PHYSICAL REVIEW D VOLUME 18, NUMBER 2 15 JULY 1978

Use of inequalities for the experimental test of a general conception of the foundations of microphysics. II

Bernard d'Espagnat
*Laboratoire de Physique Théorique et Particules Elémentaires, Orsay, France**
(Received 24 March 1977)

Some aspects of a previous article bearing the same title—hereafter referred to as I—are further developed and clarified. For reasons explained in the text, a variant of the set of assumptions used in I is also introduced. The Bell inequalities are shown to follow from that new set of assumptions as well. The derivation does not appeal to the concept of a density of probability in a supplementary variables space, and it is basically independent of the other modes of derivation of the same inequalities. Illustrations by means of model theories are given.

I. INTRODUCTION

Recent experimental results[1] are generally interpreted as providing substantial indications that the Bell inequalities[2] are violated in the way that quantum mechanics predicts. Hence, it is important to study how wide the field of the concepts and general ideas that a violation of these inequalities falsify actually is.

To contribute to this program was already the purpose of a previous article bearing the same title[3] and referred to as I below. Specifically, paper I based a derivation of the Bell inequalities on a set of outwardly plausible assumptions, with the purpose of using the experimental violation of the inequalities to show that at least one of the assumptions of the set must be false, independently of any interpretation of the quantum-mechanical formalism.

One of these assumptions (assumption 3 of I) is "contrafactual": It asserts that if something were changed in the instruments with which a given system S will interact in the future, that could not change the actual properties of S. Undoubtedly, this sounds plausible. However, Stapp[4] recently based a very simple derivation of the Bell inequalities essentially on assumptions of such a kind. As a result, it might seem likely that, in I, the blame for the experimental discrepancy should be entirely attributable to assumption 3, thus making the other ones unobjectionable again. One of the purposes of the present article is to show that the situation is not so simple, and, more specifically, that if the remaining assumptions are associated with a new plausible one which is *not* contrafactual, the derivation of the Bell inequalities remains possible.

Sections II and III make a few points of I more explicit. The new assumption and the corresponding derivation are then described in Sec. IV. Section V is an application to three models.

II. SUBENSEMBLES AND NONCOMPATIBLE PROPOSITIONS

In I a system Σ of two free stable spin-$\frac{1}{2}$ particles U and V lying in a state of total spin zero (owing to some previous interaction) is considered and a set $(D+A)$ of definitions (D) and assumptions (A), bearing in particular on such propositions as

$$u_i\ (v_i) \equiv \text{the spin of } U\ (V) \text{ along } \vec{e}_i \text{ is } +\hbar/2\,, \quad (1)$$

is introduced. *Propositions* are defined by referring to the existence of a specific type of instrument (here the Stern-Gerlach device) by means of which we could check their validity on any specified particle. But it is assumed (as part of A) that in *some* instances such a proposition may be true on a system even if nobody knows it is. The full $(D+A)$ set is then introduced and it is shown that in the particular case of the $U+V$ systems introduced above, the Bell inequalities follow. Since these inequalities are violated, it is concluded that at least one of the elements of $(D+A)$ must be false or meaningless.

The main intermediate step in the argument is to show what follows. If $(D+A)$ were valid, then *in the particular case of the $U+V$ systems under consideration* the following "statement S" would be true:

Statement S. Any ensemble E of N free subsystems V (the same would hold for systems U) is composed of disjoint subensembles $E_{\sigma_1,\ldots,\sigma_k}(\sigma_1,\ldots,\sigma_k=\pm 1)$, the $n(\sigma_1,\ldots,\sigma_k)$ elements of which being such that proposition $v_i\ (v_i')$ is true if $\sigma_i=+1\ (-1)$, where v_i' is the orthogonal complement of v_i and $i=1,\ldots,k$.

Statement S implies that several noncompatible propositions are simultaneously true on the same system. As such, is it to be rejected as violating either experimental facts or consistency requirements? This is our first question. Should it be answered affirmatively, then the result we are aiming at, namely a disproof of $(D+A)$, would be obtained at "a low price" since it would require no experiment and no comparison with *observable* quantum-mechanical predictions. However, statement S does *not* contradict experimental facts, since *we* cannot sort out the $E_{\sigma_1, \ldots, \sigma_k}$ from E.[5] It does not contradict *our* operational definition of a proposition either, since the latter merely means that *if* we consider just *one* proposition on a system, an instrument exists that enables us to verify its truth. Hence, contrary to appearances, statement S cannot be rejected *a priori* except on the basis of specific *interpretations* of quantum mechanics. It can be disproved only *a posteriori*, through the fact that it entails the Bell inequalities.

On the other hand, it may be noted that assertions such as the assertion

$$v_1 \wedge v_2 \wedge \cdots \wedge v_k \tag{2}$$

(which is true on $E_{+1,+1,\ldots,+1}$, $\wedge$ meaning the logical "and") are *not* propositions since no corresponding instrument exists. But this merely shows that if $(D+A)$ were valid, meaningful assertions would exist that would not fall into the class of the propositions, when the latter is operationally defined in the manner specified above. This creates no difficulty, provided that assertion (2), which is not empirically testable, is not identified with the intersection

$$v_1 \cap v_2 \cap \cdots \cap v_k \tag{3}$$

of propositions $v_1, v_2, \ldots, v_k$ *as that intersection is defined in the quantum propositional calculus*.[6,7] In that calculus the intersection $\cap a_i$ of several propositions a_i *is itself a proposition*. It can be defined either as corresponding to the intersection of the Hilbert-space linear manifolds corresponding to the a_i or, operationally, by the Jauch procedure of infinitely iterated measurements[7] (if such a procedure is considered as "operational"). The latter definition, in particular, shows that $\cap a_i$ is in principle experimentally testable, as any proposition should be. In the simple case $a_i = v_i$ such an intersection is v_1 if all the $\vec{e}_i$ coincide and $\emptyset$ otherwise.

