International Journal of Theoretical Physics, Vol. 27, No. 11, 1988

Probability and Logical Structure of Statistical Theories

Michael J. W. Hall

Received May 12, 1988

A characterization of statistical theories is given which incorporates both classical and quantum mechanics. It is shown that each statistical theory induces an associated logic and joint probability structure, and simple conditions are given for the structure to be of a classical or quantum type. This provides an alternative for the quantum logic approach to axiomatic quantum mechanics. The Bell inequalities may be derived for those statistical theories that have a classical structure and satisfy a locality condition weaker than factorizability. The relation of these inequalities to the issue of hidden variable theories for quantum mechanics is discussed and clarified.

1. INTRODUCTION

A number of physical systems, such as two-sided coins, electrons, and viral infections, behave in a *statistical* manner. Whether an electron will be detected in a certain place, or a tossed coin will land with "head" upward, or a member of some population will succumb to an influenza virus, cannot be correctly predicted in many circumstances. However, as is well known, the behavior of *many* coins, *many* electrons, and *large* population groups is often amenable to accurate predictions involving *relative frequencies*, or probabilities.

This paper is primarily concerned with characterizing the class of theories which describe statistical phenomena, and linking this characterization with (1) the quantum logic approach to axiomatic quantum mechanics [originated by Birkhoff and von Neumann (1936); see also Gudder (1979), Beltrametti and Cassinelli (1981), and Primas (1981)] and (2) the derivation and significance of the Bell inequalities [see the review by Clauser and Shimony (1978) for an exposition and further references].

[1]Department of Theoretical Physics, Research School of Physical Sciences, Australian National University, Canberra, ACT 2600, Australia.

0020-7748/88/1100-1285$06.00/0

In the next section notation suitable for describing theoretical predictions is introduced, allowing a definition of a *statistical theory.* Statistical theories include classical and quantum mechanics as examples.

It is shown in Section 3 that each statistical theory induces an associated probability and logical structure. Simple necessary and sufficient conditions are given for this structure to be of either a "classical" or a "quantum" type. This provides the basis for a "statistical theory approach" to quantum mechanics, which is contrasted with the quantum logic approach mentioned above.

Section 4 is concerned with "hidden variables" within the context of statistical theories. The concept of *covering theories*, which includes the notion of hidden variable theories, is defined, and some examples are given. In particular, it is shown that there is a covering theory with a "classical" structure for each statistical theory.

In Section 5 it is demonstrated that the generalized Bell inequalities first derived by Clauser *et al.* (1969) hold for all "classical" statistical theories which satisfy a weak locality condition. This is based on and clarifies some earlier work (Hall, 1988). The derivation is contrasted to that of Clauser and Horne (1974) and of Fine (1982), and its relation to the existence of hidden variable theories for quantum mechanics discussed.

Finally, results are summarized in a conclusion. Technical proofs are kept to Appendices for easier reading.

2. STATISTICAL THEORIES

The usefulness of a theory is determined, at least in part, by its ability to make predictions. These predictions (which may be retrospective) refer to experiments or observations on some class of systems, and in particular make statements about the results of such experiments. Notation suitable for discussing predictions will now be developed, as a prelude to characterizing those theories that make *statistical* predictions.

If the possible results of some experiment E are contained in a set R_E, then there is an associated group of *propositions* of the form, "The result of E is contained in a subset X of R_E." These propositions are verified or falsified by the performance of experiment E, and indeed fully characterize the outcome of the experiment.

It follows that predictions of the theory may be expressed in terms of results of the yes/no experiments which correspond to testing various propositions. The set of all propositions involved in predictions of the theory will be denoted by $\mathcal{P}$. Further, the class of systems on which the experiments are performed may be characterized by a set of *states*, denoted by S, where each state is a description of a member of the class of systems. The result of testing proposition $A \in \mathcal{P}$ on state $\lambda \in S$ will be denoted by $r(A, \lambda)$, where $r(A, \lambda) := 1$ (0) if A is verified (falsified).

Hence, if propositions $A_1, A_2, \ldots, A_N \in \mathcal{P}$ can be tested respectively on systems described by states $\lambda_1, \lambda_2, \ldots, \lambda_N \in S$, the theory is expected to make some prediction(s) about the results

$$r(A_1, \lambda_1), r(A_2, \lambda_2), \ldots, r(A_N, \lambda_N)$$

For example, a *deterministic theory* contains a mapping d from $\mathcal{P} \times S$ to the set $\{0, 1\}$ such that $r(A_i, \lambda_i)$ is predicted to be $d(A_i, \lambda_i)$ for each i.

In contrast, a *statistical theory* is defined here as a theory which contains a mapping p from $\mathcal{P} \times S$ to the interval $[0, 1]$ such that the total number of verifications for the above sequence $N_{\text{yes}} := \sum_{i=1}^{N} r(A_i, \lambda_i)$ is predicted to have the behavior

$$\frac{N_{\text{yes}}^{\text{theoretical}} - N_{\text{yes}}}{N} \to 0 \qquad \text{as} \quad N \to \infty \tag{1a}$$

where $N_{\text{yes}}^{\text{theoretical}}$ is given by the expression

$$N_{\text{yes}}^{\text{theoretical}} := \sum_{i=1}^{N} p(A_i, \lambda_i) \tag{1b}$$

Thus, deterministic theories are special cases of statistical theories, with $p(A, \lambda) \equiv d(A, \lambda)$.

If $p(A_i, \lambda_i) = p$ for each value of i, then equations (1) imply that $N_{\text{yes}} \approx pN$ for large N. Thus, for $A \in \mathcal{P}$, $\lambda \in S$, the quantity $p(A, \lambda)$ may be interpreted as "the probability of verifying proposition A on state λ."

Now, if $p(A, \lambda)$ and $p(B, \lambda)$ are equal for all states $\lambda \in S$, then the theory does not distinguish between predictions involving the propositions A or B. Hence, we may say that A and B are *equivalent* with respect to the theory, and write $A \equiv B$. Of course, one might be able to refine the state descriptions so as to discriminate between two such propositions; however, this would entail a *new* theory.

Thus, the propositions of a statistical theory may in fact be represented (up to equivalence) as *mappings* from the set of states S to the interval $[0, 1]$, where for $A \in \mathcal{P}$, $\lambda \in S$, we define

$$A(\lambda) := p(A, \lambda) \tag{2}$$

Using the symbol A to denote both a proposition and a mapping should not lead to confusion, as the remainder of the paper is primarily concerned with properties of the latter representation. Thus, for the purposes of what follows,, a statistical theory will be denoted by $(S, \mathcal{P})$, where S is the set of states and $\mathcal{P}$ is the set of mappings from S to $[0, 1]$ obtained via equation (2).

3. THE PROBABILITY AND LOGICAL STRUCTURE OF STATISTICAL THEORIES

3.1. Introduction

The definition of a statistical theory in the previous section is of a very general nature, yet it turns out that very few conditions need to be further imposed to obtain interesting structural properties. Indeed, without *any* extra assumptions, formulations of "joint probabilities," "complementary propositions," and an "implication relation" can be specified in a natural way, leading to an associated probability structure and a representative logic for each statistical theory.

The existence of an inherent structure, in which "classical" rules for manipulating probabilities and propositions [e.g., $p(a \,\&\, b)+p(a \,\&\, b')=p(a)$, and $a \,\&\, a' \equiv 0$] need not hold in general is consistent with the conceptual viewpoint that such rules are not absolute. This has similarities with the idea that the rules of Euclidean geometry do not apply *a priori* to physical space-time, and such an analogy has been explored by Accardi (1984) and Pitowsky (1986). Alternatively, one may argue that the classical rules are indeed absolute, and hence that only statistical theories with a "classical" structure can be of fundamental interest.

The first viewpoint supports the quantum logic approach to axiomatic quantum mechanics, in which propositions are constrained to satisfy some, but not all, of the properties of Boolean logic (see, e.g., Beltrametti and Cassinelli, 1981). The second viewpoint, when confronted with quantum mechanics, leads to a conflict with notions of local causality in the form of the Bell inequalities (see Section 5).

In the following sections, the probability and logical structure inherent to statistical theories are investigated, and simple necessary and sufficient conditions are given for statistical theories to exhibit classical and quantum-like behavior. The latter case leads to a "statistical theory" approach to axiomatic quantum mechanics, which is contrasted with the quantum logic approach mentioned above.

3.2. Potential Propositions

Now, an aim of this paper is to associate with any two propositions A, B of a statistical theory $(S, \mathcal{P})$ a quantity $(A \wedge B)(\lambda)$ for each $\lambda \in S$, which may be interpreted as a "joint probability." Clearly, then, a mapping $A \wedge B$: $S \to [0, 1]$ must be specified. Desirable properties of such a mapping are discussed in Section 3.3; for now it is noted that one may not necessarily have $A \wedge B \in \mathcal{P}$. This motivates the definition of the set $\mathcal{PP}$ of *potential* *propositions* of $(S, \mathscr{P})$ as the set of mappings from S to $[0, 1]$, i.e.,

$$\mathscr{P}\mathscr{P} := \{A \mid A: S \to [0, 1]\} \tag{3}$$

It immediately follows that $\mathscr{P} \subseteq \mathscr{P}\mathscr{P}$.

Two special elements of $\mathscr{P}\mathscr{P}$, $\boldsymbol{0}$ and $\boldsymbol{1}$, are defined by

$$\boldsymbol{0}(\lambda) := 0, \qquad \boldsymbol{1}(\lambda) := 1 \qquad \forall \lambda \in S \tag{4}$$

and for each $A \in \mathscr{P}\mathscr{P}$, the *complement* A' of A is defined by

$$A'(\lambda) := 1 - A(\lambda) \qquad \forall \lambda \in S \tag{5}$$

For convenience it will be assumed that $\boldsymbol{0}, \boldsymbol{1} \in \mathscr{P}$ for the remainder of the paper. These propositions can be interpreted as corresponding to yes/no experiments which always give a no and a yes result, respectively.

In some cases, the complementary proposition A' may also be interpreted physically for $A \in \mathscr{P}$. First, define the experimental proposition $\tilde{A}$ such that $\tilde{A}$ is tested on state $\lambda \in S$ by testing A on λ and reversing the yes/no results (see Section 2). Thus, $r(\tilde{A}, \lambda) = 1 - r(A, \lambda)$. From equations (1) and (2) it follows that if $\tilde{A}_i \in \mathscr{P}$ for each i, then for N large, $\sum_{i=1}^{N} \tilde{A}_i(\lambda_i) \approx \sum_{i=1}^{N} A_i'(\lambda_i)$. Hence it is consistent for propositions $\tilde{A}$ and A' to be equivalent with respect to $(S, \mathscr{P})$, and in such a case the latter proposition has a simple interpretation.

