

# A Collection of Probabilistic Hidden-Variable Theorems and Counterexamples\*

Patrick Suppes<sup>†</sup>      J. Acacio de Barros<sup>‡</sup>

Gary Oas<sup>§</sup>

*Ventura Hall, Stanford University,  
Stanford, California 94305-4115*

## Abstract

Key words: hidden-variables, Bell's theorem, GHZ theorem, locality, correlations

The purpose of this article is to formulate a number of probabilistic hidden-variable theorems, to provide proofs in some cases, and counterexamples to some conjectured relationships. The first theorem is the fundamental one. It asserts the general equivalence of the existence of a hidden variable and the existence of a joint probability distribution of the observable quantities, whether finite or continuous.

PACS numbers: 03.65.Bz, 02.50.Cw, 02.50.Kd

---

\*It is a pleasure to dedicate this article to Giuliano Toraldo di Francia on the occasion of his 80th birthday.

<sup>†</sup>E-mail: suppes@ockham.stanford.edu. To whom correspondence should be addressed.

<sup>‡</sup>Permanent Address: Physics Department, Federal University at Juiz de Fora, 36036-330 Juiz de Fora, MG Brazil. E-mail: acacio@fisica.ufjf.br

<sup>§</sup>E-mail: oas@ockham.stanford.edu.

The literature on hidden variables in quantum mechanics is now enormous, and it may seem there is little that is new that can be said. Not everything in the present article is new, but several things are. We have tried to collect together a variety of results that go beyond the standard Clauser-Horne-Shimony-Holt form of the Bell inequalities for four observables.

First, we state, and sketch the proof, of the fundamental theorem of the collection we consider: there is a factoring hidden variable for a finite set of finite or continuous observables, i.e., random variables in the language of probability theory, if and only if the observables have a joint probability distribution. The physically important aspect of this theorem is that under very general conditions the existence of a hidden variable can be reduced completely to the relationship between the observables alone, namely, the problem of determining whether or not they have a joint probability distribution compatible with the given data, e.g., means, variances and correlations of the observables.

We emphasize that although most of the literature is restricted to no more than second-order moments such as covariances and correlations, there is no necessity to make such a restriction. It is in fact violated in the fourth-order moment that arises in the well-known Greenberger, Horne and Zeilinger [5] three- and four- particle configurations providing new Gedanken experiments on hidden variables. For our probabilistic proof of an abstract GHZ result, see Theorem 9.

As is familiar, Bell's results on hidden variables were mostly restricted to  $\pm 1$  observables, such as spin or polarization. But there is nothing essential about this restriction. Our general results cover any finite or continuous observables (Theorem 1). We also state a useful theorem (Theorem 7) on functions of random variables, and give a partial corollary (Theorem 8) showing how such general probabilistic results are implicit in the reduction of higher spin cases to two-valued random variables in the physics literature. At the end we give various results on hidden variables for Gaussian observables and formulate as the final theorem a nonlinear inequality that is necessary and sufficient for three Gaussian random variables to have a joint distribution compatible with their given means, variances and correlations.

**Factorization** In the literature on hidden variables, what we call the principle of factorization is sometimes baptized as a principle of locality. The terminology is not really critical, but the meaning is. We have in mind a quite general principle for random variables, continuous or discrete, which is the following. Let  $\mathbf{X}_1, \dots, \mathbf{X}_n$  be random variables, then a necessary and sufficient condition that there is a random variable  $\lambda$ , which is intended to be the hidden variable, such that  $\mathbf{X}_1, \dots, \mathbf{X}_n$  are conditionally independent given  $\lambda$ , is that there exists a joint probability distribution of  $\mathbf{X}_1, \dots, \mathbf{X}_n$ , without consideration of  $\lambda$ . This is our first theorem, which is the general fundamental theorem relating hidden variables and joint probability distributions of observable random variables.

**Theorem 1** (*Suppes & Zanotti [13] Holland & Rosenbaum [7]*) *Let  $n$  random variables  $X_1, \dots, X_n$ , finite or continuous, be given. Then there exists a hidden variable  $\lambda$  such that there is a joint probability distribution  $F$  of  $(\mathbf{X}_1, \dots, \mathbf{X}_n, \lambda)$  with the properties*

- (i)  $F(x_1, \dots, x_n | \lambda) = P(\mathbf{X}_1 \leq x_1, \dots, \mathbf{X}_n \leq x_n | \lambda = \lambda)$
- (ii) *Conditional independence holds, i.e., for all  $x_1, \dots, x_n, \lambda$ ,*

$$F(x_1, \dots, x_n | \lambda) = \prod_{j=1}^n F_j(x_j | \lambda),$$

*if and only if there is a joint probability distribution of  $\mathbf{X}_1, \dots, \mathbf{X}_n$ . Moreover,  $\lambda$  may be constructed so as to be deterministic, i.e., the conditional variance given  $\lambda$  of each  $\mathbf{X}_i$  is zero.*

To be completely explicit in the notation

$$F_j(x_j | \lambda) = P(\mathbf{X}_j \leq x_j | \lambda = \lambda). \tag{1}$$

*Idea of the proof.* Consider three  $\pm 1$  random variables  $\mathbf{X}$ ,  $\mathbf{Y}$  and  $\mathbf{Z}$ . There are 8 possible joint outcomes  $(\pm 1, \pm 1, \pm 1)$ . Let  $p_{ijk}$  be the probability of outcome  $(i, j, k)$ . Assign this probability to the value  $\lambda_{ijk}$  of the hidden variable  $\lambda$  we construct. Then the probability of the quadruple  $(i, j, k, \lambda_{ijk})$  is just  $p_{ijk}$  and the conditional probabilities are deterministic, i.e.,

$$P(\mathbf{X} = i, \mathbf{Y} = j, \mathbf{Z} = k | \lambda_{ijk}) = 1,$$

and factorization is immediate, i.e.,

$$P(\mathbf{X} = i, \mathbf{Y} = j, \mathbf{Z} = k \mid \lambda_{ijk}) = P(\mathbf{X} = i \mid \lambda_{ijk})P(\mathbf{Y} = j \mid \lambda_{ijk})P(\mathbf{Z} = k \mid \lambda_{ijk}).$$

Extending this line of argument to the general case proves the joint probability distribution of the observables is sufficient for existence of the factoring hidden variable. From the formulation of Theorem 1 necessity is obvious, since the joint distribution of  $(\mathbf{X}_1, \dots, \mathbf{X}_n)$  is a marginal distribution of the larger distribution  $(\mathbf{X}_1, \dots, \mathbf{X}_n, \boldsymbol{\lambda})$ .

It is obvious that the construction of  $\boldsymbol{\lambda}$  is purely mathematical. It has in itself no physical content. In fact, the proof itself is very simple. All the real mathematical difficulties are to be found in giving workable criteria for observables to have a joint probability distribution. As we remark in more detail later, we still do not have good criteria in the form of inequalities for necessary and possibly sufficient conditions for a joint distribution of three random variables with  $n > 2$  finite values, as in higher spin cases.

When additional physical assumptions are imposed on the hidden variable  $\boldsymbol{\lambda}$ , then the physical content of  $\boldsymbol{\lambda}$  goes beyond the joint distribution of the observables. A simple example is embodied in the following theorem about two hidden variables. We impose an additional condition of symmetry on the conditional expectations, and then a hidden variable exists only if the correlation of the two observables is nonnegative, a strong additional restriction on the joint distribution. The proof of this theorem is found in the article cited with its statement.

