Foundations of Physics, Vol. 33, No. 10, October 2003 (© 2003)

Nonlocal Hidden-Variable Theories and Quantum Mechanics: An Incompatibility Theorem

A. J. Leggett[1]

Received February 28, 2003; revised May 29, 2003

It is argued that among possible nonlocal hidden-variable theories a particular class (called here "crypto-nonlocal" or CN) is relatively plausible on physical grounds. CN theories have the property that (for example) the two photons emitted in an atomic cascade process are indistinguishable in their individual statistical properties from photons emitted singly, and that in the latter case the effects of nonlocality are unobservable. It is demonstrated that all CN theories are constrained by inequalities which are violated by the quantum-mechanical predictions; these inequalities bear no simple relation to Bell's inequalities, and an explicit example is constructed of a CN theory which violates the latter. It is also shown that while existing experiments cannot rule out general CN theories, they do rule out (subject to a few caveats such as the usual ones concerning the well-known "loopholes") the subclass in which the photon polarizations are linear.

KEY WORDS: quantum mechanics; hidden-variable theories; nonlocality.

1. INTRODUCTION

Bell's celebrated theorem[(1)] states that, in a situation like that considered by Einstein *et al.*,[(2)] which involves the correlation of measurements on two spatially separated systems which have interacted in the past, no *local* hidden-variable theory (or more generally, no objective local theory) can predict experimental results identical to those given by standard quantum mechanics. Over the past thirty years a very large number of experiments have been conducted with the aim of testing the predictions of quantum mechanics against those of local hidden-variable theories, and while to the

[1] Department of Physics, University of Illinois, 1110 West Green Street, Urbana, Illinois 61801-3080; e-mail: aleggett@uiuc.edu

best of my knowledge no single existing experiment has simultaneously blocked all of the so-called "loopholes" (detector efficiency, random choice of setting, etc.), each one of those loopholes has been blocked in at least one experiment (cf., e.g., Weihs *et al.*[3]). Thus, to maintain a *local* hidden-variable theory in the face of the existing experiments would appear to require belief in a very peculiar conspiracy of nature.

In this paper I argue that, among possible *nonlocal* hidden-variable theories, a certain class of theories which I call "crypto-nonlocal" (CN) is *a priori* relatively plausible on physical grounds, and then demonstrate that any theory of this class must likewise give experimental predictions in conflict with those of quantum mechanics. I also show that, while existing experimental results are not yet adequate to rule out the whole class of such theories, they do rule out a certain sub-class of them. For simplicity of exposition I restrict consideration here to the case of primary experimental interest, namely the emission of two photons, whose (linear or elliptical) polarization is to be measured, by an atom in a cascade process. In this case the defining characteristic of a CN theory is, very crudely speaking, that the cascade process can be described in terms of the emission of pairs of photons each of which "behaves like" a photon emitted in a standard one-photon radiation process, even though the response of (all) photons to a detector involves nonlocal effects. More precisely, the ensemble of pairs of photons emitted in the cascade process can be regarded as a disjoint union of subensembles corresponding to the emission of two photons, each of which has a definite polarization vector and behaves (statistically) exactly as if it had been emitted in a single-photon process. (This is made more precise in Sec. 2.) Evidently, the most general CN theory would allow the "constituent" photons to be elliptically polarized; it is useful for purposes of exposition to define a subclass (let us call it subclass L) in which they are constrained to be linearly polarized.

The reader might well ask why the whole subject of nonlocal hidden-variable theories is of any interest. In my view, the point of considering such theories is not so much that they are in themselves a particularly plausible picture of physical reality, but that by investigating their consequences one may attain a deeper insight into the nature of the quantum-mechanical "weirdness" which Bell's theorem exposes. In particular, I believe that the results of the present investigation provide quantitative backing for a point of view which I believe is by now certainly well accepted at the qualitative level, namely that the incompatibility of the predictions of objective local theories with those of quantum mechanics has relatively little to do with locality and much to do with objectivity.

In Sec. 2 I give a more precise definition of the concept of crypto-nonlocal theory, and introduce the necessary notation (which follows that of Bell's[1] original paper as far as possible). In Sec. 3 I consider an idealized situation in which polarizers and detectors are 100% efficient, etc., and show that under those conditions the subclass L of CN theories makes predictions incompatible with those of quantum mechanics; in Sec. 4 I extend the proof to the whole class of CN theories. In Sec. 5 I give an explicit example of a (subclass-L) CN theory which violates Bell's inequalities and hence is not trivially excluded by existing experiments (though see below). In Sec. 6 I briefly discuss the experimental situation and show that while existing experiments cannot exclude general CN theories, they can (subject to a few subsidiary assumptions) exclude subclass-L theories. Section 7 discusses the significance of the results.

2. "CRYPTO-NONLOCAL" THEORIES

Consider a source S (in practice an atom, or more precisely, an ensemble of atoms) which emits two photons 1 and 2 in different directions in the course of a cascade process. The photons 1 and 2 impinge respectively on polarizers P_1 and P_2 and, if transmitted, on detectors D_1 and D_2. We will use the term "station 1" to denote the combination of the polarizer P_1 and the detector D_1 and similarly for "station 2." We define, as is conventional in the discussion of Bell's theorem, a variable A which takes the value $+1$ (-1) according as the detector D_1 does (does not) register the arrival of a photon; similarly, the variable B takes values $+1$ (-1) according as D_2 does or does not register. The output of possible theories of the emission and detection process (including, of course, the standard quantum-mechanical theory) is, inter alia, a prediction of the correlation $\langle AB \rangle$, which, if everything else is held constant, is expected to be a function of the way in which the polarizers P_1 and P_2 are set.

In the context of the present discussion it is necessary to be rather explicit as to what we mean by "polarizer" (analyzer). For the sake of simplicity of exposition let us for the moment assume that the polarizers (and detectors) subtend zero solid angle at the source and are located on the positive and negative z-axes. Moreover let us for the present consider only ideally efficient linear polarizers (for the more general case see below, and Sec. 4). Then, most generally, a (linear) polarizer is some physical object (e.g., a calcite crystal) whose orientation is characterized by some real unit vector $\mathbf{c}$ which can lie in any direction in the xy-plane and which has the following property: If behind the polarizer we place a detector, and in front of it a suitably specified source of light (which in general might include other "polarizers") and if, holding everything else fixed, we rotate the polarizer (i.e., the vector $\mathbf{c}$) in the xy-plane, then the number of counts recorded in the detector, divided by the number recorded when the polarizer is absent, is equal to the quantity $(\mathbf{e}\cdot\mathbf{c})^2$, where $\mathbf{e}$ is a fixed real vector lying in the xy-plane. If we were being less self-conscious, we should of course express this result by saying "the probability of a photon with linear polarization $\mathbf{e}$ passing a detector set with its transmission axis in direction $\mathbf{c}$ is $(\mathbf{e}\cdot\mathbf{c})^2$." Note that this thought-experiment therefore also provides an implicit definition of what we mean by "a beam (ensemble) of photons with (linear) polarization $\mathbf{e}$." In the language introduced above, we can say that for such a beam $\langle A\rangle$, the ensemble average of the quantity A, is proportional to $2(\mathbf{e}\cdot\mathbf{c})^2-1$ (it is equal to it if and only if the detector is 100% efficient).

The above discussion is easily extended to the case of less than ideal efficiency, and general (elliptical) polarization. For nonideal polarizers the number of counts recorded in the detector, divided by the number when the polarizer is absent, is given by $\epsilon_m+(\epsilon_M-\epsilon_m)(\mathbf{e}\cdot\mathbf{c})^2$, when ϵ_M and ϵ_m are respectively the maximum and minimum transmittances of the polarizer. This statement is a definition of ϵ_M and ϵ_m as well as (implicitly) of the notion of "a beam of photons with polarization $\mathbf{e}$;" it should be carefully noted that since the definition is framed directly in terms of the number of counts recorded in the detector, we do *not* have to assume that the probability of a photon being counted is independent of whether or not it has passed a polarizer, and moreover the quantities ϵ_M and ϵ_m are, strictly speaking, properties of the whole arrangement (polarizer plus detector) and not of the polarizer alone. However, it is of course an experimental fact that a very large class of detectors yields, for a given polarizer, the same value of the transmittances.

