Foundations of Physics, Vol. 21, No. 12, 1991

Existence of Hidden Variables Having Only Upper Probabilities[1]

Patrick Suppes[2] and Mario Zanotti[2]

Received October 9, 1991

We prove the existence of hidden variables, or, what we call generalized common causes, for finite sequences of pairwise correlated random variables that do not have a joint probability distribution. The hidden variables constructed have upper probability distributions that are nonmonotonic. The theorem applies directly to quantum mechanical correlations that do not satisfy the Bell inequalities.

Great attention has been given to the exact formulation of locality conditions in the existence of hidden variables for quantum mechanical configurations that do not satisfy Bell's inequalities. In contrast, almost no attention has been given to examining what happens if the requirements for a probability measure are relaxed. The purpose of this paper is to show that local hidden variables do exist if the requirement of a probability measure is weakened to that of having only an upper probability measure. In the past several decades, upper and lower measures have been brought into the theory of statistical inference and the measurement of beliefs rather extensively, but in almost all cases considered in statistical theory or the belief framework, it is assumed that a probability measure exists, or in many cases, that an entire family of probability measures is given such that the sup and inf of the family defines the upper and lower measures. Sometimes still more is required, namely, that the upper–lower measure be a capacity (in the sense of Choquet) of infinite order. This is, for example,

[1] It is a pleasure to dedicate this paper to Karl Popper in celebration of this 90th birthday. The first author has known Popper for more than three decades, and has profited much from their discussion of many different topics, among which have been the foundations of probability and the foundations of quantum mechanics, both central to the present paper.

[2] Department of Philosophy, Stanford University, Stanford, California.

true of Dempster's theory of statistical inference. For an extensive recent analysis of the various theories of imprecise statistical models or belief functions using upper and lower probabilities, Walley (1) is an excellent reference.

In contrast and as is well known, in standard quantum mechanical cases the Bell inequalities cannot be satisfied and this implies that there exists no probability measure. As far as we know, the present application is almost unique in being an example in which upper probabilities can be defined in a natural way and yet there is no physically possible probability measure. Why we ignore lower probabilities is explained in detail at the end of Sec. 2, but the basis is easy to state. The upper probabilities in the situations we consider are nonmonotonic and for such upper probabilities the standard definition of lower probabilities does not work.

From a more general philosophical standpoint it is clear that views about hidden variables are very much conditioned on having a complete probabilistic account of the phenomena considered. Almost certainly, most physicists and philosophers will not find the existence of upper probabilities, without corresponding probabilities, physically appealing. Certainly as we ourselves have pointed out in earlier articles on upper and lower probabilities (Suppes and Zanotti[(2,3)]), the natural questions about conditional upper and lower probabilities and independence are much more complicated than in the standard probabilistic framework. On the other hand, there has been very extensive consideration of weakening classical logic to various special forms of quantum logic. There may be for the Bell-type situations equally good arguments for keeping the logic standard, but weakening the probability requirements in the direction we develop in this paper.

To illustrate the development of ideas, in the next section we consider the standard elementary example of three random variables that do not have a joint probability distribution, namely three random variables with values ± 1 and with the correlation of any pair being -1. We construct an upper measure for these three random variables and also show the existence of a hidden variable—what we call a *generalized* random variable—with respect to the upper measure. In the following section we prove a theorem about the existence of generalized common causes, i.e., hidden variables, for finite sequences of pairwise correlated random variables. In Sec. 3 we apply the same techniques to a standard example of quantum mechanics that does not satisfy the Bell inequalities. We also generalize this quantum mechanical example to a theorem about such cases.

1. THREE RANDOM VARIABLES WITH MAXIMUM NEGATIVE CORRELATIONS

As already indicated, we consider in this section three random variables $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3$ with values ± 1 and expectations

$$E(\mathbf{X}_1) = E(\mathbf{X}_2) = E(\mathbf{X}_3) = 0$$

$$\mathrm{Cov}(\mathbf{X}_i, \mathbf{X}_j) = -1, \qquad i \neq j$$

We can express the covariance and correlation just in terms of the expectation because the standard deviations are one and the expectations are zero:

$$E(\mathbf{X}_i \mathbf{X}_j) = -1, \qquad i \neq j$$

We use the notation

$$p_{i\bar{j}} = P(\mathbf{X}_i = 1, \mathbf{X}_j = -1), \text{ etc.}$$

So

$$p_{i\bar{j}} = p_{\bar{i}j} = \tfrac{1}{2}, \qquad i \neq j$$

$$p_{ij} = p_{\bar{i}\bar{j}} = 0$$

This implies, to fit the correlations,

$$p^*_{i\bar{j}} = \tfrac{1}{2}, \qquad p^*_{\bar{i}j} = \tfrac{1}{2}$$

$$p^*_{ij} = 0, \qquad p^*_{\bar{i}\bar{j}} = 0$$

In the previous four equations we have used standard notation for the upper probabilities, a superscript star. More generally, we have the following axioms on upper probability, embodied in a definition.

Definition 1. Let Ω be a nonempty set, $\mathscr{F}$ a Boolean algebra on Ω, and P^* a real-valued function on $\mathscr{F}$. Then $\Omega = (\Omega, \mathscr{F}, P^*)$ is an *upper probability space* if and only if for every A and B in $\mathscr{F}$

1. $0 \leqslant P^*(A) \leqslant 1$;
2. $P^*(\varnothing) = 0$ and $P^*(\Omega) = 1$;
3. If $A \cap B = \varnothing$, then $P^*(A \cup B) \leqslant P^*(A) + P^*(B)$.

Axiom 3 on finite subadditivity could be strengthened to σ-subadditivity, but we are not concerned with that issue here.

