International Journal of Theoretical Physics, Vol. 42, No. 10, October 2003 (© 2003)

Quantum Logic, Probability, and Information: The Relation With the Bell Inequalities

Emilio Santos[1]

Received June 30, 2003

We study the relation between the Bell inequalities—characteristic of noncontextual hidden variables theories of quantum mechanics—with quantum logic, quantum probability, and quantum information. The emphasis is on clarity and simplicity, although sometimes this implies a lack of mathematical rigor which, I hope, could be resolved without difficulty by the reader.

KEY WORDS: quantum logic; probability; Bell inequalities.

1. HIDDEN VARIABLES THEORIES

The question of hidden variables in quantum mechanics aroused soon after the formulation of the theory during the years 1925–26. It was explicitly mentioned in the book by von Neumann in 1932 (von Neumann, 1995), where he derived a celebrated no hidden variables theorem. From that time many books and articles have been devoted to the subject. Nevertheless, there is no sharp definition of hidden variables (HV) theory which is widely accepted. I propose the following:

Definition 1.1. HV is a theory physically equivalent to quantum mechanics (that is giving the same predictions for all experiments) which has the formal structure of classical statistical mechanics.

The definition may be illustrated in the following table (Table I) giving the correspondence of concepts in experiments, standard quantum theory, and a possible HV theory:

The parameter (or parameters) λ is usually called the *hidden variable*. Two observables, A and B, which are associated to commuting operators, $\hat{A}$ and $\hat{B}$, are said *compatible*. The correlation may be extended to more than two compatible observables. It is easy to see that the latter equality implies the equality of the joint probability distributions of compatible observables. In fact, it is enough to

[1] Departamento de Física, Universidad de Cantabria, Santander, Spain; e-mail: santose@unicam.es.

0020-7748/03/1000-2545/0 © 2003 Plenum Publishing Corporation

Table I. Correspondence of Concepts

Empirical	Quantum theory	HV theory
Physical system state	Hilbert space H vector $\|\Psi \in H$	Phase space Λ probability density $\rho(\lambda)$
Observable A expectation value	Self-adjoint operator $\hat{A}$ $\langle\Psi\|\hat{A}\|\Psi\rangle$	Function $A(\lambda) = \int A(\lambda)\rho(\lambda)\, d\lambda$
Correlation	If $\hat{A}\hat{B} = \hat{B}\hat{A}$, $\langle\Psi\|\hat{A}\hat{B}\|\Psi\rangle = \int A(\lambda)B(\lambda)\rho(\lambda)\, d\lambda$	

substitute $\exp(i\xi\hat{A})$ for $\hat{A}$ and $\exp[i\xi A(\lambda)]$ for $A(\lambda)$ in the equality, and similar for B, in order to show the equality of the characteristic function of the joint probability distribution. On the other hand, it is well known that quantum mechanics does not provide joint distributions of observables not compatible (the associated operators noncommuting). For the sake of clarity, in the table we have considered only quantum pure states. The most general states are associated to density operators, $\hat{\rho}$ and the quantum expectation value and correlation should be written, respectively

$$Tr(\hat{\rho}\hat{A}), \qquad Tr(\hat{\rho}\hat{A}\hat{B}).$$

To make clear what is the content of the theorems against HV theories, discussed later, I propose the following:

Definition 1.2. A simple experiment consists of the preparation of a state of a physical system, followed by the evolution of the system and finishing with the measurement of a set of compatible observables.

Definition 1.3. A composite experiment consists of several simple experiments with the same preparation and the same later evolution, but measuring different sets of compatible observables in each simple experiment.

