Quantum probability and the foundations of quantum theory

Luigi Accardi
Centro Matematico V.Volterra
Dipartimento di di Matematica
Universitá di Roma II

Indice

1 Introduction

The point of view advocated, in the last years, by quantum probability about the foundations of quantum mechanics, is based on the investigation of the mathematical consequences of a deep and elementary idea developed by the founding fathers of quantum mechanics and accepted nowadays as a truism by most physicists, namely: *one should be careful when applying the rules derived from the experience of macroscopic physics to experiments which are mutually incompatible in the sense of quantum mechanics.*This statement,

like Dirichlet pidgeonhole principle, looks quite trivial at first sight but some of its consequences are far from evident. It might therefore happen that one accepts the above statement as self-evident but quivers at some other statements which are necessary, if not so evident, corollaries of it.In order to

concretize the general premise above, let us come to the point and sum up, without proofs the main conclusions of the quantum probabilistic approach to the foundations of quantum theory.
(1.) The basic empirical data, against which we check the predictions of

our theories, are conditional probabilities. The conditional probability of an event A given an event C is a model independent notion which can be experimentally estimated by preparing an ensenble of systems for which the event C is realized and measuring on it the relative frequency of A. Since the preparation C preceeds in time the measurement of A, the events C and A can be incompatible in the sense of quantum theory. Given these transition probabilities, all the other relevant quantities: mean values, fluctuations, correlations, ... , can be obtained by application of standard matfhematical formulae.
(2.) If $A, B, C, ...$ are events and the transition probabilities $P(A \mid B), P(B \mid C), P(C \mid A)$, ... are estimated in mutually incompatible experiments, then the unrestricted application of the rules of classical probability to these statistical data is unwarranted. The existence of a single classical probabilistic model for these data (which is equivalent to the possibility of unrestricted application of these data of the rules of classical probability) cannot be postulated a priori, but must be proved by means of a mathematical theorem. The mathematical technique to prove theorems of this kind consists in evaluating the *Statistical Invariant*, which guarantee the existence (or non existence) of the model uniquely in terms of the experimental data $P(A|B)$,...

(3.) All the existing "proofs" of the statement that, if a system is in a superposition state with respect to an observable A, then the observable A does not actually assume any of its values, are based on the uncritical application of the rules of the classical probabilistic model to statistical data obtained from mutually incompatible experiments. As such they are wrong. Therefore there is no rational, nor mathematical, nor experimental, nor theoretical support for the statement "in a superposition state with respect to the observable A, the observable A cannot actually assume any of its values". Once this statement is rejected, there is no ground to claim that terms like "collapse of the wave packet" refer to an actual change of the physical state of the system.

(4.) If one accepts that, in a superposition state for the observable A, A cannot actually assume any of its values, and that only the act of measurements "collapses" the physical state of the system so that one and only one of the values A is assumed, then one cannot escape paradoxes of EPR type (a measurement here collapses a particle there) or of the type of the 2-slit experiment (an electron not looked at is spread throughout the available space).

(5.) Conversely, if one proves that the statement: "in a superposition state with respect to A, the observable A cannot actually assume any of its values" is unwarranted, then all the so called paradoxes of quantum theory, constructed by means of variations on the theme "collapse of the wave packet" are swept away.

Quantum probability has provided such a proof.

(6.) The fact that quantum mechanics introduced a new kind of probability calculus was recognized since the early days of the theory. The new elements introduced by quantum probability in this debate are:

(i) the individuation of the single axiom of classical probability theory whose breakdown, in the quantum domain, can be proved uniquely in terms of the experimental data.

(ii) the proof of the fact that a consistent physical interpretation of the new probability calculus can be developed entirely within the conceptual framework of classical physics. There is no need at all to introduce the notion of superposition as a new physical state of matter (unobservable in principle). Contrarily to what stated by Feymann, Heisenberg, Dyson, and many others..., no contradiction between theory and experiments can be deduced from the statement that neutrons not looked at pass through one and only one slit, that spins not looked at assume one and only one of their possible

values, etc...
(7.) An equivalent formulation of the statement (ii) of point (6.) above is: the breakdown of the applicability of classical probability theory to the description of some microscopic phenomena does not imply the necessity of postulating a non classical behaviour of the microsystems in absence of observations of them. In fact, even for macroscopic systems, the applicability of the classical probabilistic rules cannot be postulated a priori, but should be checked through the comparison of the experimental data with the statistical invariants. The existence of macroscopic physical systems, to which the applicability of the classical laws of probability fails, has been investigated by D.Aerts who has obtained interesting partial results in this direction.
(8.) Quantum probability gives a mathematical proof of the fact that no contradiction with the experiments can be deduced from the statement "electrons not looked at behave exactly as they do when they are looked at" (in particular, they pass through one and only one slit, their spin assumes one and only one value,...). With that it allows to realize the criterium stated by Heisenberg, Bohr,... according to which a good physical theory should not commit itself on situations which are in principle unobservable (e.g. the statement that a neutron not looked at cannot pass through one and only one slit). Any interpretation which accepts that, in a superposition state with respect to the observable A, this observable cannot actually assume any of its values, violates this criterium.
(9.) The only conclusion one can draw from the 2-slit experiment is that a set of 3 statistical data (conditional probabilities), obtained in 3 mutually incompatible experiments, cannot be described within a single classical probabilistic model. There is no need to introduce a new physical state, in which the position observable has no definite value in order to explain the apparent contradiction arising from this experiment. Exactly the same conclusion applies to the experiments related to the Bell inequalities. Also in that case one has three different, mutually incompatible experiments which produce statistical data that cannot be described within a single kolmogorovian model. Bell' s statistical data concern correlations and a composite system; the 2-slit experiment concerns conditional probabilities and individual particles. The connection between the two situations escaped Bell' s analysis and he was led, like his predecessors and exactly for the same reasons (the implicit postulate that, if the usual physical properties such as locality, objective reality , separability, ... , were true, then the laws of classical probability should be applicable) to believe that a physical property (locality in his

case) was responsible of the apparent paradox. It is precisely the validity of this implicit postulate that the quantum probabilistic analysis disproves by means of mathematical and experimental arguments.

One should give Bell the credit of having, with his analysis, brought back to measurable and observable statements the debate on the foundations of quantum theory, which was sinking into purely formal and mathematical problems. However the limits of his analysis consist in not having recognized that, from the scientific point of view, the famous Bell' s inequalities contained nothing new with respect to the old 2-slit experiment.

Quantum probability vindicates the merit of having individuated precisely in this point the root of all the problems concerning the interpretation of quantum mechanics and, more generally, to have realized that no new unplausible physical properties (such as nonlocality, non reality, ...) need to be introduced to explain the difficulties arising in the interpretation of the experimental data arising in quantum theory, because these are all related to the unwarranted application of a single axiom of probability theory: the Bayes axiom (which in fact, in the usual texts on probability is not even introduced as an axiom, but as the definition of conditional probability). Why this axiom is not a tautology on relative frequencies (as implicitly believed for several centuries) and which deep and far from obvious assumptions it hides, even independently on the experimental results of quantum theory, is explained in [3].

