PATRICK SUPPES AND MARIO ZANOTTI

ON THE DETERMINISM OF HIDDEN VARIABLE THEORIES WITH STRICT CORRELATION AND CONDITIONAL STATISTICAL INDEPENDENCE OF OBSERVABLES

I. INTRODUCTION

The main purpose of this note is to prove a lemma about random variables, and then to apply this lemma to the characterization of local theories of hidden variables by Bell (1964, 1966) and Wigner (1970), which are focused around Bell's inequality. We use the results of the lemma in two different ways. The first is to show that the assumptions of Bell and Wigner can be weakened to conditional statistical independence rather than conditional determinism because determinism follows from conditional independence and the other assumptions that are made about systems of two spin-$\frac{1}{2}$ particles.

The second direction is to question the attempt of Clauser and Horne (1974) to derive a Bell-type inequality for local stochastic theories of hidden variables which use an assumption of conditional statistical independence for observables.

The main thrust of our analysis obviously arises from the probabilistic lemma we prove. Roughly speaking, this lemma asserts that if two random variables have strict correlation, that is, the absolute value of the correlation is one, and it is in addition assumed that their expectations are conditionally independent when a third random variable λ is given, then the conditional variance of X and Y given λ is zero. In other words, given the hidden variable λ the observables X and Y are strictly determined. The lemma itself, of course, depends on no assumptions about quantum mechanics. It may be regarded as a limitation on any theories that assume both strict correlation between observables and their conditional independence on the basis of some prior or hidden variable.

II. PROBABILISTIC LEMMA ABOUT DETERMINISM

In the statement of the lemma we use standard notation for the expectation (E), covariance (Cov), variance (Var), and standard deviation (σ)

P. Suppes (ed.), Logic and Probability in Quantum Mechanics, 445–455.

of random variables. We use both Var for variance and σ for the standard deviation for compactness of notation.

In most physical discussions of these matters it is assumed that the random variables in question have continuous densities but for the general proof we give here no such assumption is necessary. Finiteness of expectations as indicated in the statement of the lemma is all that is required. The second clause of the lemma just expresses the fact that the correlation is strict, that is, in other notation the absolute value of the correlation $\rho(X, Y)=1$. The conclusion that the conditional variances of the random variables X and Y are zero is a conclusion that holds with probability one, which is the strongest result of a deterministic kind we would expect in a probabilistic setting.

LEMMA. *Let X, Y, and λ be three random variables such that*

(i) $\mathrm{E}(XY \mid \lambda)=\mathrm{E}(X \mid \lambda)\,\mathrm{E}(Y \mid \lambda)$;
(ii) $|\mathrm{Cov}(X, Y)|=\sigma(X)\,\sigma(Y)$,
(iii) $\sigma(X)>0$ *and* $\sigma(Y)>0$,
(iv) *the expectations in* (i) *and* (ii) *are finite;*

then with probability one

$$\sigma(X \mid \lambda)=\sigma(Y \mid \lambda)=0.$$

Proof. Note first that by (ii)

$$Y=a+bX\,\mathrm{sign}(\mathrm{Cov}(X, Y)) \tag{1}$$

with probability one where a, b are real numbers with $b>0$, and therefore

$$\sigma(Y)=b\sigma(X), \tag{2}$$

$$\mathrm{E}(Y \mid \lambda)=a+b\mathrm{E}(X \mid \lambda)\,\mathrm{sign}(\mathrm{Cov}(X, Y)). \tag{3}$$

Thus, (ii) may be expressed as

$$\mathrm{Cov}(X, Y)=b\,\mathrm{Var}(X)\,\mathrm{sign}(\mathrm{Cov}(X, Y)). \tag{4}$$

We next note the (relatively well-known) fact that

$$\mathrm{Var}(X)=\mathrm{E}(\mathrm{Var}(X \mid \lambda))+\mathrm{Var}(\mathrm{E}(X \mid \lambda)). \tag{5}$$