With these definitions we may state the following theorem:

Theorem 1.4. *For any simple experiment there exists a HV theory.*

Proof: The essential part of the proof is to show that for any state $|\Psi\rangle$ and two compatible observables A, B the expectation may be obtained in the form

$$\langle\Psi|\hat{A}\hat{B}\Psi\rangle = \int A(\lambda)B(\lambda)\rho(\lambda)\, d\lambda. \tag{1.1}$$

For simplicity we consider just two observables, but the generalization to any finite number is trivial. To proceed with the proof we recall that there exists a complete set of orthonormal vectors which are simultaneous eigenvectors of two commuting self-adjoint operators. Let us label $|\lambda\rangle$ one of the common eigenvectors of $\hat{A}$ and

$\hat{B}$. Complete means that

$$\int |\lambda\rangle\langle\lambda \mid \, d\lambda = 1, \tag{1.2}$$

which leads to

$$\begin{aligned}\langle\Psi|\hat{A}\hat{B}|\Psi\rangle &= \int d\lambda\, d\lambda' d\lambda'' \langle\Psi \mid \lambda\rangle\langle\lambda \mid \hat{A} \mid \lambda'\rangle\langle\lambda' \mid \hat{B} \mid \lambda''\rangle\langle\lambda'' \mid \Psi\rangle \\ &= \int d\lambda \langle\Psi \mid \lambda\rangle\langle\lambda|\hat{A}|\lambda\rangle\langle\lambda|\hat{B}|\lambda\rangle\langle\lambda \mid \Psi\rangle \\ &= \int d\lambda \mid \langle\Psi|\lambda\rangle|^2\langle\lambda|\hat{A}|\lambda\rangle\langle\lambda|\hat{B}|\lambda\rangle. \end{aligned} \tag{1.3}$$

This has the structure of the right side of Eq. (1.1) provided we identify $\langle\lambda|\hat{A}|\lambda\rangle$ with the function $A(\lambda)$ and $|\langle\Psi \mid \lambda\rangle|^2$ with the density $\rho(\lambda)$. Indeed, the density is positive and normalized (the latter because Eq. (1.2)). The second equality of Eq. (1.3) follows from the equality

$$\langle\lambda|\hat{A}|\lambda'\rangle = \langle\lambda|\hat{A}|\lambda\rangle\delta(\lambda - \lambda'), \tag{1.4}$$

δ being Dirac's delta, which is a consequence of $|\lambda\rangle$ and $|\lambda'\rangle$ being eigenvectors of $\hat{A}$.

We see that hidden variables are always possible, a fact made clear by J. S. Bell in 1966 (Bell, 1966, 1987). However, some families of HV theories are excluded, for instance those in which expectations fulfill linear relations of the form

$$\langle\Psi \mid \hat{A} + \hat{B} \mid \Psi\rangle = \int [A(\lambda) + B(\lambda)]\rho(\lambda)\, d\lambda, \tag{1.5}$$

The impossibility of such HV theories is the content of von Neumann's theorem mentioned above (von Neumann, 1955). Assumption (1.5) is unphysical, as pointed out by Bell (1966, 1987), which shows that von Neumann's theorem is not very relevant. More physical requirements are nonocontextuality and locality which we discuss in the following section. □

2. NONCONTEXTUALITY, LOCALITY, AND THE BELL INEQUALITIES

Definition 2.1. A HV theory is noncontextual if there exists a joint probability distribution for all observables of the system (even if some of them are not compatible.)

In particular this implies that the marginal for the variable A in the joint distribution of the compatible observables A and B is the same as the marginal for A in the joint distribution of the compatible observables A and C, even if B

and C are not compatible. For this reason noncontextuality is sometimes stated saying that the result of measuring A does not depend on the context (in particular, the result is the same whether we measure A together with B or we measure A together with C; remember that A, B, C cannot be measured simultaneously, that is with the same experimental set up). This property is true in quantum mechanics, but the existence of a joint distribution is a stronger constraint. What is required is the existence of some function of all the observables, $p(A, B, C \ldots)$, which fulfills the mathematical properties of a joint probability distribution and it is such that the marginals for every subset of compatible observables is the same given by quantum mechanics. The said distribution is just a mathematical object (it cannot be measured if some of the observables are not compatible) but their mere existence puts constraints which may be tested empirically.