Let it be stressed that the existence of meaningful but not empirically testable assertions, such as (2), is a logical possibility. In fact, the hidden-variables models provide elementary examples. Let us, for instance, consider the Bell-Clauser model for one spin-$\frac{1}{2}$ particle.[2,8] It exactly reproduces all the observable predictions of quantum mechanics for such a spin, hence it also reproduces the structure of the quantum-mechanical lattice of propositions as regards spin-$\frac{1}{2}$ systems. In particular, the Jauch procedure applied to $v_1 \cap v_2$ gives $\emptyset$ whenever $\vec{e}_1 \neq \vec{e}_2$. Nevertheless, since the theory is deterministic, the result we would obtain *if* we measured v_1 is determined by the hidden variables (which are "local"[2,9]) in any particular case, and so is—simultaneously—the result we would obtain *if* we measured v_2 instead. Let $v_1 \wedge v_2$ be the assertion "the hidden variables of the particular particle under consideration have values such that they would induce the result of a measurement of v_1 to be "yes" *if that* measurement were performed *and* that they would induce the result of a measurement of v_2 to be "yes" *if that* measurement were performed." As soon as we consider the local-hidden-variables assumptions as *not a priori meaningless* we must grant that $v_1 \wedge v_2$ is meaningful (and not necessarily false). However, it is not empirically testable because *we* cannot sort out from an ensemble the particles for which it is true. Hence, meaningful assertions that are not empirically testable may exist in some models, which is what was to be shown. The proof cannot be objected to on the basis of the von Neumann argument against hidden variables[10] since that argument makes use of a supplementary assumption[11] that we do not introduce. It cannot be objected to on the basis of what are sometimes called the "Gleason troubles"[11] either, since the relevant Hilbert space has dimensionality 2.

Remark. The explicit interpretation of assertion $v_1 \wedge v_2$ given in the model just described makes it apparent that $E_{+1,+1}$ is the intersection of the ensembles $E_{+1,\cdot}$ and $E_{\cdot,+1}$ corresponding to v_1 and v_2, respectively, in the very sense that the notion of intersection has in the classical theory of sets. This accounts for the use made in I of the classical theory of sets for combining the $E_{\sigma_1, \ldots, \sigma_k}$. In particular, it is the reason why the equalities

$$\sum_{\sigma_1, \sigma_2, \sigma_3} n(\sigma_1, \sigma_2, \sigma_3) = N \tag{4}$$

and

$$M(i,j) = N^{-1} \sum_{\sigma_1, \sigma_2, \sigma_3} \sigma_i \sigma_j n(\sigma_1, \sigma_2, \sigma_3) \tag{5}$$

could be written there.

III. ROLE OF THE CONCEPT OF ISOLATION

Let us consider the following statements:

$p_1(S)$ = the system S is free (or "isolated") as regards spins;

p_2 = the Bell inequalities are satisfied.

When p_1 or p_2 concern exclusively a system $S \in \{U, V\}$, where U and V are the two particles considered in Sec. II we write respectively $p_1(U, V)$ or $p_2(U, V)$.

It has been pointed out[12] that what is proved in I is not

$$(D+A) \Rightarrow p_2(U, V), \tag{6}$$

but

$$(D+A)\wedge p_1(U,V)\Rightarrow p_2(U,V) \quad (7)$$

or equivalently

$$p_2'(U,V)\Rightarrow(D+A)'\vee p_1'(U,V)\,, \quad (7')$$

where the notations $\Rightarrow$, $\vee$, $\wedge$, a, and a' mean logical implication, conjunction, disjunction, statement a is true, and statement a is false, respectively. In other words, this means that the proof given in I formally depends on the assumption that U and V are isolated during their flight.

On the other hand, in the context of the derivation actually performed in I a definition of $p_1(S)$ (i.e., of "S is free" as regards spins) that would be both completely precise and fully specified once and for all is not requested. In order for (7) to hold true it is sufficient that the meaning of $p_1(S)$ should be specified in such a way that when $V(U)$ is a stable spin-$\frac{1}{2}$ particle (or a photon) and $p_1(V)$ $(p_1(U))$ is true then the proposition (1) (or the analogous one with polarization) is persistent (in the sense defined empirically in Sec. IV below) as well as its orthogonal complement. Now the fact that a proposition such as proposition v_1 defined by (1) is persistent under some specified general conditions G of an operational type is a fact that can be checked by independent experiments bearing on particles with spin $\frac{1}{2}$ (or on photons) submitted to such conditions G. Any set of general conditions G for which the result of such experiments is positive[13] can be used for a partial specification of the meaning of $p_1(V)$, simply by asserting that whenever these conditions are realized $p_1(V)$ is true by definition. Such a partial definition is of a stipulative type. Within its range it will usually coincide with the more intuitive definitions of isolation very briefly sketched in I but this is by no means necessary for the argument.

It is then clear that the term $p_1(U,V)$ on the left-hand side of (7) is not an *assumption* but an *experimental specification*. It merely asserts that if we want to disprove $(D+A)$ by observing violations of the Bell inequalities we must choose the general operational conditions G under which the experiment is to be performed in some specific way. Namely, we must choose them so that between the times t_0 (corresponding to the end of the emission process) and t_1 (at which a measurement begins), measured in the laboratory referential, they belong to the class of those for which the experiments mentioned above have shown that propositions of the type of the proposition v_i defined by (1) are persistent.