**Theorem 2** (*Suppes & Zanotti [12]*) *Let  $\mathbf{X}$  and  $\mathbf{Y}$  be two-valued random variables, for definiteness, with possible values 1 and  $-1$ , and with positive variances, i.e.,  $\sigma(\mathbf{X}), \sigma(\mathbf{Y}) > 0$ . In addition, let  $\mathbf{X}$  and  $\mathbf{Y}$  be exchangeable, i.e.,*

$$P(\mathbf{X} = 1, \mathbf{Y} = -1) = P(\mathbf{X} = -1, \mathbf{Y} = 1).$$

*Then a necessary and sufficient condition that there exist a hidden variable  $\boldsymbol{\lambda}$  such that*

$$E(\mathbf{X}\mathbf{Y} \mid \boldsymbol{\lambda} = \lambda) = E(\mathbf{X} \mid \boldsymbol{\lambda} = \lambda)E(\mathbf{Y} \mid \boldsymbol{\lambda} = \lambda)$$

*and*

$$E(\mathbf{X} \mid \boldsymbol{\lambda} = \lambda) = E(\mathbf{Y} \mid \boldsymbol{\lambda} = \lambda)$$

*for every value  $\lambda$  (except possibly on a set of measure zero) is that the correlation of  $\mathbf{X}$  and  $\mathbf{Y}$  be nonnegative.*

The informal statement of Theorems 1 and 2, which we call the *Factorization Theorems*, is that the necessary and sufficient condition for the existence of a factorizing hidden variable  $\lambda$  is just the existence of a joint probability distribution of the given random variables  $\mathbf{X}_i$ .

Often, in physics, as in the present paper, we are interested only in the means, variances and covariances – what is called the second-order probability theory, because we consider only second-order moments. We say that a hidden variable  $\lambda$  satisfies the *Second-Order Factorization Condition* with respect to the random variables  $\mathbf{X}_1, \dots, \mathbf{X}_n$  whose two first moments exist if and only if

$$(a) E(\mathbf{X}_1 \cdots \mathbf{X}_n | \lambda) = E(\mathbf{X}_1 | \lambda) \cdots E(\mathbf{X}_n | \lambda),$$

$$(b) E(\mathbf{X}_1^2 \cdots \mathbf{X}_n^2 | \lambda) = E(\mathbf{X}_1^2 | \lambda) \cdots E(\mathbf{X}_n^2 | \lambda).$$

We then have as an immediate consequence of Theorem 1 the following.

**Theorem 3** *Let  $n$  random variables discrete or continuous be given. If there is a joint probability distribution of  $\mathbf{X}_1, \dots, \mathbf{X}_n$ , then there is a deterministic hidden variable  $\lambda$  such that  $\lambda$  satisfies the Second-Order Factorization Condition with respect to  $\mathbf{X}_1, \dots, \mathbf{X}_n$ .*

**Locality.** The next systematic concept we want to discuss is locality. We mean by locality what we think John Bell meant by locality in the following quotation from his well-known 1966 paper [2].

It is the requirement of locality, or more precisely that the result of a measurement on one system be unaffected by operations on a distant system with which it has interacted in the past, that creates the essential difficulty. ... The vital assumption is that the result  $B$  for particle 2 does not depend on the setting  $\mathbf{a}$ , of the magnet for particle 1, nor  $A$  on  $\mathbf{b}$ .

Although Theorems 1 and 2 are stated at an abstract level without any reference to space-time or other physical considerations, there is an implicit hypothesis of locality in their statements. To make the locality hypothesis explicit, we need to use additional concepts. For each random variable  $\mathbf{X}_i$ , we

introduce a vector  $\mathbf{M}_i$  of parameters for the local apparatus (in space-time) used to measure the values of random variable  $\mathbf{X}_i$ .

**Definition 1 (Locality Condition I)**

$$E(\mathbf{X}_i^k | \mathbf{M}_i, \mathbf{M}_j, \boldsymbol{\lambda}) = E(\mathbf{X}_i^k | \mathbf{M}_i, \boldsymbol{\lambda}),$$

where  $k = 1, 2$ , corresponding to the first two moments of  $\mathbf{X}_i$ ,  $i \neq j$ , and  $1 \leq i, j \leq n$ .

Note that we consider only  $\mathbf{M}_j$  on the supposition that in a given experimental run, only the correlation of  $\mathbf{X}_i$  with  $\mathbf{X}_j$  is being studied. Extension to more variables, as considered in Theorem 7, is obvious. In many experiments the direction of the measuring apparatus is the most important parameter that is a component of  $\mathbf{M}_i$ .

**Definition 2 (Locality Condition II: Noncontextuality)** *The distribution of  $\boldsymbol{\lambda}$  is independent of the parameter values  $\mathbf{M}_i$  and  $\mathbf{M}_j$ , i.e., for all functions  $g$  for which the expectation  $E(g(\boldsymbol{\lambda}))$  and  $E(g(\boldsymbol{\lambda}) | \mathbf{M}_i, \mathbf{M}_j)$  are finite,*

$$E(g(\boldsymbol{\lambda})) = E(g(\boldsymbol{\lambda}) | \mathbf{M}_i, \mathbf{M}_j).$$

Here we follow [11]. In terms of Theorem 3, locality in the sense of Condition I is required to satisfy the hypothesis of a fixed mean and variance for each  $\mathbf{X}_i$ . If experimental observation of  $\mathbf{X}_i$  when coupled with  $\mathbf{X}_j$  was different from what was observed when coupled with  $\mathbf{X}_{j'}$ , then the hypothesis of constant means and variances would be violated. The restriction of Locality Condition II must be satisfied in the construction of  $\boldsymbol{\lambda}$  and it is easy to check that it is.

We embody these remarks in Theorem 4.

**Theorem 4** *Let  $n$  random variables  $\mathbf{X}_1, \dots, \mathbf{X}_n$  be given satisfying the hypothesis of Theorem 2. Let  $\mathbf{M}_i$  be the vector of local parameters for measuring  $\mathbf{X}_i$ , and let each  $\mathbf{X}_i$  satisfy Locality Condition I. Then there is a hidden variable  $\boldsymbol{\lambda}$  satisfying Locality Condition II and the Second-Order Factorization Condition if there is a joint probability distribution of  $\mathbf{X}_1, \dots, \mathbf{X}_n$ .*

**Inequalities for three random variables.** The next theorem states two conditions equivalent to an inequality condition given in [13] for three random variables having just two values.