The case of general (elliptical) polarization is handled by allowing the vectors $\mathbf{c}$ and $\mathbf{e}$ to be complex quantities, and replacing $(\mathbf{e}\cdot\mathbf{c})^2$ by $|(\mathbf{e}^*\cdot\mathbf{c})|^2$.

Let us now return to the consideration of the correlation observed in a two-photon cascade process. To simplify the discussion I shall assume for the rest of this and the next section that not only the polarizers but also the detectors are ideally efficient, that they are located on the $\pm z$ axis and subtend zero angle at the source, and also (except when otherwise stated) that all photon polarizations are linear, and hence specified by real vectors lying in the xy-plane. Then the settings of the (linear) polarizers P_1 and P_2 are specified by real unit vectors $\mathbf{a}$ and $\mathbf{b}$ in the xy-plane; the vector $\mathbf{a}$ corresponds to the axis of maximum transmittance (what we called $\mathbf{c}$ above) for polarizer P_1, and $\mathbf{b}$ has a similar meaning for P_2. Then the observed correlation $\langle AB\rangle$ of the counts in the two detectors will be a function of $\mathbf{a}$ and $\mathbf{b}$; let us write, as is conventional

$$\langle AB\rangle \equiv P(\mathbf{a},\mathbf{b}). \tag{2.1}$$

For a given type of cascade process quantum mechanics makes unambiguous predictions for $P(\mathbf{a}, \mathbf{b})$ (see, for example, Belinfante[4]). In particular, for a $0^+ \to 1^- \to 0^+$ cascade (the type used in most experiments on Ca) the prediction is

$$P_{QM}(\mathbf{a}, \mathbf{b}) = 2(\mathbf{a} \cdot \mathbf{b})^2 - 1 \equiv \cos 2\phi, \tag{2.2}$$

when ϕ is the angle between **a** and **b**, while for a $1^- \to 1^- \to 0^+$ cascade (the type used in experiments on Hg) the sign of the right-hand side of (2.2) is reversed.

Let us now turn to a discussion of possible hidden-variable theories of the emission and detection process. A general hidden-variable theory may be characterized by the following properties:

1. Each pair of photons emitted in the cascade of a given atom is characterized by a unique value of some (possibly very complicated) set of "hidden" variables which we schematically label λ.
2. In a given type of cascade process occurring under given physical conditions at the source, the ensemble of pairs of emitted photons is determined by a unique, reproducible statistical distribution of the values of λ which we describe by a normalized distribution function $\rho(\lambda)$. The form of the function $\rho(\lambda)$ depends only on conditions in the neighborhood of the source, and in particular is independent both of the polarizer settings **a**, **b** and of the outputs of D_1 and D_2.
3. For a given pair, the value of the quantities A defined above (i.e., whether or not the photon "1" in question is counted in the detector) is determined by the values of **a**, **b**, and λ, and possibly also by the value of B; similarly, the value of B is determined by **a**, **b**, λ and possibly also by the value of A. Conditions (1)–(3) imply for the measured correlation $P(\mathbf{a} \cdot \mathbf{b})$ the result

$$P(\mathbf{a}, \mathbf{b}) = \int_\wedge dp(\lambda)\, \rho(\lambda)\, A(\mathbf{a}, \mathbf{b}, \lambda : B)\, B(\mathbf{a}, \mathbf{b}, \lambda : A), \tag{2.3}$$

where $\wedge$ denotes the complete space of λ.

Local hidden-variable theories must satisfy also two further conditions:

4.

$$A(\mathbf{a}, \mathbf{b}, \lambda : B) = A(\mathbf{a}, \mathbf{b}, \lambda), \qquad B(\mathbf{a}, \mathbf{b}, \lambda : A) = B(\mathbf{a}, \mathbf{b}, \lambda), \tag{2.4}$$

i.e., the outcome of the measurement of A is independent of the *outcome* at the distant station 2 and vice versa ("outcome-independence," cf. Jarrett[5]).

5.

$$A(\mathbf{a}, \mathbf{b}, \lambda) = A(\mathbf{a}, \lambda), \qquad B(\mathbf{a}, \mathbf{b}, \lambda) = B(\mathbf{b}, \lambda), \tag{2.5}$$

i.e., the outcome of the measurement of A is independent of the *setting* **b** at the different station 2 and vice versa ("setting-independence").

As is well known, any theory which satisfied all the conditions (1)–(5) must predict inequalities for $P(\mathbf{a}, \mathbf{b})$ which are violated by quantum mechanics (Bell,[1] Clauser *et al.*[6]). Moreover, these inequalities are almost certainly violated by existing experimental results (Clauser and Shimony,[7] Weihs *et al.*[3]). Thus it seems very unlikely that a theory which is to give agreement with experiment can maintain all of conditions (1)–(5).

For present purposes I define a "nonlocal hidden-variable theory" as one in which conditions (1)–(4) are maintained (with one proviso; see below) but condition (5) is relaxed. To be sure, it is by no means obvious that this is the most plausible modification one could make, or indeed that it is not somewhat artificial; cf. Sec. 7. There has been some discussion of nonlocal theories, defined as above, in the literature; in particular, Garuccio and Selleri[8] show that if one imposes various (alternative) conditions then such a theory will still satisfy Bell's inequality and related ones. These conditions are formulated directly in terms of the effect of the nonlocality on the correlations, and although they are formally perhaps the simplest one could think of it is not clear whether they have any intuitive physical justification. On the other hand, if one allows the nonlocality in the detection process, i.e., the function $A(\mathbf{a}, \mathbf{b}, \lambda)$, $B(\mathbf{a}, \mathbf{b}, \lambda)$, to be totally arbitrary in character one gets nothing useful (in particular, it is always possible to reproduce the results of quantum mechanics by suitable choices of the functions A and B (Bell[1]). I shall therefore seek a physical motivation for restricting the class of nonlocal theories in a nontrivial way. To obtain the desired motivation, let us first enquire why it is that to many physicists *all* nonlocal hidden-variable theories are *a priori* implausible. Apart from an ingrained prejudice in favor of a local description (which, however, is outraged to at least an equal extent by quantum mechanics itself) the following consideration is probably at least subconsciously important: From the point of view of the system which performs measurements on photon 1 (that is, the polarizer P_1 and detector D_1) the polarizer P_2 is a physical object which is part of the distant environment. But in physics we are normally accustomed to require some positive reason before we accept a particular part of the environment as relevant to the outcome of an experiment. Now the polarizer P_2 is nothing more than (e.g.) a calcite crystal, and nothing in our experience of physics indicates that the orientation of distant calcite crystals is either more or less likely to affect the outcome of an experiment than, say, the position of the keys in the experimenter's pocket or the time shown by the clock on the wall; in particular we know no special causal influence propagating from P_2 (as distinct from the rest of the environment, which for brevity we label E) to P_1, and to assume any such special influence would seem to attach a special importance to the fact that P_2, as distinct from E, has been set up with the particular purpose of measuring correlations with P_1. Such a principle would seem either to call into question our normal ideas about causality and the so-called "arrow of time" (cf. Costa de Beauregard[9]) or introduce a bizarrely anthropomorphic element into physics. Such consequences are too radical to be easily stomached by most physicists. It is probably this line of thinking which sometimes leads people to describe theories which allow the variable A to depend on the setting of the distant polarizer $\mathbf{b}$ as well as on $\mathbf{a}$ and λ as "conspiratorial."

If one wishes to meet this objection while at the same time keeping the consequent radical revisions of basic concepts to a minimum, then the natural way to do it is to deny that there is in fact anything "special" about P_2 as distinct from the rest of the environment, E. In other words, we suppose that nonlocality is actually a quite everyday and universal feature of the world; we should in fact strictly speaking write not $A = A(\mathbf{a}, \mathbf{b}, \lambda : B)$ but

$$A = A(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}, \ldots, \lambda : B), \tag{2.5a}$$

where $\mathbf{c}, \mathbf{d}, \ldots$ are quantities which schematically describe the behavior of E. Such a statement, moreover, should naturally apply not only to measurements of photon polarizations but to any kind of measurement whatsoever.