To be perfectly clear about our notation, note that

$$p^*_{i\bar{j}} = P^*(\mathbf{X}_i = 1, \mathbf{X}_j = -1)$$

Since "mixed" $i\bar{j}$ or $\bar{i}j$ never occur in p^*_{123} or $p^*_{\bar{1}\bar{2}\bar{3}}$, we may set

$$p^*_{123} = p^*_{\bar{1}\bar{2}\bar{3}} = 0$$

By symmetry and to satisfy subadditivity—e.g., $p^*_{1\bar{2}} \leqslant p^*_{1\bar{2}3} + p^*_{1\bar{2}\bar{3}}$, since

$$p^*_{i\bar{j}} = p^*_{\bar{i}j} = \tfrac{1}{2}, \qquad \text{for} \quad i \neq j$$

we set the remaining 6 triples at 1/4:

$$p^*_{12\bar{3}} = p^*_{1\bar{2}3} = p^*_{\bar{1}23} = p^*_{1\bar{2}\bar{3}} = p^*_{\bar{1}2\bar{3}} = p^*_{\bar{1}\bar{2}3} = \tfrac{1}{4} \tag{1}$$

Notice that P^* is nonmonotonic for $p^*_{12\bar{3}} > p_{12\cdot} = 0$. We examine this phenomenon in detail in Sec. 2.

We now define in the expected manner the upper expectation of a random variable which we express here for $\mathbf{X}_1$:

$$\begin{aligned} E^*(\mathbf{X}_1) &= \sum x p^*(x) \\ &= 1(p^*_{12\bar{3}} + p^*_{1\bar{2}3} + p^*_{1\bar{2}\bar{3}}) + (-1)(p^*_{\bar{1}23} + p^*_{\bar{1}\bar{2}3} + p^*_{\bar{1}2\bar{3}}) \\ &= 0 \end{aligned} \tag{2}$$

By symmetry

$$E^*(\mathbf{X}_i) = E(\mathbf{X}_i) = 0, \qquad \text{for} \quad i = 1, 2, 3 \tag{3}$$

It is also obvious how we define the upper expectation of the product of two random variables:

$$\begin{aligned} E^*(\mathbf{X}_i\mathbf{X}_j) &= \sum x_i x_j\, p^*(x_i x_j), \qquad i \neq j \\ &= (-1)\, p^*_{i\bar{j}} + (-1)\, p^*_{\bar{i}j} \\ &= -\tfrac{1}{2} + -\tfrac{1}{2} \\ &= -1 \end{aligned}$$

So the correlations are preserved:

$$E^*(\mathbf{X}_i\mathbf{X}_j) = E(\mathbf{X}_i\mathbf{X}_j) = -1, \qquad i \neq j \tag{4}$$

Hidden variable. We now turn to the existence of a hidden variable for the three random variables. Given

$$E(\mathbf{X}_i\mathbf{X}_j) = -1, \qquad 1 \leqslant i < j \leqslant 3 \tag{5}$$

we want to find a hidden variable $\boldsymbol{\lambda}$ so that

$$E^*(\mathbf{X}_i\mathbf{X}_j \mid \boldsymbol{\lambda} = \lambda) = E^*(\mathbf{X}_i \mid \boldsymbol{\lambda} = \lambda)\, E^*(\mathbf{X}_j \mid \boldsymbol{\lambda} = \lambda) \tag{6}$$

where the upper conditional expectation for any random variable $\mathbf{X}$ is defined in the obvious conditional way:

$$E^*(\mathbf{X} \mid \boldsymbol{\lambda} = \lambda) = \sum x P^*(\mathbf{X} = x \mid \boldsymbol{\lambda} = \lambda) \tag{7}$$

with, of course, upper conditional probability having the usual definition

$$P^*(\mathbf{X} = x \mid \boldsymbol{\lambda} = \lambda) = P^*(\mathbf{X} = x, \boldsymbol{\lambda} = \lambda)/P^*(\boldsymbol{\lambda} = \lambda) \tag{8}$$

for all values λ.

Moreover, we also want to satisfy

$$E(\mathbf{X}_i\mathbf{X}_j \mid \boldsymbol{\lambda} = \lambda) = E^*(\mathbf{X}_i\mathbf{X}_j \mid \boldsymbol{\lambda} = \lambda) \tag{9}$$

There is no uniqueness requirement on the hidden variable $\boldsymbol{\lambda}$. Many different characterizations will work. The most transparent construction in our judgment is the deterministic one that mirrors as closely as possible the simultaneous possible values of the three random variables $\mathbf{X}_1$, $\mathbf{X}_2$ and $\mathbf{X}_3$. So we choose

$$\boldsymbol{\lambda} = (\pm 1, \pm 1, \pm 1)$$

where the ith coordinate of the vector corresponds to the values of $\mathbf{X}_i$. When we consider the correlation of $\mathbf{X}_1$ and $\mathbf{X}_2$, for instance, we write $(1, 1, \cdot) = \{(1, 1, 1), (1, 1, -1)\}$.

Then

$$p^*_{1\bar{2}} = P^*(\mathbf{X}_1 = 1, \mathbf{X}_2 = -1, \boldsymbol{\lambda} \in (1, -1, \cdot)) \tag{10}$$

since $\boldsymbol{\lambda}$ is deterministic.

Since $p_{1\bar{2}} = p^*_{1\bar{2}}$, we have

$$P(\mathbf{X}_1 = 1, \mathbf{X}_2 = -1 \mid \boldsymbol{\lambda} \in (1, -1, \cdot)) = 1 \tag{11}$$

and

$$P(\boldsymbol{\lambda} \in (1, -1, \cdot)) = P^*(\boldsymbol{\lambda} \in (1, -1, \cdot)) = \tfrac{1}{2}$$

but, of course, we must have

$$P^*(\boldsymbol{\lambda} = (1, -1, 1)) = P^*(\boldsymbol{\lambda} = (1, -1, -1)) = \tfrac{1}{4}$$

for the upper probabilities of the values of $\boldsymbol{\lambda}$ mirror the values of $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3$, taken together.