**Theorem 5** *Let three random variables  $\mathbf{X}$ ,  $\mathbf{Y}$  and  $\mathbf{Z}$  be given with values  $\pm 1$  satisfying the symmetry condition  $E(\mathbf{X}) = E(\mathbf{Y}) = E(\mathbf{Z}) = 0$  and with covariances  $E(\mathbf{XY})$ ,  $E(\mathbf{YZ})$  and  $E(\mathbf{XZ})$  given. Then the following three conditions are equivalent.*

- (i) *There is a hidden variable  $\lambda$  with respect to  $\mathbf{X}$ ,  $\mathbf{Y}$  and  $\mathbf{Z}$  satisfying Locality Condition II and the Second-Order Factorization Condition holds.*
- (ii) *There is a joint probability distribution of the random variables  $\mathbf{X}$ ,  $\mathbf{Y}$ , and  $\mathbf{Z}$  compatible with the given means and expectations.*
- (iii) *The random variables  $\mathbf{X}$ ,  $\mathbf{Y}$  and  $\mathbf{Z}$  satisfy the following inequalities.*

$$-1 \leq E(\mathbf{XY}) + E(\mathbf{YZ}) + E(\mathbf{XZ}) \leq 1 + 2\text{Min}(E(\mathbf{XY}), E(\mathbf{YZ}), E(\mathbf{XZ})).$$

There are several remarks to be made about this theorem, especially the inequalities given in (iii). For discussion we introduce the standard correlation, and its standard notation, for two random variables  $\mathbf{X}$  and  $\mathbf{Y}$  whose variances are not zero:

$$\rho(X, Y) = \frac{E(\mathbf{XY}) - E(\mathbf{X})E(\mathbf{Y})}{\sigma(\mathbf{X})\sigma(\mathbf{Y})},$$

where  $\sigma(X)$ ,  $\sigma(Y)$  are the standard deviations of  $\mathbf{X}$  and  $\mathbf{Y}$ , i.e., the square roots of the variances:

$$\sigma(X) = \sqrt{\sigma^2(\mathbf{X})}$$

and

$$\sigma^2(\mathbf{X}) = \text{Var}(\mathbf{X}) = E(\mathbf{X}^2) - E(\mathbf{X})^2.$$

First, the explicit correlation notation  $\rho(\mathbf{X}, \mathbf{Y})$  is not standard in physics, but is necessary here for comparing various theorems. The notation adopted throughout this article conforms fairly closely to what is standard in mathematical statistics.

Physicists use less general notation, because they often assume certain symmetry conditions are satisfied, e.g.,  $E(\mathbf{X}) = E(\mathbf{Y}) = E(\mathbf{Z}) = 0$ . To make these relations explicit, keeping in mind the earlier definition of  $\rho(\mathbf{X}, \mathbf{Y})$ , we have:

(i) Covariance of  $\mathbf{X}$  and  $\mathbf{Y} = \text{Cov}(\mathbf{X}, \mathbf{Y}) = E(\mathbf{XY}) - E(\mathbf{X})E(\mathbf{Y})$ ,

(ii) If  $E(\mathbf{X}) = E(\mathbf{Y}) = 0$ , then, clearly,

$$\text{Cov}(\mathbf{X}, \mathbf{Y}) = E(\mathbf{XY})$$

(iii) If  $\mathbf{X}$  and  $\mathbf{Y}$  are random variables whose only values are  $\pm 1$  and  $E(\mathbf{X}) = E(\mathbf{Y}) = 0$ , then

$$\text{Var}(\mathbf{X}) = \text{Var}(\mathbf{Y}) = 1,$$

(iv) If hypothesis of (iii) is satisfied

$$\rho(\mathbf{X}, \mathbf{Y}) = E(\mathbf{XY}),$$

which is why in the physics literature  $E(\mathbf{XY})$ , with or without a comma between  $\mathbf{X}$  and  $\mathbf{Y}$ , so commonly occurs. The statistical terminology for  $E(\mathbf{XY})$  is bivariate product moment  $\mu_{11}$ , which we shall often simply call the bivariate product moment, without further notation.

Note that with the special symmetry conditions that  $E(\mathbf{X}) = E(\mathbf{Y}) = E(\mathbf{Z}) = 0$ , the inequalities (iii) of Theorem 5 for  $\pm 1$  random variables can be written

$$-1 \leq \rho(\mathbf{X}, \mathbf{Y}) + \rho(\mathbf{Y}, \mathbf{Z}) + \rho(\mathbf{X}, \mathbf{Z}) \leq 1 + 2\text{Min}(\rho(\mathbf{X}, \mathbf{Y})\rho(\mathbf{X}, \mathbf{Z})\rho(\mathbf{Y}, \mathbf{Z})). \quad (2)$$

**Three Counterexamples.** To show how special (iii) of Theorem 5, or the equivalent (2) written in terms of correlation, is, because of the strong symmetry assumptions, we now give three different examples that do not satisfy these inequalities. The first is for  $\pm 1$  random variables that do not have expectations equal to zero. For this case neither the correlations nor covariances have linear inequalities, only the moments  $E(\mathbf{XY})$ . The second case is for random variables with values  $-1, 0, 1$  and zero expectations. An example is given which is satisfied by the covariances but not the correlations. The third case is for random variables with values  $-2, 0, 2$  and zero expectations. The inequalities of (iii) are not satisfied by the covariances, which in this case are equal to the expectations  $E(\mathbf{XY})$ .

First, for the general case of  $\pm 1$  random variables we have

$$E(\mathbf{X}) = x_0, \quad E(\mathbf{Y}) = y_0, \quad E(\mathbf{Z}) = z_0$$



and

$$-1 < x_0, y_0, z_0 < 1,$$

and it is straightforward to derive the analogue of (iii) of Theorem 5 for the bivariate product moments, as well as the corresponding correlations, but the expressions are more complicated for the correlations. We only give part of the details here. We generalize on the derivation given in [13]. We need to consider in detail the eight probabilities  $p_{ijk}$  for  $i, j, k = \pm 1$ . When referring to the marginals we use a dot for the missing random variable. For example,

$$\begin{aligned} p_{11\cdot} &= P(\mathbf{X} = 1, \mathbf{Y} = 1) \\ p_{0\cdot 1} &= P(\mathbf{X} = -1, \mathbf{Z} = 1) \end{aligned}$$

(For ease of typography we use 0 rather than -1 as a subscript.)

We note immediately the following equations:

$$E(\mathbf{XY}) = p_{11\cdot} - p_{10\cdot} - p_{01\cdot} + p_{00\cdot}$$

$$p_{10\cdot} + p_{01\cdot} = \frac{1 - E(\mathbf{XY})}{2}$$

and correspondingly,

$$p_{10\cdot} + p_{01\cdot} = \frac{1 - E(\mathbf{YZ})}{2}$$

$$p_{1\cdot 0} + p_{0\cdot 1} = \frac{1 - E(\mathbf{XZ})}{2}$$

$$p_{1\cdot} = p_{10\cdot} + p_{11\cdot} = \frac{x_0 + 1}{2}$$

$$p_{\cdot 1} = p_{11\cdot} + p_{01\cdot} = \frac{y_0 + 1}{2}$$

$$p_{\cdot 1} = p_{11\cdot} + p_{01\cdot} = \frac{z_0 + 1}{2}$$

From these equations we easily derive

$$p_{10} = \frac{x_0 - y_0}{4} + \frac{1 - E(\mathbf{XY})}{4}$$

$$p_{11\cdot} = \frac{1}{4} + \frac{x_0 + y_0 + E(\mathbf{XY})}{4},$$

and similar expressions for  $p_{10}$ ,  $p_{1\cdot 1}$ , etc. Using these equations, we may then derive

$$p_{110} = \frac{1}{4} + \frac{x_0 + y_0 + E(\mathbf{XY})}{4} - p_{111}$$

$$p_{101} = \frac{1}{4} + \frac{x_0 + z_0 + E(\mathbf{XZ})}{4} - p_{111}$$

$$p_{011} = \frac{1}{4} + \frac{y_0 + z_0 + E(\mathbf{YZ})}{4} - p_{111}$$

$$p_{100} = p_{111} - \frac{y_0 + z_0}{4} - \frac{E(\mathbf{XY})}{4} - \frac{E(\mathbf{XZ})}{4}$$

$$p_{010} = p_{111} - \frac{x_0 + z_0}{4} - \frac{E(\mathbf{XY})}{4} - \frac{E(\mathbf{YZ})}{4}$$

$$p_{001} = p_{111} - \frac{x_0 + y_0}{4} - \frac{E(\mathbf{XZ})}{4} - \frac{E(\mathbf{YZ})}{4}$$

$$p_{000} = \frac{1}{4} - \frac{x_0 + z_0}{4} + \frac{E(\mathbf{XZ})}{4} - \frac{y_0 + z_0}{4} + \frac{E(\mathbf{YZ})}{4} - \frac{x_0 + y_0}{4} + \frac{E(\mathbf{XY})}{4} - p_{111}$$

so

$$1 + E(\mathbf{XY}) + E(\mathbf{YZ}) + E(\mathbf{XZ}) - 2(x_0 + y_0 + z_0) \geq 4p_{111}$$