It is not difficult to see that the existence of a joint distribution for the observables $A, B, C, \ldots$ is equivalent to the existence of a positive normalized function, $\rho(\lambda)$ of a variable or set of variables, λ, and functions $A(\lambda), B(\lambda), C(\lambda) \ldots$ However a joint probability distribution cannot be obtained with the construction of Eq. (1.3) if the observables are not compatible. This is because a complete orthonormal set of simultaneous eigenvectors of $\hat{A}, \hat{B}, \hat{C}, \ldots$ may not exist if the operators do not commute pairwise. What may be obtained are several HV theories, one for each simple experiment. For instance, let us consider a composite experiment consisting of two simple ones. In the first, where we measure A and B, a HV theory should provide the functions $\rho_1(\lambda), A_1(\lambda), B_1(\lambda)$. In the second, where we measure A and C, a HV theory would give $\rho_2(\lambda), A_2(\lambda), C_2(\lambda)$. The two HV theories together might be called a HV theory for the composite experiment. It would be noncontextual if $\rho_1(\lambda) = \rho_2(\lambda)$ and $A_1(\lambda) = A_2(\lambda)$, if this does not happen it should be *contextual*.

The impossibility of noncontextual theories is established by the following

Theorem 2.2. *Noncontextual HV theories do not exist for all (composite) experiments.*

This is usually called Kochen–Specker theorem (Kochen and Specker, 1967; reprinted in Hooker, 1975, 1979) after the authors who proved it in 1967. However the theorem had been actually proved 1 year earlier by Bell (1966, 1987) and it is a rather direct consequence of a theorem proved by Gleason (1957, reprinted in Hooker, 1957, 1979). We shall give here a proof inspired in the celebrated theorem of Bell against local hidden variables (Bell, 1964, reprinted in Bell, 1987).

Proof: It is enough to exhibit a particular type of composite experiment where the quantum predictions are incompatible with the existence of a joint probability distribution for all observables. We consider four dichotomic observables, A, B, C, and D, each of which may take the values 0 or 1. We assume that A and C are not compatible, and B and D are also not compatible, the remaining pairs being compatible. The corresponding operators will be projectors, i.e., $\hat{A}^2 = \hat{A}$, etc., all pairs commuting except

$$[\hat{A}, \hat{C}] \neq 0, [\hat{B}, \hat{D}] \neq 0. \tag{2.1}$$

Let us label p_A the probability of $A = 1$, p_{AB} the probability that $A = B = 1$, etc. The existence of a joint distribution means that there are 15 positive quantities

$$\begin{aligned} &p_A, p_B, p_C, p_D, p_{AB}, p_{AC}, p_{AD}, p_{BC}, p_{BD}, p_{CD}, \\ &p_{ABC}, p_{ABD}, p_{ACD}, p_{BCD}, p_{ABCD}, \end{aligned} \tag{2.2}$$

which should fulfill the relations

$$0 \leq p_{ABCD} \leq p_{ABC} \leq p_{AB} \leq p_A \leq 1, \tag{2.3}$$

and those obtained by all permutations of the labels. Only eight of these quantities may be measured (and they are predicted by quantum mechanics), namely

$$p_A, p_B, p_C, p_D, p_{AB}, p_{AD}, p_{BC}, p_{CD}. \tag{2.4}$$

The remaining seven quantities cannot be measured, the corresponding observables not being compatible, and quantum mechanics gives no value for them.

The question is whether there exist seven quantities fulfilling all constraints of the type (2.3) which added to the eight measurable ones provide the desired joint probability distribution (2.2). As will be discussed in Section 4, a necessary condition is the fulfillment of the following inequality:

$$p_B + p_C \geq p_{AB} + p_{BC} + p_{CD} - p_{DA}, \tag{2.5}$$

and the other three obtained by permutations involving the measurable quantities (2.4). The inequality (2.5) is called a Bell inequality (Bell, 1964) and, in this form, it was derived by Clauser and Horne (1974). The rest of the proof consists of showing that there are states and observables for which quantum mechanics violates the inequalities, which may be seen elsewhere, e.g., in clauser and Horne, 1974.

