a single chain since such a combination would lead to $(\vec{B}^-(1),\vec{C}^-(2))\rightarrow(\vec{B}^-(1),\vec{C}^+(1))$, with the conclusion $(\vec{B}^-(1),\vec{C}^+(1))$ violating the IDP theorem.

These arguments show that the nonlocal correlations in Eq. (16) are quantitatively the same as the local correlations in Eq. (12). Thus, the Bell inequalities (16) in system S are violated without any involvement of nonconventional, nonclassical nonlocalities. The cause of the violations is directly connected with the existence of the IDP theorem, which is valid for every local $S(1)$ and $S(2)$ system.

Quantitatively, the violation of Bell's inequalities in our EPR type of classical system, the pair $(S(1),S(2))$, is identical to that predicted for the quantum singlet S-wave state of two $\frac{1}{2}$ spins placed in the same plane. The reason, as we will now show, is that in the quantum case we can construct chains of arguments similar to those we have in the classical case; these chains lead to the same formulas for the probabilities present in Bell's inequalities.

Two spins are located in two unconnected places, 1 and 2, and three directions, $\vec{A}$, $\vec{B}$, $\vec{C}$, common to both places are chosen. Let the researcher install the axis of the analyzer at place 1 in direction $\vec{B}$ and the axis of the analyzer at place 2 in direction $\vec{C}$, and then select the results of measurements $s_B(1)=-\frac{1}{2}$, $s_C(2)=-\frac{1}{2}$. Two different possible chains of arguments, similar to those in our classical example, arise from such results.

(1) Taking into account that the full spin of the system equals zero, it follows from the observation $s_B(1)=-\frac{1}{2}$ that $s_B(2)=+\frac{1}{2}$—the certain value. From the local noncommutativity of operators $\hat{s}_B(2)$ and $\hat{s}_C(2)$, it then follows that the value of $s_C(2)$ cannot be certain; the joint probability of the observed values $s_C(2)=-\frac{1}{2}$ and $s_B(1)=-\frac{1}{2}$ must be $\frac{1}{2}\sin^2(\Delta\theta_{BC}/2)$.

(2) From the observation $s_C(2)=-\frac{1}{2}$ follows the certain value $s_C(1)=+\frac{1}{2}$ and then the same joint probability as above. The same holds for other pairs of axes. Note that due to noncommutativity (IDP in the classical example), it is impossible, in principle, to combine any two possible chains of arguments into a single chain.

These arguments not only are similar to those in classical $(S(1),S(2))$, with the same quantitative results, but also similarly do not include the concept of some unknown, nonclassical type of nonlocality. Notwithstanding all these similarities, the two systems are fundamentally different. The major difference is, of course, that in the classical example we have an isotropic distribution of probabilities of mutually opposite "correct" orientations, $p(\vec{X}^\pm(1))=p(\vec{X}^\mp(2))=\frac{1}{2}$, while in the quantum example we have a probability amplitude $\Psi(\uparrow,\downarrow) = -\Psi(\downarrow,\uparrow) = 1/\sqrt{2}$, corresponding to an isotropic distribution of mutually opposite spin directions. So the analogy is not complete, nor is the possibility of completing it clear at this point. Nonetheless, the fundamental difference between our classical and quantum prepared states leaves untouched the quantitative identity of the violation of Bell's inequalities in both systems and the similarity of the arguments proving it. For the arguments in both cases take as a starting point a measurement (a decision in the classical example) *already made* in one of two places 1, 2. In both cases this assumption means that the initially prepared state has already collapsed in all space: in $(S(1),S(2))$, it is the ''collapse'' of the prepared probability distribution, and in the quantum example, the collapse of the prepared probability amplitude. And in both systems, the newly prepared states into which the initial states have collapsed are similar to one another, unlike those initial states.

However, the quantum mechanics concept of wave-function collapse in all space simultaneously, which lies behind the scenes in our arguments, is not completely understood and is not similar to the classical ''collapse'' of probabilities. While taking this into account, we conclude from the similarity of the arguments in all other respects (as laid out above) that the violation of Bell's inequalities in quantum mechanics does not require the existence of some special, unknown quantum nonlocality. Since, in the classical $(S(1),S(2))$ system, there are no such nonlocalities and the violation of Bell's inequalities is caused by the locally applied IDP theorem, the corresponding violation in quantum mechanics may be interpreted as a result of local noncommutativity combined with nonlocalities of the classical type.

The author thanks the Open Society Institute for supporting this work and Gennady Berman, Kurt Gottfried, and Sidney Orlov for useful discussions.

[1] A. Einstein, B. Podolski, and N. Rosen, Phys. Rev. **47**, 777 (1935).
[2] J. S. Bell, Physics (Long Island City, N.Y.) **1**, 195 (1964).
[3] Yuri F. Orlov, Phys. Rev. Lett. **82**, 243 (1999).
[4] I. Kanter, Phys. Rev. Lett. **64**, 332 (1990).
[5] Niels Bohr, in *A Centenary Volume*, edited by A. P. French and P. J. Kennedy (Harvard University Press, Cambridge, MA, 1985).
[6] Y. F. Orlov, Ann. Phys. (N.Y.) **234**, 245 (1994).
[7] B. D'Espagnat, Sci. Am. **241**(5), 158 (1979).