We now have the same strategy as in (1), for assigning upper probabilities of 1/4 to each pair generated by terms like that of (9). Thus

$$\begin{aligned} p^*(1, \bar{2}, (1, -1, \cdot)) = \tfrac{1}{2} \leqslant p^*&(1, \bar{2}, 3, (1, -1, 1)) \\ &+ p^*(1, \bar{2}, \bar{3}, (1, -1, -1)) = \tfrac{1}{4} + \tfrac{1}{4} \end{aligned}$$

The conditional expectations for the pairs $\mathbf{X}_i\mathbf{X}_j$, $i \neq j$, then all have the expected deterministic values, for

$$E^*(X_iX_j \mid \boldsymbol{\lambda} = \lambda) = E(X_iX_j \mid \boldsymbol{\lambda} = \lambda)$$

More important is the factorization. For example,

$$\begin{aligned} E(X_1X_2 \mid \boldsymbol{\lambda} \in (1, -1, \cdot)) &= E(X_1 \mid \boldsymbol{\lambda} \in (1, -1, \cdot))\, E(X_2 \mid \boldsymbol{\lambda} \in (1, -1, \cdot)) \\ &= 1 \cdot (-1) \\ &= -1 \end{aligned}$$

(Notice that we have written E, since here $E^* = E$ in the restricted cases considered.) Because the hidden variable $\boldsymbol{\lambda}$ is deterministic, all aspects of the upper probabilities are entirely reflected in the upper probabilities for the values of $\boldsymbol{\lambda}$. Thus

$$\begin{aligned} p^*(\pm 1, \pm 1, \pm 1) = \tfrac{1}{4} \text{ except } p^*(1, 1, 1) = p^*(-1, -1, -1) = 0 \\ p^*(1, -1, \cdot) = \tfrac{1}{2}, \text{ etc.} \\ p^*(1, \cdot, \cdot) = \tfrac{1}{2}, \text{ etc.} \end{aligned}$$

The hidden variable constructed above mirrors quite directly the upper probability measure on the space of possible outcomes of the three random variables $\mathbf{X}_1$, $\mathbf{X}_2$ and $\mathbf{X}_3$. As can be seen from the construction, if we only consider a pair, say $\mathbf{X}_1$ and $\mathbf{X}_2$ with correlation -1, then a hidden variable that is a random variable can be constructed and no generalization to an upper measure is needed. Additional requirements of symmetry with respect to $\boldsymbol{\lambda}$ can change the situation, as we now show.

Symmetry with respect to $\boldsymbol{\lambda}$. In an earlier paper, we proved the following theorem that gives a necessary and sufficient condition for two-valued exchangeable random variables to have "identical" causes in the sense of conditional expectation.

Theorem (Suppes and Zanotti[4]). Let **X** and **Y** be two-valued random variables, for definiteness, with possible values 1 and -1, and with positive variances, i.e., $\sigma(X), \sigma(Y) > 0$. In addition, let **X** and **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 $E(\mathbf{XY} \mid \boldsymbol{\lambda}=\lambda) = E(\mathbf{X} \mid \boldsymbol{\lambda}=\lambda)\, E(\mathbf{Y} \mid \boldsymbol{\lambda}=\lambda)$ and $E(\mathbf{X} \mid \boldsymbol{\lambda}=\lambda) = E(\mathbf{Y} \mid \boldsymbol{\lambda}=\lambda)$ for every value λ (except possibly on a set of measure zero) is that the correlation of **X** and **Y** be nonnegative.

Random variables $\mathbf{X}_1$ and $\mathbf{X}_2$ as defined earlier with correlation -1 satisfy the hypothesis of this theorem, and so they can have no hidden variable within standard probability theory satisfying the symmetry condition of causality:

$$E(\mathbf{X}_1 \mid \boldsymbol{\lambda}=\lambda) = E(\mathbf{X}_2 \mid \boldsymbol{\lambda}=\lambda) \tag{12}$$

On the other hand, we can find a hidden variable with an upper probability measure satisfying (12), as well as the standard factorization, but now in terms of *upper* conditional expectations, which we defined earlier.

The specification of p^* is as follows, where the subscript 3 refers to values of $\boldsymbol{\lambda}$, and, as before, 3 indicates the value $\boldsymbol{\lambda}=1$ and $\bar{3}$ the value $\boldsymbol{\lambda}=-1$, with only two values of $\boldsymbol{\lambda}$ being needed for this analysis:

$$p^*_{ijk} = \tfrac{1}{4}$$

for all 8 combinations of ijk.

Also

$$p_{1\cdot\cdot} = p_{\cdot 2 \cdot} = p_{\cdot\cdot 3} = \tfrac{1}{2}$$

$$p_{12\cdot} = p_{\bar{1}\bar{2}\cdot} = 0$$

$$p_{1\bar{2}\cdot} = p_{\bar{1}2\cdot} = \tfrac{1}{2}$$

The rest of the specification to satisfy the constraints of the situation are obvious, although $\boldsymbol{\lambda}$ is a stochastic rather than deterministic hidden variable, so it easy to compute that

$$E^*(\mathbf{X}_1 \mid \boldsymbol{\lambda}=\pm 1) = E^*(\mathbf{X}_2 \mid \boldsymbol{\lambda}=\pm 1) = 0$$

$$E^*(\mathbf{X}_1\mathbf{X}_2 \mid \boldsymbol{\lambda}=\pm 1) = 0$$

which satisfy the condition required for symmetry in "the" cause of a negative correlation. But as can be seen from the value of p^*_{ijk}, we must construct p^* to be nonmonotonic.

2. GENERALIZED RANDOM VARIABLES AS GENERALIZED COMMON CAUSES

The hidden variable λ of the previous section suggests at once a generalization of the standard notion of a random variable. As the example suggests, it will be convenient for a generalized random variable to take as values vectors of real numbers of a given dimension, rather than simply real numbers. But this is not the real point of our use of the term *generalized*; it is rather the generalization from a random variable's having a probability to having an upper probability.

Let $\mathbf{\Omega} = (\Omega, \mathscr{F}, P^*)$ be an upper probability space and let λ be a function from Ω to Re^k such that for every vector $(b_1,..., b_k)$ the set

$$\{\omega : \omega \in \Omega \;\&\; \lambda_i(\omega) \leqslant b_i,\ i = 1,..., k\}$$

is in $\mathscr{F}$. Then λ is a *generalized random variable* (*with respect to* $\mathbf{\Omega}$).