Instead of observables taking the values 0 or 1, we might use observables taking the values -1 or $+1$. They are trivially related to the previous ones by

$$a = 2A - 1, b = 2B - 1, \text{ etc.} \tag{2.6}$$

and the inequality (2.5) takes the form of Clauser–Horne–Shimony–Holt (CHSH: 1974)

$$|\langle ab\rangle + \langle bc\rangle + \langle cd\rangle - \langle ad\rangle| \leq 2. \tag{2.7}$$

Therefore this, CHSH, and the Clauser–Horne inequalities (2.5) are equivalent.

An important class of HV theories are *local HV theories*. The concept of local applies to EPR experiments. We call EPR (Einstein *et al.*, 1935; Wheeler and Zurek, 1983) an experiment where we prepare locally a system which is later divided in two subsystems, each of which moves in a different direction. Measurements on each subsystem are later made at space-like separation (in the sense of relativity theory). □

Definition 2.3. A HV theory is local if, for any EPR experiment where we may measure one of several observables, A_i, of the first subsystem and one of several observables, B_j, on the second, there exist a joint probability distribution for all the observables $\{A_i, B_j; i, j = 1, 2, \ldots\}$.

The impossibility of local HV theories is established by the celebrated Bell's theorem of 1964 (Bell, 1964).

Theorem 2.4. *Local HV theories do not exist for all (EPR) experiments.*

Proof: The proof is the same as for noncontextual HV theories, but considering an EPR experiment. That is, the observables A, C belong to one subsystem and B, D to the other subsystem. In particular, this guarantees that the pairs $\{A, B\}$, $\{A, D\}$, $\{C, B\}$, and $\{C, D\}$ are compatible because they belong to space-like separated regions (the condition that space-like separated observables are compatible is called *microcausality* in quantum field theory).

The class of local theories is wider than that of noncontextual HV theories because the constraints in their definition are weaker. Indeed, in local theories the existence of a joint distribution is only required for EPR experiments, but noncontextual theories assume it for all experiments. Consequently the empirical disproof is easier for noncontextual theories than for local theories. In the former it is enough to perform a composite experiment where the measurements are made locally, the latter requires measurements at space-like separation.

The fact that the proofs of both theorems are very similar has been a source of misunderstanding, like the assertion that locality is not needed in order to prove Bell's theorem. I hope that in our presentation the point is clear enough. □

3. QUANTUM INFORMATION

The amount of information is quantified with the concept of *entropy*. In classical physics, if we have a continuous random variable, λ, with a probability distribution $\rho(\lambda)$, the entropy, S^{c}, as defined by Shannon is

$$S^{\mathrm{C}} = -\int \rho(\lambda) \log \rho(\lambda)\, d\lambda. \tag{3.1}$$

The quantum entropy was defined by von Neumann in terms of the density operator, $\hat{\rho}$, with an expression which looks similar to the one, namely

$$S^{\mathrm{Q}} = -Tr(\hat{\rho}\ \log \hat{\rho}). \tag{3.2}$$

In both cases $S \geq 0$ and the entropy increases with the lack of information, so that the pure states (maximal information) corresponds to $S = 0$.

There are two other properties which hold true for both classical and quantum entropy:

Concavity. $\lambda S(\rho_a) + (1-\lambda)S(\rho_b) \leq S(\lambda\rho_a + (1-\lambda)\rho_b), 0 \leq \lambda \leq 1,$ where ρ_a stands for either the classical probability density, $\rho_a(\lambda)$, or the quantum density operator, $\hat{\rho}_a$, and similarly ρ_b for a different probability density or density operator of the same system.