3.3. The Associated Probability Structure and Representative Logic of a Statistical Theory

In this section, the mappings $A \wedge B$ and $A \vee B \in \mathscr{P}\mathscr{P}$, called the *join* and *meet* of propositions A, $B \in \mathscr{P}\mathscr{P}$, will be defined, as well as the partial ordering relation $A \Rightarrow B$, called *implication.* Properties of these definitions are stated, with proofs given in Appendix A.

To motivate these definitions, it is first suggested that reasonable properties to be satisfied for the quantity $(A \wedge B)(\lambda)$ to behave like a "joint probability" include $A \wedge B \equiv B \wedge A$, $A \wedge (B \wedge C) \equiv (A \wedge B) \wedge C$, and $(A \wedge B)(\lambda) \leq A(\lambda)$ for each $\lambda \in S$. The latter property can be written more succinctly as $A \wedge B \Rightarrow A$, where $\Rightarrow$ is the natural partial ordering on $\mathscr{P}\mathscr{P}$, given by

$$A \Rightarrow B \quad \text{iff} \quad A(\lambda) \leq B(\lambda) \qquad \forall \lambda \in S \tag{6}$$

Condition (6) may be considered as generalizing equation (3c) of Birkhoff and von Neumann (1936) to a definition of implication for all statistical theories.

It follows from condition (6) that the implication relation satisfies the following properties for all A, B, $C \in \mathcal{PP}$:

$$A \Rightarrow A \quad \text{(reflexivity)} \tag{7a}$$

$$\text{if} \quad A \Rightarrow B,\, B \Rightarrow C, \quad \text{then } A \Rightarrow C \quad \text{(transitivity)} \tag{7b}$$

$$\text{if} \quad A \Rightarrow B,\, B \Rightarrow A, \quad \text{then } A \equiv B \tag{7c}$$

$$\text{if} \quad A \Rightarrow B, \quad \text{then } B' \Rightarrow A' \tag{7d}$$

$$\boldsymbol{0} \Rightarrow A \Rightarrow \boldsymbol{1} \tag{7e}$$

The quantities $(A \wedge B)(\lambda)$ and $(A \vee B)(\lambda)$ are now defined for A, $B \in \mathcal{PP}$, $\lambda \in S$, by

$$(A \wedge B)(\lambda) := \sup\{X(\lambda) \mid X \in \mathcal{P}, X \Rightarrow A, X \Rightarrow B\} \tag{8a}$$

$$(A \vee B)(\lambda) := \inf\{X(\lambda) \mid X \in \mathcal{P}, A \Rightarrow X, B \Rightarrow X\} \tag{8b}$$

The existence of $A \wedge B$ and $A \vee B$ is assured by the earlier assumptions $\boldsymbol{0} \in \mathcal{P}$ and $\boldsymbol{1} \in \mathcal{P}$, respectively.

It is shown in Appendix A that the reasonable properties suggested earlier in this section for $A \wedge B$ are indeed satisfied by definition (8a), i.e., for A, B, $C \in \mathcal{PP}$ one has

$$A \wedge B \equiv B \wedge A \tag{9a}$$

$$A \wedge (B \wedge C) \equiv (A \wedge B) \wedge C \tag{9b}$$

$$A \wedge B \Rightarrow A \tag{9c}$$

It is also shown that $A \wedge B$ has properties similar to a greatest lower bound for A and B, namely

$$(X \Rightarrow A \wedge B) \quad \text{iff} \quad (X \Rightarrow A, X \Rightarrow B) \qquad \forall X \in \mathcal{P} \tag{9d}$$

for all A, $B \in \mathcal{PP}$.

The mapping $A \vee B$ satisfies analogous properties, also proved in Appendix A. Thus, for A, B, $C \in \mathcal{PP}$ one has

$$A \vee B \equiv B \vee A \tag{10a}$$

$$A \vee (B \vee C) \equiv (A \vee B) \vee C \tag{10b}$$

$$A \Rightarrow A \vee B \tag{10c}$$

$$(A \vee B \Rightarrow X) \quad \text{iff} \quad (X \Rightarrow A, X \Rightarrow B) \qquad \forall X \in \mathcal{P} \tag{10d}$$

where (10d) demonstrates that $A \vee B$ behaves like a least upper bound for A and B. From relations (9) and (10) follow such properties as

$$A \wedge \boldsymbol{0} \equiv \boldsymbol{0}, \quad A \vee \boldsymbol{1} \equiv \boldsymbol{1}, \quad A \wedge A \equiv A \wedge \boldsymbol{1} \equiv A \equiv A \vee \boldsymbol{0} \equiv A \vee A, \quad \forall A \in \mathcal{P} \tag{11}$$

Relations (9)-(11) indicate that definitions (8a) and (8b) give a reasonable, abstract formulation of a probability structure for statistical theories. The *representative logic* of the statistical theory $(S, \mathcal{P})$ is defined to be the partially complemented poset $(\mathcal{P}, \Rightarrow, ')$. For elements A, B of the poset, the greatest lower bound of A and B exists only if $A \wedge B \in \mathcal{P}$, and is then given by $A \wedge B$, from condition (9d). Similarly, from (10d) the least upper bound exists only if $A \vee B \in \mathcal{P}$, and is then given by $A \vee B$. The representative logic is fully complemented if $A' \in \mathcal{P}$, $\forall A \in \mathcal{P}$.

Finally, definitions (8a), (8b) may be extended consistently as follows, to give the join and meet of any countable sequence $A_1, A_2, A_3, \ldots \in \mathcal{P}\mathcal{P}$.

$$\left(\bigwedge_i A_i\right)(\lambda) := \sup\left\{X(\lambda) \mid X \in \mathcal{P}, X \Rightarrow \bigcap_i A_i\right\} \tag{12a}$$

$$\left(\bigvee_i A_i\right)(\lambda) := \inf\left\{X(\lambda) \mid X \in \mathcal{P}, \bigcup_i A_i \Rightarrow X\right\} \tag{12b}$$

where the propositions $\bigcap_i A_i, \bigcup_i A_i \in \mathcal{P}\mathcal{P}$ are defined by

$$\left(\bigcap_i A_i\right)(\lambda) := \inf\{A_i(\lambda)\}, \qquad \left(\bigcup_i A_i\right)(\lambda) := \sup\{A_i(\lambda)\} \tag{13}$$

It is shown in Appendix A that

$$\text{if} \quad A' \in \mathcal{P}, \forall A \in \mathcal{P}, \quad \text{then } \bigvee_i A_i \equiv \left(\bigwedge_i A_i'\right)' \tag{14}$$

i.e., de Morgan's law holds for those statistical theories in which the representative logic is fully complemented.

3.4. Classical and Regular Statistical Theories

In the last subsection the probability and logical structure of statistical theories was investigated, and the fundamental structural properties noted [(7), (9), (10), (11), and (14)]. In this subsection, classical and regular statistical theories will be defined and discussed and examples of each provided.

First, a statistical theory $(S, \mathcal{P})$ is defined to be *classical* when the following conditions are satisfied for all $A, B \in \mathcal{P}$, $\lambda \in S$:

$$A' \in \mathcal{P} \tag{15a}$$

$$A \wedge B \in \mathcal{P} \tag{15b}$$

$$(A \wedge B)(\lambda) + (A' \wedge B)(\lambda) = B(\lambda) \tag{15c}$$

Conditions (15a), (15b) imply that $\mathcal{P}$ is closed under complementation and join of propositions, and hence, from equation (14), it is also closed under the meet of propositions, i.e., $A \vee B \in \mathcal{P}$, $\forall A, B \in \mathcal{P}$. Condition (15c) may be recognized as a fundamental law of classical probability (see, e.g., Kolmogorov, 1950), especially if rewritten in the form $p(A \wedge B, \lambda) + p(A' \wedge B, \lambda) = p(B, \lambda)$, using equation (2).

To motivate the term "classical" for statistical theories which satisfy conditions (15), it is first convenient to define *generalized probability measures.* A generalized probability measure on $(S, \mathscr{P})$ is a mapping $m: \mathscr{P}\mathscr{P} \to [0, 1]$ such that

$$\text{if} \quad A_1, A_2, A_3, \ldots \in \mathscr{P} \quad \text{satisfy} \quad A_i \Rightarrow A_j' \quad \forall i \neq j,$$

$$\text{then} \quad m\left(\bigvee_i A_i\right) = \sum_i m(A_i) \tag{16a}$$

$$m(\boldsymbol{1}) = 1 \tag{16b}$$

Conditions (16a), (16b) generalize the definition of a probability measure by Beltrametti and Cassinelli (1981). If the representative logic of $(S, \mathscr{P})$ is *Boolean* (see Appendix B), then the relations $A_i \Rightarrow A_j'$ and $A_i \wedge A_j \equiv \boldsymbol{0}$ are equivalent, and equation (16a) then implies that m is additive for mutually exclusive propositions, which is the basic tenet of classical probability theory (Kolmogorov, 1950). Note that from $\boldsymbol{0} \Rightarrow \boldsymbol{0}'$ and (16a) it follows that $m(\boldsymbol{0}) = 0$, while equation (16b) is a normalization condition.

It is shown in Appendix B that a statistical theory $(S, \mathscr{P})$ is classical if and only if (1) the representative logic is Boolean, and (2) the mappings $m_\lambda: \mathscr{P}\mathscr{P} \to [0, 1]$ defined by $m_\lambda(A) := A(\lambda)$ are generalized probability measures on $(S, \mathscr{P})$ for all states $\lambda \in S$. Thus, classical statistical theories are just those statistical theories which have a "classical" probability and logical structure. The significance of the Bell inequalities for such statistical theories will be discussed in Section 5.