Extending the earlier example, we can then prove that for any finite set of two-valued correlated random variables there exists an underlying generalized random variable that makes each of the correlated pairs conditionally independent. Put in more general philosophical language, we prove the existence of a generalized common cause.

This result extends our earlier result[5] that a joint probability of all the random variables is necessary and sufficient for the existence of a common cause. Of course, the extension is at the expense of using the weaker concept of an upper probability measure.

Before stating the theorem, we introduce some concepts and notation that are needed. Let Ω be the space on which the random variables $\mathbf{X}_1,..., \mathbf{X}_n$ are defined, and let $\mathscr{F}$ be the given algebra of events, i.e., subsets of Ω. We require that $\mathscr{F}$ contain as a subalgebra $\mathscr{F}^*$, the algebra of cylinder sets of Ω defined by the sequence of values of the random variables $\mathbf{X}_1,...,\mathbf{X}_n$. First, we characterize any index set (of dimension no greater than n) by two disjoint sets of positive integers I and J, with the largest integer in either set being no greater than n. The set I is the set of indices i for which $\mathbf{X}_i = 1$ and the set J is the set of indices i for which $\mathbf{X}_i = -1$. So a pair (I, J) uniquely defines an index set, an element of $\mathscr{F}^*$, relative to $\mathbf{X}_1,...,\mathbf{X}_n$.

Note that if $I \cup J = \varnothing$, then $(I, J) = \Omega$, and if $|I \cup J| = n$, i.e., the cardinality of the set $I \cup J$ is n, then (I, J) is an atom of $\mathscr{F}^*$. To illustrate the notation concretely, if $n = 3$, then $(\{1\}, \{3\}) = (1, \cdot, -1)$, in our earlier notation.

It is easy to show explicitly that the algebra $\mathscr{F}^*$ of cylinder sets results from taking the closure under union and complementation of the index sets. So if (I, J) is an index, then its complement is a cylinder set, and if (I_1, J_1) and (I_2, J_2) are two index sets, then $(I_1, J_1) \cup (I_2, J_2)$ is a cylinder set.

What is central to the theorem is to begin by assuming only pairwise probability functions of the random variables and then to construct an upper measure P^* on $\mathscr{F}^*$. The measure constructed is in general not unique.

Theorem 1 (Generalized Common Causes). Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be two-valued (± 1) random variables whose common domain is a space Ω with an algebra $\mathscr{F}$ of events that includes the subalgebra $\mathscr{F}^*$ of cylinder sets of dimension n defined above. Also, let pairwise probability functions P_{ij}, $1 \leqslant i < j \leqslant n$, compatible with the single functions P_i, $1 \leqslant i \leqslant n$, be given. Then there exists an upper probability space $\mathbf{\Omega} = (\Omega, \mathscr{F}^*, P^*)$ and a generalized random variable $\boldsymbol{\lambda}$ on Ω to the set of n-dimensional vectors whose components are ± 1 such that for $1 \leqslant i < j \leqslant n$ and every value λ of $\boldsymbol{\lambda}$.

(i) $P^*(\mathbf{X}_i = \pm 1, \mathbf{X}_j = \pm 1) = P_{ij}(\mathbf{X}_i = \pm 1, \mathbf{X}_j = \pm 1)$;

(ii) $P^*(\mathbf{X}_1 = \lambda_1, \ldots, \mathbf{X}_n = \lambda_n) = P^*(\boldsymbol{\lambda}_1 = \lambda_1, \ldots, \boldsymbol{\lambda}_n = \lambda_n)$;

(iii) $\boldsymbol{\lambda}$ is *deterministic*, i.e.,

$$P(\mathbf{X}_i = 1 \mid \boldsymbol{\lambda}_i = 1) = 1$$

and

$$P(\mathbf{X}_i = -1 \mid \boldsymbol{\lambda}_i = -1) = 1$$

(iv) $E(X_i X_j \mid \boldsymbol{\lambda} = \lambda) = E(X_i \mid \boldsymbol{\lambda} = \lambda)\, E(X_i \mid \boldsymbol{\lambda} = \lambda)$.

Proof. We first set $P^*(\Omega) = 1$. As shown earlier, a pair (I, J) of mutually exclusive sets of positive integers uniquely defines an index set. In this notation, by the hypothesis of the theorem we are given for a singleton or pair set—here the cardinality of $I \cup J$ is 1 or 2—the probability, so we set

$$P^*((I, J)) = P_{ij}((I, J)), \qquad 1 \leqslant |I \cup J| \leqslant 2 \tag{13}$$

Consider now any index set (I, J) such that $|I \cup J| > 2$. We define recursively

$$P^*((I, J)) = \tfrac{1}{2} \max_{I', J'}(P^*(I', J), P^*(I, J')) \tag{14}$$

where, if $I \neq \varnothing$, then $I' \subseteq I$, $|I'| = |I| - 1$; if $J \neq \varnothing$, then $J' \subseteq J$, $|J'| = |J| - 1$; if $I = \varnothing$, then $I' = \varnothing$; and if $J = \varnothing$, then $J' = \varnothing$.

We now prove that P^* is subadditive on the index sets. We assert that for any $i \notin I \cup J$ with $|I \cup J| < n$

$$P^*((I, J)) \leqslant P^*((I \cup \{i\}, J)) + P^*((I, J \cup \{i\})) \tag{15}$$

(Of course if $|I \cup J| = n$, then (I, J) is an atom.)