Additivity. $S(\rho_{12}) \leq S(\rho_1) + S(\rho_1)$, where ρ_{12} stands for either the classical probability density, $\rho_{12}(\lambda_1, \rho_2)$, or the quantum density operator, $\hat{P}_{12}$, the subindex 1 (2) referring to the first (second) subsystem of a composite system, and we have

$$\rho_1(\lambda_1) = \int \rho_{12}(\lambda_1, \lambda_2)\, d\lambda_2,\ \hat{\rho}_1 = Tr_2\hat{\rho}_{12}. \tag{3.3}$$

There is, however, a property which dramatically distinguish classical from quantum entropy. In fact, in the case of a system consisting of two subsystems, the classical, Shannon's, entropy fulfills

$$S^{\mathrm{C}}(\rho_{12}) \geq \max(S^{\mathrm{C}}(\rho_1), S^{\mathrm{C}}(\rho_2)), \tag{3.4}$$

while the quantum entropy fulfills the weaker triangle inequality

$$S^{\mathrm{Q}}(\hat{\rho}_{12}) \geq |S^{\mathrm{Q}}(\hat{\rho}_1) - S^{\mathrm{Q}}(\hat{\rho}_2)|. \tag{3.5}$$

In my opinion, the fact that the quantum entropy does not fulfill an inequality similar to (3.4) is highly paradoxical, I would even say bizarre. In fact, (3.5) allows for the possibility that both $S^{\mathrm{Q}}(\hat{\rho}_1)$ and $S^{\mathrm{Q}}(\hat{\rho}_2)$ are positive while $S^{\mathrm{Q}}(\hat{\rho}_{12})$ is zero. This should be interpreted saying that we have complete information about a composite system whilst we have incomplete information about every subsystem. This contrast with the classical, and intuitive idea that full information about the whole *means* that we have complete information about every part. In my view, this is indicative that the concept of "complete" information in quantum theory is not the same as in classical physics, and the different meanings of completeness has been the source of misunderstandings about the interpretation of quantum theory, for example, in the debate between Einstein and Bohr.

The violation of an inequality similar to (3.4) is closely related to the violation of the Bell inequality. But in order to establish the connection it is necessary to introduce the concept of *linear entropy*. Actually, although the definitions of entropy (3.1) and (3.2) are standard and in some sense an optimum, it is possible to give alternative definitions of entropy which fulfill the essential properties of concavity and additivity. The most simple are the linear entropies

$$S^{\mathrm{CL}} = 1 - \int \rho(\lambda)^2 \, d\lambda, \qquad S^{\mathrm{QL}} = 1 - Tr(\hat{\rho}^2). \tag{3.6}$$

The desired connection between linear entropy and the Bell inequalities has been studied by several authors in the last few years. For instance Horodecki *et al.* (1996) proved that the inequality (3.4) is a sufficient condition for the Bell inequalities. A slightly stronger result may be stated as follows:

Theorem 3.1. *The inequality*

$$S^{\mathrm{QL}}(\hat{\rho}_{12}) \geq \frac{1}{2}[S^{\mathrm{QL}}(\hat{\rho}_1) + S^{\mathrm{QL}}(\hat{\rho}_2)],$$

is a sufficient condition for the Bell inequalities (2.5) or (2.7).

Proof: We consider observables $\{a, b\}$ for the first particle and $\{c, d\}$ for the second, all of which may take values 1 or -1, and the associated operators, $\hat{A}$, $\hat{B}$, $\hat{c}$, and $\hat{d}$. We define the Bell operator, $\hat{B}$, by

$$\hat{B} = \hat{A} \otimes \hat{B} + \hat{c} \otimes \hat{B} + \hat{c} \otimes \hat{d} - \hat{A} \otimes \hat{d}. \tag{3.7}$$