A *regular* statistical theory is defined to be one for which $(S, \mathscr{P})$ satisfies the conditions

$$A' \in \mathscr{P} \tag{17a}$$

$$\text{if} \quad A \Rightarrow B, \quad \text{then } A' \wedge B \in \mathscr{P} \tag{17b}$$

$$\text{if} \quad A \Rightarrow B, \quad \text{then } (A \wedge B)(\lambda) + (A' \wedge B)(\lambda) = B(\lambda) \tag{17c}$$

for all A, $B \in \mathscr{P}$, $\lambda \in S$. Comparison with conditions (15) shows that all classical statistical theories are contained in the set of regular statistical theories. It is demonstrated in Appendix C that a statistical theory $(S, \mathscr{P})$ is regular if and only if (1) the representative logic is orthocomplemented, orthocomplete, and orthomodular, and (2) the m_λ as defined above are generalized probability measures for all $\lambda \in S$. These conditions are related to the quantum logic approach to axiomatic quantum mechanics, and indeed conditions (17) provide a basis for an alternative "statistical theory" approach, outlined in Section 3.5.

To conclude this section, generic examples of classical and regular statistical theories are given.

Example 3.1. Let S be the phase space of some system of classical dynamics and $\mathcal{P}$ be the set of propositions corresponding to whether the state of the system is contained within some subset of S. Thus, for $X \subseteq S$, the proposition $A_X \in \mathcal{P}$ is defined by $A_X(\lambda) := 1\ (0)$ for $\lambda \in X$ $(\lambda \notin X)$. From (5) it follows that $A'_X \equiv A_{X^c} \in \mathcal{P}$, where $X^c = S \backslash X$, while from relations (6) and (8) it follows that $A_X \Rightarrow A_Y$ is equivalent to $X \subseteq Y$, and hence that $A_X \wedge A_Y \equiv A_{X \cap Y} \in \mathcal{P}$. Thus, $(S, \mathcal{P})$ satisfies conditions (15a), (15b). Further, if $\lambda \in X$, then either $\lambda \in X \cap Y$ or $\lambda \in X \cap Y^c$; while if $\lambda \notin X$, then $\lambda \notin X \cap Y$ and $\lambda \notin X \cap Y^c$. It follows that $A_{X \cap Y}(\lambda) + A_{X \cap Y^c}(\lambda) = A_X(\lambda)$ for all $\lambda \in S$, and therefore condition (15c) is also satisfied. Hence $(S, \mathcal{P})$ is a classical statistical theory.

Example 3.2. Let S be the set of unit vectors of a separable Hilbert space H, and let $\mathcal{P}$ be indexed by the set of closed subspaces of H, where if $\mathbf{E}$ denotes the projection onto subspace E, then $A_E(\psi) := (\psi, \mathbf{E}\psi)$ for all $\psi \in S$, $A_E \in \mathcal{P}$. Thus, $(S, \mathcal{P})$ describes a quantum mechanical system. From (5) one has $A'_E \equiv A_{E^\perp} \in \mathcal{P}$ where $E^\perp$ is the orthogonal complement of E in H. Further, from relations (6) and (8) it follows that the relations $A_{E_1} \Rightarrow A_{E_2}$ and $E_1 \subseteq E_2$ are equivalent and hence that $A_{E_1} \wedge A_{E_2} \equiv A_{E_1 \cap E_2} \in \mathcal{P}$. Thus, conditions (17a), (17b) are satisfied by $(S, \mathcal{P})$. Finally, if $E_1 \subseteq E_2$, then E_2 may be expressed as the direct sum of subspaces $E_1 \cap E_2$ and $E_1^\perp \cap E_2$, and so from the properties of the inner product on H it follows that condition (17c) is also satisfied. Hence, $(S, \mathcal{P})$ is a regular statistical theory. Note that (15c) is *not* satisfied (e.g., let E_1, E_2 be distinct, nonorthogonal, one-dimensional subspaces of H, and choose $\lambda \in E_1$), and hence $(S, \mathcal{P})$ is not a classical statistical theory.

3.5. Comparison with Quantum Logic Approach

The aim of the quantum logic approach is to derive the Hilbert space structure of quantum mechanics from a small number of axioms. The approach (see, e.g., Gudder, 1979) begins with a set of propositions $\mathcal{P}$ which has a number of conditions imposed on it, including the existence of a partial ordering, an orthocomplementation, orthomodularity, and orthocompleteness (see Appendix C for definitions). States are then defined independently, essentially as mappings from $\mathcal{P} \to [0, 1]$ which satisfy conditions (16). Thus, defining $A(\lambda) := \lambda(A)$ for state λ, it is seen that the quantum logic approach leads to a statistical theory $(S, \mathcal{P})$.

In contrast, the "statistical theory" approach recognizes that quantum mechanics is a statistical theory, and *begins* from this point, i.e., with the existence of a statistical theory $(S, \mathcal{P})$. A natural complementation [definition (5)] and a natural partial ordering [definition (6)] are used to *obtain* a probability and logical structure. In this sense equations (4)-(12) are "free"; they arise from the contemplation of statistical theories in general. To arrive at an orthocomplemented, orthomodular, orthocomplete poset with the states corresponding to generalized probability measures, only three relatively simple conditions [conditions (17)] need be satisfied. These are to be compared with the seven conditions needed in the quantum logic approach (see Appendix C).

A further technical advantage of the "statistical theory" approach is that the quantities $(A \wedge B)(\lambda)$ and $(A \vee B)(\lambda)$ are always well defined, and hence can be manipulated, whereas in the quantum logic approach they can only be defined as greatest lower bounds and least upper bounds, respectively, which need not exist in general.

Finally, in the quantum logic approach there is interest in the case where the set of states is *order-determining*, i.e. where the relations $\lambda(A) \leq \lambda(B)$, $\forall \lambda \in S$, and $A \Rightarrow B$ are equivalent. The order-determining property is guaranteed in the "statistical theory" approach through definition (6).

4. COVERING THEORIES

In Section 3.4 it was shown (Example 3.2) that the Hilbert space structure of quantum mechanics leads to a *nonclassical* statistical theory. As discussed in Section 3.1, such a result may be interpreted as indicating either (1) nonclassical structures are required in general to describe the physical world, or (2) quantum mechanics is an inadequate description of the physical world.

In the latter case one would be led to search for a *classical* statistical theory which describes all the phenomena explained by quantum mechanics. More generally, one may ask whether the predictions of a given statistical theory may all be reproduced by another statistical theory, and what relations there might be between the probability and logical structures of these theories.

Accordingly, a statistical theory $(\bar{S}, \bar{\mathscr{P}})$ is defined to be a *covering* theory for a statistical theory $(S, \mathscr{P})$ if and only if there exist mappings $\alpha: \mathscr{P} \to \bar{\mathscr{P}}$ and $\beta: S \to \bar{S}$ such that

$$A \text{ and } \bar{A} \text{ correspond to the same yes/no experiment} \tag{18a}$$

$$A(\lambda) = \bar{A}(\bar{\lambda}), \qquad \forall \lambda \in S, \quad A \in \mathscr{P} \tag{18b}$$

where $\bar{A} := \alpha(A)$ and $\bar{\lambda} := \beta(\lambda)$. Thus, $(\bar{S}, \bar{\mathscr{P}})$ reproduces all the predictions of $(S, \mathscr{P})$, and in general may yield further predictions also.

It follows from (6) and (18b) that the implication relations $A \Rightarrow B$ and $\bar{A} \Rightarrow \bar{B}$ are not equivalent. In particular, the former relation may hold without the latter being satisfied. Thus, $(S, \mathcal{P})$ and $(\bar{S}, \bar{\mathcal{P}})$ in general have different probability and logical structures. Also, the relation $A \equiv B$ does not imply $\bar{A} \equiv \bar{B}$ in general, i.e., while A and B may be tested by the same experiment for $(S, \mathcal{P})$, they may correspond to distinct experiments for $(\bar{S}, \bar{\mathcal{P}})$. Since the propositions and states of $(S, \mathcal{P})$ are obtained as mappings (α^{-1} and β^{-1}) of corresponding quantities of the covering theory, the latter may in some cases be thought of as forming the basis of or underlying the former, i.e., as a "hidden variable" theory. However, such an interpretation will be appropriate only for certain types of mappings.

Four interesting examples of covering theories are now given to conclude this section. The fourth example demonstrates the existence of a *classical* covering theory for each statistical theory.

Example 4.1. Let $S = \{\lambda_1, \lambda_2\}$ and $\mathcal{P} = \{\mathbf{0}, A, A', \mathbf{1}\}$, where $A(\lambda_1) = 0$ and $A(\lambda_2) = 1/2$. It follows that $A \wedge A' \equiv A \not\equiv \mathbf{0}$, and hence $(S, \mathcal{P})$ does not have a classical structure. However, if $\bar{S} = \{\bar{\lambda}_1, \bar{\lambda}_2, \bar{\lambda}_3\}$ and $\bar{\mathcal{P}} = \{\mathbf{0}, \bar{A}, \bar{A}', \mathbf{1}\}$, with $\bar{A}(\bar{\lambda}_1) = A(\lambda_1)$, $\bar{A}(\bar{\lambda}_2) = A(\lambda_2)$, and $\bar{A}(\bar{\lambda}_3) = 1$, then $(\bar{S}, \bar{\mathcal{P}})$ is a covering theory for $(S, \mathcal{P})$ and is in fact a classical statistical theory.

Example 4.2. For a given statistical theory $(S, \mathcal{P})$, define $(S^c, \mathcal{P})$ by

$$S^c := \left\{ \rho\colon S \to [0, \infty) \;\middle|\; \sum_{\lambda \in S} \rho(\lambda) = 1 \right\}$$

$$A(\rho) := \sum_{\lambda \in S} \rho(\lambda) A(\lambda), \qquad \text{for all} \quad A \in \mathcal{P} \quad \rho \in S^c$$

A state $\rho \in S^c$ can be interpreted as being obtained by choosing a member of an ensemble of systems described by the distribution $\rho(\lambda)$. For $\lambda \in S$, the state $\rho_\lambda \in S^c$ defined by $\rho_\lambda(\lambda') := \delta_{\lambda,\lambda'}$ satisfies $A(\rho_\lambda) = A(\lambda)$ for all $A \in \mathcal{P}$, $\lambda \in S$, and hence $(S^c, \mathcal{P})$ is a covering theory for $(S, \mathcal{P})$. Further, if $\rho_1, \rho_2 \in S^c$, then $a\rho_1 + (1-a)\rho_2 \in S^c$ for $0 \le a \le 1$. Therefore, $(S^c, \mathcal{P})$ may be called the *convex-state* covering theory corresponding to $(S, \mathcal{P})$. One may show that the implication relation $A \Rightarrow B$ is equivalent for both $(S, \mathcal{P})$ and $(S^c, \mathcal{P})$, and hence that their representative logics are identical. For the case of Example 3.2 of Section 3.4, S^c corresponds to the set of density operators on the Hilbert space H, and a state $\mathbf{W} \in S^c$ is interpreted as a mixture of states $\psi \in S$. For the case of Example 3.1 of Section 3.4, S^c is the set of ensembles of classical systems, and includes the equilibrium ensembles of statistical mechanics.