An example of a case in which $p_1(U,V)$ is *not* valid is the one in which U and/or V interact during their flight with atoms or gas molecules by way of a spin-dependent interaction, even if such an interaction is assumed to be local. An experimental test of the Bell inequalities performed under such conditions would not (in the present stage of our knowledge) constitute a relevant test for $(D+A)$. On the other hand, it would still—by the way of the generalized Bell inequalities[14,2]—be a test of the local-hidden-variables hypothesis.[2,9] This is one illustration (there are others) of the fact that it is somewhat more difficult to disprove $(D+A)$ in full generality than just to disprove p_2.

Remark 1. The fact that a proposition is persistent may depend on the referential. However, such an observation does not make assumption 2 of I less credible; if a proposition a is persistent on a stable system in a given referential, then assumption 2 is a natural one in *that* referential.

Remark 2. The question of the *reproducibility* of the general conditions G deserves a comment. If we operate on one particle only—V for instance—then in order to check that v_i is persistent under some such conditions we must measure v_i at a time t_1, measure it again at a time t_2, and verify that both measurements give always (i.e., for any V) the same result. But then the presence of the very instruments used for these measurements should in principle be considered as a part of the general conditions G. This creates a difficulty since in principle it restricts the case in which $p_1(V)$ is given a meaning to those in which the instruments in question are really present, while the applications described in I of that concept are to a case in which *no* instrument interacts with V.

Such a difficulty is alleviated to a fair extent by considering both particles U and V. Since these particles are strictly correlated, the first of the two measurements considered above can be replaced by a measurement made on U. The corresponding instrument can then be placed arbitrarily far away from the region of interest, simply by setting the source of the $U+V$ system far enough away. Similarly, t_2 may be taken arbitrarily large and the second instrument, acting on V, can correspondingly be placed very far from the region of space in which we want to define the concept $p_1(V)$, that V is free as regards spin. It is then a natural convention to exclude both instruments from the general experimental conditions G that are to be exactly reproduced whenever we want to be sure that V is "free" in the region in question.

Unless otherwise specified, such a convention is adopted in what follows. In other words, the general experimental conditions G merely refer to the nature of the transversed material, the presence or absence of a magnetic field, and so on, in a given region R of space. It is assumed that the persistency of v_i has been checked on many systems V by several methods including the indirect one just described. Then if the results are always positive, we say that by definition, under conditions G, any system V produced by a similar source is *free* (or *isolated*) in R, as regards spin.

IV. DESCRIPTION OF THE ARGUMENT

This section describes the present status of our proposed method of derivation of the Bell inequalities. For the reader's convenience it incorporates enough of the material already presented in I so as to make a rereading of I unnecessary.

The method makes use of the concepts of *systems* and *propositions*. As regards the former we assume that we can use it in the usual way. In particular, if a system is composed of two noninteracting parts U and V that have interacted in the past (a case that repeatedly occurs in the following argument), then U and V are themselves to be considered as "systems."

As regards the notion of proposition, we require first of all that it should be an assertion (a) bearing on a system, and (b) that can be either "yes" (true), "no" (false), or undefined. Moreover, we require also, as stated in Sec. II, that such a concept should be more restrictive than the vague concept of a mere assertion in that a proposition *must refer to the existence of a specific class of instruments* by means of which it could be measured. In that respect it may be said that we define propositions operationally. On the other hand, the foregoing requirement merely means that in order to know whether an assertion a about a system S is a proposition or not we must (a) consider a independently of any other assertion that may or may not be formulated simultaneously about S, and (b) inquire whether or not an instrument A *exists* that *would* enable us to measure a at time t *if* we could make S interact with A at time t or immediately afterward. Note that it is *not* required that we should have the possibility to *actually* do so on the particular system S that we consider. It suffices that we can perform such measurements on systems of the same type as system S.

Let a' denote again the orthogonal complement to a.

Let us now proceed to formulate the definitions and assumptions needed.

Assumption 1. It is meaningful to associate to any proposition a the concept of a family $F(a)$ of systems, $F(a)$ being defined by the following condition. The systems S that belong to $F(a)$ are all those that are such that if a were measured on S by any method the result yes would be obtained with certainty.[15,16]

Obviously the definition does *not* imply that if S does not belong to $F(a)$ it belong to $F(a')$.

Definition 1. If a system S belongs to $F(a)$, a is *true* on S.

Definition 2. Let a system S be considered between times t_1 and t_2 in a given referential in which all the instruments are at rest. Let $t_1 < t_a < t_b < t_2$. a is said to be *persistent* on S between t_a and t_b if whenever a if found true on S at time t_a it is also true at time t_b.

Definition 3. Some systems $U+V$ are produced in such a way that measurements of the spin components $S^{(U)}(\vec{e}_i)$ and $S^{(V)}(\vec{e}_i)$ of U and V along any common direction $\vec{e}_i$ always give opposite results. Such systems are said to be "in a state of total spin zero."

Definition 4. In the case of a V particle that constitutes a part of a spin-zero $U+V$ system, V is said to be *free as regards spin* between times t_1 and t_2 if during that time interval it is subjected to general experimental conditions G under which it has previously been checked, for any i, on other systems that the proposition "$S^{(V)}(\vec{e}_i)=m_i$" is persistent (the precise nature of G and of the check is described in remark 2 of Sec. III).

Assumption 2. Again let $t_1 < t_a < t_b < t_2$ and let particle V be free as regards spin during the time interval (t_1, t_2). Then if v_i is true at time t_b, it is also true at time t_a (as regards the meaning of that assertion see again footnote 16).