At this point I shall rather arbitrarily assert assumption (4) (outcome independence). The reason for doing it this not so much that it is particularly "natural" (after all, the outcome at the distant station is just one more variable characterizing the overall "environment"!) but is a purely practical one; if one relaxes (4) it appears unlikely (though I have no rigorous proof) that one can prove anything useful at all, and in particular it appears very likely that one can reproduce the quantum-mechanical results for an arbitrary experiment. Assuming that one has indeed invoked assumption (4), Eq. (2.5a) is modified to

$$A = A(\mathbf{a}; \mathbf{b}, \mathbf{c}, \mathbf{d}, \ldots, \lambda), \tag{2.5b}$$

where of course the list **c**, **d**,... does not include the distant outcome B. (Of course, in the single-photon case to be discussed below, since in general there *is* no "distant station," the dependence not only on B but also on **b** trivially vanishes.)

Clearly, Eq. (2.5b) must be supplemented by extremely strong constraints if it is not to violate our normal, "common-sense" expectations that the results of measurements are reproducible and do not depend on arbitrary parts of the environment. The most natural supplementary postulate is that the forms of the distribution of hidden variables, $\rho(\lambda)$, which we are likely to meet with in ordinary life are just such as to guarantee this result, i.e.,

$$\bar{A}(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d},...) \equiv \int \rho(\lambda)\, A(\mathbf{a}; \mathbf{b}, \mathbf{c}, \mathbf{d},..., \lambda)\, d\lambda = \bar{A}(\mathbf{a}). \tag{2.6}$$

Let us in particular consider the ensemble which we would normally characterize as "a beam of photons with polarization **e**." In a hidden-variable theory this will be characterized by some distribution $\rho(\lambda)$ of the hidden variables, and we then postulate, in accordance with the discussion at the beginning of this section, that it has the property (for ideal detector efficiency)

$$\bar{A}_{\mathbf{e}}(a; b, c, d,...) = 2(\mathbf{e}\cdot\mathbf{a})^2 - 1 \equiv \bar{A}(\mathbf{e}, \mathbf{a}), \tag{2.7}$$

i.e.,

$$\int \rho_{\mathbf{e}}(\lambda)\, A(\mathbf{a} : \mathbf{b}, \mathbf{c}, \mathbf{d},..., \lambda)\, d\lambda = 2(\mathbf{e}\cdot\mathbf{a})^2 - 1. \tag{2.8}$$

It is, of course, an experimental fact that such ensembles can be prepared, and in the present context it is not an entirely trivial remark that, in principle at least, they can be prepared by direct manipulation of atomic states involved, without the need of inserting an intermediate polarizer between the source and P_1. (For example, if one's source is a solid with crystal-field anisotropy, photons of a given frequency will have a polarization vector along a given axis). To the extent that we employ only linear polarizers, all our experimental results on single beams can be explained by assuming that the photon beams we deal with are mixtures of "pure" beams with the property (2.7).

Let us now consider the case in which two photons (or, more precisely, an ensemble of photon pairs) are emitted and may be detected by apparatus 1 and 2 with polarizer settings **a** and **b**, as above. (From now on we drop the explicit reference to the rest of the environment, since it is not relevant to our argument; all the expectation values discussed below are assumed independent of **c**, **d**,...). We may note in passing that the photons examined in a "single-beam" experiment very often are indeed emitted in conjunction with other photons of different frequency and polarization, although of course we do not usually go to the trouble of inserting polarizer P_2 and detector D_2 unless we are interested in measuring correlations in the second beam. We assume that we may (but, of course need not) measure any or all of the expectation values $\bar{A}$, $\bar{B}$, and $\overline{AB}$, the last being a correlation between the responses of the two detectors. Let us consider three cases, describing them as we normally would (again we assume ideal detector efficiency):

(1) Emission of two photons of definite polarizations **u**, **v**, by two different (and by assumption "uncorrelated") atoms. In a hidden-variable theory we would describe this by some distribution $\rho_{\mathbf{u},\mathbf{v}}(\lambda)$ of the hidden variables, and the natural assumption, which is (presumably!) consistent with existing experiments is

$$\bar{A}(\mathbf{u}, \mathbf{v}, \mathbf{a}, \mathbf{b}) \equiv \int \rho_{\mathbf{uv}}(\lambda)\, A(\mathbf{ab}\lambda)\, d\lambda = 2(\mathbf{u}\cdot\mathbf{a})^2 - 1 = \bar{A}(\mathbf{u}, \mathbf{a}), \tag{2.9a}$$

$$\bar{B}(\mathbf{u}, \mathbf{v}, \mathbf{a}, \mathbf{b}) \equiv \int \rho_{\mathbf{uv}}(\lambda)\, B(\mathbf{ab}\lambda)\, d\lambda = 2(\mathbf{v}\cdot\mathbf{b})^2 - 1 = \bar{B}(\mathbf{v}, \mathbf{b}), \tag{2.9b}$$

$$\overline{AB}(\mathbf{u}, \mathbf{v}; \mathbf{a}, \mathbf{b}) = \bar{A}(\mathbf{u}, \mathbf{a})\, \vec{B}(\mathbf{v}, \mathbf{b}), \tag{2.10}$$

where

$$\overline{AB}(\mathbf{u}, \mathbf{v}; \mathbf{a}, \mathbf{b}) \equiv \int \rho_{\mathbf{uv}}(\lambda)\, A(\mathbf{ab}\lambda)\, B(\mathbf{ab}\lambda)\, d\lambda, \tag{2.11}$$

i.e., there is no correlation between the counts in detectors D_1 and D_2.

(2) Emission of two photons of definite polarization **u**, **v** by the same atom (e.g., in a cascade process in a solid). Since the photons emitted in this kind of situation seem, as far as we know, to behave individually exactly like those emitted incoherently (case 1), it is natural to postulate Eq. (2.9) again in this case. However, it is *not* immediately obvious that we should also postulate (2.10); this is a matter for experiment, and I suspect that there have been rather few experiments in which (2.10) has been tested (if indeed

there have been any). However that may be, let us go on to case (3).

(3) Emission of two photons of "indefinite" polarization. This is precisely the situation which occurs in cascade processes in atoms of the type discussed in this paper. Within the context of a hidden-variable theory (as distinct from quantum mechanics) it is natural to regard the total ensemble as the disjoint union of subensembles corresponding to case (2), where we constrain the subensemble averages to obey (2.9) but *not* necessarily (2.10).

The above considerations lead us naturally to define a certain class of nonlocal hidden variable theories, which I shall call "crypto-nonlocal" (CN), as follows. (I change the notation here slightly: $\rho_{\mathbf{uv}}(\lambda) \to g_{\mathbf{uv}}(\lambda)$.) The ensemble of pairs of photons is a disjoint union of subensembles characterized by distribution functions $g_{uv}(\lambda)$ *which have the property (2.9)* (but *not*, in general (2.10)). That is, we can write

$$\rho(\lambda) = \iint F(\mathbf{u}, \mathbf{v})\, g_{\mathbf{uv}}(\lambda)\, d\mathbf{u}\, d\mathbf{v}, \tag{2.12}$$

$$\int_\Lambda g_{\mathbf{uv}}(\lambda)\, d\lambda = \iint F(\mathbf{u}, \mathbf{v})\, d\mathbf{u}\, d\mathbf{v} = 1, \tag{2.13}$$

$$g_{\mathbf{uv}}(\lambda) \geqslant 0, \qquad F(\mathbf{u}, \mathbf{v}) \geqslant 0, \tag{2.14}$$

and so

$$P(\mathbf{a}, \mathbf{b}) \equiv \langle AB \rangle = \iint F(\mathbf{u}, \mathbf{v})\, \overline{AB}(\mathbf{u}, \mathbf{v}; \mathbf{a}, \mathbf{b})\, d\mathbf{u}\, d\mathbf{v}, \tag{2.15}$$

where $\overline{AB}(\mathbf{u}, \mathbf{v}; \mathbf{a}, \mathbf{b})$ is given by (2.11) (but in general does not satisfy (2.10)). Here and in the following I use pointed brackets to indicate the average over the whole ensemble and bars for the subensemble averages.