Example 4.3. For a given statistical theory $(S, \mathcal{P})$, define $(S^c, \mathcal{P}^c)$ by

$$\mathcal{P}^c = \left\{ \mu\colon \mathcal{P} \to [0, \infty) \;\middle|\; \sum_{A \in \mathcal{P}} \mu(A) = 1 \right\}$$

S^c as in Example 4.2 above, and $\mu(\rho) := \sum_{A \in \mathcal{P}} \mu(A) A(\rho)$ for all $\mu \in \mathcal{P}^c$,

$\rho \in S^c$. Defining $\mu_A \in \mathscr{P}^c$ by $\mu_A(B) = \delta_{A,B}$ and $\rho_\lambda \in S^c$ as in the previous example, it follows that $\mu_A(\rho_\lambda) = A(\lambda)$, $\forall A \in \mathscr{P}$, $\lambda \in S$. Further, $\mathscr{P}^c$ and S^c are both convex sets, and hence $(S^c, \mathscr{P}^c)$ may be called the *completely convex* covering theory corresponding to $(S, \mathscr{P})$. In general, $(S, \mathscr{P})$ and $(S^c, \mathscr{P}^c)$ have different representative logics. For the case of Example 3.2 of Section 3.4, $\mathscr{P}^c$ corresponds to the set of *effects* encountered in the operation-effect formulation of quantum mechanics (Kraus, 1983).

Example 4.4. Let $(S, \mathscr{P})$ be a statistical theory and consider a set of propositions $\mathscr{P}^d$, the elements of which are indexed by the mappings from S to subsets of the interval $[0, 1)$, i.e.,

$$\mathscr{P}^d := \{\alpha_F \mid F: S \to \{M \mid M \subseteq [0, 1)\}\}$$

The *deterministic theory* $(S \times [0, 1), \mathscr{P}^d)$ may then be defined by

$$\alpha_F(\lambda, x) := \begin{cases} 1, & x \in F(\lambda) \\ 0, & x \notin F(\lambda) \end{cases}$$

for all $\alpha_F \in \mathscr{P}^d$, $\lambda \in S$, and $x \in [0, 1)$. Thus, the state $(\lambda, x) \in S \times [0, 1)$ fully determines the outcome of testing any proposition $\alpha_F \in \mathscr{P}^d$. It is not difficult to show that $(S \times [0, 1), \mathscr{P}^d)$ is *classical.* First, from (5) one may show that $\alpha'_F \equiv \alpha_{F^c} \in \mathscr{P}^d$, where $F^c(\lambda) := [0, 1) \backslash F(\lambda)$. Second, from relations (6) and (8) it follows that the relation $\alpha_F \Rightarrow \alpha_G$ is equivalent to $F(\lambda) \subseteq G(\lambda)$, $\forall \lambda \in S$, and hence that $\alpha_F \wedge \alpha_G \equiv \alpha_{F \cap G} \in \mathscr{P}^d$, where $(F \cap G)(\lambda) := F(\lambda) \cap G(\lambda)$. Third, if $x \in G(\lambda)$, then either $x \in (F \cap G)(\lambda)$ or $x \in (F^c \cap G)(\lambda)$; while if $x \notin G(\lambda)$, then $x \notin (F \cap G)(\lambda)$ and $x \notin (F^c \cap G)(\lambda)$; hence $(\alpha_F \wedge \alpha_G)(\lambda, x) + (\alpha'_F \wedge \alpha_G)(\lambda, x) = \alpha_G(\lambda, x)$ for all $\lambda \in S$, $x \in [0, 1)$. Thus, conditions (15a)-(15c) are satisfied by $(S \times [0, 1), \mathscr{P}^d)$, so that it is indeed a classical theory.

Suppose now that an ensemble of states is constructed from elements of $S \times [0, 1)$, corresponding to some fixed $\lambda \in S$ and a distribution $\rho(x)$ over $[0, 1)$. The probability of a proposition $\alpha_F \in \mathscr{P}^d$ being verified on a member of this ensemble is then

$$\alpha_F(\lambda, \rho) := \int_0^1 \rho(x) \alpha_F(\lambda, x) = \int_{F(\lambda)} \rho(x)\, dx$$

This motivates the definition of the statistical theory $(S^d, \mathscr{P}^d)$, where

$$S^d := S \times [0, 1)^c = S \times \left\{ \rho: [0, 1) \to [0, \infty) \,\middle|\, \int_0^1 \rho(x)\, dx = 1 \right\}$$

and with $\alpha_F(\lambda, \rho)$ as above. Thus, $(S^d, \mathscr{P}^d)$ has a simple ensemble interpretation in terms of the underlying deterministic theory $(S \times [0, 1), \mathscr{P}^d)$. Moreover, the implication relations for these two theories are equivalent [consider the case $\rho_x(x') := \delta_{xx'}$], and it follows that conditions (15a)-(15c) are satisfied by $(S^d, \mathscr{P}^d)$, i.e., the latter is a *classical statistical theory.*

Finally, for proposition $A \in \mathcal{P}$, define a proposition $\alpha_{F_A} \in \mathcal{P}^d$ such that

$$\int_{F_A(\lambda)} dx = A(\lambda)$$

for all $\lambda \in S$, e.g., let $F_A(\lambda) := [0, A(\lambda))$. It follows that for the uniform distribution, $\rho_0(x) \equiv 1$, one has

$$\alpha_{F_A}(\lambda, \rho_0) = \int_0^1 \alpha_{F_A}(\lambda, x)\, dx = \int_{F_A(\lambda)} dx = A(\lambda)$$

Hence, $(S^d, \mathcal{P}^d)$ is a covering theory for $(S, \mathcal{P})$, where $A \in \mathcal{P}$ and $\lambda \in S$ correspond to $\alpha_{F_A} \in \mathcal{P}^d$ and $(\lambda, \rho_0) \in S^d$, respectively. In view of the above, $(S^d, \mathcal{P}^d)$ provides what may be called the *deterministic-picture classical covering theory* for $(S, \mathcal{P})$.

It should be noted that there is a wide range of choice for the mapping $A \to \alpha_{F_A}$; however, this range may be narrowed by imposing certain physical requirements. For example, suppose that a group of propositions $A_1, A_2, A_3, \ldots \in \mathcal{P}$ satisfy $\sum_i A_i(\lambda) = 1$ for all $\lambda \in S$, and correspond to the possible outcomes $r_1, r_2, r_3, \ldots$, of some experiment E (i.e., A_i is verified if and only if r_i is obtained). If it is required that the corresponding propositions $\alpha_{F_1}, \alpha_{F_2}, \alpha_{F_3}, \ldots \in \mathcal{P}^d$ are also tested by experiment E, then it is physically plausible, and consistent with equations (1), to postulate that $\sum_i \alpha_{F_i}(\lambda, \rho) = 1$ for all $(\lambda, \rho) \in S^d$.

The choice $F_i(\lambda) = [0, A_i(\lambda))$ does not satisfy this summation condition; however, a suitable choice *does* exist, given by

$$F_i(\lambda) := \left[\sum_{j=1}^{i-1} A_j(\lambda), \sum_{j=1}^{i} A_j(\lambda)\right)$$

This latter choice generalizes a hidden variable model for quantum mechanics due to Bell (1966). Bell's criticism of his own model, namely that it is very artificial, also applies here—in particular, the mapping $A \to \alpha_{F_A}$ depends crucially on the ordering of the experimental outcomes $r_1, r_2, r_3, \ldots$. It must be emphasized that this criticism in no way affects the significance of the example, i.e., there exists a classical covering theory for each statistical theory.

5. THE BELL INEQUALITIES

5.1. Introduction

The viewpoint that only classical statistical theories are of fundamental interest (see Section 3.1) is supported by Example 4.4, since any nonclassical statistical theory may be replaced by a classical covering theory which describes the same physical phenomena. However, it will now be shown in the case of quantum mechanics that no such covering theory can satisfy a certain locality condition. The demonstration relies upon a derivation of the well-known Bell inequalities (Clauser *et al.*, 1969; Clauser and Horne, 1974; Clauser and Shimony, 1978; Fine, 1982), and is significant in that the locality condition used is of a weaker nature than the standard factorizability condition.

Coincidence experiments are defined and discussed in the next section, and the locality condition is given in Section 5.3. Both "formal" and "physical" versions of the Bell inequalities are then derived in Section 5.4, following on from and clarifying some earlier work (Hall, 1988) which lacked the conceptual basis of statistical theories. Finally, the relevance of these results to quantum mechanics is discussed in Sections 5.5 and 5.6.

5.2. Coincidence Experiments

The concept of a coincidence experiment is formulated here within the context of statistical theories.

For the statistical theory $(S, \mathscr{P})$ consider the case of a system described by state $\lambda \in S$, for which the procedures for *two* yes/no experiments have been carried out, corresponding to propositions $A, B \in \mathscr{P}$. For example, A may refer to a range of position values for a free particle at some time, while B refers to a range of momentum values at a later time. A second example is the simultaneous measurement of two propositions for a quantum mechanical system (Example 3.2, Section 3.4), where the corresponding projections commute.

One may then consider that a single *joint experiment*, labeled by $[A, B]$, has been performed, the result of which lies in the set {(yes, yes), (no, yes), (yes, no), (no, no)}. Clearly $[A, B]$ is not a yes/no experiment, and so cannot be directly discussed within a statistical theory framework. However, one may define the propositions $A.B$, $A'.B$, $A.B'$, and $A'.B'$ as being respectively verified if the result of $[A, B]$ is (yes, yes), (no, yes), (yes, no), and (no, no), and falsified otherwise. If these propositions are elements of $\mathscr{P}$, i.e., probabilities of their verification are predicted by the theory for each state $\lambda \in S$, then $[A, B]$ may be called a joint experiment of the theory.