Remark 1. Assumption 2 could easily be generalized to system other than the V (or U) parts of such spin-zero $U+V$ systems. But as regards such V (or U) parts it has an immediate consequence which is that if it is true, then $S^{(U)}(\vec{e}_i)$ (or $S^{(V)}(\vec{e}_i)$) cannot be affected by the physical interaction of U (or of V) with the instrument that measures it. In other words, if a measurement gives a definite value for $S^{(U)}(\vec{e}_i)$, then $S^{(U)}(\vec{e}_i)$ must have had that definite value even before the measurement, whenever conditions G prevail.

The proof is that if $S^{(U)}(\vec{e}_i)$ is measured at a time t' and is found to be equal to a certain value $m_i = \pm 1$ in $\hbar/2$ units then at $t_b > t'$ $S^{V}(\vec{e}_i)$ has the definite value $-m_i$. Since $S^{(V)}(\vec{e}_i)$ is persistent, assumption 2 implies that $S^{(V)}(\vec{e}_i) = -m_i$ also at a time $t_a < t'$. The strict correlation between $S^{(V)}(\vec{e}_i)$ and $S^{(U)}(\vec{e}_i)$ then implies that at time t' $S^{(U)}(\vec{e}_i)=m_i$. Q.E.D.

Such a result is not compatible with the usual interpretation of quantum mechanics. Is this a sufficient proof that assumption 2 has to be discarded? The answer is no, for reasons similar to those already developed in Sec. II. In fact, for observable consequences to be reached, a third assumption must be introduced, such as assumption 3 of I. However, for the reasons given in Sec. I, we prefer to introduce here a new "assumption 3" instead, which can also be given the name of "inductive causality."

Assumption 3 (inductive causality). Let a be a statement that bears on a property of a system and let it be the case that the validity of a at time t on a system S can be inferred from a knowledge of the fact that S has a certain property b at a *later* time t_1 $(t_1 > t)$. Let E be an ensemble of systems S considered at time t. Let E_1 be an unbiased sample of E, selected at a time $t' < t$ by some assistant, ignorant of what we shall do with that sample. Let then a measurement of b be made at a time immediately before t_1 on every element of E_1. Then the assumption is: If the result of that measurement is positive in every case, a was valid at time t, not only on every element of E_1 but also on every element of E.

Remark 2. That assumption closely parallels our intuitive notion that "the cause must be anterior to the effect," that "therefore" a cannot have been induced in the elements of E_1 by the very act of measurement, and that "therefore" it must have been true already on every element of E_1, hence, by induction, on every element of E.

Derivation of the Bell inequalities. The adjunction of assumption 3 to the list makes it possible to derive the Bell inequalities.[17]

Let F be a source emitting a large number N of spin-zero $U+V$ systems during a short time interval centered on time t_0. Let $\mathcal{S}$ be the composite system constituted by all these N $U+V$ systems and let S be the composite system constituted by all the corresponding particles U. By imagining a large number of replicas we may build up an ensemble E of systems S; let us identify it, at a time $t > t_0$, with ensemble E considered in assumption 3. Similarly, let the property b considered in assumption 3 be identified with a property b_i that S may indeed have and which is described as: The U parts of S constitute two distinct families $F(u_i)$ and $F(u_i')$, in which $S^{(U)}(\vec{e}_i)$ has the definite values $+1$ and -1, respectively. Finally, let a be the statement expressed by the same sentence as the one describing b, the difference being, as stated in assumption 3, that a and b are not considered at the same time and do not bear on the same ensembles of systems S.[18]

Let us now assume that at some time $t' > t_0$ $(t' \leq t)$ an assistant selects an unbiased sample E_1 of E (as stipulated in assumption 3). Let $t_1 > t$ and let us assume further that immediately before t_1 all the elements of E_1 are made to interact with an experimental device that measures the $S^{(U)}(\vec{e}_i)$ of all the constituting particles U. As we know, property b is then possessed at t_1 by every element of E_1. On the other hand, we know from the foregoing remark 1 (this is where the V parts, the spin-zero hypothesis, and all our other specific assumptions intervene) that the validity of statement a on a system S at time t can be infered from a knowledge of the fact that at time $t_1 > t$ the same system S possesses property b. The conditions for applying assumption 3 are therefore satisfied; and assumption 3 then tells us that a was valid at time t on every element of E, that is on all the systems S, including those that never interact with any measuring instrument.

Let us now consider not just one but four subensembles, E_1, E_2, E_3, and $\hat{E}$, of E, all constituting unbiased samples. Immediately before t_1 let the elements of E_1 be subjected to a measurement of b_1, as described above, and similarly let E_2 and E_3 be subjected respectively to measurements of b_2 and b_3. The foregoing reasoning can be repeated as regards E_1, E_2, and E_3. It shows in particular that at time t the elements of $\hat{E}$ (on which no measurement is assumed to be ever made) are such that statements a_1, a_2, and a_3 are valid on them.

Let S be an element of $\hat{E}$. The reasoning above shows that at time t S is simultaneously composed of two different (and disjoint) families $F(u_1)$ and $F(u_1')$, of two different (and disjoint) families $F(u_2)$ and $F(u_2')$ and of two different (and disjoint) families $F(u_3)$ and $F(u_3')$. Let U be a component of S and let $\sigma_i = +1$ or -1 according to whether U belongs to $F(u_i)$ or to $F(u_i')$. Let $n(\sigma_1, \sigma_2, \sigma_3)$ be the number of components of S on which σ_1, σ_2, and σ_3 have the specified values shown ($\sigma_i = \pm 1$). Let $n(\sigma_1, \sigma_2, .)$ be the number of components of S on which σ_1 and σ_2 have the specified values shown, σ_3 being arbitrary and let $n(\sigma_1, ., \sigma_3)$ and $n(., \sigma_2, \sigma_3)$ be defined similarly. Both common sense and the *classical* theory of set (which is the only one in existence) independently inform us that

$$n(\sigma_1, \sigma_2, .) = n(\sigma_1, \sigma_2, +1) + n(\sigma_1, \sigma_2, -1) \equiv \sum_{\sigma_3} n(\sigma_1, \sigma_2, \sigma_3) \qquad (8)$$

and that

$$\sum_{\sigma_1, \sigma_2} n(\sigma_1, \sigma_2, .) = \sum_{\sigma_1, \sigma_3} n(\sigma_1, ., \sigma_3) = \sum_{\sigma_2, \sigma_3} n(., \sigma_2, \sigma_3) = N, \qquad (9)$$

where N is the number of particles U composing S.