And as a generalization of the left-hand inequality of (iii) of Theorem 5, we then have

$$E(\mathbf{XY}) + E(\mathbf{YZ}) + E(\mathbf{XZ}) - 2(x_0 + y_0 + z_0) \geq -1. \quad (3)$$

This result is much simpler than the corresponding one for correlation. We have at once

$$\rho(\mathbf{X}, \mathbf{Y}) = \frac{E(\mathbf{XY}) - x_0y_0}{\sqrt{1 - x_0^2}\sqrt{1 - y_0^2}},$$

and so

$$E(\mathbf{XY}) = \sqrt{1 - x_0^2}\sqrt{1 - y_0^2}\rho(\mathbf{X}, \mathbf{Y}) + x_0y_0.$$

Substituting the right-hand side for  $E(\mathbf{XY})$ , and the corresponding expressions for  $E(\mathbf{YZ})$  and  $E(\mathbf{XZ})$  yields a rather complicated inequality in terms of correlation, which we shall not write out here.

The next remark is that (iii) is not necessary for the correlations of three-valued random variables with expectations equal to zero. Let the three values be 1, 0,  $-1$ . Here is a counterexample where each of the three correlations is  $-\frac{1}{2}$ , and thus with a sum equal to  $-\frac{3}{2}$ , violating (2).

There is a joint probability distribution with the following values. Let  $p(x, y, z)$  be the probability of a given triple of values, e.g.,  $(1, -1, 0)$ . Then, of course, we must have for all  $x, y$  and  $z$

$$p(x, y, z) \geq 0 \text{ and } \sum_{x,y,z} p(x, y, z) = 1,$$

where  $x, y$  and  $z$  each have the three values 1, 0,  $-1$ . So, let

$$p(-1, 0, 1) = p(1, -1, 0) = p(0, 1, -1) = p(1, 0, -1) = p(-1, 1, 0) = p(0, -1, 1) = \frac{1}{6}$$

and the other 21  $p(x, y, z) = 0$ . Then it is easy to show that in this model  $E(\mathbf{X}) = E(\mathbf{Y}) = E(\mathbf{Z}) = 0$ ,  $\text{Var}(\mathbf{X}) = \text{Var}(\mathbf{Y}) = \text{Var}(\mathbf{Z}) = \frac{2}{3}$ , and  $\text{Cov}(\mathbf{XY}) = \text{Cov}(\mathbf{YZ}) = \text{Cov}(\mathbf{XZ}) = -\frac{1}{3}$ , so that the correlations are

$$\rho(\mathbf{X}, \mathbf{Y}) = \rho(\mathbf{Y}, \mathbf{Z}) = \rho(\mathbf{X}, \mathbf{Z}) = -\frac{1}{2}.$$

Note that in the example just given the covariances for the three-valued random variables, with the joint distribution as stated, do satisfy (iii) of Theorem 5.

For the third promised case, it is easy to construct a counterexample for covariances of three-valued random variables with values  $-2, 0, 2$  and expectations zero. We use the same distribution for these new values:  $p(-2, 0, 2) = p(2, -2, 0) = p(0, 2, -2) = p(2, 0, -2) = p(-2, 2, 0) = p(0, -2, 2) = \frac{1}{6}$ . It is easy to see at once that

$$\text{Cov}(\mathbf{X}, \mathbf{Y}) = \text{Cov}(\mathbf{Y}, \mathbf{Z}) = \text{Cov}(\mathbf{X}, \mathbf{Z}) = -\frac{4}{3},$$

and so (iii) of Theorem 5 is not satisfied by these covariances.

It is a somewhat depressing mathematical fact that even for three random variables with  $n$ -values and expectations equal to zero, a separate investigation seems to be needed for each  $n$  to find necessary and sufficient conditions to have a joint probability distribution compatible with given means, variances and covariances or correlations. A more general recursive result would be highly desirable, but seems not to be known. Such results are pertinent to the study of multi-valued spin phenomena, the discussion of which we continue after the next theorem.

**Bell's original inequality.** We now return to Theorem 5 for another look at the inequalities (iii), which assume  $E(\mathbf{X}) = E(\mathbf{Y}) = E(\mathbf{Z}) = 0$ . How do these inequalities relate to Bell's well-known inequality [1], written in terms of the bivariate product moments,

$$1 + E(\mathbf{YZ}) \geq |E(\mathbf{XY}) - E(\mathbf{XZ})| \quad (4)$$

Bell's inequality is in fact neither necessary nor sufficient for the existence of a joint probability distribution of the random variables  $\mathbf{X}, \mathbf{Y}$  and  $\mathbf{Z}$  with values  $\pm 1$  and expectations equal to zero. That it is not sufficient is easily seen from letting all three covariances equal  $-\frac{1}{2}$ . Then the inequality is satisfied, for

$$1 - \frac{1}{2} \geq |-\frac{1}{2} - (-\frac{1}{2})|$$

i.e.,

$$\frac{1}{2} \geq 0,$$

but, as is clear from (iii) there can be no joint distribution with the three covariances equal to  $-\frac{1}{2}$ , for

$$-\frac{1}{2} + -\frac{1}{2} + -\frac{1}{2} < -1.$$

Secondly, Bell's inequality is not necessary. Let  $E(\mathbf{XY}) = \frac{1}{2}$ ,  $E(\mathbf{XZ}) = -\frac{1}{2}$ , and  $E(\mathbf{YZ}) = -\frac{1}{2}$ , then (4) is violated, because

$$1 - \frac{1}{2} < \left| \frac{1}{2} - \left(-\frac{1}{2}\right) \right|,$$

but (iii) is satisfied, and so there is a joint distribution:

$$-1 \leq \frac{1}{2} - \frac{1}{2} - \frac{1}{2} \leq 1 + 2\text{Min}\left(\frac{1}{2}, -\frac{1}{2}, -\frac{1}{2}\right),$$

i.e.,

$$-1 \leq -\frac{1}{2} \leq 0.$$

Bell derived his inequality for certain cases satisfied by a local hidden-variable theory, but violated by the quantum mechanical covariance equal to  $-\cos\theta_{ij}$ . In particular, let  $\theta_{\mathbf{XY}} = 30^\circ$ ,  $\theta_{\mathbf{XZ}} = 60^\circ$ ,  $\theta_{\mathbf{YZ}} = 30^\circ$ , so, geometrically  $\mathbf{Y}$  bisects  $\mathbf{X}$  and  $\mathbf{Z}$ . Then

$$\left| -\frac{1}{2} - \left(-\frac{\sqrt{3}}{2}\right) \right| > 1 - \frac{\sqrt{3}}{2}.$$

**Bell's Inequalities in the CHSH form.** The next theorem states two conditions equivalent to Bell's Inequalities for random variables with just two values. This form is due to Clauser et al., [3]. The equivalence of (ii) and (iii) was proved by Fine [4].