In general, even if a joint experiment $[A, B]$ exists for $A, B \in \mathscr{P}$, it need not be a joint experiment of the theory. For example, consider the case of the position/momentum measurement mentioned above, where $(S, \mathscr{P})$ is the "quantum mechanical" statistical theory from Section 3.4 (Example 3.2). If the experiment $[A, B]$ is suitably chosen (essentially such that the projection postulate is applicable), then it follows for $\psi \in S$ that $(A.B)(\psi) = (\psi, \mathbf{ABA}\psi)$, where $\mathbf{A}$ and $\mathbf{B}$ are projections onto the appropriate ranges of position and momentum, respectively. Here, since the operator **ABA** is *not* a projection onto a subspace of H, then $A.B \notin \mathcal{P}$, and hence $[A, B]$ is not a joint experiment of $(S, \mathcal{P})$. It may be noted, however, that **ABA** is an *effect* (Kraus, 1983), and hence, for the completely convex covering theory $(S^c, \mathcal{P}^c)$ (Example 4.3, Section 4), $[A, B]$ *is* a joint experiment of the theory. Explicitly, for state $\mathbf{W} \in S^c$, one has

$$(A.B)(\mathbf{W}) = \mathrm{tr}[\mathbf{WABA}], \qquad (A'.B)(\mathbf{W}) = \mathrm{tr}[\mathbf{WA'BA'}], \text{ etc,}$$

where $\mathbf{A}' := \mathbf{1} - \mathbf{A}$.

It is perhaps tempting to identify $A.B$ with the joint proposition $A \wedge B$; however, this turns out to be incorrect in general, as will be seen below. The locality condition given in the next section provides a sufficient criterion for such an identification to be made in the case of a class of *regular* statistical theories.

5.3. Local Statistical Theories

Let $[A, B]$ be a joint experiment of some statistical theory $(S, \mathcal{P})$, and for a proposition $X \in \mathcal{P}$ let $[X]$ denote the corresponding yes/no experiment by which X may be tested. Then, the experiments $[A]$, $[B]$, and $[A, B]$ correspond to three physically distinct cases, even though the latter is a physical combination of the first two. In particular, results of experiments $[A]$ and $[B]$ performed singly may not be simply related to the result of $[A, B]$.

Consider now the situation where experiments $[A]$ and $[B]$ involve, respectively, regions R_{I} and R_{II} of space-time which are nonoverlapping. Predictions of $(S, \mathcal{P})$ may now be compared for N systems described by state λ in two cases: (1) $[A, B]$ is performed for each system; and (2) $[B]$ is performed for each system. In each case, consider the relative frequency of verifications in region R_{II} for proposition B. For sufficiently large N, this is predicted in case 1 to be $(A.B)(\lambda) + (A'.B)(\lambda)$, and in case 2 to be $B(\lambda)$. If these quantities are not equal, then an experiment in region R_{II} can discriminate between cases 1 and 2, i.e., it can be determined within region R_{II} whether or not proposition A was tested in region R_{I}.

As an example, let $[A, B]$ be the position/momentum measurement discussed in the previous section, for the completely convex theory $(S^c, \mathcal{P}^c)$. It follows that the relative frequencies for cases 1 and 2 above are given respectively by

$$(A.B)(\mathbf{W}) + (A'.B)(\mathbf{W}) = \mathrm{tr}[\mathbf{WABA}] + \mathrm{tr}[\mathbf{WA'BA}]$$

and

$$B(\mathbf{W}) = \mathrm{tr}[\mathbf{WB}]$$

Since **A** and **B** are noncommuting projections, these relative frequencies cannot in general be equal. Thus, the statistical behavior of the momentum measurement $[B]$ is influenced by the existence/nonexistence of an earlier position measurement $[A]$.

The above example demonstrates how the past may influence the future. In particular, an experimenter in $R_{\rm I}$ could in principle signal to the future region $R_{\rm II}$ by either performing or not performing experiment $[A]$ for a large number of systems.

The principle of local causality is roughly that signals may *only* be sent from the past to the future. Thus, for Newtonian space-time it implies that no signals can be sent between spatially separated regions with the same time coordinate; while for Einsteinian space-time, no signals can be sent between spacelike-separated regions. The latter has the stronger experimental significance, as in practice all measurements extend over a period of time.

In the light of the above paragraphs, a *local* statistical theory is defined to be one for which, *if* $[A, B]$ *is a joint experiment of the theory such that* $[A]$ *and* $[B]$ *are performed in spacelike-separated regions* $R_{\rm I}$ *and* $R_{\rm II}$ *respectively, then*

$$(A.B)(\lambda)+(A'.B)(\lambda)=B(\lambda) \tag{19a}$$

$$(A.B')(\lambda)+(A'.B')(\lambda)=B'(\lambda) \tag{19b}$$

$$(A.B)(\lambda)+(A.B')(\lambda)=A(\lambda) \tag{19c}$$

$$(A'.B)(\lambda)+(A'.B')(\lambda)=A'(\lambda) \tag{19d}$$

for all states $\lambda \in S$. Equations (19a) and (19b) preclude the sending of a signal from $R_{\rm I}$ to $R_{\rm II}$ via the mechanism discussed above; while equations (19c) and (19d) preclude the sending of such a signal from $R_{\rm II}$ to $R_{\rm I}$.

It follows from relations (6), (9d), and (19) that $A.B \Rightarrow A \wedge B$, $A'.B \Rightarrow A' \wedge B$, etc., for local statistical theories. One may then show (Appendix D) *for local, regular statistical theories satisfying* $A \wedge B \in \mathcal{P}$ *for all* $A, B \in \mathcal{P}$ *that*

$$A.B \equiv A \wedge B, \qquad A.B' \equiv A \wedge B', \qquad A.B' \equiv A \wedge B', \qquad A'.B' \equiv A' \wedge B' \tag{20}$$

in the case where propositions A, B *are to be tested in spacelike-separated regions.*

If the "quantum mechanical" statistical theory of Section 3.4 (Example 3.2) is constrained to be local, equations (19) and (20) then imply that

$$A_{E_1 \cap E_2}(\psi)+A_{E_1^{\perp} \cap E_2}(\psi)=A_{E_2}(\psi), \quad \text{etc.}$$

for all $\psi \in H$, where A_{E_1}, A_{E_2} are to be tested in spacelike-separated regions. It follows from the relation $A_E(\psi)=(\psi, \mathbf{E}\psi)$ that $[\mathbf{E}_1, \mathbf{E}_2] \equiv 0$; i.e., *for* *quantum mechanics to be a local statistical theory, the projections corresponding to propositions tested in spacelike-separated regions must commute.*

5.4. Derivation of Bell Inequalities

Let $(S, \mathscr{P})$ be a classical statistical theory, i.e., conditions (15a)-(15c) are satisfied. One may then show (Appendix E) for all $A_{\rm I}$, $B_{\rm I}$, $A_{\rm II}$, $B_{\rm II} \in \mathscr{P}$ and $\lambda \in S$ that

$$\begin{aligned} -1 \leq (A_{\rm I} \wedge A_{\rm II})(\lambda) + (A_{\rm I} \wedge B_{\rm II})(\lambda) + (B_{\rm I} \wedge B_{\rm II})(\lambda) \\ - (B_{\rm I} \wedge A_{\rm II})(\lambda) - A_{\rm I}(\lambda) - B_{\rm II}(\lambda) \leq 0 \end{aligned} \tag{21}$$

Further, if $(S, \mathscr{P})$ is a *local* classical statistical theory and $[A_{\rm I}, A_{\rm II}]$, $[A_{\rm I}, B_{\rm II}]$, $[B_{\rm I}, A_{\rm II}]$, and $[B_{\rm I}, B_{\rm II}]$ are joint experiments of $(S, \mathscr{P})$ such that the individual components of each are performed in spacelike-separated regions, then (20) and (21) yield

$$\begin{aligned} -1 \leq (A_{\rm I}.A_{\rm II})(\lambda) + (A_{\rm I}.B_{\rm II})(\lambda) + (B_{\rm I}.B_{\rm II})(\lambda) \\ - (B_{\rm I}.A_{\rm II})(\lambda) - A_{\rm I}(\lambda) - B_{\rm II}(\lambda) \leq 0 \end{aligned} \tag{22}$$

for all states $\lambda \in S$.

Condition (21) is equivalent to equation (8) of Fine (1982) [replace $A_{\rm I}$, $B_{\rm I}$, $A_{\rm II}$, and $B_{\rm II}$ by A, A', B, and B', respectively; $(A_{\rm I} \wedge A_{\rm II})(\lambda)$ by $P(AB)$, etc], derived by him as a property of classical distribution functions. Note that this condition connects the predictions of the theory for experiments of the type $[X \wedge Y]$, $[X]$, and $[Y]$, where the first is not necessarily physically related to the last two. In particular, conditions (20) need not hold. This is a consequence of the fact that $X \wedge Y$ is only formally connected to propositions $X, Y \in \mathscr{P}$ in general, via definition (8a). Hence, the contents of (21) may be referred to as the *formal* Bell inequalities. By contrast, the contents of (22) may be called the *physical* Bell inequalities, as this condition connects predictions for physically related experiments of the type $[X, Y]$, $[X]$, and $[Y]$.

The physical Bell inequalities (22) may be compared with the equation immediately preceding equation (4) of Clauser and Horne (1974) [replace $A_{\rm I}$, $B_{\rm I}$, $A_{\rm II}$, and $B_{\rm II}$ by a, a', b, and b' respectively, $(A_{\rm I}.A_{\rm II})(\lambda)$ by $p_{12}(\lambda, a, b)$, $B_{\rm I}(\lambda)$ by $p_1(\lambda, a')$, etc.], or equivalently with equation (3.18) of Clauser and Shimony (1978). The Clauser–Horne result was obtained as a consequence of certain factorizability assumptions, which in the notation of this paper may be stated as

$$(A.B)(\lambda) = A(\lambda)B(\lambda) \tag{23a}$$

$$(A'.B)(\lambda) = A'(\lambda)B(\lambda) \tag{23b}$$

$$(A.B')(\lambda) = A(\lambda)B'(\lambda) \tag{23c}$$

$$(A'.B')(\lambda) = A'(\lambda)B'(\lambda) \tag{23d}$$

for all states $\lambda \in S$, where $[A, B]$ is a joint experiment of the theory such that $[A]$ and $[B]$ are performed in spacelike-separated regions. Equations (23a)-(23d) regarded as a locality condition, express the statistical independence of measurement results in the two regions, *provided the quantities* $(A.B)(\lambda)$, $(A.B')(\lambda)$, *etc., are joint probabilities.* This is a further assumption, valid only if (20) holds. In the present derivation of the physical Bell inequalities, condition (20) is *derived* within the framework of statistical theories. Further, the result relies not on (23), but on (19), where the latter is implied by but does not imply (23) and is therefore a *weaker* formulation of locality.