The numbers such as $n(\sigma_1, \sigma_2, .)$ are not directly measurable. But they can be measured indirectly, by making use of the strict spin correlation between U and V. Let us consider again the composite system $\mathcal{S}$, the U particles of which compose S.

If we consider three such systems $\mathcal{S}$, all of them having N $U+V$ components and if N is very large, then, disregarding small statistical fluctuations, the numbers $n(\sigma_1, \sigma_2, .)$ are the same on the three $\mathcal{S}$.[19] Moreover, they are equal to the numbers $\mathcal{S}(\sigma_1, -\sigma_2, .)$ of the $U+V$ components of $\mathcal{S}$, the U part of which has $S^U(\vec{e}_1)=\sigma_1$ and the V part of which has $S^{(V)}(\vec{e}_2)=-\sigma_2$. Similar remarks hold true as regards $n(\sigma_1, ., \sigma_3)$ and $n(., \sigma_2, \sigma_3)$. Hence, we can measure $n(\sigma_1, \sigma_2, .)$ by measuring $\mathcal{S}(\sigma_1, \sigma_2, .)$ on one of our three composite systems $\mathcal{S}$, that is, by making a correlation experiment. And similarly, we can measure $n(\sigma_1, ., \sigma_3)$ and $n(., \sigma_2, \sigma_3)$ by operating on the two remaining $\mathcal{S}$.

Let then $P(i,j)$ be the mean value of the quantity

$$\frac{2}{\hbar} S^{(U)}(\vec{e}_i) S^{(U)}(\vec{e}_j), \tag{10}$$

$$P(1,2) = N^{-1} \sum_{\sigma_1, \sigma_2} \sigma_1\sigma_2 \mathcal{S}(\sigma_1, \sigma_2, .) \tag{11}$$

$$= -N^{-1} \sum_{\sigma_1, \sigma_2} \sigma_1\sigma_2 n(\sigma_1, \sigma_2, .)$$

$$= -N^{-1} \sum_{\sigma_1, \sigma_2, \sigma_3} \sigma_1\sigma_2 n(\sigma_1, \sigma_2, \sigma_3) \tag{12}$$

and similarly as regards $P(1,3)$ and $P(2,3)$, so that

$$B \equiv P(1,2) + P(1,3) + P(2,3)$$

$$= -N^{-1} \sum_{\sigma_1, \sigma_2, \sigma_3} (\sigma_1\sigma_2 + \sigma_2\sigma_3 + \sigma_3\sigma_1) n(\sigma_1, \sigma_2, \sigma_3) \tag{13}$$

$$= -\frac{N^{-1}}{2} \sum_{\sigma_1, \sigma_2, \sigma_3} [(\sigma_1 + \sigma_2 + \sigma_3)^2 - 3] n(\sigma_1, \sigma_2, \sigma_3). \tag{14}$$

Hence, since $(\sigma_1 + \sigma_2 + \sigma_3)^2 = 9$ or 1,

$$B \leq N^{-1} \sum_{\sigma_1, \sigma_2, \sigma_3} n(\sigma_1, \sigma_2, \sigma_3) = 1. \tag{15}$$

This is one of the generalized Bell inequalities, which is thus proved as a consequence of assumptions 1, 2, and 3. The other Bell inequalities can be proved in the same way.

V. ILLUSTRATIONS

The assumptions on which the foregoing derivation rests do not exactly coincide with those on which any of the other existing derivations are based. In order to grasp more accurately the role of these various assumptions it is appropriate to investigate three model theories that violate the Bell inequalities, but which do so for reasons that, according to the preceding analysis, are different.

Theory 1: Naive quantum mechanics

Theory 1 is just the interpretation of quantum theory that attributes an objective existence to the state vector. In that it coincides with Schrödinger's first conceptions about wave mechanics. In theory 1 the reduction of the state vector is objective and the state vector constitutes a complete representation of reality. Therefore a spin-zero $U+V$ system does not have any definite value of $S^{(V)}(\vec{e}_i)$ in that theory. But when a measurement of $S^{(U)}(\vec{e}_i)$ is made, the overall state vector is thereby reduced and $S^{(V)}(\vec{e}_i)$ suddenly acquires a value. Since V is then *free* according to our *definition 4* it follows that assumption 2 is violated in the present theory. Hence, according to the analysis of Sec. IV the violation of the Bell inequalities is to be attributed to the violation of assumption 2 by the theory in question.

Theory 2: de Broglie-Bohm-Bell (dBBB) model

As is well known, de Broglie[20] proposed in 1927 a hidden-variables model that was developed independently by Bohm.[21] More recently, Bell[11,22] described one possible way of introducing spin in that model while preserving the observable predictions of quantum mechanics. Bell also discussed, in that theory, the correlations between the spin components $S^{(U)}(\vec{e}_i)$ and $S^{(V)}(\vec{e}_j)$ of a two-particle system $U+V$ along one common direction $\vec{e}_i = \vec{e}_j$. Here it is necessary for our purpose to write down explicitly the equations for the general case $\vec{e}_i \neq \vec{e}_j$. This should make it possible to locate exactly the reason why the proof of Sec. IV does not apply to that model.