It is easy to see that

$$Tr\hat{B} = 0, \qquad Tr(\hat{B}^2) = 16, \tag{3.8}$$

and that the Bell inequality (2.7) is violated if

$$|\beta| > 2, \qquad \beta \equiv Tr(\hat{B}\hat{\rho}_{12}), \tag{3.9}$$

(while quantum mechanics predicts just $|\beta| \leq 2\sqrt{2}$). Now from the obvious inequality

$$Tr\left(\hat{\rho}_{12} - \frac{1}{2}\hat{\rho}_1 \otimes \hat{I}_2 - \frac{1}{2}\hat{I}_1 \otimes \hat{\rho}_2 + \frac{1}{4}\hat{I}_1 \otimes \hat{I}_2 + \lambda\hat{B}\right)^2 \geq 0, \qquad \forall\lambda \in R, \tag{3.10}$$

where $\hat{I}_1(\hat{I}_2)$ is the identity operators for the first (second) particle, we get after some algebra

$$2Tr\left(\hat{\rho}_{12}^2\right) - Tr\left(\hat{\rho}_1^2\right) - Tr\left(\hat{\rho}_2^2\right) \geq \frac{1}{8}(\beta^2 - 4). \tag{3.11}$$

Hence the inequality (3.11) implies $|\beta| \leq 2$, which proves the theorem. □

4. QUANTUM LOGIC AND QUANTUM PROBABILITY

The concept of quantum logic was introduced by Birckhoff and von Neumann in 1936 (Birckhoff and von Neumann, 1936). Their starting point was the association of propositions with projection operators. They postulated that the proposition associated to the projector $\hat{P}$ is true (or false) if the vector $|\Psi\rangle$ is an eigenvector of $\hat{P}$ (or $\hat{I} - \hat{P}$). This may seem a natural assumption, but gives rise to a trivalent logic where propositions may be, besides true or false, also undefined (if $|\Psi\rangle$ is neither an eigenvector of $\hat{P}$ nor an eigenvector of $\hat{I} - \hat{P}$.) As projectors are associated to closed subspaces of the Hilbert space, we see that propositions are associated to such subspaces.

From this assumption it is straightforward to define a partial ordering amongst propositions. We say that, for two propositions A and B we have A $\leq$ B if the subspace associated to B contains that associated to A. Hence it is straightforward to define the binary operations "meet", $\curlywedge$ and "join," $\curlyvee$, of propositions and it follows that the propositions form a lattice. The lattice is *orthocomplemented* (for each proposition A there exist another one, A, which is true if and only if the first is false) and *complete* (there exist the sure proposition, I, corresponding to the whole Hilbert space and its negation, Φ, the absurd proposition). All this is similar to what happens with the classical propositions, which is a Boolean algebra, that is a *distributive* lattice. But the quantum lattice is not Boolean (distributive) at a difference with the classical one. As a conclusion the authors claim that the non-Boolean character of the lattice of propositions is the essential characteristic of quantum theory. The details may be seen in the original article (Birckhoff and von Newman, 1936).

In the 70 years elapsed since the work of Birkhoff and von Neumann many articles and several books have been devoted to the subject of quantum logic (see e.g., in Hooker, 1975, 1979), in many cases starting from different definitions of quantum propositions. Also some criticisms have aroused in the sense that "quantum logic" is not a true logic, but just a propositional calculus.

It is straightforward to define a probability distribution (or "state") on a lattice as follows

Definition 4.1. If $\mathcal{L}$ is a lattice, a probability distribution is a mapping $p : \mathcal{L} \to [0, 1]$ with the axioms

(1) $p(\Phi) = 0$, $p(I) = 1$, where $\Phi(I)$ is the absurd (sure) proposition,
(2) If $\{A_i\}$ is a sequence such that $A_i \leq A'_j$, A' being the negation of A, for all pairs $i \neq j$, then $\sum_i p(A_i) = p(\curlyvee A_i)$,
(3) For any sequence $\{A_i\}$, $p(A_i) = 1 \quad \forall_i \Rightarrow p(\curlywedge A_i) = 1$.