**Theorem 6 (Bell's Inequalities)** *Let  $n$  random variables be given satisfying the locality hypothesis of Theorem 4. Let  $n = 4$ , the number of random variables, let each  $\mathbf{X}_i$  be discrete with values  $\pm 1$ , let the symmetry condition  $E(\mathbf{X}_i) = 0$ ,  $i = 1, \dots, 4$  be satisfied, let  $\mathbf{X}_1 = \mathbf{A}$ ,  $\mathbf{X}_2 = \mathbf{A}'$ ,  $\mathbf{X}_3 = \mathbf{B}$ ,  $\mathbf{X}_4 = \mathbf{B}'$ , with the covariances  $E(\mathbf{AB})$ ,  $E(\mathbf{AB}')$ ,  $E(\mathbf{A'B})$  and  $E(\mathbf{A'B}')$  given. Then the following three conditions are equivalent.*

- (i) *There is a hidden variable  $\lambda$  satisfying Locality Condition II and equation (a) of the Second-Order Factorization Condition holds.*
- (ii) *There is a joint probability distribution of the random variables  $\mathbf{A}$ ,  $\mathbf{A}'$ ,  $\mathbf{B}$  and  $\mathbf{B}'$  compatible with the given means and covariances.*
- (iii) *The random variables  $\mathbf{A}$ ,  $\mathbf{A}'$ ,  $\mathbf{B}$  and  $\mathbf{B}'$  satisfy Bell's inequalities in the CHSH form*

$$\begin{aligned}
-2 &\leq E(\mathbf{AB}) + E(\mathbf{AB}') + E(\mathbf{A'B}) - E(\mathbf{A'B}') \leq 2 \\
-2 &\leq E(\mathbf{AB}) + E(\mathbf{AB}') - E(\mathbf{A'B}) + E(\mathbf{A'B}') \leq 2 \\
-2 &\leq E(\mathbf{AB}) - E(\mathbf{AB}') + E(\mathbf{A'B}) + E(\mathbf{A'B}') \leq 2 \\
-2 &\leq -E(\mathbf{AB}) + E(\mathbf{AB}') + E(\mathbf{A'B}) + E(\mathbf{A'B}') \leq 2
\end{aligned}$$

It is worth emphasizing that in contrast to Bell's original inequality (4), the CHSH inequalities with four random variables give necessary and sufficient conditions for the existence of a joint probability distribution.

It will now be shown that the CHSH inequalities remain valid for three-valued random variables, (spin 1 particles). Consider a spin-1 particle with the 3 state observables,  $A(a, \lambda) = +1, 0, -1$ ,  $B(b, \lambda) = +1, 0, -1$ .  $\lambda$  is a hidden variable having a normalized probability density,  $\rho(\lambda)$ . The expectation of these observables is defined as,

$$E(a, b) = \int AB\rho(\lambda)d\lambda.$$

We have suppressed the variable dependence on  $A$  and  $B$  for clarity. (Note that in this discussion we follow the notation of physicists, especially as used by Bell, rather than the standard notation of mathematical statistics for expectations, including covariances.) Consider the following difference,

$$|E(a, b) - E(a, b')| = \left| \int A[B - B']\rho(\lambda)d\lambda \right|.$$

Since the density  $\rho > 0$  and  $|A| = 1, 0$  we have the following inequality,

$$\begin{aligned}
|E(a, b) - E(a, b')| &\leq \int |A[B - B']|\rho(\lambda)d\lambda, \\
&\leq \int |[B - B']|\rho(\lambda)d\lambda.
\end{aligned}$$

Similarly we have the following inequality,

$$\begin{aligned} |E(a', b) + E(a', b')| &= \left| \int A'[B + B']\rho(\lambda)d\lambda \right|, \\ &\leq \int |[B + B']|\rho(\lambda)d\lambda. \end{aligned}$$

Adding the two expressions we arrive at the following inequality,

$$|E(a, b) - E(a, b')| + |E(a', b) + E(a', b')| \leq \int [||B - B'| + |B + B'|]\rho(\lambda)d\lambda.$$

The term in square brackets is equal to 2 in all cases except when  $B$  and  $B'$  are both equal to zero, in which case the right-hand side vanishes. With this and the normalization condition for the hidden variable density we have the same inequality as the spin- $\frac{1}{2}$  CHSH inequality,

$$|E(a, b) - E(a, b')| + |E(a', b) + E(a', b')| \leq 2.$$

Note that we could create a stronger inequality by adding the function  $2(|E(a, b)| - 1)(|E(a, b')| - 1)$  to the left-hand side.

**Higher Spin Cases.** For higher spins we can proceed analogously and derive the following inequality which must be satisfied for spin  $j$  particles,

$$|E(a, b) - E(a, b')| + |E(a', b) + E(a', b')| \leq 2j.$$

If we define normalized observables,  $\frac{A(a, \lambda)}{j}$  the original CHSH inequality will need to be satisfied for local hidden variable theories, although stronger inequalities could be constructed.

In Peres' work on higher spin particles the observable is defined by a mapping from the,  $2j + 1$ -state,  $J_z$  operator to a two-state operator [10]. Under this mapping it was shown that Bell's inequality is violated for certain parameter settings of the detectors.

The mapping from many values to  $\pm 1$ , as used by Peres and others is justified probabilistically by the following theorem, which provides a way of avoiding deriving separate inequalities for each of the higher spin cases ( $n > 2$ ).

**Theorem 7** *Let  $\mathbf{X}_1, \dots, \mathbf{X}_n$  be  $n$  random variables with joint probability distribution  $F(x_1, \dots, x_n)$ . Let  $f_1, \dots, f_k$  be finite-valued measurable functions of the random variables  $\mathbf{X}_1, \dots, \mathbf{X}_n$ , with  $y_1 = f_1(x_1, \dots, x_n), \dots, y_k = f_k(x_1, \dots, x_n)$ . Then there is a function  $G(y_1, \dots, y_k)$ , unique up to sets of measure zero, that determines the joint probability distribution of the random variables  $\mathbf{Y}_1, \dots, \mathbf{Y}_k$  that are functions of  $\mathbf{X}_1, \dots, \mathbf{X}_n$ .*

*Idea of the proof:* We only sketch the proof for a simple finite case to avoid technical details, for the underlying idea is very intuitive.