Let us recall that the model theory in question is based on a somewhat simplified dynamics in which the motion of a spin-$\frac{1}{2}$ particle along its axis Ox of propagation may be considered as classical, and in which a wave packet initially centered around $z=0$ is shifted to a wave packet of the same width and same direction of propagation but centered around $z = m h(t)$ (where $m = \pm 1 = S_z$) if, from time $-\infty$ to time t, it traverses a magnetic field $H(x)$, where $h(t)$ is a *constant* whenever $H_z(x) = 0$ at the place where the particle is at time t. In addition, the model incorporates supplementary variables Z_i ("hidden" variables) that make it deterministic and that give directly the result of the measurement. Specifically, to each spin-$\frac{1}{2}$ particle there corresponds a vector $\vec{Z}$ in the plane perpendicular to Ox and measurement of a spin component along $\vec{e}_i$ means observing whether $\vec{Z}(\vec{e}_i)$ is positive or negative after the particle has traversed a Stern-Gerlach magnet oriented along $\vec{e}_i$. Hence, the $\vec{Z}$ may be considered as the instrument coordinates. In application of the general theory[20,21] the derivatives of the components of $\vec{Z}$ are proportional to the corresponding currents. In the case of two spin-$\frac{1}{2}$ particles U and V each traversing one Stern-Gerlach magnet this leads to the equations[22,23]

$$\dot{Z}_i = g_i(t)\frac{\sum_{m,n}(-1)^k|a_{m,n}|^2|\varphi_1(Z_1-(-1)^m h_1(t))|^2|\varphi_2(Z_2-(-1)^n h_2(t))|^2}{\sum_{m,n}|a_{m,n}|^2|\varphi_1(Z_1-(-1)^m h_1(t))|^2|\varphi_2(Z_2-(-1)^n h_2(t))|^2}, \quad (16)$$

where $i=1, 2$, $k=m$ if $i=1$ and $k=n$ if $i=2$, and where Z_1, h_1 and Z_2, h_2 are respectively the Z coordinates and the shifts at time t of the trajectories of particles 1 and 2 in the directions $\vec{e}_1, \vec{e}_2$ of the two magnets. The function $g_i(t)$ in Eq. (16) is different from zero and positive only during the (short) time interval when particle i traverses the corresponding magnet, $h_i(t)$ and $g_i(t)$ are related by

$$h_i(t) = \int_{-\infty}^{t} g_i(t')dt', \quad (17)$$

and, finally, the coefficients $a_{m,n}$ $(m, n=1, 2)$ are those corresponding to the initial quantum state

$$\psi = \sum_{m,n} a_{m,n} u_m\varphi_1(z_1) \otimes v_n\varphi_2(z_2) \quad (18)$$

of the two-particles system where $u_m\varphi_1(z_1)$ and $v_n\varphi_2(z_2)$ are the two-component spinor wave functions of the particles, where u_m (v_n) are the eigenvectors of the spin component of U (V) along direction $\vec{e}_1$ $(\vec{e}_2)$ (notice that the spins of U and V are thus quantized along different directions).

Let us consider the case in which the two particles U and V are in a state of total spin zero. With the foregoing definition of u_m and v_n this gives

$$a_{11} = a_{22} = \sin\theta, \qquad a_{12} = -a_{21} = \cos\theta. \quad (19)$$

Let the two measurements be well separated in time (as well as in space) and let the one on U take place first. Let $(t_1, t_1+\Delta t_1)$ be the corresponding time interval. $h_2(t)$ and $g_2(t)$ are zero during that time so that Eq. (16) reduces to

$$\dot{Z}_1 = g_1(t)\frac{-|\varphi_1(Z_1+h_1(t))|^2 + |\varphi_1(Z_1-h_1(t))|^2}{|\varphi_1(Z_1+h_1(t))|^2 + |\varphi_1(Z_1-h_1(t))|^2}, \quad (20)$$

$$\dot{Z}_2 = 0. \quad (21)$$

After that first measurement is completed, $h_1(t)$ takes the constant value H_1 given by

$$H_1 = \int_{t_1}^{t_1+\Delta t_1} g_1(t')dt', \quad (22)$$

so that as regards the second measurement $[g_1(t)=0]$, Eq. (16) reduces to

$$\dot{Z}_1 = 0, \quad (23)$$

$$\dot{Z}_2 = g_2(t)\frac{N}{D}, \quad (24)$$

where

$$N = |\varphi_1(Z_1+H_1)|^2[-\sin^2\theta|\varphi_2(Z_2+h_2(t))|^2 + \cos^2\theta|\varphi_2(Z_2-h_2(t))|^2] + |\varphi_1(Z_1-H_1)|^2[-\cos^2\theta|\varphi_2(Z_2+h_2(t))|^2 + \sin^2\theta|\varphi_2(Z_2-h_2(t))|^2] \quad (25)$$

and where D is obtained from N by deleting the minus signs before $\sin^2\theta$ and $\cos^2\theta$.

Let the wave packets φ be strongly peaked around zero. Equation (20) then shows that if Z_1 is initially very small and positive (negative) it follows the behavior of $h(t)$ $(-h(t))$, that is, it takes essentially the value $+H_1$ $(-H_1)$ after the first measurement. Under such conditions either $\varphi_1(Z_1+H_1)$ or $\varphi_1(Z_1-H_1)$ is negligibly small so that Eqs. (24) and (25) simplify. In the case $\theta=0$ they show[22] that after the second measurement the value of Z_2/H_2 is necessarily opposite to that of Z_1/H_1. A comparison with the behavior of Z_1 shows, therefore, that the behavior of the instrument coordinate Z_2 is quite different from what it would have been if the first measurement had not been made.