5.5. The Bell Inequalities and Quantum Mechanics

In this section it is shown that the physical Bell inequalities (22) provide a necessary condition for a statistical theory $(S, \mathscr{P})$ to admit a local, classical covering theory. Since, as is well known, quantum mechanics makes some predictions which violate the physical Bell inequalities, it follows in particular that *there are no local, classical covering theories for quantum mechanics.*

To validate the remarks of the above paragraph, suppose that $(\bar{S}, \bar{\mathscr{P}})$ is a covering theory for some statistical theory $(S, \mathscr{P})$. Hence, for $A, B \in \mathscr{P}$ there exist propositions $\bar{A}, \bar{B} \in \bar{\mathscr{P}}$ such that experiments $[A]$, $[B]$ are identical to $[\bar{A}]$, $[\bar{B}]$ respectively [condition (18a)]. In particular, if $[A]$ and $[B]$ may both be performed for a system described by state $\lambda \in S$, then $[\bar{A}]$ and $[\bar{B}]$ may be performed on the same system and hence the joint experiments $[A, B]$, $[\bar{A}, \bar{B}]$ are also identical. Thus, if $[A, B]$ is a joint experiment *of the theory* $(S, \mathscr{P})$ (see Section 5.2), so that $A.B \in \mathscr{P}$, then conditions (18a), (18b) imply that

$$(A.B)(\lambda) = (\bar{A}.\bar{B})(\bar{\lambda}) = \overline{A.B}(\bar{\lambda}) \tag{24}$$

for all states $\lambda \in S$. Equations (24) imply that if the physical Bell inequalities hold for $(\bar{S}, \bar{\mathscr{P}})$, then they must also hold for $(S, \mathscr{P})$, and the statements of the above paragraph immediately follow.

Note that there is no result analogous to equation (24) which yields a similar significance for the *formal* Bell inequalities (21). Indeed, Example 4.1 of Section 4 demonstrates that the quantities $(\bar{A} \wedge \bar{B})(\bar{\lambda})$ and $\overline{A \wedge B}(\bar{\lambda})$ are not equal in general. Thus the possibility remains of a classical covering theory for quantum mechanics (although such a theory must be nonlocal), and in fact Example 4.4 of Section 4 realizes this possibility.

5.6. Modified Statistical Theories

If quantum mechanics cannot be embedded within the framework of a local, classical statistical theory, a question arises as to whether an acceptable modification of statistical theories exists such that both "local" and "classical" concepts can be retained for quantum mechanics.

Perhaps the simplest such modification to the definition of a statistical theory (Section 2) is to suppose that propositions cannot in general be tested for *all* states in S. Thus, given a state $\lambda \in S$, then only *some* propositions (e.g., spin in certain directions) may be tested. This may be easily interpreted as implying that the experimenter does not have free will in choosing which experiment to conduct.

For such theories, the physical Bell inequalities (22) hold only for those states on which it is possible to perform any one of $[A_{\mathrm{I}}, A_{\mathrm{II}}]$, $[A_{\mathrm{I}}, B_{\mathrm{II}}]$, $[B_{\mathrm{I}}, A_{\mathrm{II}}]$, $[B_{\mathrm{I}}, B_{\mathrm{II}}]$, $[A_{\mathrm{II}}]$, and $[B_{\mathrm{I}}]$. However, there may be no such states! An explicit example of "escaping" the Bell inequalities in this manner has been given by Shimony *et al.* (1985).

6. SUMMARY

The basic tool used in this paper is the concept of a *statistical theory*, as characterized in Section 2. The existence of an associated probability and logical structure for each such theory (Section 3.3) then provides a basis for the definitions of *classical* and *regular* statistical theories (Section 3.4). The former definition identifies the class of theories which have classical probability and logical structures, while the latter definition distinguishes the class of theories regulated by those properties associated with the quantum logic approach to axiomatic quantum mechanics (Section 3.5). The class of regular statistical theories contains the class of classical theories.

The notion of a statistical theory which "explains" the same set of physical phenomena as another theory leads to the concept of a *covering* theory (Section 4). Logical properties are not generally preserved in the transition from a statistical theory to its covering theory, and in particular there exists a classical covering theory for each statistical theory (Example 4.4).

Whereas classical and regular theories are defined in a manner designed to make their probability and logical structure explicit, i.e., in a formal manner, the definition of a *local* statistical theory rests on the physical concept of coincidence experiments (Section 5.2). The latter concept provides a setting for imposing restrictions on signaling between spacelike-separated regions of space-time, leading to conditions for a statistical theory to be local (Section 5.3).

The *formal* and the *physical* Bell inequalities are shown to apply to classical and to local, classical statistical theories, respectively (Section 5.4). Violation of the latter inequalities by quantum mechanics leads to the result that there are no local, classical covering theories for quantum mechanics (Section 5.5).

I conclude that the framework of statistical theories is well suited for discussing certain aspects of the foundations of quantum mechanics. In particular, seemingly disparate features of the quantum logic approach are united in the alternative "statistical theory approach" and the role of locality in the derivation and significance of the Bell inequalities is elucidated by this approach.

APPENDIX A

In this Appendix, relations (9), (10), and (14) of §3.3 are derived from definitions (8), (12), and (13). A lemma to be used in Appendices B and C is also proved.

Relations (9) are derived first. Equation (9a) follows immediately from the symmetry of definition (8a). Also, defining $A \cap B \in \mathscr{P}\mathscr{P}$ by $(A \cap B)(\lambda) = \min\{A(\lambda), B(\lambda)\}$ [see definition (13)], it follows from (6) that

$$(X \Rightarrow A \cap B) \quad \text{iff} \quad (X \Rightarrow A, X \Rightarrow B) \qquad \forall X \in \mathscr{P}\mathscr{P} \tag{A1}$$

Hence, (8a) may be written as

$$\begin{aligned}(A \wedge B)(\lambda) &= \sup\{X(\lambda) | X \in \mathscr{P}, X \Rightarrow A \cap B\} \\ &\leq \sup\{X(\lambda) | X \in \mathscr{P}\mathscr{P}, X \Rightarrow A \cap B\} \\ &= (A \cap B)(\lambda)\end{aligned}$$

and thus $(A \wedge B) \Rightarrow (A \cap B)$. Relations (A1) and (7b) then imply (9c), and further from (A1) it follows that if $X \Rightarrow A \wedge B$, then $X \Rightarrow A$ and $X \Rightarrow B$, for all $X \in \mathscr{P}\mathscr{P}$. But for $X \in \mathscr{P}$ such that $X \Rightarrow A$, $X \Rightarrow B$, one has

$$\begin{aligned}X(\lambda) &\leq \sup\{Y(\lambda) | Y \in \mathscr{P}, Y \Rightarrow A, Y \Rightarrow B\} \\ &= (A \wedge B)(\lambda)\end{aligned}$$

for all $\lambda \in S$, and hence (9d) is proved. Finally, (9d) may be used to give

$$\begin{aligned}((A \wedge B) \wedge C)(\lambda) &= \sup\{X(\lambda) | X \in \mathscr{P}, X \Rightarrow A \wedge B, X \Rightarrow C\} \\ &= \sup\{X(\lambda) | X \in \mathscr{P}, X \Rightarrow A, X \Rightarrow B, X \Rightarrow C\} \\ &= \sup\{X(\lambda) | X \in \mathscr{P}, X \Rightarrow A, X \Rightarrow B \wedge C\} \\ &= (A \wedge (B \wedge C))(\lambda)\end{aligned}$$

thus proving equation (9b). This last result also leads to the consistency of definitions (8a) and (12a).

Relations (10) may be proved in an entirely analogous manner to the above, using definition (8b).

Relation (14) may be derived as follows. Suppose that $X' \in \mathcal{P}$, $\forall X \in \mathcal{P}$. Then, using (12), (13), and (7d), one has

$$\begin{aligned}
\Big(\bigwedge_i A_i'\Big)'(\lambda) &= 1 - \Big(\bigwedge_i A_i'\Big)(\lambda) \\
&= 1 - \sup\Big\{X(\lambda) | X \in \mathcal{P}, X \Rightarrow \bigcap_i A_i'\Big\} \\
&= \inf\Big\{1 - X(\lambda) | X \in \mathcal{P}, X \Rightarrow \bigcap_i A_i'\Big\} \\
&= \inf\Big\{X'(\lambda) | X \in \mathcal{P}, X \Rightarrow \bigcap_i A_i'\Big\} \\
&= \inf\Big\{X'(\lambda) | X \in \mathcal{P}, \Big(\bigcap_i A_i'\Big)' \Rightarrow X'\Big\}
\end{aligned}$$

But

$$\begin{aligned}
\Big(\bigcap_i A_i'\Big)'(\lambda) &= 1 - \Big(\bigcap_i A_i'\Big)(\lambda) \\
&= 1 - \inf\{1 - A_i(\lambda)\} \\
&= \sup\{A_i(\lambda)\} \\
&= \bigcup_i A_i(\lambda)
\end{aligned}$$

i.e., $(\bigcap_i A_i')' \equiv \bigcup_i A_i$. Also, if $X \in \mathcal{P}$, then $X' \in \mathcal{P}$, and vice versa. Hence,

$$\begin{aligned}
\Big(\bigwedge_i A_i'\Big)'(\lambda) &= \inf\Big\{X'(\lambda) | X' \in \mathcal{P}, \bigcup_i A_i \Rightarrow X'\Big\} \\
&= \inf\Big\{X(\lambda) | X \in \mathcal{P}, \bigcup_i A_i \Rightarrow X\Big\} \\
&= \Big(\bigvee_i A_i\Big)(\lambda)
\end{aligned}$$

Finally, a useful lemma will be proved.