Let  $\mathbf{X}$ ,  $\mathbf{Y}$  and  $\mathbf{Z}$  be  $\pm 1$  random variables with a joint distribution. Let  $\mathbf{A}$  and  $\mathbf{B}$  be random variables that are functions of  $\mathbf{X}$ ,  $\mathbf{Y}$  and  $\mathbf{Z}$ . In particular, let

$$\begin{aligned} \mathbf{A} &= f(\mathbf{X}, \mathbf{Y}) = \mathbf{X} + \mathbf{Y} \\ \mathbf{B} &= f(\mathbf{Y}, \mathbf{Z}) = \mathbf{Y} + \mathbf{Z}. \end{aligned}$$

Then it is easy to see that range of values of  $\mathbf{A}$  and  $\mathbf{B}$  is  $\{-2, 0, 2\}$ . More importantly, the joint distribution of  $\mathbf{A}$  and  $\mathbf{B}$  is easily computed from the joint distribution of  $\mathbf{X}, \mathbf{Y}$  and  $\mathbf{Z}$ . Of the nine possible triples of values for the joint distribution, we show four, the remaining five are very similar:

$$\begin{aligned} P(\mathbf{A} = -2 \ \& \ \mathbf{B} = -2) &= P(\mathbf{X} = -1 \ \& \ \mathbf{Y} = -1 \ \& \ \mathbf{Z} = -1) \\ P(\mathbf{A} = -2 \ \& \ \mathbf{B} = 0) &= P(\mathbf{X} = -1 \ \& \ \mathbf{Y} = -1 \ \& \ \mathbf{Z} = 1) \\ P(\mathbf{A} = -2 \ \& \ \mathbf{B} = 2) &= 0 \\ P(\mathbf{A} = 0 \ \& \ \mathbf{B} = 0) &= P((\mathbf{X} = -1 \ \& \ \mathbf{Y} = 1 \ \& \ \mathbf{Z} = -1) \ \text{or} \\ &\quad (\mathbf{X} = 1 \ \& \ \mathbf{Y} = -1 \ \& \ \mathbf{Z} = 1)) \end{aligned}$$

The following partial converse of Theorem 7 is really what is implicit in the reduction of higher spin cases to just two values, rather than Theorem 7 itself. For simplicity of formulation we restrict the statement of the theorem to four random variables, using the familiar notation of Theorem 6, and also restrict the functions to functions of a single random variable, with the additional constraint that the functions have only the values  $\pm 1$ .

**Theorem 8** *Let  $\mathbf{A}$ ,  $\mathbf{B}$ ,  $\mathbf{A}'$ ,  $\mathbf{B}'$  be random variables with means, variances and covariances given, but with no assumption of a joint distribution. Let*



$f_{\mathbf{A}}, f_{\mathbf{B}}, f_{\mathbf{A}'}, f_{\mathbf{B}'}$  be finite-valued measurable functions of the respective random variables and let the functions have only the values  $\pm 1$ . If there is no joint distribution of  $f_{\mathbf{A}}(\mathbf{A}), f_{\mathbf{B}}(\mathbf{B}), f_{\mathbf{A}'}(\mathbf{A}'),$  and  $f_{\mathbf{B}'}(\mathbf{B}')$  compatible with the means, variances and covariances of the functional random variables, then there is no joint distribution of  $\mathbf{A}, \mathbf{B}, \mathbf{A}', \mathbf{B}'$  compatible with the given means, variances and covariances.

**GHZ Probabilistic Theorem.** Changing the focus, we now consider an abstract version of the GHZ gedanken experiment. All arguments known to us, in particular GHZ's [5] own argument, the more extended one in [6] and Mermin's [9] proceed by assuming the existence of a deterministic hidden variable and then deriving a contradiction. It follows immediately from Theorem 1 that the nonexistence of a hidden variable is equivalent to the nonexistence of a joint probability distribution for the given observable random variables. The next theorem states this purely probabilistic GHZ result, and, more importantly, the proof is purely in terms of the observables, with no consideration of possible hidden variables.

**Theorem 9 (Abstract GHZ version)** . Let  $\mathbf{A}_{\varphi_1}, \mathbf{B}_{\varphi_2}, \mathbf{C}_{\varphi_3}, \mathbf{D}_{\varphi_4}$  be an infinite family of  $\pm 1$  random variables, with  $\varphi_i$  a periodic angle or phase,  $0 \leq \varphi_i \leq 2\pi,$  and let the following condition hold:

$$E(\mathbf{A}_{\varphi_1} \mathbf{B}_{\varphi_2} \mathbf{C}_{\varphi_3} \mathbf{D}_{\varphi_4}) = -\cos(\varphi_1 + \varphi_2 - \varphi_3 - \varphi_4) \quad (5)$$

Then the finite subset of random variables  $\mathbf{A}_0, \mathbf{B}_0, \mathbf{C}_0, \mathbf{D}_0, \mathbf{A}_\pi, \mathbf{A}_{\frac{\pi}{2}}, \mathbf{C}_{\frac{\pi}{2}}, \mathbf{D}_{\frac{\pi}{2}}$  does not have a joint probability distribution.

*Proof:* We note first, as an immediate consequence of (5),

- (i) if  $\varphi_1 + \varphi_2 - \varphi_3 - \varphi_4 = 0$  then  $E(\mathbf{A}_{\varphi_1} \mathbf{B}_{\varphi_2} \mathbf{C}_{\varphi_3} \mathbf{D}_{\varphi_4}) = -1,$
- (ii) if  $\varphi_1 + \varphi_2 - \varphi_3 - \varphi_4 = \pi$  then  $E(\mathbf{A}_{\varphi_1} \mathbf{B}_{\varphi_2} \mathbf{C}_{\varphi_3} \mathbf{D}_{\varphi_4}) = 1.$

The proof proceeds by deriving a contradiction from the supposition of the existence of a joint probability distribution. Because conditional probabilities are used repeatedly, we must check the given condition in each such probability has positive probability. Let  $s_i, i = 1, \dots, 4$  be  $+1$  or  $-1$ . One of

the 16 products of the four signs must have positive probability, in the sense that

$$P(\mathbf{A}_0 = s_1, \mathbf{B}_0 = s_2, \mathbf{C}_0 = s_3, \mathbf{D}_0 = s_4) > 0. \quad (6)$$

(We do not need to know whether each  $s_i$  is  $+1$  or  $-1$ .) Then since the angles sum to 0, the product

$$s_1 s_2 s_3 s_4 = -1. \quad (7)$$

We also can infer at once from (5) and (ii)

$$P(\mathbf{A}_\pi = s_2 s_3 s_4 \mid \mathbf{B}_0 = s_2, \mathbf{C}_0 = s_3, \mathbf{D}_0 = s_4) = 1, \quad (8)$$

since (5) ensures that the condition in (8) has positive probability. Using (i) now, by a similar argument

$$P(\mathbf{A}_0 \mathbf{C}_0 = -s_2 s_4 \mid \mathbf{B}_0 = s_2, \mathbf{D}_0 = s_4) = 1, \quad (9)$$

and from (5) and familiar facts about probability-1 propositions (see Lemma 1 of the Appendix), we may add  $\mathbf{C}_0 = s_3$  to the condition (9) to obtain

$$P(\mathbf{A}_0 \mathbf{C}_0 = -s_2 s_4 \mid \mathbf{B}_0 = s_2, \mathbf{C}_0 \mathbf{D}_0 = s_3 s_4) = 1. \quad (10)$$

Using (i) and (5) again

$$P(\mathbf{A}_{\frac{\pi}{2}} \mathbf{C}_{\frac{\pi}{2}} = -s_2 s_4 \mid \mathbf{B}_0 = s_2, \mathbf{C}_0 \mathbf{D}_0 = s_3 s_4) = 1 \quad (11)$$

And so, using Lemma 2 of the Appendix and (10) and (11), we infer

$$P(\mathbf{A}_0 \mathbf{C}_0 = \mathbf{A}_{\frac{\pi}{2}} \mathbf{C}_{\frac{\pi}{2}} \mid \mathbf{B}_0 = s_2, \mathbf{C}_0 \mathbf{D}_0 = s_3 s_4) = 1. \quad (12)$$