In the general case $\theta\neq 0$ the time evolution of Z_2 is complicated. It depends partly on the initial value of Z_2 and partly on whether $Z_1=\pm H_1$; the relative "weight" of these two determinations being dependent on θ, that is, on the orientation of the magnet that interacted with U beforehand.

Let us now locate the point that makes the derivation of the Bell inequalities described in Sec. IV inapplicable to the dBBB model theory.

Considering again a spin-zero $U+V$ system on which a measurement of $S^{(U)}(\vec{e})$ is made at time t_1 by means of a Stern-Gerlach magnet, we could first argue as follows: Let us assume that initially both Z_1 and Z_2 have very small *positive* values. If $S^{(V)}(\vec{e})$ were measured at $t_a < t_1$, the foregoing calculation shows that a positive result would be obtained with certainty. On the other hand, the same calculation shows that if that measurement were made at $t_b > t_1$, a negative result would be obtained with certainty. In other words (see definition 1) v is true before t_1 and false after t_1; hence assumption (2) is violated.

But, in fact, the reasons why the proof of Sec. IV does not apply here are even more radical. If we allow for the possibility of indirect measurements, as we should, then assumption 1 (and definition 1) cannot even be *applied* to v since in the present case a direct method of measurement and an indirect one [a measurement of $S^{(U)}(\vec{e})$] would give opposite results.

Theory 3: A model inspired by Bohr's approach

Bohr repeatedly stressed (a) that he did not consider the wave-packet reduction as an objective process and (b) that he considered quantum mechanics as a complete theory. These two apparently contradictory statements are reconciled if the view is accepted according to which microsystems never possess any definite properties. Those we attribute to them at a given time t are mere conventions, the convenience of which depend on the experimental setup.

Should we incorporate in that setup the instruments with which the system will interact later than t? Several of Bohr's statements and, in particular, his answer to Einstein, Podolsky, and Rosen criticism seem to be based on a positive answer to that question (see, e.g., Ref. 24 for an unfolding of that point). The list of the "properties" of a system of which we may conveniently think as having definite (though unknown) values is determined by our choice of our instruments of observation. It changes if we change the latter.

Because of the purely conventional nature of the attribution of properties to microsystems, Bohr's theoretical framework distinctly contradicts assumption 1 of Sec. IV and the violation of the Bell inequalities by conventional quantum mechanics can, in the spirit of the foregoing analysis, be attributed to that fact. On the other hand, for the sake of illustrating the interplay of the assumptions of Sec. IV, we may also consider a naive and semi-realistic model inspired by Bohr's approach. In that model, the microsystems would really possess *some* definite properties, some of which could be unknown, and the list of the latter (i.e., of those that are definite *and* unknown) would be determined by the nature of the instruments with which the system will later interact.

In such a model theory let us consider again the system constituted by (a) a source F emitting spin-zero $U+V$ systems, (b) these $U+V$ systems, and (c) a Stern-Gerlach magnet oriented along $\vec{e}_i$ and interacting with U at a time t_1. According to the foregoing rules, $S^{(U)}(\vec{e}_i)$ and $S^{(V)}(\vec{e}_i)$ have definite values all the time after the emission time t_0 on every U and V system and these values are merely revealed by the measurement device. After t_0 the proposition v_i is persistent and V is free as regards spin. Assumptions 1 and 2 are satisfied. Why is it then that the model violates the Bell inequalities (as it must since it reproduces the quantum-mechanical predictions)? The reason is that it violates assumption 3 (inductive causality), as is obvious if at time $t < t_1$ we consider the ensemble E_1 of "the V particles just considered" as a subensemble of an ensemble E of V particles similarly produced but with no magnet present.

VI. DISCUSSION

One feature of the present derivation is that it makes no use of the concept of a probability density in a space of hidden variables.[25] In fact, it does not *assume* the existence of hidden variables. But it derives it for the particular case of the spin-zero $U+V$ systems, and only for that case, from the specific assumptions made [these variables are then the definite values of $S^{(U)}(\vec{e}_i)$ and $S^{(V)}(\vec{e}_i)$ for every i]. In that respect the derivation is similar to, in particular, the very first derivation of the Bell inequalities.[2] The difference is that the present derivation is not based on a principle of "locality" or "separability." In fact, it is based on assumptions that sound plausible only because of our implicit belief in some kind of a principle of locality, so that the difference may appear slight. It is significant nevertheless. The principle of locality commonly used for deriving the Bell inequalities can only be expressed in terms of absences of "influences," and its relationships with the principle of microcausality (so basic to field theory) is therefore not entirely clear. It may be hoped that by multiplying the investigations on the possible origins of the Bell inequalities some new insight will eventually be gained on these matters. Since such problems as quark confinement focus again the attention of the physicists on possible action-at-a-distance physical effects, it is not strictly inconceivable that such investigations should be prolonged in more conventional physics.

Finally, a comment is in order as regards the relationship of the present derivation with that of Stapp. In spite of the replacement of assumption 3 of I, which was of a contrafactual nature, by the present *inductive causality* assumption, which is not, it could be argued that the derivation described here in Sec. IV is still based, at least implicitly, on some other contrafactual assumptions (see footnote 16). And since the Stapp derivation is so simple, the question could then still be asked whether, compared to the latter, the derivation of Sec. IV contributes any original information.