We prove (5) by observing that

$$\mathrm{E}(\mathrm{Var}(X \mid \lambda))=\mathrm{E}(X^2 \mid \lambda)-\mathrm{E}(\mathrm{E}^2(X \mid \lambda)), \tag{6}$$

and

$$\text{(7)} \qquad \mathrm{Var}(\mathrm{E}(X \mid \lambda)) = \mathrm{E}(\mathrm{E}^2(X \mid \lambda)) - (\mathrm{E}(\mathrm{E}(X \mid \lambda)))^2.$$

So adding (6) and (7) we get

$$\mathrm{E}(\mathrm{E}(X^2 \mid \lambda)) - (\mathrm{E}(\mathrm{E}(X \mid \lambda)))^2 = \mathrm{E}(X^2) - \mathrm{E}^2(X) = \mathrm{Var}(X).$$

We next observe that

$$\text{(8)} \qquad \mathrm{Cov}(X, Y) = \mathrm{E}(\mathrm{E}(XY \mid \lambda)) - \mathrm{E}(\mathrm{E}(X \mid \lambda))\, \mathrm{E}(\mathrm{E}(Y \mid \lambda)).$$

Combining (8), (i), and (3) we have

$$\text{(9)} \qquad \begin{aligned} \mathrm{Cov}(X, Y) &= \mathrm{E}(\mathrm{E}(X \mid \lambda)\, \mathrm{E}(Y \mid \lambda)) - \mathrm{E}(\mathrm{E}(X \mid \lambda))\, \mathrm{E}(\mathrm{E}(Y \mid \lambda)) \\ &= \mathrm{E}(\mathrm{E}(X \mid \lambda)\, (a + b\mathrm{E}(X \mid \lambda)\, \mathrm{sign}(\mathrm{Cov}(X, Y)))) \\ &\quad - \mathrm{E}(\mathrm{E}(X \mid \lambda))\, \mathrm{E}(a + b\mathrm{E}(X \mid \lambda)\, \mathrm{sign}(\mathrm{Cov}(X, Y))) \\ &= b(\mathrm{E}(\mathrm{E}^2(X \mid \lambda) - \mathrm{E}^2(\mathrm{E}(X \mid \lambda)))\, \mathrm{sign}(\mathrm{Cov}(X, Y)) \\ &= b\, \mathrm{Var}(\mathrm{E}(X \mid \lambda))\, \mathrm{sign}(\mathrm{Cov}(X, Y)). \end{aligned}$$

So by (4) and (5)

$$\mathrm{E}(\mathrm{Var}(X \mid \lambda)) = 0,$$

and since $\mathrm{Var}(X \mid \lambda) \geqslant 0$, with probability one $\mathrm{Var}(X \mid \lambda) = 0$. By obvious symmetry of argument, $\mathrm{Var}(Y \mid \lambda) = 0$. Q.E.D.

III. AXIOMS FOR SYSTEMS OF TWO SPIN-$\frac{1}{2}$ PARTICLES

Consider a system of two spin-$\frac{1}{2}$ particles initially in the singlet state. Measurements are made of the components of spin for each particle; in particular, let apparatus I measure one particle and apparatus II the other.

There are a number of natural physical assumptions made by Wigner (and Bell at least implicitly), e.g., axial symmetry. These will come out in the axioms. It is to be emphasized that the axioms given here are for this special situation of pairs of spin-$\frac{1}{2}$ particles formed at the source in the singlet spin state with one particle moving to the measuring apparatus I, in one direction, and the other particle to measuring apparatus II in the opposite direction.

The point of the axiomatization is to permit an explicit analysis of just what assumptions are involved in deriving a contradiction between local deterministic theories of hidden variables and the standard quantum mechanical theoretical results for this situation.

Bell states his assumptions basically in terms of expectations; Wigner uses probabilities. Since the random variables whose conditional expectations are the focus of the axioms are two-valued, there is no essential difference between the two approaches. On the other hand, by far the common practice in quantum mechanics is to consider expectations rather than probabilities, and this is the course we have chosen.

Using explicit random-variable notation, the axioms are stated in the spirit of modern probability theory; no additional physical assumptions are left implicit, to be used as needed.