By an argument just like that of (9) - (12), we also infer

$$P(\mathbf{A}_0 \mathbf{D}_0 = \mathbf{A}_{\frac{\pi}{2}} \mathbf{D}_{\frac{\pi}{2}} \mid \mathbf{B}_0 = s_2, \mathbf{C}_0 \mathbf{D}_0 = s_3 s_4) = 1 \quad (13)$$

Dividing the equation of (12) by that of (13), we get

$$P\left(\frac{\mathbf{C}_0}{\mathbf{D}_0} = \frac{\mathbf{C}_{\frac{\pi}{2}}}{\mathbf{D}_{\frac{\pi}{2}}} \mid \mathbf{B}_0 = s_2, \mathbf{C}_0 \mathbf{D}_0 = s_3 s_4\right) = 1, \quad (14)$$

and since the random variables have only values  $+1$  and  $-1$ , we may rewrite (14) as:

$$P(\mathbf{C}_0 \mathbf{D}_0 = \mathbf{C}_{\frac{\pi}{2}} \mathbf{D}_{\frac{\pi}{2}} \mid \mathbf{B}_0 = s_2, \mathbf{C}_0 \mathbf{D}_0 = s_3 s_4) = 1 \quad (15)$$

From (15) and Lemma 3 of the Appendix we get

$$P(\mathbf{C}_{\frac{\pi}{2}}\mathbf{D}_{\frac{\pi}{2}} = s_3s_4 \mid \mathbf{B}_0 = s_2, \mathbf{C}_0\mathbf{D}_0 = s_3s_4) = 1, \quad (16)$$

and so immediately we may infer from (5) and (16)

$$P(\mathbf{B}_0 = s_2, \mathbf{C}_{\frac{\pi}{2}}\mathbf{D}_{\frac{\pi}{2}} = s_3s_4) > 0, \quad (17)$$

Then from (i) and (17)

$$P(\mathbf{A}_\pi = -s_2s_3s_4 \mid \mathbf{B}_0 = s_2, \mathbf{C}_{\frac{\pi}{2}}\mathbf{D}_{\frac{\pi}{2}} = s_3s_4) = 1, \quad (18)$$

and finally from (15) and (18) and Lemma 5 of the Appendix

$$P(\mathbf{A}_\pi = -s_2s_3s_4 \mid \mathbf{B}_0 = s_2, \mathbf{C}_0\mathbf{D}_0 = s_3s_4) = 1. \quad (19)$$

⊥ Obviously, (8) and (19) together yield the desired contradiction.

**Gaussian random variables.** A fundamental second-order theorem about finite sequences of continuous random variables is the following:

**Theorem 10** *Let  $n$  continuous random variables be given, let their means, variances and covariances all exist and be finite, with all the variances nonzero. Then a necessary and sufficient condition that a joint Gaussian probability distribution of the  $n$  random variables exists, compatible with the given means, variances and covariances, is that the eigenvalues of the correlation matrix be nonnegative.*

A thorough discussion and proof of this theorem can be found in Loève [8]. It is important to note that the hypothesis of this theorem is that each pair of the random variables has enough postulated for there to exist a unique bivariate Gaussian distribution with the given pair of means and variances and the covariance of the pair. Moreover, if, as required for a joint distribution of all  $n$  variables, the eigenvalues of the correlation matrix are all nonnegative, then there is a unique Gaussian joint distribution of the  $n$  random variables.

We formulate the next theorem to include cases like Bell's inequalities when not all the correlations or covariances are given.

**Theorem 11** *Let  $n$  continuous random variables be given such that they satisfy the locality hypothesis of Theorem 4, let their means and variances exist and be finite, with all the variances nonzero, and let  $m \leq n(n - 1)/2$  covariances be given and be finite. Then the following two conditions are equivalent.*

- (i) *There is a joint Gaussian probability distribution of the  $n$  random variables compatible with the given means, variances and covariances.*
- (ii) *Given the  $m \leq n(n - 1)/2$  covariances, there are real numbers that may be assigned to the missing correlations so that the completed correlation matrix has eigenvalues that are all nonnegative.*

*Moreover, (i) or (ii) implies that there is a hidden variable  $\lambda$  satisfying Locality Condition II and the Second-Order Factorization Condition.*

The proof of Theorem 11 follows directly from Theorem 10.

Using Theorem 10, we can also derive a nonlinear inequality necessary and sufficient for three Gaussian random variables to have a joint distribution. In the statement of the theorem  $\rho(\mathbf{X}, \mathbf{Y})$  is the correlation of  $\mathbf{X}$  and  $\mathbf{Y}$ .

**Theorem 12** *Let  $\mathbf{X}, \mathbf{Y}$  and  $\mathbf{Z}$  be three Gaussian random variables whose means, variances and correlations are given, and whose variances are nonzero. Then there exists a joint Gaussian distribution of  $\mathbf{X}, \mathbf{Y}$  and  $\mathbf{Z}$  (necessarily unique) compatible with the given means, variances and correlations if and only if*

$$\rho(\mathbf{X}, \mathbf{Y})^2 + \rho(\mathbf{X}, \mathbf{Z})^2 + \rho(\mathbf{Y}, \mathbf{Z})^2 \leq 2\rho(\mathbf{X}, \mathbf{Y})\rho(\mathbf{Y}, \mathbf{Z})\rho(\mathbf{X}, \mathbf{Z}) + 1.$$

The proof comes directly from the determinant of the correlation matrix. For a matrix to be non-negative definite the determinant of the entire matrix and all principal minors must be greater than or equal to zero,

$$\text{Det} \begin{pmatrix} 1 & \rho(\mathbf{X}, \mathbf{Y}) & \rho(\mathbf{X}, \mathbf{Z}) \\ \rho(\mathbf{X}, \mathbf{Y}) & 1 & \rho(\mathbf{Y}, \mathbf{Z}) \\ \rho(\mathbf{X}, \mathbf{Z}) & \rho(\mathbf{Y}, \mathbf{Z}) & 1 \end{pmatrix} \geq 0. \tag{20}$$

Including the conditions for the minors we have,

$$\begin{aligned} \rho(\mathbf{X}, \mathbf{Y})^2 + \rho(\mathbf{X}, \mathbf{Z})^2 + \rho(\mathbf{Y}, \mathbf{Z})^2 - 2\rho(\mathbf{X}, \mathbf{Y})\rho(\mathbf{X}, \mathbf{Z})\rho(\mathbf{Y}, \mathbf{Z}) &\leq 1 \\ \rho(\mathbf{X}, \mathbf{Y})^2 &\leq 1 \\ \rho(\mathbf{Y}, \mathbf{Z})^2 &\leq 1. \end{aligned} \quad (21)$$

The last three inequalities are automatically satisfied since the correlations are bounded by  $\pm 1$ .

**Simultaneous observations and joint distributions.** When observations are simultaneous and the environment is stable and stationary, so that with repeated simultaneous observations satisfactory frequency data can be obtained, then there exists a joint distribution of all of the random variables representing the simultaneous observations. Note what we can then conclude from the above: in all such cases there must be, therefore, a factorizing hidden variable because of the existence of the joint probability distribution. From this consideration alone, it follows that any of the quantum mechanical examples that violate Bell's inequalities or other criteria for hidden variables must be such that not all the observations in question can be made simultaneously. The extension of this criterion of simultaneity to a satisfactory relativistic criterion is straightforward.