It should be pointed out that assumption (2.9) actually constrains the value of the ensemble averages $\langle A \rangle$, $\langle B \rangle$; since any sum of expressions of the form $(\mathbf{u} \cdot \mathbf{a})^2$ with different $\mathbf{u}$ can be written as $\alpha(\mathbf{e} \cdot \mathbf{a})^2 + \beta$, where $\mathbf{e}$ is a fixed unit vector, it follows that there exists some α' and $\mathbf{e}$ such that

$$\langle A \rangle = \alpha'[2(\mathbf{e} \cdot \mathbf{a})^2 - \beta']. \tag{2.16}$$

This constraint is of course automatically satisfied by the experimental distributions with which we are familiar (given, as always, ideal detectors); α' may of course be zero.

In the above discussion we have explicitly assumed that the subensembles described by $g_{\mathbf{uv}}(\lambda)$ correspond to *linear* photon polarizations **u**, **v**. This is clearly insufficiently general; in fact, what we have just defined is the "subclass L" of crypto-nonlocal hidden-variable theories. To obtain a general CN theory, it is necessary to allow the subensembles to describe general (elliptical) photon polarizations. This can be done simply by allowing the vectors $g_{\mathbf{u}}$ and **v** to be complex unit vectors in the xy-plane, and generalizing the right-hand side of (e.g.) Eq. (2.9a) to read $2\,|\mathbf{u}^* \cdot \mathbf{a}|^2 - 1$, where the vectors **ab** are now also in general complex, corresponding to polarizers set to accept a given elliptical polarization.

The fundamental result of this paper is that once the subensemble distribution functions $g_{\mathbf{uv}}(\lambda) \equiv P_{\mathbf{uv}}(\lambda)$ are constrained to satisfy (2.9) (or its generalization), the values of $P(\mathbf{a}, \mathbf{b})$ predicted by any nonlocal hidden-variable theory of the type described are incompatible with those predicted by quantum mechanics. This is proved for subclass L in the next section and for the general case in Sec. 4.

3. INCOMPATIBILITY OF SUBCLASS—L THEORIES WITH QUANTUM MECHANICS

The subclass (L) of nonlocal separable hidden-variable theories corresponding to linear photon polarizations is defined by the postulates (Eqs. (2.13), (2.11), and (2.9)):

$$P(\mathbf{a}, \mathbf{b}) = \iint F(\mathbf{u}, \mathbf{v})\, \overline{AB}(\mathbf{u}, \mathbf{v}; \mathbf{a}, \mathbf{b})\, d\mathbf{u}\, d\mathbf{v}, \tag{3.1}$$

$$\overline{AB}(\mathbf{u}, \mathbf{v}; \mathbf{a}, \mathbf{b}) \equiv \int g_{\mathbf{uv}}(\lambda)\, A(\mathbf{a}, \mathbf{b}, \lambda)\, B(\mathbf{a}, \mathbf{b}, \lambda)\, d\lambda, \tag{3.2}$$

$$\bar{A}(\mathbf{u}, \mathbf{v}; \mathbf{a}, \mathbf{b}) \equiv \int g_{uv}(\lambda)\, A(\mathbf{a}, \mathbf{b}, \lambda)\, d\lambda = 2(\mathbf{u} \cdot \mathbf{a})^2 - 1, \tag{3.3a}$$

$$\bar{B}(\mathbf{u}, \mathbf{v}; \mathbf{a}, \mathbf{b}) \equiv \int g_{uv}(\lambda)\, B(\mathbf{a}, \mathbf{b}, \lambda)\, d\lambda = 2(\mathbf{v} \cdot \mathbf{b})^2 - 1, \tag{3.3b}$$

where **u**, **v**, **a**, **b** are real vectors lying in the same plane and where the weight functions $g_{\mathbf{uv}}(\lambda)$ and $F(\mathbf{u}, \mathbf{v})$ are positive and satisfy the normalization conditions

$$\int g_{\mathbf{uv}}(\lambda)\, d\lambda = \iint F(\mathbf{u}, \mathbf{v})\, d\mathbf{u}\, d\mathbf{v} = 1, \tag{3.4}$$

and $A(\mathbf{a}, \mathbf{b}, \lambda)$ and $B(\mathbf{a}, \mathbf{b}, \lambda)$ are functions taking the values ± 1. We will show in this section that Eqs. (3.1)–(3.4) predict inequalities for $P(\mathbf{a}, \mathbf{b})$ which are violated by the quantum-mechanical expressions, e.g., Eq. (2.2).

Although the theorem to be proved is of course independent of the precise definition of the integrals $d\mathbf{u}$ and $d\mathbf{v}$, it is convenient to have a standard definition. Since $\mathbf{u}$ and $\mathbf{v}$ are real unit vectors in the xy-plane, they are characterized by a single angle, say θ_u and θ_v, relative to some standard reference axis. We therefore define

$$d\mathbf{u} \equiv \frac{d\theta_u}{2\pi}, \qquad d\mathbf{v} \equiv \frac{d\theta_v}{2\pi}, \tag{3.5}$$

where the factor $(2\pi)^{-1}$ is included for convenience.

The proof is based on the following simple observation. Let C, D be quantities which can take the values ± 1 only and let $\bar{C}$, $\bar{D}$ be their averages with respect to some positive normalized weight function; let $\overline{CD}$ similarly be the average of this product. Then, by explicitly invoking the fact that the number of cases corresponding to a particular outcome (e.g., $C = +1$, $D = -1$) cannot be negative, we can easily demonstrate the inequalities:

$$-1 + |\bar{C} + \bar{D}| \leqslant \overline{CD} \leqslant 1 - |\bar{C} - \bar{D}|. \tag{3.6}$$

We apply this result to the variables A and B, with weight function $g_{\mathbf{uv}}(\lambda)$ as in Eqs. (3.2) and (3.3). We then insert the resulting inequalities in (3.1); since $F(\mathbf{u}, \mathbf{v})$ is by hypothesis positive, we get

$$\begin{aligned} &-1 + 2 \iint d\mathbf{u}\, d\mathbf{v}\, F(\mathbf{u}, \mathbf{v})\, |(\mathbf{u}\cdot\mathbf{a})^2 + (\mathbf{v}\cdot\mathbf{b})^2 - 1| \\ &\qquad \leqslant P(\mathbf{a}, \mathbf{b}) \leqslant 1 - 2 \iint d\mathbf{u}\, d\mathbf{v}\, F(\mathbf{u}\cdot\mathbf{v})\, |(\mathbf{u}\cdot\mathbf{a})^2 - (\mathbf{v}\cdot\mathbf{b})^2|. \end{aligned} \tag{3.7}$$

Let us now introduce angles θ_a and θ_b characterizing the orientation of $\mathbf{a}$ and $\mathbf{b}$, and further define

$$\begin{aligned} \xi &\equiv \frac{\theta_a + \theta_b}{2}, \qquad \varphi \equiv \theta_a - \theta_b, \\ \Psi &\equiv \frac{\theta_u + \theta_v}{2}, \qquad \chi \equiv \theta_u - \theta_v. \end{aligned} \tag{3.8}$$

Expressing the inequalities (3.7) in terms of these angles and using standard trigonometric identities, we obtain (writing $P(\mathbf{a}, \mathbf{b}) \equiv P(\xi, \varphi)$)

$$-1+2\iint \frac{d\Psi}{2\pi}\int \frac{d\chi}{2\pi}\, F(\Psi, \chi)\,|\cos 2(\xi-\Psi)|\,|\cos(\varphi-\chi)|$$

$$\leqslant P(\xi, \varphi) \leqslant 1-2\int \frac{d\Psi}{2\pi}\int \frac{d\chi}{2\pi}\, F(\Psi, \chi)\,|\sin 2(\xi-\Psi)|\,|\sin(\varphi-\chi)|. \quad (3.9)$$

Now let us integrate the inequalities (3.9) over ξ and then perform the Ψ-integration. We use the fact that

$$\int |\cos 2(\xi-\Psi)|\,\frac{d\xi}{2\pi} = \int |\sin 2(\xi-\Psi)|\,\frac{d\xi}{2\pi} = \frac{2}{\pi}, \quad (3.10)$$

and define

$$\int P(\xi, \varphi)\,\frac{d\xi}{2\pi} \equiv \bar{P}(\varphi), \quad (3.11)$$

$$\int F(\Psi, \chi)\,\frac{d\Psi}{2\pi} \equiv \rho(\chi), \quad (3.12)$$

so that from (3.4), $\rho(\chi)$ is a normalized weight function:

$$\int \rho(\chi)\,\frac{d\chi}{2\pi} = 1. \quad (3.13)$$

Then we have

$$-1+\frac{4}{\pi}\int \frac{d\chi}{2\pi}\,\rho(\chi)\,|\cos(\varphi-\chi)| \leqslant \bar{P}(\varphi) \leqslant 1-\frac{4}{\pi}\int \frac{d\chi}{2\pi}\,\rho(\chi)\,|\sin(\varphi-\chi)|. \quad (3.14)$$