Acknowledgments The author wants to thank Rudolph Kalman for several remarks on the topics dealt with in the present paper . He also acknowledges support from Grants AFOSR 870249, ONR N00014-86-K-0538, DAAL03-89-0018 through the Center for Mathematical System Theory, University of Florida.

2 Probabilistic Models

When speaking of the probability of an event A we assume that a set of conditions C , under which this probability is evaluated, have been fixed . This probability will be denoted $P(A \mid C)$ and called the conditional probability of A given C. Sometimes it is useful to interpret C as the preparation of an experiment done to verify the occurrence of A . The number $P(A \mid C)$ is a model independent quantity: it has a meaning independently on how it has

been evaluated (by means of relative frequencies, bets, opinions, coherence properties, ...) and independently on the mathematical model one uses to describe the set of events and their probabilities. Sometimes the notion of time ordering will be made explicit by writing $P(A_t \mid C_s)$ where s and t ($s < t$) are real numbers interpreted respectively as time of occurrence of the preparation C and of the event A .

The basic heuristic models, both for events and preparations, that we shall have in mind in the following are given by statements of the type: - at a given time the values of a given set of observable lie in a given set of numerical intervals. Thus, for example, the expression $P(A \in I \mid C \in J)$ will denote the probability that the values of the observable A lie in interval I at time t when measured on a system prepared so that at time $s < t$ the observable C lied in the interval J. When J consists of a single point, say J = c , the conditional probability

$$P(A \in I \mid C = c)$$

will also be called a **transition probability**. We assume that the probabilities $P(A \in I \mid C \in J)$ have a meaning whenever I and J are arbitrary Borel subsets of the real line. If A is an observable and $I \subseteq \mathbf{R}$ is an interval, the symbol $[A \in I]$ denotes the event that the measurement of A yields a result in I. These events have a natural structure of a Boolean σ-algebra, induced by the Borel σ-algebra of $\mathbf{R}$. In the following our starting point will be a quadruple of the form:

$$\left\{ T \ , \ \{A(x)\}_{x \in T} \ , \ \mathcal{F}_o \ , \ \{P(A(x) \in I \mid C)\} \right\} \tag{1}$$

where

- T is a set

- $\{A(x) \ : \ x \in T\}$ is a family of observables

- $\mathcal{F}_o$ is a set, called the set of preparation conditions

- For each $x \in T$, $C \in \mathcal{F}_o$ and for each interval $I \subseteq R$ the number

$$P(A(x) \in I \mid C) \tag{2}$$

is the conditional probability of the event $[A(x) \in I]$ given the preparation condition C. All these data have to be considered as **model independent** , for example they might be experimental data. We are going to describe several different mathematical models for these model independent data and the basic problem we are going to discuss is the following:

Can one discriminate among the different matematical models, describing the data (1) , uniquely in terms of the data themselves ?

By a mathematical model for the data (1) , we mean a correspondence which associates to each event $[A(x) \in I]$ (or observable A(x)) and to each preparation condition C a mathematical object in such a way that the model independent probabilities $P(A(x) \in I \mid C)$ are described by some formulas relating among themselves the mathematical objects associated to the events (or the observables) and the preparations. Definition (3) in the following will make this notion precise. Two such models are called **stochastically equivalent** if they describe the same set of conditional probabilities. Note that this notion of stochastic equivalence is very weak : it only requires that the two models describe the same experimental data -while their mathematical structures can be quite different . For example any set of three 2×2 bistochastic matrices which admits a Kolmogorovian model admits also a complex Hilbert space model (cf. [2]). Hence , for the description of three 2×2 bistochastic matrices, the two models are stochastically equivalent in the sense specified above , even if they are far from being isomorphic as mathematical structures. **Remark (1)**. When the condition C and the observable A(x) are fixed the map

$$P(A(x) \in I \mid C) = P_{A(x)}(I \mid C) \qquad ; \; I \subseteq \mathbf{R}$$

is a classical probability measure whose numerical characteristics (mean, variance, momenta, characteristic function, etc. ...) can be defined in the usual way. In particular one defines the mean , or expectation value, of A(x) given C

$$E(A(x) \mid C) = \int_{\mathbf{R}} \lambda dP_{A(x)}(\lambda \mid C)$$

the momenta of order k of A(x) given C

$$E(A(x)^k \mid C) = \int_{\mathbf{R}} \lambda^k dP_{A(x)}(\lambda \mid C)$$

the variance of A(x) given C

$$Var(A(x) \mid C) = E(A(x)^2 \mid C) - E(A(x) \mid C)^2$$

and the characteristic function of A(x) given C

$$\int_{\mathbf{R}} e^{i\lambda p} dP_{A(x)}(\lambda \mid C)$$

Thus, for a single observable (or more generally for a set of compatible observables) and a single conditioning all these models are reduced to the usual one . Only in presence of several observables or several conditionings the nonclassical statistical features arise.

Definition 1 *In the above notations, an observable $A(x)(x \in T)$ is called* **discrete** *if there exists a finite or countable subset $S(x) \subseteq R$ such that, for every $C \in \mathcal{F}_o$ one has*

$$\sum_{a \in S(x)} P(A(x) = a \mid C) = 1 \tag{3}$$

If moreover, for each value $a \in S(x)$, there exists a preparation $C \in \mathcal{F}_o$ such that

$$P(A(x) = a \mid C) > 0 \tag{4}$$

then the set S(x) is called the set of values of the discrete observable A(x).

We shall consider as primitive the notion of observable. Heuristically an observable is defined by a set of measurement procedures and by the set of its values. Usually the values of an observable are real numbers, but sometimes it is convenient to consider a set of (compatible) observables as a single vector valued observable. Two observables A(x), A(y) are called **compatible** if they can be simultaneusly measured. This is a model independent, and therefore vaguely defined, notion. In the class of models of von Neumann type (cf. the the following Definition (3)) , this property will correspond to the fact that the algebras $\mathcal{A}(x)$ and $\mathcal{A}(y)$, generated by A(x) and A(y), commute. An observable compatible with any other observable is called **universally compatible** or also a **superselection observable** . An observable A(x) is called a (measurable, continuous,...) **function** of a set

$$\{A(y) \ , \ y \in F \subseteq T\}$$

of compatible observables if there exists a (measurable, continuous,...) function

$$f : \mathbf{R}^F \longrightarrow \mathbf{R}$$

such that for every measurement of A(x) with result a there is a simultaneous measurement of all the $A(y)$ with result $a(y)$ $(y \in F)$ such that $a = f(\{a(y)\}_{y \in F})$. An observable A(x) is called **maximal** if another observable B can be compatible with A(x) if and only if B is expressible as a funtion of A(x) and of the universally compatible observables (i.e. the result of each measurement of B can be obtained from a simultaneous measurement of A(x) and of a set of universally compatible observables).