Lemma. Suppose $(S, \mathcal{P})$ satisfies the following conditions. (i) If $A \Rightarrow B'$, then $(A' \wedge B')' \in \mathcal{P}$, and (ii) if $A \Rightarrow B'$, then $(A' \wedge B')'(\lambda) = A(\lambda) + B(\lambda)$; for all $A, B \in \mathscr{P}$, $\lambda \in S$. Then, for any countable sequence $A_1, A_2, A_3, \ldots \in \mathscr{P}$ such that $A_i \Rightarrow A_j'$, $\forall i \neq j$, one has

$$\left(\bigwedge_i A_i'\right)'(\lambda) = \sum_i A_i(\lambda), \qquad \forall \lambda \in S$$

Proof. First note that for a sequence $A, B \in \mathscr{P}$ such that $A \Rightarrow B'$, the result follows by condition (ii) of the lemma. Now consider a finite sequence $A_1, \ldots, A_n \in \mathscr{P}$, $n \geq 2$, satisfying $A_i \Rightarrow A_j'$, $\forall i \neq j$; and assume that $(\bigwedge_{i=1}^{n-1} A_i')' \in \mathscr{P}$, and

$$\left(\bigwedge_{i=1}^{n-1} A_i'\right)'(\lambda) = \sum_{i=1}^{n-1} A_i(\lambda), \qquad \forall \lambda \in S$$

Then $A_n \Rightarrow A_i'$ for $i = 1, 2, \ldots, n-1$, and hence

$$A_n(\lambda) \leq \min\{A_1'(\lambda), \ldots, A_{n-1}'(\lambda)\}, \qquad \forall \lambda \in S$$

i.e., $A_n \Rightarrow \bigcap_{i=1}^{n-1} A_i'$, from (6) and (13), and so $A_n \Rightarrow \bigwedge_{i=1}^{n-1} A_i'$ from (12a). Hence, from condition (i) of the lemma,

$$\left(\bigwedge_{i=1}^{n} A_i'\right)' = \left(A_n' \wedge \left(\bigwedge_{i=1}^{n-1} A_i'\right)\right)' \in \mathscr{P}$$

and from condition (ii)

$$\left(\bigwedge_{i=1}^{n} A_i'\right)'(\lambda) = A_n(\lambda) + \left(\bigwedge_{i=1}^{n-1} A_i'\right)'(\lambda)$$

$$= \sum_{i=1}^{n} A_n(\lambda), \qquad \forall \lambda \in S$$

Thus, the lemma has been proved inductively for finite sequences.

Consider, finally, the case of a countably infinite sequence, and define

$$T_n(\lambda) := \left(\bigwedge_{i=1}^{n} A_i'\right)'(\lambda)$$

Then the limit of the sequence $\{T_n(\lambda)\}$ as $n \to \infty$ exists from (12a). But it has been shown that $T_n(\lambda) = \sum_{i=1}^{n} A_i(\lambda)$ for all finite values of n, and hence in the limit $n \to \infty$ this must also hold, i.e.,

$$\left(\bigwedge_{i=1}^{\infty} A_i'\right)'(\lambda) = \sum_{i=1}^{\infty} A_i(\lambda)$$

and the lemma is proved. ■

APPENDIX B

It is proved in this Appendix that a statistical theory is classical if and only if (1) the representative logic is Boolean, and (2) the mappings

$m_\lambda: \mathscr{P}\mathscr{P} \to [0, 1]$ given by $m_\lambda(A) := A(\lambda)$ are generalized probability measures for all $\lambda \in S$.

Now, for a partially complemented poset $(\mathscr{P}, \Rightarrow, ')$ to be a Boolean logic (Gudder, 1979; Beltrametti and Cassinelli, 1981), one requires $\forall A, B, C \in \mathscr{P}$ that

$$A \wedge B \in \mathscr{P}, \qquad A \vee B \in \mathscr{P} \tag{B1}$$

$$A \wedge (B \vee C) \equiv (A \wedge B) \vee (A \wedge C) \tag{B2}$$

$$A \vee (B \wedge C) \equiv (A \vee B) \wedge (A \vee C) \tag{B3}$$

$$A' \in \mathscr{P} \tag{B4}$$

$$\text{if } A \Rightarrow B, \qquad \text{then } B' \Rightarrow A' \tag{B5}$$

$$A \wedge A' \equiv \boldsymbol{0}, \qquad A \vee A' \equiv \boldsymbol{1} \tag{B6}$$

Conditions (B1)-(B3) ensure the poset is a distributive lattice, and conditions (B4)-(B6) ensure that the partial complementation is an orthocomplementation. Further, if condition (2) is to be satisfied, then from (16) one has, for each $\lambda \in S$,

$$\text{if } A_1, A_2, \ldots \in \mathscr{P} \text{ satisfy } A_i \Rightarrow A_j' \quad \forall i \neq j,$$

$$\text{then } \left(\bigvee_i A_i \right)(\lambda) = \sum_i A_i(\lambda) \tag{B7}$$

$$\boldsymbol{1}(\lambda) = 1 \tag{B8}$$

It will now be shown that (B1)-(B8) are equivalent to conditions (15a)-(15c) of the text.

First, suppose that (B1)-(B8) are satisfied. Then, (15a) and (15b) follow from (B4) and (B1), respectively. Further, (9c) and (10c) imply that $A \wedge B \Rightarrow A \Rightarrow A \vee B'$, i.e., $A \wedge B \Rightarrow A \vee B' \equiv (A' \wedge B)'$, using (B4) and (14). Then, using (B7), (B3), (B6), and (11), we have

$$\begin{aligned}(A \wedge B)(\lambda) + (A' \wedge B)(\lambda) &= ((A \wedge B) \vee (A' \wedge B)(\lambda) \\ &= ((A \vee A') \wedge B)(\lambda) \\ &= (\boldsymbol{1} \wedge B)(\lambda) \\ &= B(\lambda)\end{aligned}$$

for all $\lambda \in S$, and hence (15c) is obtained.

It remains to be shown that (B1)-(B8) follow from (15a)-(15c). But (B5) and (B8) hold automatically, from (7d) and (4), respectively. Hence, only (B1)-(B4), (B6), and (B7) must be checked.

Now, (B4) follows immediately from (15a), and then (14), (15a) and (15b) imply (B1). Also, if $B \equiv A$ is substituted into (15c), then (11) implies that $(A \wedge A')(\lambda) = 0$, $\forall \lambda \in S$, and hence that $A \vee A' \equiv (A \wedge A')' \equiv \boldsymbol{0}' \equiv 1$, i.e., (B6) holds. Further, (15a) and (15b) imply $(A' \wedge B')' \in \mathscr{P}$ for all $A, B \in \mathscr{P}$, and from (15c),

$$\begin{aligned}(A' \wedge B')'(\lambda) &= 1 - (A' \wedge B')(\lambda) \\ &= 1 - (B'(\lambda) - (A \wedge B')(\lambda) \\ &= B(\lambda) + (A \wedge B')(\lambda)\end{aligned}$$

Thus, if $A \Rightarrow B'$, then $A \wedge B' \equiv A$, and so $(A' \wedge B')'(\lambda) \equiv A(\lambda) + B(\lambda)$. It follows that the lemma proved in Appendix A is applicable, i.e., if $A_1, A_2, \ldots \in \mathscr{P}$ satisfy $A_i \Rightarrow A_j'$, $\forall i \neq j$, then

$$\left(\bigwedge_i A_i'\right)'(\lambda) = \sum_i A_i(\lambda), \qquad \forall \lambda \in S$$

But (15a) and (14) imply that $(\bigwedge_i A_i')'(\lambda) = \bigvee_i A_i(\lambda)$, and hence (B7) holds.

To prove (B2) and (B3), first note that repeated use of (15c) gives

$$\begin{aligned}(A' \wedge B' \wedge C')(\lambda) &= (A' \wedge B')(\lambda) - (A' \wedge B' \wedge C)(\lambda) \\ &= A'(\lambda) - (A' \wedge B)(\lambda) - (A' \wedge C)(\lambda) + (A' \wedge B \wedge C)(\lambda) \\ &= 1 - A(\lambda) - B(\lambda) + (A \wedge B)(\lambda) - C(\lambda) + (A \wedge C)(\lambda) \\ &\quad + (B \wedge C)(\lambda) - (A \wedge B \wedge C)(\lambda)\end{aligned}$$

Then, (15a) and (14) yield

$$\begin{aligned}(A \vee B \vee C)(\lambda) &= A(\lambda) + B(\lambda) + C(\lambda) - (A \wedge B)(\lambda) - (B \wedge C)(\lambda) \\ &\quad - (C \wedge A)(\lambda) + (A \wedge B \wedge C)(\lambda)\end{aligned} \tag{B9}$$

Substituting $C \equiv \boldsymbol{0}$ in (B9), one obtains

$$(A \vee B)(\lambda) = A(\lambda) + B(\lambda) - (A \wedge B)(\lambda)$$

This result together with (B9) then gives

$$\begin{aligned}(A \wedge (B \vee C))(\lambda) &= A(\lambda) + (B \vee C)(\lambda) - (A \vee B \vee C)(\lambda) \\ &= A(\lambda) + B(\lambda) + C(\lambda) - (B \wedge C)(\lambda) \\ &\quad - (A \vee B \vee C)(\lambda) \\ &= (A \wedge B)(\lambda) + (A \wedge C)(\lambda) - (A \wedge B \wedge C)(\lambda) \\ &= ((A \wedge B) \vee (A \wedge C))(\lambda)\end{aligned}$$

thus proving (B2). Finally,

$$\begin{aligned} A \vee (B \wedge C) &\equiv (A' \wedge (B \wedge C)')' \\ &\equiv (A' \wedge (B' \vee C'))' \\ &\equiv ((A' \wedge B') \vee (A' \wedge C'))' \\ &\equiv (A' \wedge B')' \wedge (A' \wedge C')' \\ &\equiv (A \vee B) \wedge (A \vee C) \end{aligned}$$

and so (B3) also holds.

APPENDIX C

It is proved in this Appendix that a statistical theory is regular if and only if (1) the representative logic is an orthocomplemented, orthocomplete, and orthomodular poset, and (2) the mappings $m_\lambda: \mathscr{P}\mathscr{P} \to [0, 1]$ given by $m_\lambda(A) := A(\lambda)$ are generalized probability measures for all $\lambda \in S$.

Now, for the partially complemented poset $(\mathscr{P}, \Rightarrow, ')$ to satisfy condition (1), it is required for all A, B, $C \in \mathscr{P}$ that

$$\text{if } A \Rightarrow B, \qquad \text{then } A \vee B' \in \mathscr{P} \tag{C1}$$

$$\text{if } A \Rightarrow B, \qquad \text{then } A \vee (A' \wedge B) \equiv B \tag{C2}$$

$$A' \in \mathscr{P} \tag{C3}$$

$$\text{if } A \Rightarrow B, \qquad \text{then } B' \Rightarrow A' \tag{C4}$$

$$A \wedge A' \equiv \mathbf{0}, \qquad A \vee A' \equiv \mathbf{1} \tag{C5}$$

Conditions (C1) and (C2) ensure orthocompleteness and orthomodularity, respectively, while (C3)-(C5) imply that the partial complementation is an orthocomplementation (Gudder, 1979; Beltrametti and Cassinelli, 1981). Further, for condition (2) above to be satisfied, then from (16) it follows that for all $\lambda \in S$ one must have

$$\text{if } A_1, A_2, \ldots \in \mathscr{P} \text{ satisfy } A_i \Rightarrow A_j' \quad \forall i \neq j,$$

$$\text{then } \left(\bigvee_i A_i \right)(\lambda) = \sum_i A_i(\lambda) \tag{C6}$$

$$\mathbf{1}(\lambda) = 1 \tag{C7}$$

It will now be shown that (C1)-(C7) are equivalent to conditions (17a)-(17c) of the text.