Now we add and subtract the inequalities (3.14) for $\bar{P}(\varphi)$ and $\bar{P}(\varphi')$, and use the fact that

$$|\sin(\varphi-\chi)|+|\sin(\varphi'-\chi)| \geqslant |\sin(\varphi-\varphi')|, \quad (3.15a)$$

$$|\cos(\varphi-\chi)|+|\cos(\varphi'-\chi)| \geqslant |\sin(\varphi-\varphi')|, \quad (3.15b)$$

$$|\sin(\varphi-\chi)|+|\cos(\varphi'-\chi)| \geqslant |\cos(\varphi-\varphi')|. \quad (3.15c)$$

In this way we obtain our final inequalities for a subclass-L theory:

$$|\bar{P}(\varphi)+\bar{P}(\varphi')| \leqslant 2-\frac{4}{\pi}|\sin(\varphi-\varphi')|, \tag{3.16a}$$

$$|\bar{P}(\varphi)-\bar{P}(\varphi')| \leqslant 2-\frac{4}{\pi}|\cos(\varphi-\varphi')|, \tag{3.16b}$$

where $\bar{P}(\varphi)$ is defined by Eq. (3.14). The inequalities (3.16) are clearly violated by the quantum-mechanical expression (2.2) (for which $\bar{P}(\varphi)\equiv P(\varphi)=\cos 2\varphi$); e.g., (3.16a) is violated for $\varphi=0$ and φ' small. Thus the incompatibility is proved.

4. INCOMPATIBILITY OF GENERAL CNHV THEORIES WITH QUANTUM MECHANICS

It is clear that the arguments of the last section fail as soon as we allow the pair-wise emitted photons to have general (elliptical) polarizations. In fact, if we simply consider an ensemble which is the disjoint union of two subensembles, in one of which both photons are right-circularly polarized and in the other both left-circularly polarized, then not only do we reproduce the quantum-mechanical predictions for correlations of circular polarization (for a $0^+\to 0^+$ transition) but the result of the generalization of the arguments of Sec. 3 for correlations of linear polarization is completely vacuous. Thus it is clear that any useful generalization of these arguments must refer to measurements of "nontrivially elliptical" (i.e., neither circular nor linear) components of polarization.

While it would be perfectly possible to carry out the required generalization explicitly in terms of (complex) polarization in the transverse plane, one's grasp of the geometrical structure underlying the argument is much assisted by performing the well-known mapping of this problem on to that of a spin-$\frac{1}{2}$ particle. In such a mapping a general state of linear polarization corresponds to an eigenstate of $\tilde{\boldsymbol{\sigma}}\cdot\hat{\mathbf{n}}$ where the (real) unit vector $\hat{\mathbf{n}}$ is restricted to lie in the xy-plane, while arbitrary elliptical polarization corresponds to an arbitrary direction of $\hat{\mathbf{n}}$ (circular polarization corresponding to $\hat{\mathbf{n}}=\pm\hat{\mathbf{z}}$. The natural definition of the quantity A of Sec. 2 is now the value of the projection $\tilde{\boldsymbol{\sigma}}\cdot\hat{\mathbf{a}}(=\pm 1)$, so that the quantum-mechanical expectation value of $\langle AB\rangle$ analogous to (2.2) is now $\cos\varphi(\equiv\mathbf{a}\cdot\mathbf{b})$ rather than $\cos 2\varphi$; similarly, for a "subclass-L" CNHV theory (in which all the relevant spins are constrained to lie in the xy-plane) all angles in the argument of Sec. 3 are simply replaced by half-angles, so that (e.g.) the analog of Eq. (3.16a) is

$$|\bar{P}(\varphi)+\bar{P}(\varphi')| \leqslant 2-\frac{4}{\pi}|\sin(\varphi-\varphi')/2|, \tag{4.1}$$

and so on.

Consider now a general CNHV theory in the "spin-$\frac{1}{2}$" representation, so that the "spins" **u, v** of the pairwise emitted particles (as well as the measurement axes **a, b**) are allowed to be in arbitrary directions. Quite generally, the analog of Eq. (3.7) in the spin-$\frac{1}{2}$ representation is

$$-1+2\iint d\mathbf{u}\,d\mathbf{v}\,F(\mathbf{u},\mathbf{v})\,|(\mathbf{u}\cdot\mathbf{a})+(\mathbf{v}\cdot\mathbf{b})|$$

$$\leqslant P(\mathbf{a},\mathbf{b}) \leqslant 1-2\iint d\mathbf{u}\,d\mathbf{v}\,F(\mathbf{u},\mathbf{v})\,|\mathbf{u}\cdot\mathbf{a}-\mathbf{v}\cdot\mathbf{b}|. \tag{4.2}$$

Let us now choose a specific plane (which we can without loss of generality take to be the xy-plane) and constrain the real unit vectors **a**, **b** to lie in this plane. Let $\lambda_z \equiv (1-(\mathbf{u}\cdot\hat{\mathbf{z}})^2)^{\frac{1}{2}}$ and $\mu_s \equiv (1-(\mathbf{v}\cdot\mathbf{z})^2)^{\frac{1}{2}}$ be the magnitudes of the projections of **u** and **v** respectively on this plane. Then it is clear that for any particular values of the λ_z and μ_z we can generalize each step of the argument of Sec. 3; for example, the generalization of Eq. (3.14) consists in the replacements

$$|\cos(\varphi-\chi)| \rightarrow (\Sigma_z^2\cos^2\{(\varphi-\chi)/2\}+\Delta_z^2\sin^2\{(\varphi-\chi)/2\})^{\frac{1}{2}}, \tag{4.3a}$$

$$|\sin(\varphi-\chi)| \rightarrow (\Sigma_z^2\sin^2\{(\varphi-\chi)/2\}+\Delta_z^2\cos^2\{(\varphi-\chi)/2\})^{\frac{1}{2}}, \tag{4.3b}$$

where Σ_z and Δ_z are defined by

$$\Sigma_z \equiv \tfrac{1}{2}(\lambda_z+\mu_z), \qquad \Delta_z \equiv \tfrac{1}{2}(\lambda_z-\mu_z). \tag{4.4}$$

Using (3.15a–c) (with the obvious "half-angle" replacements) plus the easily proved result that if for positive quantities a, b, c, d, p, q, the inequalities $a+c \geqslant p$ and $b+d \geqslant q$ are satisfied, then $(a^2+b^2)^{\frac{1}{2}}+(c^2+d^2)^{\frac{1}{2}} \geqslant (p^2+q^2)^{\frac{1}{2}}$, we find that the contribution (call it $\bar{P}(\varphi:\lambda_z,\mu_z)$) of the subensemble of pairs with the specified values of λ_z, μ_z to the correlation $\bar{P}(\varphi)$ satisfies the inequalities (the generalization of (3.16a, b))

$$|\bar{P}(\varphi:\lambda_z,\mu_z)+\bar{P}(\varphi':\lambda_z,\mu_z)|\leqslant W(\lambda_z,\mu_z)\left(2-\frac{2}{\pi}(\lambda_z^2+\mu_z^2)^{\frac{1}{2}}|\sin(\varphi-\varphi')/2|\right),\tag{4.5a}$$

$$|\bar{P}(\varphi:\lambda_z,\mu_z)-\bar{P}(\varphi':\lambda_z,\mu_z)|\leqslant W(\lambda_z,\mu_z)\left(2-\frac{2}{\pi}(\lambda_z^2+\mu_z^2)^{\frac{1}{2}}|\cos(\varphi-\varphi')/2|\right),\tag{4.5b}$$

where the quantity

$$W(\lambda_z,\mu_z)\equiv\int\frac{d\theta_u}{2\pi}\int\frac{d\theta_v}{2\pi}F(\mathbf{u},\mathbf{v})\tag{4.6}$$

is the "weight" of this subensemble in the complete ensemble. It is now obvious (since the statement $|a|\leqslant b$ is simply equivalent to the pair of statements $a\leqslant b$, $-a\leqslant b$!) that we can integrate (4.5) over λ_z and μ_z to obtain the final result for the experimentally measurable correlations $\bar{P}(\varphi)$

$$|\bar{P}(\varphi)+\bar{P}(\varphi')|\leqslant 2-\frac{2}{\pi}\overline{(\lambda_z^2+\mu_z^2)^{\frac{1}{2}}}|\sin(\varphi-\varphi')/2|,\tag{4.7a}$$

$$|\bar{P}(\varphi)-\bar{P}(\varphi')|\leqslant 2-\frac{2}{\pi}\overline{(\lambda_z^2+\mu_z^2)^{\frac{1}{2}}}|\cos(\varphi-\varphi')/2|,\tag{4.7b}$$

where the bar indicates an average over the whole ensemble of emitted pairs.