Definition 2 *The preparation C is called* **pure** *with respect to the family of observables $\{A(x) \ : \ x \in T\}$ if there exists a discrete maximal observable B and a value b of B such that for all $x \in T$ and all intervals $I \subseteq \mathbf{R}$ one has:*

$$P(A(x) \in I \mid B = b) = P(A(x) \in I \mid C) \tag{5}$$

A preparation pure with respect to all the observables of a given theory is called **pure**.

All the probabilistic models we are going to consider fit into the abstract scheme described by the following definition :

Definition 3 *A probabilistic model for the quadruple (1) is defined by a triple:*

$$\Big\{\mathcal{A} \ , \ \{\varphi_C \ : \ C \in \mathcal{F}_o\} \ , \ \{e_{A(x)}(\ \cdot\ )\}\Big\} \tag{6}$$

where

- *$\mathcal{A}$ is a $*$-algebra over the complex numbers.*
- *For each preparation condition $C \in \mathcal{F}_o$, φ_C is a state on $\mathcal{A}$*
- *For each $x \in T$*

$$I \subseteq \mathbf{R} \mapsto e_{A(x)}(I) \in \mathcal{A} \tag{7}$$

is an $\mathcal{A}$-valued measure on $\mathbf{R}$

and where, for all x in T, C in $\mathcal{F}_o$ and $I \subseteq \mathbf{R}$ one has

$$P(A(x) \in I \mid C) = \varphi_C(e_{A(x)}(I)) \tag{8}$$

The $e_{A(x)}(I)$ correspond, in the mathematical model, to the events $[A(x) \in I]$. If they are orthogonal projections of $\mathcal{A}$, one speaks of a **von Neumann**

model *; if they are positive elements of $\mathcal{A}$, one speaks of a* **Ludwig type model** *. In both these type of models the $\mathcal{A}$-valued measure (82) is supposed to be countably additive.*

In case of models of von Neumann type, given $x \in T$, the commutative $*$-algebra generated by the family of all $e_{A(x)}(I)$ with $I \subseteq \mathbf{R}$ is called the algebra generated by the observable A(x) and denoted $\mathcal{A}(x)$. For probabilistic models not included in Definition 1 above. As we see in this scheme

the roles of the preparations and of the events are quite asymmetric. In several important applications however the conditions of the problem imply a complete symmetry between the notions of preparation and event in the sense that every preparation is a non null event and every non null event is a preparation . To deal with these situations the abstract model discussed above has to be particularized into a finer structure.

3 The standard quantum model

In the same conditions as in Section (11.)

Definition 4 *In the class of models defined by Definition (10.3) the* **standard quantum model** *for the conditional probabilities (10.2) is characterized by:*

$$\mathcal{A} = \mathcal{B}(H) \tag{9}$$

$$e_{A(x)}(I) = \chi_I(\bar{A}(x)) \tag{10}$$

$$P(A(x) \in I \mid C) = \varphi_C(e_{A(x)}(I)) = Tr(W_C \cdot e_{A(x)}(I)) \tag{11}$$

where

i) H is a complex separable Hilbert space with scalar product denoted $<$. , $>$ and antilinear in the first variable and $\mathcal{B}(H)$ denotes the algebra of all the bounded operators on H.

ii) For each $x \in T$, $\bar{A}(x)$ is a self-adjoint operator on H and

$$\bar{A}(x) = \int_R \lambda de_{A(x)}(\lambda) \tag{12}$$

is the spectral decomposition of $\bar{A}(x)$.

iii) Tr denotes the trace on H, i.e.

$$Tr(X) = \sum < e_i,\ Xe_i > \qquad ; \qquad (e_i) \text{ an o.n. basis of H} \qquad (13)$$

W_C *is a* **density matrix** *on H, i.e. a positive trace class operator satisfying* $Tr(W_C) = 1$.

In the sequel we shall use the following equivalent notations :

$$Tr(W_C \cdot e_{A(x)}(I)) = E(\chi_I(A(x)) \mid C) = E(\chi_I(A(x)) \mid W_C) = \varphi_C(e_{A(x)}(I))$$

where $\varphi_C(\ \cdot\ ) = Tr(W_C \cdot\ )$. The operator $\bar{A}(x)$ associated to the observable

A(x) is called itself an observable and denoted with the same symbol unless confusion can arise . In the literature on the mathematical foundations of quantum mechanics it is frequently postulated that there exists a one-to-one correspondence between physical observables and self-adjoint operators . In practice the self-adjoint operators to which one can associate a well defined measurement procedure are very few and conversely there are observable quantities (time, temperature,...) which do not correspond to any self-adjoint operator in the usual (nonrelativistic) quantum models. Since the triple

$$\{H\ ,\ \{\bar{A}(x)\}_{x\in T}\ ,\ W_C\}$$

uniquely determines the model, in the following we shall frequently call such a triple a standard quantum model for the probabilities (1).

Definition 5 *A* **pure quantum model** *for the conditional probabilities (3) is a standard quantum model* $\{H\ ,\ \{\bar{A}(x)\}_{x\in T}\ ,\ W_C\}$ *whose density operator has rank one , i.e. such that there exists a unit vector* $\psi_C \in H$ *satisfying*

$$W_C = \mid \psi_C >< \psi_C \mid = P(\psi_C)$$

where $\mid \psi_C >< \psi_C \mid = P(\psi_C)$ *denotes the rank one projection along the direction of* ψ_C *i.e. , the linear operator on H defined by*

$$P_\psi\ :\ \phi \in \mathcal{H} \longrightarrow < \psi, \phi > \psi = P_\psi(\phi) \in \mathcal{H} \qquad ; \qquad \psi \in \mathcal{H},\ \parallel \psi \parallel = 1 \quad (14)$$

Remark (3). We will sometimes use Dirac's notation, according to which a vector ψ in H is denoted by $\mid \psi >$ (and called a **ket vector**); the corresponding linear functional on H ($\phi \in H \rightarrow < \psi, \phi >$) is denoted $< \psi \mid$ (and

called a **bra vector**) , i.e. the bra-vectors are elements of the conjugate Hilbert space $\bar{H}$ of H . In these notations the rank one projection P_ψ defined by (76), is also denoted $\mid \psi >< \psi \mid$. Note that, for a pure conditioning, one has

$$P(A(x) \in I \mid C) =< \psi_C, e_{A(x)}(I) \cdot \psi_C >=< \psi_C, \chi_I(\bar{A}(x)) \cdot \psi_C > \qquad (15)$$

Remark (4). In terms of partial Boolean algebras these models can be described as follows:

$\mathcal{F}$ is the partial Boolean algebra of othogonal projectors on the Hilbert space H. $\mathcal{F}_o$ is the set of density matrices on H , in the case of a standard quantum model, and the set of all rank one projectors on H in the case of a pure quantum model. All the events considered in Quantum Theory can be reduced to the statement that the results of the measurements of certain observables are number lying in certain intervals of the real line, whose width is determined by the precision of the measurement. In the mathematical model an observable is represented by a selfadjoint operator A acting on H and the event that the result of the measure of A is in the interval I by the spectral projector $e_A(I)$. In particular, if I contains a single simple eigenvalue of the operator A, say λ, corresponding to the eigenvector ϕ in the sense that:

$$A\phi = \lambda\phi \qquad (16)$$

then

$$e_A(I) =\mid \phi >< \phi \mid= \ P^A_\phi \qquad (17)$$

If A is an observable satisfying (83), (84) for some interval I, B is any observable and J is any interval of $\mathbf{R}$, then the probability

$$Pr(B \in J \mid A = \lambda) = P(e_B(J) \mid P^A_\phi) = Tr(e_B(J) \cdot P^A_\phi) =< \phi, e_B(J) \cdot \phi > \qquad (18)$$

is interpreted as the conditional probability that the value of B is in the interval J if the system is prepared so that the value of A is λ. If moreover

$$e_B(J) = P^B_\psi \qquad ; \qquad B\psi = \mu\psi \ ; \ \mu \in \mathbf{R}$$

then the conditional probability

$$Prob(B = \mu \mid A = \lambda) = tr(P^B_\psi \cdot P^A_\phi) =\mid< \psi, \phi >\mid^2= \qquad (19)$$

is called a **transition probability**. Notice that, in view of (86), the quantum transition probabilities satisfy the symmetry condition

$$Prob(B = \mu \mid A = \lambda) = Prob(A = \lambda \mid B = \mu) \tag{20}$$

which holds whenever λ and μ are simple discrete eigenvalues of A and B respectively. **Remark (5)**. When considered as elements of $\mathcal{F}$, rank one

projectors on H are called **pure** or **atomic events** ; when considered as elements of $\mathcal{F}_o$ they are called **pure preparations** or, more frequently, **pure states**. Rank one projectors lay in the intersection between the preparations and the events and , since they are in one-to-one correspondence with the one-dimensional sub-spaces of H, another way of looking at pure states is as rays in H, i.e. elements of the projective space over H (the set of equivalence classes in $H - \{0\}$ for the relation $\phi \sim \psi \Longleftrightarrow \exists \lambda \in \mathbf{C} : \phi = \lambda\psi$).
Often, by abuse of language, one uses the term **state** for a density matrix W_C or, in case $W_C =\mid \psi_C >< \psi_C \mid$, for the vector ψ_C itself. In the former case one speaks of a **mixed state** or a **mixture** , in the latter of a **pure** or **vector state** or, if H is realized as the complex L^2 space over some measure space, of a **wave function**. The term **pure** refers here to the fact that the state φ_C , corresponding to a rank one density matrix, is extremal in the convex set of all the states on the algebra $\mathcal{B}(H)$. This terminology agrees with the more general one to be introduced in Section (5).

4 Transitions among discrete observables: the Kolmogorov and the Renyi models

The notations and the assumptions are the same as in Section (2). Let S be a finite or countable subset of the natural integers. Assume that, for each $x \in T$ the observable A(x) is discrete, and let $\{a_n(x) : n \in S\}$ denote its set of values. We choose the set of preparations conditions to coincide with the set of all the events of the form $[A(x) = a_n(x)]$ for some $x \in T$ and $n \in S$. These events will be called **atomic events** .

Definition 6 *A* **Kolmogorovian model** *for the transition probabilities*

$$P\Big(A(y) = a_n(y) \mid A(x) = a_m(x)\Big) \qquad , \qquad m, n \in S \qquad ; \qquad a, y \in T \tag{21}$$

is a pair:

$$\{(\Omega\ ,\ \mathcal{F}\ ,\ \mu)\ ,\ \{A_n(x)\}\ (n \in S)\ (x \in T)\} \tag{22}$$

where

i)$(\Omega\ ,\ \mathcal{F}\ ,\ \mu)$ *is a probability space.*

ii) For each $x \in T$, $A_n(x) : n \in S$ *is a a measurable partition of* Ω *and for each* $x, y \in T \quad ; \quad m, n \in S$ *one has*

$$\mu(A_n(x)) > 0 \tag{23}$$

$$P(A(y) = a_n(y) \mid A(x) = a_m(x)) = \frac{\mu(A_n(y) \cap A_m(x))}{\mu(A_m(x))} \tag{24}$$

The definition of the Renyi model for the transition probabilities (1) differs from that of Kolmogorov model only in that it is not required that $\mu(\Omega) = 1$ but μ can be an arbitrary positive σ- finite measure. In particular the basic identity (3) holds in both models. In both models the observable A(x) is represented by the function $\bar{A}(x) : \Omega \longrightarrow \{a_1(x), ..., a_n(x), ...\}$ taking on $A_n(x)$ the constant value $a_n(x)$.

5 Transitions among discrete maximal observables: the pure quantum model

For discrete observables the notion of maximality can be introduced in a purely probabilistic way. We denote $P(B \mid A(y) = a_n(y); A(x) = a_m(x))$ the probability of an event B under the condition that at a certain moment the measurement of the observable A(x) has given the result $a_m(x)$ and at a subsequent time the measurement of A(y) has given the result $a_n(y)$ (in both cases the times can be included in the variables x ,y and omitted from the relations).

Definition 7 *The discrete observables* $\{A(x)\}\ ,\ x \in T\}$ *are called* **maximal** *if for each event* B *and for all* $x, y \in T\ ,\ m, n \in S$ *one has :*

$$P\Big(B \mid A(y) = a_n(y); A(x) = a_m(x)\Big) = P\Big(B \mid A(y) = a_n(y)\Big) \tag{25}$$

Condition (76) means that the information contained in the knowledge that the event $[A(x) = a_n(x)]$ *has happened can be changed, but not increased , by subsequent measurements . The existence of different sets of maximal observables is the qualitative essence of the* **Heisenberg indeterminacy principle** .

According to quantum theory, the transition probabilities (3.11) , when referred to the values of a maximal family $\{A(x)\ (x \in T)\}$ are described by the following model :

Definition 8 *A* **pure quantum model** *, or simply a* **complex Hilbert space model** *in the following , for the transition probabilities (3.11) is a pair:*

$$\{H,\ \{\psi_n(x)\ :\ n \in S\ x \in T\}\} \tag{26}$$

where

i) H is a complex Hilbert space of dimension equal to the cardinality of S (i.e. the number of different values of the A(x)).

ii) for each $x \in T$ *the vectors* $(\psi_n(x) : n \in N)$ *form an orthonormal basis of H and for each* $x, y \in T$ *and* $m, n \in S$ *one has*

$$P(A(y) = a_n(y) \mid A(x) = a_m(x)) = |< \psi_n(y), \psi_m(x) >|^2 \tag{27}$$

If H is a real Hilbert space satisfying (i) , then we speak of a **real Hilbert space model** .