From (14), we infer, for $i \notin I \cup J$

$$P^*((I \cup \{i\}, J)) = \tfrac{1}{2} \max P^*((I \cup \{i\})', J), P^*((I \cup \{i\}, J'))$$

where the prime operation and max are as defined for (14). Thus

$$P^*(I \cup \{i\}, J) \geqslant \tfrac{1}{2} P^*((I, J)) \tag{16}$$

and by a similar argument

$$P^*((I, J \cup \{i\})) \geqslant \tfrac{1}{2} P^*((I, J)) \tag{17}$$

Inequality (15) follows at once from (16) and (17), so P^* is an upper measure on the index sets of $(\Omega, \mathscr{F}^*)$. We now extend P^* to all of the cylinder sets (of dimension not greater than n), i.e., to all of $(\Omega, \mathscr{F}^*)$. To avoid heavy notation we first describe informally this extension. For any cyclinder set A that is not an index set, we consider all partitions of the set in terms of index sets. For each partition we take the arithmetic sum of the upper measure P^* of the sets in the partition. We now take the min of P^* over these partitions as $P^*(A)$. In symbols

$$P^*(A) = \min_{\Pi(A)} \sum_{L_i \in \Pi(A)} P^*(L_i) \tag{18}$$

where $\Pi(A)$ is a partition of A in terms of index sets L_i. The min operation guarantees that the axiom of subadditivity will be satisfied by A, i.e., we have at once from (18)

$$P^*(A) \leqslant \sum_{L_i \in \Pi(A)} P^*(L_i) \tag{19}$$

The same argument applies to Ω, so that

$$P^*(\Omega) = P^*(A \cup \bar{A}) \leqslant P^*(A) + P^*(\bar{A})$$

and from (18) it is also clear that

$$0 \leqslant P^*(A) \leqslant 1$$

which completes the proof that P^* is an upper probability on $(\Omega, \mathscr{F}^*)$. Equation (13) guarantees satisfaction of (i) of the theorem.

The hidden variable λ, which is a function from Ω to the set of n-dimensional vectors whose components are ± 1, is defined to be deterministic, i.e., for $1 \leqslant i \leqslant n$:

$$P(\mathbf{X}_i = 1 \mid \lambda_i = 1) = 1$$

and

$$P(\mathbf{X}_i = -1 \mid \lambda_i = -1) = 1$$

so the upper probability function for λ is the same as that of $\mathbf{X}_1, \ldots, \mathbf{X}_n$, which establishes (ii) and (iii).

Finally, since λ is deterministic, the conditional independence of $\mathbf{X}_i$ and $\mathbf{X}_j$, $i \neq j$, given $\lambda = \lambda$, follows at once, which establishes (iv) and completes the proof of the theorem.

Status of Monotonicity. A familiar strong axiom for upper measures in the context of statistical inference is the monotonicity axiom:

$$\text{If } A \subseteq B, \text{ then } P^*(A) \leqslant P^*(B)$$

This axiom is fundamental for Walley,[1] Dempster, Shafer, and others. Unfortunately, it is violated already by the example given in the preceding section of three random variables having pairwise correlations of -1. When we say "violated," we mean it is easy to show that no upper measure compatible with the given pairwise probability distributions can satisfy the axiom of monotonicity. Intuitively speaking, satisfaction of this axiom in the case of three two-valued variables implies there is a probability measure dominated by the upper measure and compatible with the pairwise distributions, so that an upper measure that is monotonic cannot be constructed for cases where no probability distribution exists. The following two theorems summarize these facts.

Theorem 2 (Monotonicity Implies Probability). Let $\mathbf{X}_1$, $\mathbf{X}_2$ and $\mathbf{X}_3$ be two-valued (± 1) random variables with $E(\mathbf{X}_i) = 0$, $i = 1, 2, 3$, such that there is a monotonic upper probability function compatible with the given correlations $E(\mathbf{X}_i\mathbf{X}_j)$, $1 \leqslant i < j \leqslant 3$. Then there exists a joint probability function of $\mathbf{X}_1$, $\mathbf{X}_2$ and $\mathbf{X}_3$.

Proof. A necessary and sufficient condition for the existence of a joint probability distribution of three random variables satisfying the hypothesis of the theorem is given in Ref. 5:

$$\begin{aligned} -1 &\leqslant E(\mathbf{X}_1\mathbf{X}_2) + E(\mathbf{X}_2\mathbf{X}_3) + E(\mathbf{X}_1\mathbf{X}_3) \\ &\leqslant 1 + 2\min(E(\mathbf{X}_1\mathbf{X}_2), E(\mathbf{X}_2\mathbf{X}_3), E(\mathbf{X}_1\mathbf{X}_3)) \end{aligned} \tag{20}$$

For brevity of notation we write $\rho_{ij} = E(X_iX_j)$, and we prove that when P^* is monotonic, (20) follows.

Since $E(X_i) = 0$, it is obvious that

$$p_{ij} - p_{i\bar{j}} = \frac{\rho_{ij}}{2}$$

$$p_{ij} + p_{i\bar{j}} = \frac{1}{2}$$

so

$$\begin{aligned} 4p_{ij} &= 1 + \rho_{ij} \\ 4p_{i\bar{j}} &= 1 - \rho_{i\bar{j}} \end{aligned} \tag{21}$$

By subadditivity we then have

$$4p_{1\bar{2}\cdot} = 1 - \rho_{12} \leqslant 4p^*_{1\bar{2}3} + 4p^*_{1\bar{2}\bar{3}} \tag{22}$$

And by monotonicity

$$\begin{aligned} 4p^*_{1\bar{2}3} &\leqslant 4p_{1\cdot 3} = 1 + \rho_{13} \\ 4p^*_{1\bar{2}\bar{3}} &\leqslant 4p_{\cdot\bar{2}\bar{3}} = 1 + \rho_{23} \end{aligned} \tag{23}$$

From (22) and (23)

$$1 - \rho_{12} \leqslant 1 + \rho_{23} + 1 + \rho_{13}$$

and so

$$-1 \leqslant \rho_{12} + \rho_{23} + \rho_{13}$$

Second, without loss of generality, we may assume

$$\rho_{12} = \min(\rho_{12}, \rho_{23}, \rho_{13}) \tag{24}$$

But using (21) again and monotonicity

$$\begin{aligned} 1 + \rho_{13} &\leqslant 4p^*_{123} + 4p^*_{\bar{1}\bar{2}3} \\ &\leqslant 1 + \rho_{12} + 1 - \rho_{23} \end{aligned}$$

so

$$\rho_{23} + \rho_{13} \leqslant 1 + \rho_{12} \tag{25}$$

but (24) and (25) imply

$$\rho_{12} + \rho_{23} + \rho_{13} \leqslant 1 + 2\min(\rho_{12}, \rho_{23}, \rho_{13})$$

which completes the proof.