The crucial point, now, is that the inequalities (4.7) must hold for an arbitrary plane. Let us then (for example) consider the inequality (4.7a), set $\varphi'\equiv 0$ and take the vector **a** and **b** defining φ to be first rotated relative to one another around the z-axis and then around the x-axis: let us take the angles of rotation to be equal but for clarity denote them by φ_z and φ_x respectively. Since by the above inequalities we have $(\lambda_z^2+\mu_z^2)^{\frac{1}{2}}+(\lambda_x^2+\mu_x^2)^{\frac{1}{2}}\geqslant\sqrt{2}$, we find by adding the relevant forms of (4.7a) the inequality

$$\bar{P}(\varphi_x)+\bar{P}(\varphi_z)\leqslant 2-\frac{2\sqrt{2}}{\pi}\{|\sin\varphi/2|\},\tag{4.8}$$

where $\varphi\equiv\varphi_x\equiv\varphi_z$. (Needless to say, more general inequalities can be proved, but are not needed for our purposes.)

Finally, in order to transcribe the result (4.8) back into the "polarization" language, it is necessary to define an angle corresponding to the "ellipticity difference" between two possible elliptical polarizations corresponding to the two analyzers. For our purposes it is sufficient to assume that the relevant polarizations both have the form $\cos\theta_j\,\mathbf{a}+i\sin\theta_j\,\mathbf{b}$ ($j=1,2$), where **a** and **b** are a particular pair of real mutually orthogonal unit vectors lying in the xy-plane; then the angle in question is just $\varphi_{el}\equiv\theta_1-\theta_2$. As in Sec. 3, we continue to define φ, for a case in which the polarizations are *real*, as the angle between them. Then the transcription of the inequality (4.8) is

$$\bar{P}(\varphi=\chi)+\bar{P}(\varphi_{el}=\chi)\leqslant 2-\frac{2\sqrt{2}}{\pi}\,|\sin\chi|, \tag{4.9}$$

when as previously the bar over P indicates an average over the relevant "center-of-mass" angle (thus, in the case of elliptical polarizations, over the variable $\frac{1}{2}(\theta_1+\theta_2)$). On the other hand, the quantum-mechanical prediction for the LHS of (4.9) for a $0^+\to 1^-\to 0^+$ transition is

$$\bar{P}_{\mathrm{QM}}(\varphi=\chi)+\bar{P}(\varphi_u=\chi)=2\cos 2\chi, \tag{4.10}$$

which is clearly incompatible with (4.9) over a finite range of small but nonzero χ. This completes the proof that no theory of the CNHV class can give experimental predictions consistent with those of quantum mechanics.

5. AN EXPLICIT EXAMPLE OF A NONTRIVIAL CN THEORY

The above results on general CN theories naturally say nothing about whether or not the theory in question satisfies Bell's inequalities. However, if it does, then the whole exercise is somewhat pointless since any such theory is already made very difficult to maintain by existing experiments. It is therefore important to demonstrate the existence of "nontrivial" CN theories, that is, theories which violate Bell's inequalities and hence cannot be mimicked by any local hidden-variable theory. It is immediately clear that a necessary (but by no means sufficient) condition for a CN theory to be nontrivial is that the subensemble averages fail to satisfy the condition $\overline{AB}=\bar{A}\cdot\bar{B}$; for if they do satisfy it, then the predictions are identical to those given by a quantum-mechanical mixture of states with definite photon polarizations, and the results of such a theory can in turn be reproduced by a local hidden-variable theory. In this section I construct an explicit example of a nontrivial (subclass-L) CN theory; like most such examples, it has a somewhat artificial appearance.

As the set of hidden variables we take unit vectors $\boldsymbol{\lambda}$, $\boldsymbol{\lambda}'$, $\boldsymbol{\mu}$, and $\mathbf{v}$ lying in the xy-plane. We define angle variables as follows:

$$\chi \equiv 2\cos^{-1} \boldsymbol{\lambda}\cdot\boldsymbol{\mu}, \qquad \chi' \equiv 2\cos^{-1} \boldsymbol{\lambda}'\cdot\mathbf{v}, \tag{5.1a}$$

$$\theta \equiv 2\cos^{-1} \boldsymbol{\mu}\cdot\mathbf{a}, \qquad \theta' \equiv 2\cos^{-1} \mathbf{v}\cdot\mathbf{b}, \tag{5.1b}$$

$$\xi \equiv 2\cos^{-1} \boldsymbol{\lambda}\cdot\mathbf{a}, \qquad \xi' \equiv 2\cos^{-1} \boldsymbol{\lambda}'\cdot\mathbf{b}, \tag{5.1c}$$

with sign conventions defined so that

$$\xi \equiv \theta + \chi, \qquad \xi' \equiv \theta' + \chi'. \tag{5.2}$$

As before the angle between the polarizer settings $\mathbf{a}$ and $\mathbf{b}$ is denoted by ϕ. The normalized distribution function $g_{\mathbf{uv}}(\lambda)$ for the subensemble characterized by $\mathbf{u}$ and $\mathbf{v}$ is

$$\begin{aligned} g_{uv}(\lambda) &= \tfrac{1}{2}\,\delta(\boldsymbol{\mu}-\mathbf{u})\,\delta(\mathbf{v}-\mathbf{v})\,\delta(\chi-\chi')\cos\chi, \qquad |\chi| \leqslant \pi/2 \\ &= 0, \qquad |\chi| > \pi/2, \end{aligned} \tag{5.3}$$

in an obvious notation. (In this section, in contrast to Sec. 3, we define the integrals over unit vectors to correspond to (e.g.) $\int d\theta$ rather than $\int d\theta/2\pi$.) We also write

$$F(\mathbf{u}, \mathbf{v}) = (2\pi)^{-1}\,\delta(\mathbf{u}-\mathbf{v}). \tag{5.4}$$

We first write down a simple *local* hidden-variable theory for this model, denoting the relevant values of $A(\mathbf{a}, \mathbf{b}, \lambda)$ and $B(\mathbf{a}, \mathbf{b}, \lambda)$ by $A_0(\mathbf{a}, \lambda)$, $B_0(\mathbf{b}, \lambda)$. We simply put

$$\begin{aligned} A(\mathbf{a}, \mathbf{b}, \lambda) &\equiv A_0(\mathbf{a}, \lambda) = \operatorname{sgn}(\pi/2-\xi), \\ B(\mathbf{a}, \mathbf{b}, \lambda) &\equiv B_0(\mathbf{b}, \lambda) = \operatorname{sgn}(\pi/2-\xi'). \end{aligned} \tag{5.5}$$

Substituting (6.5) and (6.3) into the definitions of A, etc., it is easy to show that

$$\bar{A}(\mathbf{u}, \mathbf{a}) = \cos\theta = 2(\mathbf{u}\cdot\mathbf{a})^2 - 1, \tag{5.6a}$$

$$\bar{B}(\mathbf{v}, \mathbf{b}) = \cos\theta' = 2(\mathbf{v}\cdot\mathbf{b})^2 - 1, \tag{5.6b}$$

in agreement with (2.9) and

$$\overline{AB} = 1 - |\cos\theta - \cos\theta'|. \tag{5.7}$$

Hence, from (3.1), (5.3), and (5.4), we have

$$P(\mathbf{a}, \mathbf{b}) \equiv \langle AB \rangle = 1 - \frac{4}{\pi} |\sin \phi|, \tag{5.8}$$

which evidently satisfies both Bell's inequalities and the inequalities (3.16) as of course it should.