Remark (3). Notice that (1) implies that a necessary condition for the family of transition probabilities (3.11) to be described by a quantum model for discrete maximal observables A(x) is that the symmetry condition

$$P(A(y) = a_n(y) \mid A(x) = a_m(x)) = P(A(x) = a_m(x) \mid A(y) = a_n(y)) \tag{28}$$

be satisfied for each $x, y \in T$ and $m, n \in S$. **Remark (4)**. The maximality

of the discrete observable A(x) is expressed, in the quantum model, by the fact that the correspondence between the values $a_n(x)$ of A(x) and the vectors $\psi_n(x)$ of the orthonormal basis is one to one. The statement : at time t the system is in the pure state $\psi_n(x)$ means a (idealized, instantaneous) measurement of the observable A(x) at time t has given the result $a_n(x)$. Notice that in the expression (1) of the probabilities the actual values of the

observables A(x) play no role: only the orthonormal basis $(\psi_n(x))$ is used. In this model one associates to the discrete observable A(x) the self-adjoint operator

$$\bar{A}(x) = \sum_n a_n(x) \mid \psi_n(x) >< \psi_n(x) \mid$$

and , due to the identity

$$|< \phi, \psi >|^2 = Tr(\mid \phi >< \phi \mid \cdot \mid \psi >< \psi \mid)$$

The following result shows that a pure quantum model for the transition probabilities $p_{m,n}(x,y)$ can be equivalently defined in terms of **transition amplitudes** .

Proposition 1 *Every pure quantum model* $\{H, \{\psi_n(x) : n \in S_{x\in T}\}$*for the transition probabilities (3) defines a triple*

$$\{H \ , \ \{U(x,y)\}_{x,y\in T} \ , \ (e_n)\} \tag{29}$$

such that

i) H is an Hilbert space

ii) (e_n) *is an orthonormal basis of H*

iii) For each $x, y \in T$ *, U(x,y) is a unitary operator on H satisfying:*

$$U(x,y)^{-1} = U(x,y)^* = U(y,x) \qquad ; \qquad U(x,x) = 1 \tag{30}$$

$$U(y,z)U(x,y) = U(x,z) \qquad z \in T \tag{31}$$

iv) For each $x, y \in T$ *and* $m, n \in S$ *one has:*

$$P(A(y) = a_n(y) \mid A(x) = a_m(x)) = |< e_n, U(x,y)e_m >|^2 = \mid U_{n,m}(x,y) \mid^2 \tag{32}$$

(where, by definition, $< e_n, U(x,y)e_m >= U_{n,m}(x,y)$ *. Conversely any triple (80) defines a pure quantum model for the transition probabilities (3) via the correspondence*

$$\psi_n(x) = U(x_o, y)e_n \qquad , \qquad n \in S \ , \ x_o \in T - \text{fixed arbitrarily} \tag{33}$$

Proof. For each $x, y \in T$ denote $U(x,y) : H \longrightarrow H$ the unitary operator characterized by

$$U(x,y)\psi_n(x) = \psi_n(y) \qquad ; \qquad n \in N \tag{34}$$

Then, by definition for any $x, y \in T$ (81) and (82) hold. Moreover, if $x_o \in T$ is an arbitrary fixed element of T, then denoting

$$e_n = \psi_n(x_o) \tag{35}$$

and using the unitarity of $U(x_o, y)$ and (82) one finds

$$|< e_n, U(x,y)e_m >|^2 = |< U(x_o,y)e_n, U(x_o,y)U(y,x)e_m >|^2 =$$

$$= |< \psi_n(y), U(x_o,x)e_m >|^2 = |< \psi_n(y), \psi_m(x) >|^2 \tag{36}$$

Thus (83) follows from (1). It is clear from the construction that the triple (80), with U(x,y) defined by (85) and e_n by (86), satisfies (81), (82), (83) and uniquely determines the quantum model

$$\{H \ ; \ \psi_n(x) \ ; \ n \in S \qquad , \ x \in T\}$$

Conversely, given any such a triple one can define, for each $x \in T$, the orthonormal basis

$$\psi_n(x) = U(x_o, x)e_j$$

where $x_o \in T$ is fixed once for all and independent on T. Since clearly (1) holds by assumption it follows that the resulting pair is a pure quantum model for the transition probabilities (3) in the sense of Definition 77.

Definition 9 *A family* $\{U(x,y)_{x,y \in T}\}$ *of unitary operators on a Hilbert space H satisfying conditions (81), (82) is called a family of* **transition amplitudes** *on T. The matrix elements*

$$U^{\psi}_{n,m}(x,y) = < e_n, U(x,y)e_m > = < \psi_n(x), \psi_m(y) > \tag{37}$$

are called the **transition amplitudes between the states** $\psi_n(x)$ *and* $\psi_n(y)$.

Remark (7). Note that (82) implies that

$$U^{\psi}_{n,m}(x,z) = \sum_k U^{\psi}_{n,k}(x,y)U^{\psi}_{k,m}(y,z) \tag{38}$$

which, by analogy with the classical theorem of composite (or total) probabilities will be called **the theorem of composite (or total) amplitudes** . Taking the square modulus of both sides of (89) with the simplifying notation

$$P(A(y) = a_n(y) \mid A(x) = a_m(x)) = p_{n,m}(y,x) = p_{m,n}(x,y) = \mid U_{n,m}(x,y) \mid^2 \tag{39}$$

one finds

$$p_{n,m}(x,z) = \sum_k p_{n,k}(x,y)p_{k,m}(y,z) + \sum_{k \neq k'} \bar{U}^{\psi}_{n,k}(x,y)\bar{U}^{\psi}_{k,m}(y,z)U^{\psi}_{n,k'}(x,y)U^{\psi}_{m,k'}(y,z) \tag{40}$$

On the other hand, repeating the construction of Remark (**??**) with the identifications:

$$A = \mid \psi_n(y) >< \psi_n(y) \mid \qquad ; \qquad \psi_j = \psi_j(z) \qquad ; \qquad \phi = \psi_m(x)$$

one finds:

$$E(A \mid \phi) = E(A \mid W_\phi) = p_{n,m}(x,z) - \sum_k p_{n,k}(x,y)p_{k,m}(y,z) =$$

$$= \sum_{k \neq k'} \bar{U}^{\psi}_{n,k}(x,y)\bar{U}^{\psi}_{k,m}(y,z)U^{\psi}_{n,k'}(x,y)U^{\psi}_{m,k'}(y,z) \tag{41}$$

The terms (58) are called the **interference terms** . Since in a classical model satisfying (76), one should have :

$$\sum_k p_{n,k}(x,y)p_{k,m}(y,z) =$$

$$= \sum_k P(A(x) = a_n(x) \mid A(z) = a_k(z)) \cdot P(A(z) = a_k(z) \mid A(y) = a_m(y)) =$$

$$= \sum_k P(A(x) = a_n(x) \mid A(z) = a_k(z); A(y) = a_m(y)) \cdot P(A(z) = a_k(z) \mid A(y) = a_m(y)) =$$

$$= P(A(x) = a_n(x) \mid A(y) = a_m(y))$$

it follows that:

(i) in the quantum model for the transition probabilities among discrete maximal observables the theorem of composite probabilities does not hold.