## 1 Appendix

We prove here several elementary lemmas about probability-1 statements used in the proof of Theorem 9.

**Lemma 1** *If  $P(A | B) = 1$  and  $P(BC) > 0$  then  $P(A | BC) = 1$ .*

*Proof.* Suppose, by way of contradiction, that

$$P(A | BC) < 1. \quad (22)$$

Now from (22) and the definition of conditional probability, we have at once

$$P(ABC) < P(BC). \quad (23)$$

Adding  $P(AB\overline{C})$  to both sides of (23) and simplifying we have

$$P(AB) < P(BC) + P(AB\overline{C}). \quad (24)$$

We now take conditional probabilities with respect to  $B$ , and divide both sides of (24) by  $P(B)$ , for by the hypothesis of the lemma,  $P(B) > 0$ , and thus we obtain

$$P(A | B) < P(C | B) + P(A\overline{C} | B),$$

but

$$P(C | B) + P(A\overline{C} | B) \leq 1$$

and by the hypothesis of the lemma

$$P(A | B) = 1,$$

whence we have derived the absurdity that  $1 < 1$ . Thus the lemma is established.

**Lemma 2** *Let  $\mathbf{X}$  and  $\mathbf{Y}$  be two random variables with a joint distribution, and let*

(i)  $P(A) > 0$ ,

(ii)  $P(\mathbf{X} = c | A) = 1$ ,

(iii)  $P(\mathbf{Y} = c | A) = 1$ .

*Then*

$$P(\mathbf{X} = \mathbf{Y} | A) = 1.$$

*Proof.* Let

$$B = \{\omega : \mathbf{X}(\omega) = c\}$$

$$C = \{\omega : \mathbf{Y}(\omega) = c\}$$

$$D = \{\omega : \mathbf{X}(\omega) = \mathbf{Y}(\omega)\}$$

Suppose by way of contradiction that

$$P(D | A) < 1.$$

Then

$$P(\{\omega : \mathbf{X}(\omega) \neq \mathbf{Y}(\omega)\} | A) > 0.$$

And so

$$P(\{\omega : X(\omega) \neq c \text{ or } Y(\omega) \neq c\} | A) > 0.$$

Without loss of generality, let

$$P(\{\omega : \mathbf{X}(\omega) \neq c\} | A) > 0.$$

Then

$$P(\overline{B} | A) > 0,$$

and this contradicts (ii).

We also need a sort of converse of Lemma 2.

**Lemma 3** *If  $P(A \& \mathbf{X} = c) > 0$  and  $P(\mathbf{X} = \mathbf{Y} | A \& \mathbf{X} = c) = 1$  then*

$$P(\mathbf{Y} = c | A \& \mathbf{X} = c) = 1.$$

*Proof.* By hypothesis

$$P(\mathbf{X} = \mathbf{Y} \& A \& \mathbf{X} = c) = P(A \& \mathbf{X} = c).$$

Consider now the left-hand side:

$$\begin{aligned} \{\omega : X(\omega) = Y(\omega)\} \& \{\omega : X(\omega) = c\} &= \{\omega : \mathbf{X}(\omega) = c \& \mathbf{Y}(\omega) = c\} \\ &= \{\omega : \mathbf{X}(\omega) = c\} \cap \{\omega : \mathbf{Y}(\omega) = c\}, \end{aligned}$$

and so

$$P(\mathbf{X} = \mathbf{Y} \& A \& \mathbf{X} = c) = P(\mathbf{Y} = c \& A \& \mathbf{X} = c),$$

and thus,

$$P(\mathbf{Y} = c \& A \& \mathbf{X} = c) = P(A \& \mathbf{X} = c),$$

whence

$$P(\mathbf{Y} = c | A \& \mathbf{X} = c) = 1.$$

We can also prove a kind of transitivity for conditional probabilities that are 1.

**Lemma 4** *If  $P(B) > 0$ ,  $P(C) > 0$ ,  $P(A | B) = 1$  and  $P(B | C) = 1$ , then  $P(A | C) = 1$ .*

*Proof.* By hypothesis and Lemma 1

$$P(A \mid BC) = 1,$$

so

$$P(ABC) = P(BC)$$

but by hypothesis

$$P(BC) = P(C),$$

so

$$P(ABC) = P(C),$$

and thus

$$P(AB \mid C) = 1,$$

whence

$$P(A \mid C) = 1.$$

Finally, we also use the following,

**Lemma 5** *If  $P(A \& \mathbf{Y} = d) > 0$ ,  $P(A \& \mathbf{Z} = d) > 0$ , and*

$$(i) P(\mathbf{X} = c \mid A \& \mathbf{Y} = d) = 1,$$

$$(ii) P(\mathbf{Z} = \mathbf{Y} \mid A \& \mathbf{Z} = d) = 1,$$

*then*

$$P(\mathbf{X} = c \mid A \& \mathbf{Z} = d) = 1.$$

*Proof.* By Lemma 3 and (ii)

$$P(A \& \mathbf{Y} = d \mid A \& \mathbf{Z} = d) = 1$$

So, by transitivity (Lemma 4) & (i)

$$P(\mathbf{X} = c \mid A \& \mathbf{Z} = d) = 1$$

**Acknowledgments.** We are grateful to Martin Jones for useful comments and criticisms of an earlier draft. J. A. B. acknowledges financial support from FAPEMIG (Minas Gerais State Support Agency) and support from the Laboratory for Experimental Cosmology and HEP (Lafex) of the Brazilian Centre for Physical Research (CBPF).



## References

- [1] BELL J. S., *Physics*, **1**, (1964) 195.
- [2] BELL, J. S., *Rev. Mod. Phys.* **38**, (1966) 447.
- [3] CLAUSER J. F. and SHIMONY A. , *Rep. Prog. Phys.* **41**, (1978) 1881.
- [4] FINE A. , *Phys. Rev. Lett.* **48**, (1982) 291.
- [5] GREENBERGER D. M. , HORNE M. A. and ZEILINGER A. “Going beyond Bell’s theorem”, in *Bell’s Theorem, Quantum Theory, and Conceptions of the Universe*, edited by M. KAFATOS (Kluwer Academic, Dordrecht, The Netherlands) 1989, 69–72.
- [6] GREENBERGER D. M., HORNE M. A., SHIMONY A. and ZEILINGER A., *Amer. J. Phys.* **58**, (1990) 1131.
- [7] HOLLAND P. W. and ROSENBAUM T. R. *Ann. Statist.* **14**, (1986) 1523.
- [8] LOÈVE M. *Probability Theory II*, 4th edition. (Springer Verlag, New York) 1978.
- [9] MERMIN N. D. *Am. J. Phys.* **58**, (1990), 731.
- [10] PERES A. *Phys. Rev. A*, **46**, (1992), 4413.
- [11] SUPPES P. and ZANOTTI M. in *Logic and Probability in Quantum Mechanics*, edited by P. SUPPES (Reidel, Dordrecht) 1976, 445.
- [12] SUPPES P. and ZANOTTI M. “A new proof of the impossibility of hidden variables using the principles of exchangeability and identity of conditional distributions”, in *Studies in the Foundations of Quantum Mechanics*, edited by P. SUPPES (Philosophy of Science Association, East Lansing, Michigan) 1980, 173-191.
- [13] SUPPES P. and ZANOTTI M. *Synthese* **48**, (1981) 191.