Now we introduce nonlocal effects, as follows. Let us consider for the moment only, values of **a** and **b** which are sufficiently close together, say

$$|\phi| \leqslant \pi/4. \tag{5.9}$$

Consider values of $\boldsymbol{\mu}$ (which, by (5.3) and (5.4) is identical to $\mathbf{v}$ for all cases with non-zero weight) such that the angles θ and $\theta'(\equiv \theta + 2\phi)$ have the same sign (say for definiteness positive) and are both non-zero and less than $\pi/2$. Then there exists a range of angle χ (which by (5.3) is identical to χ' for cases of interest) such that (a) the weight function $g_{uv}(\lambda)$, Eq. (5.3), is not identically zero, and (b) when we make the substitution $\chi \to -\chi$ (reflection in the μ-axis) both A_0 and B_0 (Eq. (5.5)) change sign, say from negative to positive. The limits of this range are $\pi/2$ and the greater of $\pi/2 - \theta$, $\pi/2 - \theta'$. (The reader many convince him/herself of the truth of these statements by drawing a diagram: it is recommended to draw it so that all angles correspond to the double angles defined in (5.1), when of course the angle between **a** and **b** in the diagram will be 2ϕ rather than ϕ.) Call this range Γ, and call its mirror image in the μ-axis $\tilde{\Gamma}$.

Now let us consider the effect of replacing (5.5) by the explicitly nonlocal postulates

$$B(\mathbf{a}, \mathbf{b}, \lambda) \equiv B_0(\mathbf{b}, \lambda), \tag{5.10}$$

$$A(\mathbf{a}, \mathbf{b}, \lambda) \equiv A_0(\mathbf{a}, \lambda) \qquad \text{if} \quad \chi \in \Gamma \text{ or } \tilde{\Gamma}, \tag{5.11}$$

$$A(\mathbf{a}, \mathbf{b}, \lambda) \equiv K(\mathbf{a}, \mathbf{b}, \lambda)\, A_0(\mathbf{a}, \lambda) \qquad \text{if} \quad \chi \in \Gamma \text{ or } \chi \in \tilde{\Gamma}, \tag{5.12}$$

where $K(\mathbf{a}, \mathbf{b}, \lambda) \equiv K(\mathbf{a}, \mathbf{b}, \boldsymbol{\mu} : \chi)$ is an arbitrary function taking values ± 1 subject to the condition

$$K(\mathbf{a}, \mathbf{b}, \boldsymbol{\mu} : \chi) \equiv K(\mathbf{a}, \mathbf{b}, \boldsymbol{\mu} : -\chi). \tag{5.13}$$

Since the replacement $\chi \to -\chi$ preserves the statistical weight $g_{uv}(\lambda)$, it is immediately obvious that the condition (5.13) guarantees that the sub-ensemble average $\bar{A}$ is unchanged; since from (5.10) $\bar{B}$ is obviously also unchanged, the theory defined by (5.10)–(5.13) has by construction the CN property. However, the value of $\overline{AB}$ is clearly changed from (5.7): the change is always negative, and its maximum possible value (correspond to $K(\mathbf{ab}\lambda) = -1$ for all χ in Γ) is given by

$$|\delta_{\max}(\overline{AB})(\mathbf{u}, \mathbf{v} : \mathbf{a}, \mathbf{b})| = 2\int^{\Gamma} g_{uv}(\lambda)\, d\lambda + 2\int^{\tilde{\Gamma}} g_{uv}(\lambda)\, d\lambda$$

$$= 2\int^{\Gamma} \cos\chi\, d\chi = 2(1 - \max(\cos\theta, \cos\theta')). \qquad (5.14)$$

It is easily verified that formula (5.14) holds also when θ and θ' are both negative (and $|\theta|, |\theta'| < \pi/2$). For simplicity we assume that $K(\mathbf{a}, \mathbf{b}, \boldsymbol{\mu} : \chi) \equiv +1$ when either θ or θ' is greater than $\pi/2$. Then substituting (5.14) into (3.1) and using (5.4), we find that the maximum (negative) deviation of $P(\mathbf{a}, \mathbf{b}) \equiv \langle AB \rangle$ from its value (5.8) is

$$|\delta_{\max} P(\mathbf{a}, \mathbf{b})| = \frac{1}{2} - \frac{2\,|\phi|}{\pi} - \frac{\cos 2\phi}{\pi}. \qquad (5.15)$$

Since it is clearly possible to choose $K(\mathbf{a}, \mathbf{b}, \lambda)$ in an arbitrary way to produce negative $\delta P(\mathbf{a}, \mathbf{b})$ with any magnitude less than (5.15), we can write (choosing $K(\mathbf{a}, \mathbf{b} : \lambda) \equiv K(\mathbf{a}, \mathbf{b}, \lambda)$)

$$P(\mathbf{a}, \mathbf{b}) \equiv P(\phi) \equiv \bar{P}(\phi) = 1 - \frac{4}{\pi}\,|\sin\phi| - f(\phi)\left\{\frac{1}{2} - \frac{2\,|\phi|}{\pi} - \frac{\cos 2\phi}{\pi}\right\}, \qquad (5.16)$$

where $f(\phi)$ is an arbitrary function such that $0 \leqslant f(\phi) \leqslant 1$. (We recall that in the above formula $|\phi|$ is restricted to be less than $\pi/4$: for other values we assume for simplicity that all nonlocal effects vanish, i.e., Eqs. (5.5) still hold.) It is obvious that by a suitable choice of $f(\phi)$ we can violate Bell's inequalities. For example, choose $f(\phi) = 0$ for $|\phi| < \epsilon$, $f(\phi) = 1$ for $|\phi| > \epsilon$, where ϵ is some small angle. Then Bell's original inequality, which in this case takes the form

$$1 + P(2\phi) - 2P(\phi) \geqslant 0, \qquad (5.17)$$

is clearly violated if $|\phi| < \epsilon$, $2\,|\phi| > \epsilon$ and ϵ is chosen sufficiently small. (On the other hand, it can be checked that (5.16) does satisfy the inequalities (3.16).) Thus the theory constructed above is indeed a nontrivial CN theory.

6. THE EXPERIMENTAL SITUATION

While most existing experiments on the polarization correlations of photon pairs emitted in atomic cascade decays measure the correlation of linear polarizations, there are a few (e.g., Torgerson *et al.*[10]) which examine the circular-polarization correlations. However, to the best of my knowledge there is no single experiment to date which measures the correlations in a general "nontrivially elliptical" basis. Consequently, it is impossible to use existing experimental data in conjunction with the results of Sec. 4 to exclude theories of the general CNHV class.

It is however interesting to enquire what the situation is with respect to the "subclass-L" theories discussed in Sec. 3. Since this subclass was defined primarily for pedagogical reasons and there is no particular *a priori* reason to think of it as especially plausible, there would seem no great point in spending a lot of time here on the complications (most but not all of which are analogous to those discussed in detail in the context of experimental tests for Bell's theorem by Clauser and Horne[11]), which arise in the comparison of the predictions of Sec. 3 with real experiments. I therefore simply state the following without proof:

1. The problem of imperfect detector efficiency can be handled by making "no-enhancement" assumptions similar to those of Clauser and Horne.[11]
2. The effect of imperfect polarizer efficiency can be treated exactly, and shown to result in the replacement of formulae (3.16b) by a pair of slightly weaker inequalities; of various possible alternative forms of the latter, the simplest is

$$|\bar{P}(\varphi)+\bar{P}(\varphi')| \leqslant 2-(4\kappa/\pi)\,|\sin(\varphi-\varphi')|, \tag{6.1a}$$

$$|\bar{P}(\varphi)-P(\varphi')| \leqslant 2-(4\kappa/\pi)\,|\cos(\varphi-\varphi')|, \tag{6.1b}$$

where the parameter κ is defined in terms of the transmittances of the jth polarizer for polarization parallel (perpendicular) to the nominal "transmission" axis by the formula

$$\kappa \equiv \tfrac{1}{2}\,(\varepsilon_M^1+\varepsilon_M^2-\varepsilon_m^1-\varepsilon_m^2) \quad (\lesssim 1). \tag{6.2}$$

In practice, κ is so close to 1 in recent experiments that the correction expressed by (6.1a, b) to the results of Sec. 3 is very small.