(ii) the interference terms (58) are a quantitative measure of the deviation from the validity of the theorem of the composite probabilities. Since both formulas (76) and (83) can be experimentally verified, they provide a simple experimental test for the existence of a Kolmogorovian model for the transition probabilities (3).

6 Transition probabilities, evolutions, energy

In this Section a purely probabilistic approach to the notion of "energy" in quantum theory is given and it is shown that, even in the most elementary quantum mechanical context in which this notion appears there are some physically unjustified assumptions in the way this notion is introduced. The alternative mathematical models of quantum theory which do not introduce this assumption are worth investigation both from the mathematical and the physical point of view.

In the notations of the previous Section, let us now choose T to be the real line whose elements we interpret as time. In this case the observable A(t) ($t \in T$) can be interpreted as the time t translate of a given fixed observable A. We can assume that the set of possible values of A(t) does not depend on t. Let us denote $(a_n)_{n\in S}$ this set of values. In the theory of classical Markov chains the transition probability matrices P(s,t) ($s \leq t$) satisfy the evolution equation

$$P(r,t) = P(s,t)P(r,s) \qquad r < s < t \tag{42}$$

called the Chapman-Kolmogorov equation. The analogue of this evolution equation for the transition amplitudes is the Schrödinger equation. Let

$\{H\ ,\ \{U(s,t)\}_{s,t\in R}\ ,\ (e_n)\}$ be a pure quantum model for the transition probabilities (3) in the sense of Proposition (5.5), then the equations (81), (82) imply that

$$U(s,t) = U(t,s)^{-1} \qquad ; \qquad U(t,t) = 1 \tag{43}$$

$$U(s,t)U(r,s) = U(r,t) \qquad r < s < t \tag{44}$$

Because of (77) it is sufficient to restrict one' s attention to the U(s,t) with $s < t$.

Definition 10 *A 2-parameter family U(s,t) ($s \leq t$) of unitary operators satisfying (77) and (1) will be called a* **left unitary evolution** *. If condition (1) is replaced by*

$$U(r,s)U(s,t) = U(r,t) \qquad r < s < t \tag{45}$$

then we speak of a **right unitary evolution**. *For future reference recall that, in probabilistic language, a family of random variables U(s,t) , satisfying (88) (resp. (89)), is called a left (resp. right) multiplicative functional on* $\mathbf{R}$*. Unless stated otherwise , by unitary evolution we shall mean right unitary evolution.*

Lemma 1 *Let D be a dense subspace of H such that for each $\xi \in D$ and for any $n \in \mathbf{R}$ the map $t \in [s, +\infty) \mapsto U(s,t)\xi$ is strongly differentiable with derivative $\partial_t U(s,t)\xi$. Then:*

(i) For each $s < t$ there exists an operator $H_s(t)$ symmetric on U(s,t)D such that

$$\partial_t U(s,t)\xi = iH_s(t)U(s,t)\xi \tag{46}$$

(ii) For each $r < s < t$ one has

$$H_r(t) = U(r,s)H_s(t)U(r,s)^* \tag{47}$$

on U(r,t)D.

Proof. (i) follows from (1) and differentiation of the (constant) map

$$t \mapsto \frac{d}{dt} < U(s,t)\xi, U(s,t)\eta > \qquad , \qquad \xi, \eta \in D$$

Since the U(s,t) are unitary operators, the existence of $D_t U(s,t)\xi$ in the strong sense implies the existence of $D_t U(r,t)\xi$ for any $r \leq s \leq t$ and the identity

$$\partial_t U(r,t)\xi = \partial_t U(s,t) \cdot U(s,t)\xi$$

hence, for each $\eta \in U(r,t)D$

$$H_r(t)\eta = U(r,s)H_s(t)U(r,s)^*\eta$$

and this proves (ii) . We shall refer to the result of Lemma 77 above by

saying that the right evolution U(s,t) satisfies on D the equation

$$\frac{d}{dt}U(s,t) = iH_s(t)U(s,t) \qquad ; \qquad U(s,s) = 1 \tag{48}$$

Equation (82) is called the **Schrödinger equation with time dependent hamiltonian** $H_s(t)$ or, according to the interpretation of U(s,t), the **Schrödinger equation in the interaction representation** (cf. Section ( )) .
The condition of stationarity of the transition probabilities

$$P\Big(A(t+r) = a_n \mid A(s+r) = a_m\Big) = P(A(t) = a_n \mid A(s) = a_m) \tag{49}$$

$(s < t \ , \ r \in R \ , \ m, n \in S)$ becomes in view of (83):

$$|< e_n, U^A(s+r, t+r)e_m >|^2 = |< e_n, U^A(s,t)e_m >|^2 \tag{50}$$

where $U^A(s,t))$ is the transition amplitude associated to the pair H , $(\psi_n(A,t))$ according to Proposition (5.5). It is a difficult problem to find the conditions under which the validity of (84), for all $s < t$, $r \in R$, $m, n \in S$ and for sufficiently many orthonormal bases (e_n), implies that, up to a phase

$$U^A(s+r, t+r) = U^A(s,t) \tag{51}$$

Assuming however the validity of (84) for all $s < t$ and $r \in R$, and defining

$$U_t^A = U^A(0,t) \tag{52}$$

(1) implies that one has that (U_t^A) is a 1-parameter unitary group on H. Thus, if (U_t^A) is strongly continuous, then there exists a self-adjoint operator h^A on H such that

$$\frac{d}{dt}U_t^A = ih^A U_t^A \tag{53}$$

conversely, via Stone's theorem, any self-adjoint operator h^A on H defines a 1-parameter unitary group U_t^A on H satisfying the equation (87). **Remark (3)**. Notice that the transition amplitudes $U^A(s,t)$, associated to the pure quantum model for the transition probabilities of the observables A(t) depend, a priori, on the observable A and its time translate. If we have two pure quantum models

$$\{H \ , \ \{\psi_n(A,t) \ ; \ n \in S \ , \ t \in R\} \qquad ; \qquad \{H \ , \ \{\psi_n(B,t) \ ; \ n \in S \ , \ t \in R\} \tag{54}$$

relative to two discrete maximal observables A, B and their time translates, then the corresponding transition amplitudes, defined by

$$U^X(s,t)\psi_n(X,s) = \psi_n(X,t) \tag{55}$$

$(X = A, B\ ,\ s < t\ \ n \in S)$ are related by

$$V_{AB}(t)^{-1}U^B(s,t)V_{AB}(s) = U^A(s,t) \tag{56}$$

where $V_{AB}(t)$ is the unitary operator on H defined by:

$$V_{AB}(t)\psi_n(A,t) = \psi_n(B,t) \qquad n \in S \tag{57}$$

One of the basic, and not always explicitly stated, postulates of quantum theory is that the transition amplitude $U^A(s,t)$ is in fact independent on the discrete maximal observables A. Maybe it is possible, in a relativistic and field theoretical context, to deduce the evolution postulate below from considerations of invariance with respect to a given symmetry group, but up to now this problem does not seem to have been investigated.