The proof of the following theorem on monotonicity is an immediate consequence of the proof of Theorem 2.

Theorem 3 (Nonmonotonicity). Let $\mathbf{X}_1$, $\mathbf{X}_2$ and $\mathbf{X}_3$ be two-valued (± 1) random variables with $E(X_i) = 0$, $i = 1, 2, 3$, such that there is no joint probability distribution compatible with the correlations $E(\mathbf{X}_i\mathbf{X}_j)$, $1 \leqslant i < j \leqslant 3$. Then any upper measure P^* compatible with the given correlations cannot satisfy the axiom of monotonicity.

The consequences of nonmonotonicity reach even further. In the usual statistical applications, where it is assumed that P^* is monotonic, the lower probability of any event A is defined by

$$P_*(A) = 1 - P^*(\bar{A}) \tag{26}$$

and under the assumption of monotonicity it can be proved that P_* is superadditive, i.e., if $A \cap B = \varnothing$, then

$$P_*(A) + P_*(B) \leqslant P_*(A \cup B)$$

We now prove that P_* when defined by (26) cannot be superadditive if P^* is nonmonotonic, which means that the standard relationship (26) cannot work in the environments we are concerned with.

Theorem 4. Let $(\Omega, \mathscr{F}, P^*)$ be an upper probability space such that P^* is nonmonotonic. Then the lower probability P_* defined by (26) is not superadditive.

Proof. By the hypothesis of the theorem there are events A and B such that $A \subseteq B$ but $P^*(B) < P^*(A)$. Clearly $A \neq B$ for otherwise we would have the immediate contradiction that $P^*(A) < P^*(A)$. So there is a $C \neq \varnothing$ such that $A \cap C = \varnothing$ and

$$A \cup C = B$$

Suppose now that P_* is superadditive. Then since $\bar{A} = \bar{B} \cup C$

$$P_*(\bar{B}) + P_*(C) \leqslant P_*(\bar{A})$$

so

$$1 - P^*(B) + 1 - P^*(\bar{C}) \leqslant 1 - P^*(A)$$

and thus

$$1 + P^*(A) \leqslant P^*(B) + P^*(\bar{C})$$

but $P^*(\bar{C}) \leqslant 1$ by Axiom 1 of Definition 1 for upper probability measures, whence

$$P^*(A) \leqslant P^*(B)$$

contrary to our initial hypothesis, which proves the theorem.

3. HIDDEN VARIABLES IN QUANTUM MECHANICS

We use the standard notation familiar in the Bell inequalities which we review very briefly. For definiteness, but not required, we can think of a Bell-type experiment in which we are measuring spin for particle A and for particle B. More generally, we may think of A and B as being the location of measuring equipment and we observe individual particles or a flux of particles at each of the sites. Here we will think of individual particles because the analysis is simpler. The measuring apparatus is such that along the axis connecting A and B we have axial symmetry and consequently we can describe the position of the measuring apparatus just by the angle of the apparatus A or B in the plane perpendicular to the axis. We use the notation w_A and w_B for these angles. The basic form of the locality assumption is shown in terms of the following expectation:

$$E(\mathbf{M}_A \mid w_A, w_b, \lambda) = E(\mathbf{M}_A \mid w_A, \lambda) \tag{27}$$

What this means is the expectation of the measurement $\mathbf{M}_A$ of spin of a particle in the apparatus in position A, given the two angles of measurement for apparatus A and B as well as λ, is equal to the expectation without any knowledge of the apparatus angle w_B of B. This is a reasonable causal assumption and is a way of saying that what happens at B should have no direct causal influence on what happens at A. On the other hand, we have the following theoretical result for spin, well confirmed in principle for the case where the measuring apparatuses are both set at the same angle:

$$P(\mathbf{M}_A = -1 \mid w_A = w_B = \alpha \;\&\; \mathbf{M}_B = 1) = 1 \tag{28}$$

If the angles of the apparatus are set the same, we have a deterministic result in the sense that the observation of spin at B will be the opposite at A, and conversely. Here we are letting 1 correspond to spin $\frac{1}{2}$ and -1 correspond to spin $-\frac{1}{2}$. What Bell showed is that on the assumption there exists a hidden variable, four related inequalities can be derived for settings A and A' and B and B' for the measuring apparatus. We have reduced the notation here in the following way in writing the inequalities. First, instead of writing $\mathbf{M}_A$ we write simply $\mathbf{A}$, and second, instead of writing $\mathrm{Cov}(\mathbf{A}, \mathbf{B})$ for the covariance, which in this case will be the same as the correlation, of the measurement at A and the measurement at B, we write simply $\mathbf{AB}$. With this understanding about the conventions of the notation, we then have as a consequence of the assumption of a hidden variable the following set of inequalities, which in the exact form given here are due to Clauser, Horne, Shimony, and Holt[6]:

$$\begin{aligned}
-2 &\leqslant \mathbf{AB} + \mathbf{AB'} + \mathbf{A'B} - \mathbf{A'B'} \leqslant 2 \\
-2 &\leqslant \mathbf{AB} + \mathbf{AB'} - \mathbf{A'B} + \mathbf{A'B'} \leqslant 2 \\
-2 &\leqslant \mathbf{AB} - \mathbf{AB'} + \mathbf{A'B} + \mathbf{A'B'} \leqslant 2 \\
-2 &\leqslant -\mathbf{AB} + \mathbf{AB'} + \mathbf{A'B} + \mathbf{A'B'} \leqslant 2
\end{aligned} \tag{29}$$

Quantum mechanics does not satisfy these inequalities in general. To illustrate ideas, we take as a particular case the following:

$$\mathbf{AB} - \mathbf{AB'} + \mathbf{A'B} + \mathbf{A'B'} < -2$$

We choose

$$\mathbf{AB} = \mathbf{A'B'} = -\cos 30^\circ = -\frac{\sqrt{3}}{2}$$

$$\mathbf{AB'} = -\cos 60^\circ = -\frac{1}{2}$$

$$\mathbf{A'B} = -\cos 0^\circ = -1$$

So

$$-\frac{\sqrt{3}}{2}+\frac{1}{2}-1-\frac{\sqrt{3}}{2}< -2$$

since from quantum mechanics $\mathrm{Cov}(\mathbf{AB}) = -\cos$ (angle $\mathbf{AB}$).