The random variables are these:

ω_{I} – the direction of orientation of measuring apparatus I;
ω_{II} – the direction of orientation of measuring apparatus II;
M_{I} – the spin measurement of apparatus I;
M_{II} – the spin measurement of apparatus II;
λ – the hidden variable.

The values of random variables ω_{I} and ω_{II} are direction vectors, i.e., three-dimensional vectors normed to one, and the cosine of the angle between them is the scalar product $\omega_{\mathrm{I}} \cdot \omega_{\mathrm{II}}$, which is itself a new random variable. The values of random variables M_{I} and M_{II} are $+1$ and -1, for spin $+\frac{1}{2}$ and spin $-\frac{1}{2}$, respectively. Finally, we shall assume for simplicity of notation that λ is a real-valued random variable, but it could be vector-valued and not affect any of the theory, for the essential assumptions about the hidden variable λ are minimal.

As already indicated, an axiom of determinism for the results of a spin measurement given the value of the hidden random variable λ is not required, but can be derived from the weaker axiom of statistical independence.

DEFINITION 1. *A structure* $\langle \omega_{\mathrm{I}}, \omega_{\mathrm{II}}, M_{\mathrm{I}}, M_{\mathrm{II}}, \lambda \rangle$ *of random variables is a* local hidden variable spin-$\frac{1}{2}$ system *if and only if the following axioms are satisfied:*

AXIOM 1. [*Exchangeability*] *For any bounded functions* f *and* g *the expectations* $\mathrm{E}(f(M_{\mathrm{I}})\, g(M_{\mathrm{II}}) \mid \omega_{\mathrm{I}}, \omega_{\mathrm{II}})$ *and* $\mathrm{E}(g(M_{\mathrm{I}})\, f(M_{\mathrm{II}}) \mid \omega_{\mathrm{I}}, \omega_{\mathrm{II}})$ *are* *finite and*

$$\mathrm{E}(f(M_{\mathrm{I}})\, g(M_{\mathrm{II}}) \mid \omega_{\mathrm{I}}, \omega_{\mathrm{II}}) = \mathrm{E}(g(M_{\mathrm{I}})\, f(M_{\mathrm{II}}) \mid \omega_{\mathrm{I}}, \omega_{\mathrm{II}});$$

AXIOM 2. [*Axial Symmetry*]
(i) $\mathrm{E}(M_{\mathrm{I}} \mid \omega_{\mathrm{I}}) = \mathrm{E}(M_{\mathrm{II}} \mid \omega_{\mathrm{II}}) = 0$;
(ii) *If we define for any direction vectors* w_{I} *and* w_{II} *and any real number* α

$$H(w_{\mathrm{I}}, w_{\mathrm{II}}, \alpha) = [\omega_{\mathrm{I}} = w_{\mathrm{I}}, \omega_{\mathrm{II}} = w_{\mathrm{II}}, w_{\mathrm{I}} \cdot w_{\mathrm{II}} = \alpha]$$

then

$$\mathrm{E}(M_{\mathrm{I}} M_{\mathrm{II}} \mid H(w_{\mathrm{I}}, w_{\mathrm{II}}, \alpha) = \mathrm{E}(M_{\mathrm{I}} M_{\mathrm{II}} \mid H(w'_{\mathrm{I}}, w'_{\mathrm{II}}, \alpha')) \textit{ if } \alpha = \alpha';$$

AXIOM 3. [*Opposite Measurement for Same Orientation*] *If* $\alpha = 1$ *then*

$$\mathrm{E}(M_{\mathrm{I}} M_{\mathrm{II}} \mid H(w_{\mathrm{I}}, w_{\mathrm{II}}, \alpha)) = -1;$$

AXIOM 4. [*Independence of* λ] *For all functions g for which the expectations* $\mathrm{E}(g(\lambda))$ *and* $\mathrm{E}(g(\lambda) \mid \omega_{\mathrm{I}}, \omega_{\mathrm{II}})$ *are finite,*

$$\mathrm{E}(g(\lambda)) = \mathrm{E}(g(\lambda) \mid \omega_{\mathrm{I}}, \omega_{\mathrm{II}});$$