Evolution Postulate. Denote $\mathcal{O}_o$ the family of all (discrete) maximal observables at time O and let $T = \mathcal{O}_o \times R$. The pair $(A_o, t) \in T$ will be denoted A_t . Then

P1) For each t the set of values of A_t coincides with the set of values of A_o . The common index set for the values of each A_t will be denoted S.

P2) Given a pure quantum model $\{H, \{\psi_n(A,t)\ ,\ (n \in S\ ,\ (A,t) \in T)\}\ \}$ there exists a unitary evolution $U(s,t)$ on H such that for any $s < t$, any observable $A \in \mathcal{O}_o$ and any observable B -not necessarily discrete - with spectral measure $E^B(.)$, the conditional probability that $[B_t \in I]$ ($I \subseteq R$ - a Borel set) given that $A_s = a_n$ is given by :

$$P(B_t \in I \mid A_s = a_n) =< U(s,t)^*\psi_n(A,s), E^B(I)U(s,t)^*\psi_n(A,s) > \tag{58}$$

Notice that, in case also the observable B is discrete and the set I is reduced to a single point b_n , the identity (58) becomes

$$P(B_t = b_n \mid A_s = a_n) =|< U(s,t)\psi_n(B,s), \psi_n(A,s) >|^2 \tag{59}$$

It is instructive, to understand the role of the evolution postulate, to compare this formula with (83) and (87). Recalling that in our case $T = \mathcal{O}_o \times \mathbf{R}$ according to (83) one should have

$$P\Big(B,t) = (b_n,t) \mid (A,s) = (a_m,s)\Big) =|< \psi_n(A,s), U\Big((A,s),(B,t)\Big)\psi_m(A,s) >|^2 \tag{60}$$

where we have put in (72) $x_o = (A, s)$. Thus the basic content of the Evolution Postulate in quantum theory is that, independently on the observables A,B one has

$$U((A,s),(B,t)) = U(s,t) \tag{61}$$

It is easy to build examples in which (61) is not satisfied. It is however not known if, or under which additional conditions, these constraints imply the identity (61).

Definition 11 *Assuming the evolution postulate and the time-homogeneity condition*

$$U(s+r,t+r) = U(s,t) \tag{62}$$

Then 1-parameter family

$$U_t = U(0,t) \tag{63}$$

is a strongly continuous unitary group whose generator H is called the **Hamiltonian** *of the system and is a self-adjoint operator characterized by the property :*

$$U_t = \exp itH \tag{64}$$

Remark (5). The probabilistic meaning of the Hamiltonian is illustrated by the following considerations: according to (59) and under the time homogeneity condition (62), the probability that a discrete observable A at time zero takes the value a_m and at time t the value a_n , is given by:

$$P(A_t = a_n \mid A_o = a_m) = |< \psi_n, U_t \psi_m >|^2 \tag{65}$$

where we have put

$$\psi_n = \psi_n(A,0) \tag{66}$$

Assuming that for each m the function $t \mapsto U_t\psi_m$ is twice continuously weakly differentiable and expanding the right hand side of (63) up to the second order included , one obtains, if $m \neq n$

$$P(A_t = a_n \mid A_o = a_m) = t^2 \mid< \psi_n, H\psi_m >|^2 + O_{m,n}(t^3) \tag{67}$$

and if m = n

$$P(A_t = a_m \mid A_o = a_m) = 1 - t^2 Var(H \mid \psi_m) + O_m(t^3) \tag{68}$$

where H is the Hamiltonian of the system and where , $var(H \mid \psi_m)$ is defined by (10.8). From (65) it is clear that the operator H is a measure of the rate

of transitions between the values a_m and a_n of the observable A (one often says "between the states ψ_m and ψ_n"). However it is not a rate in the usual sense, since it varies like t^2 and not like t. The operator H represents an important physical quantity called the **energy** of the system. Now let B be any observable and let U_t be given by (**??**). The self-adjoint operator

$$B(t) = U_t \cdot B \cdot U_t^* \tag{69}$$

is called the **Heisenberg evolute of B at time t**. The differential form of equation (69) i.e.

$$\frac{d}{dt}B(t) = i[H, B(t)] \qquad ; \qquad B(0) = B \tag{70}$$

is called the **Heisenberg equation of motion**. According to the evolution postulate in the form (58) and to the spectral theorem, and using the general terminology introduced in Remark (10.1), if f is any Borel function on $\mathbf{R}$ the conditional mean value of the observable $f(B)$ at time t, given that the observable A has the value a_m at time zero (or simply the mean value of B(t) in the state ψ_m), is given by

$$E(f(B(t)) \mid A(0) = a_m) =< \psi_m, f(B(t)) \cdot \psi_m > \tag{71}$$

where ψ_m is the eigenvector of A=A(0) corresponding to the eigenvalue a_m. In particular, if $B = H$, then from (**??**), (69) and (71), it follows that, for any $t \in \mathbf{R}$, for any observable A and any of its values a_m :

$$E(f(H(t)) \mid A(0) = a_m) = E(f(H(0)) \mid A(0) = a_m) = \text{constant} \tag{72}$$

which is the quantum mechanical form of the **conservation of the energy**. From (85) it follows that for short times t, the smaller the conditional

variance of the energy given that $A = a_m$, the higher the probability that the observable A will keep the value a_m after t seconds. By the Schwartz inequality

$$\mid E(H \mid \psi_m) \mid = \mid < \psi_m, H\psi_m > \mid \leq \mid\mid \psi_m \mid\mid \cdot \mid\mid H\psi_m \mid\mid = \Big(E(H^2 \mid \psi_m)\Big)^{1/2} \tag{73}$$

hence the minimum value of the conditional variance is zero and the equality in (73) (corresponding to the minimum variance) holds if and only if ψ_m and

$H\psi_m$ are proportional, i.e. for some scalar E_m (necessarily real and equal to $E(H \mid \psi_m)$) one has:

$$H\psi_m = E_m\psi_m \tag{74}$$

From (58) and (**??**) it follows that if ψ_m satisfies (91) then for each n , for each $t \in \mathbf{R}$, for each observable B, and for each Borel function f on $\mathbf{R}$ one has

$$E(f(B(t)) \mid H(0) = E_m) = E(f(B(0)) \mid H(0) = E_m) =< \psi_m, f(B(t))\cdot\psi_m >= \text{constant} \tag{75}$$

which is the dual formulation of the conservation of energy (72). For this reason a value E_m, satisfying (91) is also called a **stationary value** and the corresponding ψ_m a **stationary state** of the system.