In our opinion, the answer is that it does, because of the fact that the contrafactual assumptions used in the two derivations are of a different nature. The simplest way to grasp this is to imagine a refined species of Maxwell demons who could in any circumstances measure any property a system has, without in the least disturbing it. For such demons the assumptions of Sec. IV are purely operational, whereas those of Stapp's derivation remain contrafactual in general. In that connection, let it be stressed that assumption 3 of I would also remain contrafactual for the demons, as shown by the way in which it is used in I. This is the precise reason why it is replaced in this article by inductive causality.

ACKNOWLEDGMENTS

It is a pleasure to thank J. S. Bell, H. Ekstein, P. Moldauer, A. Shimony, and H. Stapp for highly stimulating discussions.

---

*Laboratoire associé au Centre National de la Recherche Scientifique. Postal address: Bâtiment 211, Université de Paris—Sud, 91405 Orsay, France.

[1]A review of these results and a comprehensive list of references will be published in Progress in Scientific Culture, Erice, Italy.

[2]J. S. Bell, Physics 1, 195 (1964); in *Foundations of Quantum Mechanics, Proceedings of the International School of Physics "Enrico Fermi" Course 49*, edited by B. d'Espagnat (Academic, New York, 1971).

[3]B. d'Espagnat, Phys. Rev. D 11, 1424 (1975).

[4]H. P. Stapp, Berkeley Report No. LBL 5333, 1976 (unpublished).

[5]These subensembles do not necessarily obey the rules of quantum mechanics. The assumption that among the ensembles that may be contemplated conceptually some are *nonquantum*, has been discussed in particular by J. S. Bell (Ref. 11). Within the context of measurement theory it has also been considered, in particular by P. Moldauer, who calls them "improper ensembles" [Epistem. Lett. September, 2 (1976)].

[6]G. Birkoff and J. von Neumann, Ann. Math. 37, 823 (1936).

[7]J. M. Jauch, *Foundations of Quantum Mechanics* (Addison-Wesley, Reading, Mass., 1968).

[8]J. F. Clauser, Am. J. Phys. 39, 1095 (1971).

[9]J. F. Clauser and M. A. Horne, Phys. Rev. D 10, 526 (1974).

[10]J. von Neumann, *Mathematical Foundations of Quantum Mechanics* (Princeton Univ. Press, Princeton, 1955).

[11]J. S. Bell, Rev. Mod. Phys. 38, 447 (1966).

[12]M. Mugur-Schächter, Epistem. Lett. March, 1 (1976).

[13]This implies that a system U can be identified at different times as being one and the same system.

[14]J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, Phys. Rev. Lett. 23, 880 (1969).

[15]The expression "with certainty" is often objected to on the ground that it is allegedly redundant (either the result "yes" would be obtained or it would not). However, it is not redundant in the present context. It expresses the idea that the fact of belonging to $F(a)$ is a property of S that has an existence in itself, independently of whether or not it is known by anybody and independently of whether or not it will eventually be verified by means of some measurement. Assumption 1 explicitly *assumes* that such an idea is meaningful.

[16]Assumption 1 makes use of the conditional implication (if a were measured the result *would* etc.). Therefore so does definition 1. This again might be criticized, particularly since we are led to consider factual situations in which such a measurement cannot really be done without destroying the conditions that are assumed to prevail. Implicitly, therefore, assumption 1 makes use of the idea that *some* contrafactual assertions *may* be meaningful in particular cases. Our answer to that criticism would be that these particular contrafactual assertions are unavoidable since they are an inherent part of the assumptions we want to test. See Sec. VI for further comments.

[17]In I a fourth assumption is explicitly stated to the effect that if a proposition a bears exclusively on a subsystem of a system, then if it is true on the subsystem it is also true on the system and conversely. That assumption would be a mere tautology, were it not for the fact that in some cases instruments can be conceived of that measure a indirectly, by operating on the larger system. An example of the truth of assumption 4 is: Let the larger system be made of a gun and a bullet and let the subsystem be the bullet. Let proposition a be "the speed *of the bullet* is v." If that proposition is true as regards the bullet, it is also true as regards the larger system. Such an assumption is considered here as being valid whether or not we take its content for granted (as we may well do since it goes together with the very notion of property of a system).

[18]It should be clear that the difference is significant. It is a purely experimental fact that whenever $S^U(\vec{e}_i)$ is measured on the systems S of E_1, property b is possessed by these S immediately afterwards. Ordinary induction then tells us that property b would also be possessed by any other system S if $S^{(U)}(\vec{e}_i)$ were measured on it. This trivial result is quite independent of whether or not assumption 2 is made. On the other

hand, if assumptions 2 and 3 are *not* made, there is no reason for a to be correct before the measurement in question and for systems S on which no measurement is ever made. Ordinary quantum mechanics, with no hidden variables, is a counterexample to the validity of statement a at time t since in that theory ensemble $\mathcal{S}$ is a pure case.

[19]The numbers $n(\sigma_1, \sigma_2, .)$ are (approximately) equal on all three ensembles because of the law of large numbers. The same holds true for $n(\sigma_1, ., \sigma_3)$ and for $n(., \sigma_2, \sigma_3)$.

[20]L. de Broglie, J. Phys. 5, 225 (1927).

[21]D. Bohm, Phys. Rev. 85, 166 (1952); 85, 180 (1952).

[22]J. S. Bell, contribution to Penn State Univ. Conference, 1971 [CERN Report No. TH 1424, 1971 (unpublished)].

[23]The model is idealized in that the velocities—not the accelerations—are proportional to the magnetic forces applied in the instruments.

[24]B. d'Espagnat, *Conceptual Foundation of Quantum Mechanics*, 2nd edition (Benjamin, Reading, Mass., 1976). References to original works may be found there.

[25]That feature belongs also both to Stapp's proof and to an unpublished proof by M. Kupczynski (private communication from J. S. Bell).