As we can see from the example of Sec. 2, we need only consider an upper probability for values of λ.

Here

$$\lambda=\begin{pmatrix} \mathbf{A} & \mathbf{A}' & \mathbf{B} & \mathbf{B}' \\ \pm 1, & \pm 1, & \pm 1, & \pm 1 \end{pmatrix}$$

First, we must compute the probabilities for the pairs with given correlations. So

$$p(1, \cdot, \cdot, \cdot) = p(-1, \cdot, \cdot, \cdot) = \frac{1}{2}$$

since $E(\mathbf{A}) = 0$, etc.

Now

$$\mathbf{AB} = -\frac{\sqrt{3}}{2}$$

so

$$-\frac{\sqrt{3}}{2} = p(1, \cdot, 1, \cdot) + p(-1, \cdot, -1, \cdot) - p(1, \cdot, -1, \cdot) - p(-1, \cdot, 1, \cdot)$$

But by symmetry

$$p(1, \cdot, 1, \cdot) = p(-1, \cdot, -1, \cdot)$$

and

$$p(1, \cdot, -1, \cdot) = p(-1, \cdot, 1, \cdot)$$

So solving, we obtain

$$4p(1, \cdot, 1, \cdot) - 1 = \frac{\sqrt{3}}{2}$$

and

$$p(1, \cdot, 1, \cdot) = -\frac{\sqrt{3}}{8} + \frac{1}{4}$$

$$p(1, \cdot, -1, \cdot) = \frac{\sqrt{3}}{8} + \frac{1}{4}$$

Similarly for $\mathbf{A}'\mathbf{B}' = -\sqrt{3}/2$

$$p(\cdot, 1, \cdot, 1) = -\frac{\sqrt{3}}{8} + \frac{1}{4}$$

$$p(\cdot, 1, \cdot, -1) = \frac{\sqrt{3}}{8} + \frac{1}{4}$$

Next, $\mathbf{A}\mathbf{B}' = -1/2$, so

$$4p(1, \cdot, \cdot, 1) - 1 = -\frac{1}{2}$$

$$p(1, \cdot, \cdot, 1) = \frac{1}{8}$$

$$p(1, \cdot, \cdot, -1) = \frac{3}{8}$$

Since $\mathbf{A}'\mathbf{B} = -1$

$$4p(\cdot, 1, 1, \cdot) - 1 = -1$$

$$p(\cdot, 1, 1, \cdot) = 0$$

$$p(\cdot, 1, -1, \cdot) = \frac{1}{2}$$

We are now in a position to compute the triples. We restrict ourselves to $p^*_{i \cdot jk}$ to illustrate the method. So, for example, consider the triple $p^*_{1 \cdot 11} = P^*(\mathbf{A} = 1, \mathbf{B} = 1, \mathbf{B}' = 1)$. We have at once

$$\frac{1}{8} = p^*_{1 \cdot \cdot 1} \leqslant p^*_{1 \cdot 11} + p^*_{1 \cdot -11}$$

To get as close as we can to a probability, we take

$$p^*_{1 \cdot 11} = 0 \qquad \text{and} \qquad p^*_{1 \cdot -11} = \frac{1}{8}$$

and by symmetry

$$p^*_{-1\cdot 1-1} = \frac{1}{8} \qquad \text{and} \qquad p^*_{-1\cdot -1-1} = 0$$

By similar arguments, we set

$$p^*_{1\cdot 1-1} = p^*_{-1\cdot -11} = \frac{1}{4} - \frac{\sqrt{3}}{8}$$

$$p^*_{-1\cdot 11} = p^*_{1\cdot -1-1} = \frac{1}{8} + \frac{\sqrt{3}}{8}$$

The remaining upper probabilities for this example may be found by similar lines of argument.

We prove a theorem about quantum mechanical covariances that follows directly from the theorem on generalized common causes.

Theorem 5 (Existence of Hidden Variables). Let $\mathbf{AB}$, $\mathbf{AB}'$, $\mathbf{A}'\mathbf{B}$, and $\mathbf{A}'\mathbf{B}'$ be any four quantum mechanical covariances, which will in general not satisfy the Bell inequalities. Then there is an upper probability P^* consistent with the given covariances and a generalized hidden variable $\boldsymbol{\lambda}$ with P^* such that, for every value λ of $\boldsymbol{\lambda}$,

$$E(\mathbf{AB} \mid \boldsymbol{\lambda} = \lambda) = E(\mathbf{A} \mid \boldsymbol{\lambda} = \lambda)\, E(\mathbf{B} \mid \boldsymbol{\lambda} = \lambda)$$

and similarly for $\mathbf{AB}'$, $\mathbf{A}'\mathbf{B}$, and $\mathbf{A}'\mathbf{B}'$.

Proof. Take the unspecified correlations $\mathbf{AA}'$ and $\mathbf{BB}'$ to have any values between -1 and 1. Then apply Theorem 1, with $n = 4$.

Corresponding to Theorem 2 showing that monotonicity of P^* implies existence of a probability distribution, we may derive the Bell inequalities just from existence of a monotonic upper probability. We believe this derivation is new in the literature. The similarity to Theorem 2 is evident, since satisfaction of the Bell inequalities (29) implies existence of a joint probability distribution of $\mathbf{A}$, $\mathbf{A}'$, $\mathbf{B}$, and $\mathbf{B}'$.

Theorem 6 (Monotonicity Implies Bell Inequalities). Let $\mathbf{A}$, $\mathbf{A}'$, $\mathbf{B}$, and $\mathbf{B}'$ be two-valued (± 1) random variables with expectation $E(\mathbf{A}) = E(\mathbf{A}') = E(\mathbf{B}) = E(\mathbf{B}') = 0$ such that there is a monotonic upper probability function compatible with the given correlations $\mathbf{AB}$, $\mathbf{AB}'$, $\mathbf{A}'\mathbf{B}$, and $\mathbf{A}'\mathbf{B}'$. Then the given covariances satisfy the Bell inequalities.