7 Locality and Bell's Inequality

In this Section we correct some minor algebraic errors contained in Corollary (2) and Theorem (3) of [1] (The identities (3) and (82) of this theorem are equivalent to (77) and (1) and respectively (80) and (81) only in the case of observables taking only the values ± 1. In the general case they are only a consequence of the other two. Moreover the discussion of the beads example in that paper is made obscure by quite a number of typing misprints, therefore we reproduce here a corrected version of this discussion .)

Let us recall the following Lemma from [1] .

Lemma 2 *Let $a, b, c,$ be numbers in the interval $[-1, +1]$. Then*

$$ab - bc + ac \leq 1 \tag{76}$$

Corollary (2) For any given a, b, c, d in the interval $[-1, 1]$, the following two equivalent inequalities hold :

$$|ab - bc| \leq 1 - ac \tag{77}$$

$$|ab + bc| \leq 1 + ac \tag{78}$$

and imply

$$|ab + bc| + |ad - dc| \leq 2 \tag{79}$$

If a, b, c, d are restricted to the set $\{+1, -1\}$ then (77), (1) and (3) are equivalent.

Proof. If $ab - bc = 0$, then (77) follows form Lemma 76. If $ab - bc < 0$, then (77) is equivalent to

$$bc - ab + ac \leq 1$$

and, with the substitutions

$$a' = c \qquad ; \qquad b' = a \qquad ; \qquad c' = b$$

the above inequality becomes

$$a'b' - b'c' + a'c' \leq 1$$

which again follows from Lemma 76. Thus (77) is true. Since (1) is obtained from (77) exchanging c into $-c$, it follows that (1) is equivalent to (77). Adding (1) to (77) gives (3). Finally, if a, b, c, d are restricted to the set $\{+1, -1\}$ then , letting in (3) $d = a$ one finds (77) and this proves the equivalence.

Theorem 1 *Let (Ω, F, P) be a probability space , and let A,B,C,D be any four random variables defined on Ω and taking values in the interval $[-1, 1]$. Then the following equivalent inequalities hold :*

$$|E(AB) - E(BC)| \leq 1 - E(AC) \tag{80}$$

$$|E(AB) + E(BC)| \leq 1 + E(AC) \tag{81}$$

and imply

$$| E(AB - BC) | + | E(AD - DC) | \leq 2 \tag{82}$$

where E denotes expectation with respect to the P-measure, i.e.

$$E(F) = \int_{\Omega} F dP \qquad ; \qquad F \in L^{\infty}(\Omega, \mathcal{F})$$

If A, B, C, D take values in the set $\{+1, -1\}$ then the three inequalities are equivalent .

Proof. Corollary (2) implies that

$$|AB-BC| \leq 1-AC \qquad ; |AB+BC| \leq 1+AC \qquad ; |AB-BC|+|AD+DC| \leq 2 \tag{83}$$

Thus (80), (81) and (82) follow by taking expectations of both sides of the inequalities (83) and using the inequality $|E(F)| \leq E(|F|)$. To prove their equivalence one argues as in the proof of Corollary (2).

The following example, which is a variant of one due to E. Nelson , shows that improper application of statistical locality may lead to contradictions independently on the assumption of an unique underlying Kolmogorovian model. It is interesting that Nelson has been the first author to produce an inequality of Bell type not subject to the critique of the implicit assumption of the uniqueness of the Kolmogorovian model.

Consider a set of ± 1 observables S_x with $x = a, b, c$ and the associated joint probabilities

$$P_{x,y}(S_x^1 = 1 \ ; \ S_y^2 = 1) \qquad ; \ x, y = a, b, c \tag{84}$$

Thus here neither a common preparation nor a single Kolmogorovian model is assumed. Suppose that these probabilities satisfy the **singlet condition**:

$$P_{x,x}(S_x^1 = +1 \ ; \ S_x^2 = -1) + P_{x,x}(S_x^1 = -1 \ ; \ S_x^2 = +1) = 1 \tag{85}$$

and, for each $x \in S^{(2)} \subseteq \mathbf{R}^3$, denote

$$P_x = P_{x,x}(S_x^1 = +1) \tag{86}$$

Lemma 3 *Let $x \in S^{(2)}$ and suppose that the probabilities (84) satisfy the singlet condition (85). Then the following two statements are incompatible:*

(i) There exists a neighborhood U of x such that the maps

$$x \in U \mapsto P_x \tag{87}$$

$$y \in U \mapsto P_{x,y}(S_x^1 = i \ ; \ S_y^2 = j) \tag{88}$$

$(i, j = 1 \qquad y \in S^{(2)})$ are continuous.

(ii) There exists a neighborhood of V of x such that for each $y \in V$, the statistical locality condition

$$P_{x,y}(S_x^1 = +1 \ ; \ S_y^2 = -1) = P_{x,x}(S_x^1 = +1) \cdot P_{y,y}(S_y^2 = -1) \tag{89}$$

holds.

Proof. First notice that (85) is equivalent to

$$O = P_{x,x}(S_x^1 = +1 \ ; \ S_x^2 = +1) = P_{x,x}(S_x^1 = -1 \ ; \ S_x^2 = -1)$$

so that

$$P_{x,x}(S_x^1 = +1) = P_{x,x}(S_x^1 = +1 \ ; \ S_x^2 = -1) = P_{x,x}(S_x^2 = -1) \tag{90}$$

Applying statistical locality (89) and (86), we find

$$P_{x,y}(S_x^1 = +1 \ ; \ S_y^2 = -1) + P_{x,y}(S_x^1 = -1 \ ; \ S_y^2 = +1) = P_x(1-P_y) + (1-P_x)P_y \tag{91}$$

By continuity and the singlet condition, the left hand side of (91) tends to 1, while the right hand side tends to $2P_x(1 - P_x)$ which is less than $1/2$.
Remark. The above result is intuitively obvious: (85) means total correlation between the observables S_x^1 and S_x^2, while (89) means total absence of correlations between S_x^1 and S_y^2 for every yx, no matter how near y is to x. It is clear that by postulating both relations one introduces a discontinuity in the theory. Therefore, given this conservation law, it is physically unreasonable to postulate the statistical locality property (89). The experiments of Aspect , Rapisarda ,... show that in some cases this postulate is not only physically unreasonable, but also experimentally false.

Riferimenti bibliografici

[1] Accardi L.: *Foundations of Quantum Mechanics: a quantum probabilistic approach*. in Quantum Paradoxes ; eds. G.Tarozzi, A. van der Merwe Reidel 1987

[2] Accardi L.,A. Fedullo: *On the statistical meaning of complex numbers in quantum theory*, Lettere al Nuovo Cimento **34** (1982) 161-172.

[3] Accardi L.: *The axioms of probability theory*. Conference given at the Erice school on Statistics and Probability (E.Regazzini ed.) 1989