AXIOM 5. [*Locality: Independence of Orientation of Other Measuring Apparatus*]

$$\mathrm{E}(M_{\mathrm{I}} \mid \omega_{\mathrm{I}}, \omega_{\mathrm{II}}, \lambda) = \mathrm{E}(M_{\mathrm{I}} \mid \omega_{\mathrm{I}}, \lambda),$$
$$\mathrm{E}(M_{\mathrm{II}} \mid \omega_{\mathrm{I}}, \omega_{\mathrm{II}}, \lambda) = \mathrm{E}(M_{\mathrm{II}} \mid \omega_{\mathrm{II}}, \lambda);$$

AXIOM 6. [*Statistical Independence*]

$$\mathrm{E}(M_{\mathrm{I}} M_{\mathrm{II}} \mid \omega_{\mathrm{I}}, \omega_{\mathrm{II}}, \lambda) = \mathrm{E}(M_{\mathrm{I}} \mid \omega_{\mathrm{I}}, \omega_{\mathrm{II}}, \lambda)\, \mathrm{E}(M_{\mathrm{II}} \mid \omega_{\mathrm{I}}, \omega_{\mathrm{II}}, \lambda).$$

The axioms characterize the probability of a possible outcome of measurements at I and II for two spin-$\frac{1}{2}$ particles σ_1 and σ_2 originally in the singlet state at the source. The first five axioms are implicit in the articles of Bell and Wigner. As already indicated in a general way, Axiom 6 on statistical independence, together with Axioms 1–5, implies deterministic results when the value of λ is given. Axioms 1 and 2 express a broad assumption about symmetry in the experimental measurement procedure.

In the case of Axiom 2, axial symmetry means that specific spatial orientation does not matter, only the angle $\cos^{-1}\alpha$ between the orientation w_{I} of apparatus I and the orientation w_{II} of apparatus II. Axiom 3 makes explicit a highly specific assumption (or fact perhaps) about spin-$\frac{1}{2}$ particles emitted from a singlet state. If both apparatus I and apparatus II have the same spatial orientation ($w_{\mathrm{I}} \cdot w_{\mathrm{II}} = 1$) then the measurements M_{I} and M_{II} must be opposite, i.e., the correlation $\rho(M_{\mathrm{I}}, M_{\mathrm{II}}) = -1$. Axioms 1–3 say nothing about hidden variables. Axioms 4–6 do, and the essence of what they postulate is not specific to systems of spin-$\frac{1}{2}$ particles, but rather to local hidden variable theories.

We now prove several simple theorems about the systems defined by Axioms 1–6.

THEOREM 1. [*Locality Without λ*]

$$\mathrm{E}(M_{\mathrm{I}} \mid \omega_{\mathrm{I}}, \omega_{\mathrm{II}}) = \mathrm{E}(M_{\mathrm{I}} \mid \omega_{\mathrm{I}}),$$
$$\mathrm{E}(M_{\mathrm{II}} \mid \omega_{\mathrm{I}}, \omega_{\mathrm{II}}) = \mathrm{E}(M_{\mathrm{II}} \mid \omega_{\mathrm{II}}).$$

Proof.

$$\begin{aligned} \mathrm{E}(M_{\mathrm{I}} \mid \omega_{\mathrm{I}}, \omega_{\mathrm{II}}) &= \mathrm{E}_{\lambda}(\mathrm{E}(M_{\mathrm{I}} \mid \omega_{\mathrm{I}}, \omega_{\mathrm{II}}, \lambda)) \\ &= \mathrm{E}_{\lambda}(\mathrm{E}(M_{\mathrm{I}} \mid \omega_{\mathrm{I}}, \lambda)) && \text{by Locality (Axiom 5)} \\ &= \mathrm{E}(M_{\mathrm{I}} \mid \omega_{\mathrm{I}}) && \text{by Independence of } \lambda \text{ (Axiom 4).} \end{aligned}$$

The argument is of course the same for M_{II}. Q.E.D.