Proof. Using the relations established for the proof of Theorem 2 when $E(\mathbf{X}_i) = 0$, we have first by subadditivity

$$1 - \mathbf{AB} \leqslant 4p(11 - 1\cdot) + 4(1 - 1 - 1\cdot) \tag{30}$$

Applying subadditivity again, we get

$$4p(11 - 1\cdot) \leqslant 4p(11 - 11) + 4p(11 - 1 - 1) \tag{31}$$

By monotonicity, we have at once

$$\begin{aligned} 4p(1 - 1 - 1\cdot) &\leqslant 4p(\cdot - 1 - 1\cdot) = 1 + \mathbf{A'B} \\ 4p(11 - 11) &\leqslant 4p(1 \cdot\cdot 1) = 1 + \mathbf{AB'} \\ 4p(11 - 1 - 1) &\leqslant 4p(\cdot 1 \cdot - 1) = 1 - \mathbf{A'B'} \end{aligned} \tag{32}$$

Combining (30), (31) and (32), we obtain

$$1 - \mathbf{AB} \leqslant 1 + \mathbf{AB'} + 1 + \mathbf{A'B} + 1 - \mathbf{A'B'}$$

which is equivalent to

$$-2 \leqslant \mathbf{AB} + \mathbf{AB'} + \mathbf{A'B} - \mathbf{A'B'} \tag{33}$$

Secondly, by subadditivity

$$4p(\cdot 1 \cdot - 1) = 1 - \mathbf{A'B'} \leqslant 4p^*(11 \cdot - 1) + 4p^*(-11 \cdot -1) \tag{34}$$

And again by subadditivity

$$4p^*(-11 \cdot -1) \leqslant 4p^*(-111 - 1) + 4p^*(-11 - 1 - 1) \tag{35}$$

But by monotonicity

$$\begin{aligned} 4p^*(11 \cdot - 1) &\leqslant 4p(1 \cdot\cdot - 1) = 1 - \mathbf{AB'} \\ 4p^*(-111 - 1) &\leqslant 4p(-1 \cdot 1 \cdot) = 1 - \mathbf{AB} \\ 4p^*(-11 - 1 - 1) &\leqslant 4p(\cdot 1 - 1\cdot) = 1 - \mathbf{A'B} \end{aligned} \tag{36}$$

Combining (34), (35) and (36), we have

$$1 - \mathbf{A'B'} \leqslant 1 - \mathbf{AB} + 1 - \mathbf{AB'} + 1 - \mathbf{A'B}$$

which is equivalent to

$$\mathbf{AB} + \mathbf{AB'} + \mathbf{A'B} - \mathbf{A'B'} \leqslant 2 \tag{37}$$

Inequalities (33) and (37) are the first line of the Bell inequalities (29) given earlier. By completely similar arguments we may derive the other three.

The proof of Theorem 6 also has an immediate consequence a negative theorem about monotonicity, similar to Theorem 3 for three random variables.

Theorem 7 (Nonmonotonicity). Let **AB**, **AB**′, **A**′**B**, and **A**′**B**′ be quantum mechanical covariances that do not satisfy Bell's inequalities. Then any upper measure P^* compatible with the given correlations cannot be monotonic.

4. DATA TABLES AND NONMONOTONIC UPPER PROBABILITIES

Not only in quantum mechanics but in standard statistical analysis the impossibility of the data not having a joint probability distribution implies at once a severe restriction on the kind of experiments that can be performed.

Consider first the classical general case discussed in Sec. 1 of three two-valued random variables whose pairwise correlations algebraically sum to less than -1. We can infer immediately that any experiments yielding such results could not have been ones where $\mathbf{X}_1$, $\mathbf{X}_2$, and $\mathbf{X}_3$ were simultaneously observed. If they had, the joint relative frequencies arising from the data table of the experiment would necessarily be consistent with a joint probability distribution. Put more strongly, the relative frequencies of the 8 types of tuples $(\pm 1, \pm 1, \pm 1)$ that could be observed in such an experiment constitute themselves a probability distribution.

The same remarks apply to the quantum mechanical experiments of the Bell type. Here the problem is even more troublesome, because the quantum mechanical covariances that violate the Bell inequalities and that are consistent only with a nonmonotonic upper probability are confirmed both experimentally and theoretically. So it becomes a quantum mechanical theoretical constraint that no experiment with individual particles could be performed to observe simultaneously the values of **A**, **A**′, **B**, and **B**′.

Because such simultaneous observation is impossible, we are not able to use ordinary probabilistic analysis to understand more thoroughly the phenomena being studied. For example, one of the most fundamental steps of analysis in probability is that of conditionalizing. In ordinary terms, having measured a correlation **AB**, and having also observed the correlation **AB**′, we could naturally go on to study such conditional expectations as

$$E^*(\mathbf{AB} \mid \mathbf{B}' = 1)$$

which we express here in terms of the upper expectation. But, of course, this step cannot be taken even though we have available in theory this kind of conditional expectation as defined in Eq. (7). The step cannot be taken, for no simultaneous individual observation data of **A**, **B**, and **B**′ can be collected, even in principle, consistent with quantum mechanics—we mean, of course, triples of observations of the form $(\pm 1, \pm 1, \pm 1)$.

The hidden variables we have constructed with nonmonotonic upper probabilities are indeed hidden. Whether or not anything further can be done with them in the context of quantum phenomena remains to be seen. Skepticism is warranted, for any direct application would go beyond standard quantum mechanics.

REFERENCES

1. P. Walley, *Statistical Reasoning with Imprecise Probabilities* (Chapman, 1991).
2. P. Suppes and M. Zanotti, *Synthese* **36**, 427 (1977).
3. P. Suppes and M. Zanotti, *Erkenntnis* **31**, 323 (1989).
4. P. Suppes and M. Zanotti, in *Studies in the Foundations of Quantum Mechanics*, P. Suppes, ed. (1980), p. 173.
5. P. Suppes and M. Zanotti, *Synthese* **48**, 191 (1981).
6. J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, *Phys. Rev. Lett.* **23**, 880 (1969).