First, suppose that (C1)-(C7) are satisfied by $(S, \mathscr{P})$. Hence, (17a) and (17b) follow from (C1), (C3), and (14). Also, if $A \Rightarrow B$, then $A \wedge B \equiv A$ from (9d), and $(A' \wedge B)'(\lambda) = (A \vee B')(\lambda) = A(\lambda) + B'(\lambda)$ from (C6). Hence, (17c) holds.

It remains to be shown that (C1)-(C7) follow from (17a)-(17c). Note that (C4) and (C7) hold automatically, from (7d) and (4), respectively, and hence only (C1)-(C3), (C5), and (C6) must be checked.

Now, (C3) follows immediately from (17a), and (C1) is implied by (14), (17a), and (17b). From the relation $A \Rightarrow A$, it follows from (17c) that $(A \wedge A')(\lambda) = 0$, $\forall \lambda \in S$, and thus $(A \vee A') \equiv (A \wedge A')' \equiv \mathit{1}$, i.e., (C5) holds. Also, if $A \Rightarrow B$, then $A \wedge B \equiv A$, so from (17c), $(A' \wedge B)(\lambda) = B(\lambda) - A(\lambda)$. But $A' \wedge B \Rightarrow A'$, and hence if $A \Rightarrow B$, then

$$\begin{aligned}(A' \wedge (A' \wedge B)')(\lambda) &= A'(\lambda) - (A' \wedge B)(\lambda) \\ &= A'(\lambda) - (B(\lambda) - A(\lambda)) \\ &= B'(\lambda)\end{aligned}$$

i.e., $(A \vee (A' \wedge B)) \equiv (A' \wedge (A' \wedge B)')' \equiv B$, and so (C2) holds.

Finally, (17a), (17b) and (17c), and the lemma proved in Appendix A imply for $A_1, A_2, \ldots \in \mathscr{P}$ satisfying $A_i \Rightarrow A_j'$, $\forall i \neq j$, that

$$\left(\bigvee_i A_i\right)(\lambda) = \left(\bigwedge_i A_i'\right)'(\lambda) = \sum_i A_i(\lambda)$$

and thus (C6) holds.

APPENDIX D

In this Appendix it is first shown that if $(S, \mathscr{P})$ is a regular statistical theory such that $A \wedge B \in \mathscr{P}$ for all $A, B \in \mathscr{P}$ (i.e., the representative logic is a lattice), then

$$(A \wedge B)(\lambda) + (A' \wedge B)(\lambda) \leq B(\lambda) \tag{D1}$$

for all $A, B \in \mathscr{P}$, $\lambda \in S$. Equation (20) of the text is then obtained in the case where $(S, \mathscr{P})$ is also local, as a consequence of (D1) and (19).

To prove (D1), let $C :\equiv A \wedge B$; $D :\equiv A' \wedge B$. Then $C, D \in \mathscr{P}$ by assumption, and $C \Rightarrow A \Rightarrow A \vee B' \equiv D'$, where (9c), (10c), (14), and (17a) have been used. Thus, from (7b), $C \Rightarrow D'$, and hence by (C6) of Appendix C, $(C \vee D)(\lambda) = C(\lambda) + D(\lambda)$ for all $\lambda \in S$. Finally, $C \Rightarrow B$, $D \Rightarrow B$, and so, from (6) and (10d), it follows that $(C \vee D)(\lambda) \leq B(\lambda)$ for all $\lambda \in S$, and (D1) is proved.

Suppose now that $(S, \mathscr{P})$ is local, and $[A, B]$ is a joint experiment of the theory, where $[A]$ and $[B]$ are to be tested in spacelike-separated regions.

From (6), (19a) and (19c), one then has $A.B \Rightarrow A$, $A.B \Rightarrow B$, and hence $A.B \Rightarrow A \wedge B$ from (9d). Similarly, it can be shown that $A'.B \Rightarrow A' \wedge B$, etc. Defining

$$m_\lambda := (A \wedge B)(\lambda) - (A.B)(\lambda), \qquad n_\lambda := (A' \wedge B)(\lambda) - (A'.B)(\lambda)$$

it follows from the above that m_λ, $n_\lambda \geq 0$ for all $\lambda \in S$. Thus,

$$\begin{aligned}(A \wedge B)(\lambda) + (A' \wedge B)(\lambda) &= (A.B)(\lambda) + (A'.B)(\lambda) + m_\lambda + n_\lambda \\ &= B(\lambda) + m_\lambda + n_\lambda\end{aligned}$$

using (19a). It follows from (D1) that $m_\lambda = n_\lambda = 0$ for all $\lambda \in S$, and hence $A \wedge B \equiv A.B$, $A' \wedge B \equiv A'.B$. In an analogous manner, it can also be shown that $A \wedge B' \equiv A.B'$ and $A' \wedge B' \equiv A'.B'$, proving (20) of the text.

APPENDIX E

In this Appendix, equations (21) of the text are derived for classical statistical theories. The proof is based on the derivation by Fine (1982) of an analogous result for classical distribution functions (see Section 5.4).

First, from (15a)-(15c), one has for propositions $A, B, C, D \in \mathcal{P}$, $\lambda \in S$, that

$$\begin{aligned}0 \leq (A \wedge B' \wedge C')(\lambda) &= (A \wedge B')(\lambda) - (A \wedge B' \wedge C)(\lambda) \\ &= A(\lambda) - (A \wedge B)(\lambda) - (A \wedge C)(\lambda) + (A \wedge B \wedge C)(\lambda)\end{aligned}$$

i.e., using (6) and (9c) also,

$$\begin{aligned}(A \wedge B)(\lambda) + (A \wedge C)(\lambda) - A(\lambda) &\leq (A \wedge B \wedge C)(\lambda) \\ &= (A \wedge B \wedge C \wedge D)(\lambda) + (A \wedge B \wedge C \wedge D')(\lambda) \\ &\leq (B \wedge D)(\lambda) + (C \wedge D')(\lambda) \\ &= (B \wedge D)(\lambda) + C(\lambda) - (C \wedge D)(\lambda)\end{aligned}$$

and hence

$$0 \geq (A \wedge B)(\lambda) + (A \wedge C)(\lambda) + (C \wedge D)(\lambda) - (B \wedge D)(\lambda) - A(\lambda) - C(\lambda) \tag{E1}$$

Replacing A by A' and swapping B and C in (E1), one obtains

$$\begin{aligned}0 &\geq (A' \wedge C)(\lambda) + (A' \wedge B)(\lambda) + (B \wedge D)(\lambda) - (C \wedge D)(\lambda) - A'(\lambda) - B(\lambda) \\ &= C(\lambda) - (A \wedge C)(\lambda) + B(\lambda) - (A \wedge B)(\lambda) + (B \wedge D)(\lambda) \\ &\quad - (C \wedge D)(\lambda) - 1 + A(\lambda) - B(\lambda)\end{aligned}$$

i.e.,

$$(A \wedge B)(\lambda)+(A \wedge C)(\lambda)+(C \wedge D)(\lambda)-(B \wedge D)(\lambda)-A(\lambda)-C(\lambda) \geq -1 \tag{E2}$$

Inequalities (E1) and (E2) are just the content of equations (21) of the text, where the propositions A, B, C, and D are identified with A_{I}, A_{II}, B_{II}, and B_{I}, respectively.

Note that the derivation of (E1) and (E2) relies only on the properties

$$A' \in \mathcal{P} \tag{E3}$$

$$(A \wedge B \wedge C \wedge D)(\lambda)+(A \wedge B \wedge C \wedge D')(\lambda)=(A \wedge B \wedge C)(\lambda) \tag{E4}$$

for all A, B, C, $D \in \mathcal{P}$, $\lambda \in S$. Conditions (E3) and (E4) are weaker than conditions (15a)-(15c), and hence the formal Bell inequalities (21) in fact hold for a larger class of statistical theories than the class of classical statistical theories.

ACKNOWLEDGMENT

It is a pleasure to thank Kailash Kumar for useful discussions.

REFERENCES

Accardi, L. (1984). The probabilistic roots of the quantum mechanical paradoxes, in *The Wave-Particle Dualism*, S. Diner, D. Fargue, G. Lockhak, and F. Selleri, eds., p. 297, Reidel, Dordrecht.

Bell, J. S. (1966). *Reviews of Modern Physics*, **38**, 447, § V.

Beltrametti, E. G., and Cassinelli, G. (1981). *The Logic of Quantum Mechanics*, Addison-Wesley, Reading, Massachusetts.

Birkhoff, G., and von Neumann, J. (1936). *Annals of Mathematics*, **37**, 823.

Clauser, J. F., and Horne, M. A. (1974). *Physical Review D*, **10**, 526.

Clauser, J. F., and Shimony, A. (1978). *Reports on Progress in Physics*, **41**, 1881.

Clauser, J. F., Horne, M. A., Shimony, A., and Holt, R. A. (1969). *Physical Review Letters*, **23**, 880.

Fine, A. (1982). *Physical Review Letters*, **48**, 291.

Gudder, S. P. (1979). *Stochastic Methods in Quantum Mechanics*, North-Holland, New York, Chapter 3.

Hall, M. J. W. (1988). *Foundations of Physics*, to be published.

Kolmogorov, A. (1950). *Foundations of Probability*, Chelsea, New York.

Kraus, K. (1983). States, effects and operations, in *Lecture Notes in Physics*, Vol. 190, Springer-Verlag, Berlin.

Pitowsky, I. (1986). *Journal of Mathematical Physics*, **27**, 1556.

Primas, H. (1981). Chemistry, quantum mechanics and reductionism, in *Lecture Notes in Chemistry*, Vol. 14, Springer-Verlag, Berlin.

Shimony, A., Horne, M. A., and Clauser, J. F. (1985). *Dialectica*, **39**, 97.