THEOREM 2. [*Determinism*] $\sigma(M_{\mathrm{I}} \mid \omega_{\mathrm{I}}, \lambda) = \sigma(M_{\mathrm{II}} \mid \omega_{\mathrm{II}}, \lambda) = 0$.

Proof. Immediate from Axioms 3 and 6, and the Probabilistic Lemma of Section II.

THEOREM 3. [*Symmetry*]

(i) $\mathrm{Var}(M_{\mathrm{I}} \mid \omega_{\mathrm{I}}) = \mathrm{Var}(M_{\mathrm{II}} \mid \omega_{\mathrm{II}}) = 1$,

(ii) $\mathrm{Cov}(M_{\mathrm{I}}, M_{\mathrm{II}} \mid \omega_{\mathrm{I}}, \omega_{\mathrm{II}}) = \mathrm{E}(M_{\mathrm{I}} M_{\mathrm{II}} \mid \omega_{\mathrm{I}}, \omega_{\mathrm{II}})$.

Proof. Follows directly from Axiom 2 on axial symmetry.

THEOREM 4. *For given direction vectors w_{I} and w_{II}, the joint conditional distribution of random variables M_{I} and M_{II} is determined by*

$$\mathrm{Cov}(M_{\mathrm{I}}, M_{\mathrm{II}} \mid \omega_{\mathrm{I}} = w_{\mathrm{I}}, \omega_{\mathrm{II}} = w_{\mathrm{II}}).$$

Proof. Follows from Axiom 2(i), Theorem 1, and Theorem 3.

The intuitive content of Theorem 4 is that the axioms of Definition 1 determine the joint conditional distribution of M_{I} and M_{II} up to a single parameter, which can be taken to be the conditional covariance.

THEOREM 5. *Let w_1, w_2, and w_3 be three direction vectors, and let Γ be the set of hidden variables, i.e., Γ = set of values of the random variable λ. Each λ in Γ belongs to exactly one of the eight regions $\Gamma(\pm, \pm, \pm)$, where*

$$\Gamma(\pm, \pm, \pm) = \{\lambda : \mathrm{E}(M_{\mathrm{I}} \mid \omega_{\mathrm{I}} = w_1, \lambda) = \pm 1 \ \&\ \mathrm{E}(M_{\mathrm{I}} \mid \omega_{\mathrm{I}} = w_2, \lambda) = \pm 1 \ \&\ \mathrm{E}(M_{\mathrm{I}} \mid \omega_{\mathrm{I}} = w_3, \lambda) = \pm 1\}.$$

Proof. Follows from Theorem 2 on determinism.

The statement of Theorem 5, but not the notation, makes it clear that the eight regions $\Gamma(\pm, \pm, \pm)$ are relative to the choice of the three direction vectors w_1, w_2, and w_3.

We now derive the basic inequality in terms of covariances.

THEOREM 6. [*Basic Inequality*] *Let w_1, w_2, and w_3 be three direction vectors and let for $i \neq j$, $1 \leqslant i, j \leqslant 3$,*

$$\mathrm{Cov}(w_i, w_j) = \mathrm{Cov}(M_{\mathrm{I}}, M_{\mathrm{II}} \mid \omega_{\mathrm{I}} = w_i, \omega_{\mathrm{II}} = w_j).$$

Then

$$\mathrm{Cov}(w_1, w_2) + \mathrm{Cov}(w_2, w_3) \geqslant \mathrm{Cov}(w_3, w_1) - 1.$$

Proof. Let $\alpha(+++)$ be the probability that λ lies in the region $\Gamma(+++)$, with similar notation for the other seven regions.