3. Apart from the above two complications, which are of course familiar from the Bell's theorem context, there is a third problem

which does not exist there: namely, in formula (3.16a, b) the quantity $\overline{P}(\varphi)$ is defined as the average of $\overline{P}(\varphi, \xi)$ over the "center-of-mass" coordinate ξ, and such an average is not measured in existing experiments, which typically work at one or a few values of ξ. The simplest way of dealing with this complication is to make the explicit assumption (which is certainly consistent with existing experiments) that $P(\varphi, \xi)$ is in fact independent of ξ, so that $\overline{P}(\varphi)$ in (3.16a, b) may be replaced by $P(\varphi)$. We will moreover assume that $P(-\varphi) = P(\varphi)$.

Given the assumptions listed under (1) and (3) above, and neglecting for simplicity the small difference between $\square$ and 1, we can (for example) set, in (3.16a), $\varphi = \pi/8$, $\varphi' = -\pi/8$, and in (3.16b) $\varphi = \pi/8$, $\varphi' = 3\pi/8$, and in this way attain the prediction, valid for any subclass-L CNHV theory.

$$\eta \equiv 3P(\pi/8) - P(3\pi/8) \leqslant 4(1 - \sqrt{2}/\pi) \cong 2 \cdot 2. \tag{6.3}$$

This inequality is clearly violated by (for example) the experimental result quoted in Weihs *et al.*,[3] which corresponds to $\eta = 2.73 \pm 0.02$. Given the value of κ ($\geqslant 0.96$) in this experiment, the corresponding correction increases the RHS of the inequality (6.3) only to less than 2.3, so one can conclude that with the above assumptions the experiment rules out subclass-L CNHV theories by many standard deviations.

7. DISCUSSION

In this paper I have argued that among the general class of nonlocal hidden variables of a certain subclass, which I have called "crypto-nonlocal" (CN) is relatively plausible. (The "relatively" is stressed: I am not arguing that the absolute plausibility of any nonlocal theory is particularly high!.) These theories have the advantage that (a) no special role is assigned to the distant measuring apparatus as distinct from the rest of the distant environment (b) nevertheless, the nonlocality is guaranteed, by construction, to be unobservable in the simplest experiments (those involving single beams of photons). I have then shown (Secs. 3 and 4) that such theories cannot give predictions in agreement with those of quantum mechanics (and incidentally shown (Sec. 6) that with a few plausible subsidiary assumptions existing experiments conclusively refute theories of the "subclass-L" type).

To discuss the significance of these results it is convenient to imagine that a test of quantum mechanics (QM) against general CN theories has been carried out, with one of the two results: (a) the experimental data are consistent with the predicted inequalities for a CN theory and hence violate the QM predictions (b) the experimental data agree with QM and therefore rule out the whole class of CN theories. (The third possibility, that the data agree neither with QM nor with CN theories, is of no interest for the present discussion.)

In the case of result (a), which most physicists would no doubt think a prior very unlikely, it would of course immediately become a matter of prime interest whether or not observable effects of nonlocality are confined to the very special types of situation met with in atomic cascade precesses and similar events (positron annihilation, $K\bar{K}$ production, etc.)—roughly speaking, those situations for which one can construct the "EPR paradox." (Einstein *et al.*[2]), and which in quantum mechanics have to be described by nonfactorizable two-particle wave functions. It would in fact then be of considerable interest to re-examine the data (if there is any!) on correlations in the situations labeled type (b) in Sec. 2—that is, cases where, for example, two photons are radiated in succession by the same atom, whose intermediate state is however known. In this case quantum mechanics predicts that the wave function of the photon field is a simple product, and hence that there should be *no* correlations in the joint counting rate regarded as a function of the polarizer settings: in the language of Sec. 2 we predict Eq. (2.10). In principle it would be possible to adapt existing experiments on atomic cascades to this purpose, by applying a magnetic field to the source, passing the emitted photons through a wavelength (*not* a polarization) filter and then measuring the correlations in (say) linear polarization.

In the case of the (expected) result (b), what would we have learned about quantum mechanics? In the first place, note that whereas it is possible to regard the locality assumption required for Bell's theorem as a special (and particularly plausible) case of the more general assumption of non-contextuality (cf. Sec. V of Bell,[12] and also Kochen and Specker[13]), the results proved in this paper do not rely on this property; in fact, the class of hidden-variable theories whose incompatibility with QM is proved is explicitly "contextual" (Belinfante[4]). So the relaxation of the noncontextuality condition does not necessarily allow the results of QM to be reproduced by a hidden-variable theory, provided that we replace it by some other physically plausible constraint. Secondly, the above results yield strong support to the contention made by, for example, Garuccio and Selleri[8] that the crucial element in the incompatibility between local hidden-variable theories and QM really has rather little to do with the locality condition but a lot to do with the idea of the super position principle in QM.

Finally, it should, of course, be emphasized that if one wishes to maintain some kind of objectivity principle (cf. Clauser and Horne[11]) in the face of Bell's theorem, it is by no means obvious that the most natural way to do so is to modify the postulates (1)–(4) of Sec. 2 in the way done here (that is, to reject only postulate 4). It might, for example, be thought at least as plausible *a priori* to reject the second postulate, and in particular to allow the hidden-variable distribution $\rho(\lambda)$ to depend on the settings **a** and **b** of the polarizers. Whether any nontrivial results could be obtained under this assumption is a question I have not so far investigated.

It is a pleasure to dedicate this paper to David Mermin on the occasion of his retirement, and to wish him many more happy years of (unofficial!) activity in physics.

ACKNOWLEDGMENTS

The kernel of this work was done during a semester spent at the University of Science and Technology, Kumasi, Ghana (UST), on secondment from the University of Sussex in the fall of 1976, and the present manuscript was written, with the exception of the present Sec. 4 and some minor updates and revisions, at Sussex over the next two years. I would like to thank my colleagues at the UST for their hospitality in Kumasi, and the U.K. Inter-Universities Council for Higher Education Overseas for financial support for my visit there. I am grateful to Dr. L. Allen, Dr. G. Barton, Dr. R. Kuhn, and, most particularly, the late Prof. J. S. Bell for their helpful comments at that time; and to Dipankar Home and Abner Shimony for theirs a quarter of a century later, which not only led to an improvement of the manuscript but persuaded me that it might still be worth publishing. I am particularly grateful to Abner Shimony for his comments on the near-final version. The original Sec. 4 turned out to have contained a major error, and the necessary (and nontrivial) rewriting of this section, as well as the final revision of the manuscript, was carried out during a very pleasant stay at the University of Florida, Gainesville and supported by the National Science Foundation under Grant NSF-EIA01-21568.

REFERENCES

1. J. S. Bell, *Physics* **1**, 195–200 (1965).
2. A. Einstein, B. Podolsky, and N. Rosen, *Phys. Rev.* **47**, 777–780 (1935).
3. G. Weihs, T. Jennewein, C. Simon, H. Weinfurter, and A. Zeilinger, *Phys. Rev. Lett.* **81**, 5039 (1999).
4. F. J. Belinfante, *A Survey of Hidden-Variables Theories* (Pergamon, Oxford, 1973).
5. J. P. Jarrett, *Noûs* **18**, 569 (1984).[2]

[2] Note that in the terminology of this reference condition (4) of Sec. 2 above is the "completeness" condition and (5) is "locality;" thus the class of theories considered here is "complete but not strongly local."

6. J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, *Phys. Rev. Lett.* **23**, 880–884 (1969).
7. J. F. Clauser and A. Shimony, *Rep. Prog. Phys.* **41**, 1881–1927 (1978).
8. A. Garuccio and F. Selleri, *Nuovo Cimento B* **36**, 176–185 (1976).
9. O. Costa de Beauregard, *Nuovo Cimento B* **51**, 267–279 (1979).
10. J. R. Torgerson, D. Branning, C. H. Mooken, and L. Mandel, *Phys. Rev. A* **51**, 4400 (1995).
11. J. F. Clauser and M. A. Horne, *Phys. Rev. D* **10**, 526–535 (1974).
12. J. S. Bell, *Rev. Mod. Phys.* **38**, 447 (1966).
13. S. Kochen and J. Specker, *J. Math. Mech.* **17**, 59 (1967).