Thus from the quantum logic, as defined by Birkhoff and von Neumann, we get a quantum probability, while the classical, Boolean, lattice of propositions gives the standard probability. Indeed, the above axioms are simply a generalization of axioms of probability as stated by Kolmogorov.

The connection between probability and the Bell inequality appears as follows. For any two proposition $A, B \in \mathcal{L}$ we may define a function, $d(A, B)$, as follows

$$d(A, B) = p(A \curlyvee B) - p(A \curlywedge B). \tag{4.1}$$

That function has the properties

$$0 \leq d(A, B) \leq 1, d(A, A) = 0, d(A, A') = 1, \tag{4.2}$$

and provides some measure of the "distance" between two propositions in a given state (probability distribution). The function is called a *metric (pseudometric)* if the following additional property holds (does not hold) true

$$d(A, B) = 0 \Rightarrow A = B, \tag{4.3}$$

but this property is not very relevant for our purposes. More important are the following triangle inequalities, which are (are not generally) fulfilled if the lattice is (is not) Boolean

$$|d(A, B) - d(A, C)| \leq d(B, C) \leq d(A, B) + d(A, C). \tag{4.4}$$

As the Boolean character provides the essential difference between classical and quantum theories, according to Birckhoff and von Neumann (1936), we see that the inequalities (4.4) give a criterium to distinguish both theories. The interesting point is that these inequalities are closely related to the Bell inequalities (Santos, 1986). In quantum mechanics, if we consider three compatible observables, $\{A, B, C\}$, the inequalities (4.4) hold true because the lattice of commuting observables is Boolean. On the other hand, if two of the observables, say A and B, are not compatible then their distance is not defined because quantum mechanics does not provide a joint probability of two incompatible observables (and it is assumed that they cannot be measured simultaneously). However there are inequalities, derived from (4.4), which may be violated by quantum mechanics and tested empirically. In fact, if we consider four observables $\{A, B, C, D\}$ it is easy to see that the inequalities (4.4) lead to

$$d(A, D) \leq d(A, B) + d(B, C) + d(C, D). \tag{4.5}$$

It may be realized that this is just the Bell inequality (2.5).

ACKNOWLEDGMENT

I acknowledge financial support from DGICYT, Project No. PB-98-0191 (Spain).

REFERENCES

Bell, J. S. (1964). *Physics* **1**, 195.
Bell, J. S. (1966). *Reviews of Modern Physics* **38**, 447.
Bell, J. S. (1987). *Speakable and Unspeakable in Quantum Mechanics*, Cambridge University Press, Cambridge, UK.
Birckhoff, G. and von Neumann, J. (1936). *Annals of Mathematics* **37**, 823. Reprinted in Ref. 10.
Clauser, J. F. and Horne, M. A. (1974). *Physical Review of D* **10**, 526.
Clauser, J. F., Horne, M. A., Shimony, A., and Holt, R. A. (1974). *Physical Review Letters* **23**, 880.
Einstein, A., Podolsky, B., and Rosen, N. (1935). *Physical Review* **47**, 777.
Gleason, A. M. (1957). *Journal of Mathematical Mechenics* **6**, 885.
Hooker, C. A. (ed.) (1975, 1979). *The Logico-Algebraic Approach to Quantum Mechanics, Vols. I and II*, Reidel, Dordrecht, The Netherlands.
Horodecki, R., Horodecki, P., and Horodecki, M. (1996). *Physical Review Letters A* **210**, 377.
Horodecki, R. and Horodecki, M. (1996). *Physical Review A* **54**, 1838.
Kochen, S. and Specker, E. P. (1967). *Journal of Mathematical Mechenics* **17**, 59.
Santos, E. (1986). *Physical Letters A* **115**, 363.
von Neumann, J. (1955). *Mathematical Foundations of Quantum Mechanics*, Princeton University Press, Princeton, NJ (German original 1932).
Wheeler, J. A. and Zurek, W. (eds.) (1983). *Quantum Mechanics and Measurement*, Princeton University Press, Princeton, NJ.