We note first that

$$\mathrm{Cov}(w_i, w_j) = \int_{\Gamma} \mathrm{E}(M_{\mathrm{I}} M_{\mathrm{II}} \mid \omega_{\mathrm{I}} = w_i, \omega_{\mathrm{II}} = w_j, \lambda)\, \mathrm{d}P(\lambda).$$

Because $\int P(M_{\mathrm{I}} = 1, M_{\mathrm{II}} = 1 \mid \omega_{\mathrm{I}} = w_1, \omega_{\mathrm{II}} = w_2, \lambda)\, \mathrm{d}P(\lambda) = \alpha(+-+) + \alpha(+--)$, and similarly for other terms, it is easy to show that

$$\begin{aligned}\mathrm{Cov}(w_1, w_2) = {} & \alpha(+-+) + \alpha(+--) + \alpha(-+-) + \alpha(-++) \\ & - [\alpha(+++) + \alpha(++-) + \alpha(--+) \\ & + \alpha(---)],\end{aligned}$$

$$\begin{aligned}
\mathrm{Cov}(w_2, w_3) = {} & \alpha(++-)+\alpha(-+-)+\alpha(+-+)+\alpha(--+) \\
& -[\alpha(+++)+\alpha(-++)+\alpha(+--) \\
& +\alpha(---)], \\
\mathrm{Cov}(w_3, w_1) = {} & \alpha(-++)+\alpha(--+)+\alpha(++-)+\alpha(+--) \\
& -[\alpha(+++)+\alpha(+-+)+\alpha(-+-) \\
& +\alpha(---)].
\end{aligned}$$

Then by direction substitution, and using the fact that the sum of the probability of the eight regions $\Gamma(\pm, \pm, \pm)$ is 1,

$$\begin{aligned}
& \mathrm{Cov}(w_1, w_2)+\mathrm{Cov}(w_2, w_3) \\
& \qquad = \mathrm{Cov}(w_3, w_1)-1+4\alpha(+-+)+4\alpha(-+-),
\end{aligned}$$

and since the probabilities $\alpha(+-+)$ and $\alpha(-+-)$ are nonnegative, the basic inequality follows at once.

The basic inequality of Theorem 6 is slightly different from either one given by Bell or Wigner, because of our explicit use of the standard probabilistic concept of covariance, but the result is really the same, as the next theorem shows.

From the quantum mechanical results given by Bell or Wigner it follows directly that the quantum mechanical covariances are given by the equation:

$$\mathrm{Cov}(w_i, w_j)=\sin^2\tfrac{1}{2}\theta_{ij}-\cos^2\tfrac{1}{2}\theta_{ij},$$

where θ_{ij} is the angle between direction vectors w_i and w_j We are now in a position to obtain the Bell-Wigner contradiction.

THEOREM 7. [*Bell-Wigner Contradiction*] *The quantum mechanical covariances contradict the basic inequality of Theorem 7 for some directions of measurement* w_1, w_2, *and* w_3.

Proof. To put the proof exactly in Wigner's form, we note that it follows immediately from the basic inequality of Theorem 7 that we must have

$$(1) \qquad \sin^2\tfrac{1}{2}\theta_{12}+\sin^2\tfrac{1}{2}\theta_{23}\geqslant\sin^2\tfrac{1}{2}\theta_{31}.$$

Clearly (1) is violated if w_2 bisects w_1 and w_3, e.g., if $\theta_{12}=\theta_{23}=30°$ and $\theta_{31}=60°$.

Violation of (1) is discussed in some detail by Wigner and need not be repeated here. He also points out that (1) implies Bell's original inequality.

IV. CAUSALITY AND INDEPENDENCE

In generalizing from deterministic theories of hidden variables it may seem natural to impose a condition of statistical conditional independence, as expressed in the probabilistic lemma of Section II. A good recent example of the use of such an assumption is to be found in Clauser and Horne (1974). It is our feeling that such an assumption of conditional statistical independence is too strong for a stochastic theory that is not meant to be deterministic. Clauser and Horne do not assume correlations of one between observables – in their case, counts of particles at detectors – and thus our lemma does not apply. But the kind of behavior that must be expected qualitatively can be inferred from assuming that the observables X and Y, as well as the hidden variable λ, are normally distributed. Under the assumption of a multivariate normal distribution for the three random variables, and the assumption that X and Y are conditionally independent, given λ, we can then show that the following relation holds between the correlations whose absolute values are assumed to be strictly between 0 and 1:

$$\rho(X, Y)=\rho(X, \lambda)\,\rho(Y, \lambda),$$

because the conditional correlation $\rho(X, Y|\lambda)$ is given by the same expression as the partial correlation $\rho_{XY\cdot\lambda}$ (for normally distributed random variables)

$$\rho_{XY\cdot\lambda}=\rho(X, Y|\lambda)=\frac{\rho(X, Y)-\rho(X, \lambda)\,\rho(X, \lambda)}{\sqrt{1-\rho(X, \lambda)^2}\cdot\sqrt{1-\rho(X, \lambda)^2}}.$$

Assuming symmetry of X and Y, we then get the more restricted expression:

$$\rho(X, Y)=\rho(X, \lambda)^2,$$

which shows that the correlation between X and Y is always strictly less than the correlation between X and λ or between Y and λ. What this relation shows is that if we impose conditional statistical independence then if we have a quite high correlation between observables we must have an even higher correlation between the hidden variable and the observables. Thus, in a clear sense, we must be closer to a deterministic hidden variable theory than to a deterministic theory of the observables.

This admittedly qualitative argument makes us suspicious of the use of an assumption of conditional independence between observables in formulating a hidden variable theory that is meant to be properly stochastic.

Contrary to a standard line of talk about the Einstein-Podolsky-Rosen paradox, in our judgment the absence of conditional statistical independence in a proper stochastic theory of observables does not imply a violation of causality conditions, that is, does not imply instantaneous action at a distance. A simple classical stochastic model of coin flipping will illustrate the point. Let us assume a hidden variable λ with unknown probability distribution, but let us assume as the distribution of the observable X whose values are heads or tails at detector 1 to be that the probability of heads at detector 1 is λ, the probability therefore of tails is $1-\lambda$, the probability of heads at detector 2 is $1-\lambda$, and the probability of tails is λ. Let us also assume that heads has value 1 and tails value -1, and also then a strict correlation of -1 between X and Y. This means that when heads is observed at detector 1 with probability 1, tails is observed at detector 2, and vice versa.

On the other hand, we do not have conditional statistical independence for it is obvious that it does not hold in the model. In fact, not only is the correlation -1 between X and Y, the conditional correlation, given the value of λ, between X and Y is -1. In these circumstances, the deterministic results of our probability lemma do not apply and we retain a genuine stochastic process, in particular, $\sigma(X \mid \lambda), \sigma(Y \mid \lambda) > 0$.

At the same time, it is clearly also intuitively wrong to say that because of the conditional correlation of -1 between X and Y, given λ, that there is an instantaneous action at a distance between detector 1 and detector 2. The value of λ at the source determines the probabilistic choice of heads or tails for the two detectors and if heads is sent to one detector then tails is sent to the other, just as we might think of a spin-$\frac{1}{2}$ system of particles moving out from the source. What is to arrive at the two detectors is fixed at the source, but fixed in a stochastic fashion. There is no question whatsoever of an instantaneous causal influence between the two sources.

The central point is that when the hidden parameter λ has only a stochastic relationship to the observables, then the absence of conditional statistical independence with respect to λ in no way implies instantaneous action at a distance between the two locations of the detectors or other measuring devices.

Stanford University

NOTE

[1] We are indebted to Robert Latzer for pointing out a number of errors in a far different, original draft of this section.

BIBLIOGRAPHY

Bell, J.S., 'On the Einstein Podolsky Rosen Paradox', *Physics* **1** (1964), 195–200.

Bell, J. S., 'On the Problem of Hidden Variables in Quantum Mechanics', *Reviews of Modern Physics* **38** (1966), 447–452.

Clauser, J. F. and Horne, M. A., 'Experimental Consequences of Objective Local Theories', *Physical Review D* **10** (1974), 526–535.

Wigner, E. P., 'On Hidden Variables and Quantum Mechanical Probabilities', *American Journal of Physics* **38** (1970), 1005–1009.