Journal of Statistical Physics, Vol. 56, Nos. 5/6, 1989

Are There Realistically Interpretable Local Theories?

B. d'Espagnat[1]

Received March 30, 1989

Although it rests on strongly established proofs, the statement that no realistically interpretable local theory is compatible with some experimentally testable predictions of quantum mechanics seems at first sight to be incompatible with a few general ideas and clear-cut statements occurring in recent theoretical work by Griffiths, Omnès, and Ballentine and Jarrett. It is shown here that in fact none of the developments due to these authors can be considered as a realistically interpretable local theory, so that there is no valid reason for suspecting that the existing proofs of the statement in question are all flawed.

KEY WORDS: Quantum mechanics; foundations; logic; measurements; nonlocality; Einstein–Podolsky–Rosen.

1. INTRODUCTION

Whoever considers that the aim of physics is knowledge and that negative knowledge *is* knowledge all the same must agree that a *proved* general statement, even if it is a negative one, is, when all is said and done, at least as important as any technical detail.

One of these basic general statements could only be proved within the second half of this century, with the help of Bell's theorem. For future reference let it be called *statement A*.

Statement A. No *realistically interpretable local* theory can exactly reproduce all of the verifiable predictions that follow from applying the computation rules of quantum mechanics to statistics involving pairs of particles.

Statement A derives its importance from its great generality. It does *not* depend on the axioms of quantum theory, but only on the validity of

---

[1] Laboratoire de Physique Théorique et Particules Élémentaires, Université de Paris XI, 91405 Orsay, Paris.

0022-4715/89/0900-0747$06.00/0 © 1989 Plenum Publishing Corporation

some of its *predictions*, so that it will remain true within any future theoretical development provided only that the predictions in question remain fully corroborated by the experimental findings. Of course, its meaning depends on that of the words "local" and "realistically interpretable" which occur in it, so that it is true only if these words are themselves given definite, specific meanings. On the other hand, the meanings (of these words) that make statement A provably true have nothing strange or unusual. Indeed, they more or less correspond to the *minimal* set of ideas that we are accustomed to associate with the words in question in normal language, so that statement A could be falsified only at the price of giving to some of the words occurring in it meanings that are opposite to those these words commonly have in our languages.

At least this is the main thesis of this article. That it should still be necessary to argue in favor of the validity of statement A may at first sight seem surprising, since among the proofs of it that have been given there are some at least[1–3] that are quite explicit and rigorous (I systematically described and discussed these in a detailed review article[4] to which the reader is referred). But it is a fact that notwithstanding the existence of these proofs, doubt regarding the validity of statement A has recently been induced in some physicists by the theories of Griffiths[5,6] and Omnès[7] as well as by the introduction by Ballentine and Jarrett[8] of a weak conception of locality that they called "simple locality." It is the purpose of the present paper to show that such doubts are unfounded.

This will of course necessitate detailed examinations of the relevant features of both Griffith's and Omnès' theories and of the "simple locality" concept. In Section 2 a paradox that Griffiths spotted in his theory is examined and is found to be of a *logical* nature. In Section 3 the question is investigated whether an interpretation of Omnès' theory exists in which the paradox in question vanishes, and it is found that such an interpretation requires describing the physical properties of microsystems by means of assertions that are valid merely as a result of some conventions we make and may change at whim. In Section 4 it is pointed out that in view of the above stated facts, Griffiths' theory is not a counterexample to *statement A*. In Section 5 the same is proved regarding Omnès' theory. In Section 6 the fact that neither Griffiths' nor Omnès' theory admits of counterfactual definitions of the properties of systems is commented upon, and so is the fact that a similarity exists on this point between Omnès' theory and a consistent interpretation of Niels Bohr's approach.

In Section 7 some quantitative observations are made concerning the notions of identity of systems and of completeness of descriptions. These serve, in Section 8, to discuss the "simple locality" concept. It is shown that requiring "simple locality" alone is not enough for making a theory of which it is required that it be realistically interpretable, obey relativity requirements. It is concluded (Section 9) that none of the examined theories and developments give any indications that the previously existing proofs of statement A are flawed.

Let it be stressed that regarding the Griffiths' and Omnès' theories, the scope of this article is limited to the above-defined question. In particular, a detailed scrutiny of Omnès' proposals concerning measurement theory would also be interesting, but will not be attempted here.

2. GRIFFITHS' PARADOX

Following Griffiths,[6] let us consider once again the well-known example of two spin-1/2 particles a and b which are initially, at time t_0, in a correlated singlet spin state but which do not interact with each other at any time $t > t_0$. Particle a (b) interacts with some apparatus C_a (C_b) which measures its spin polarization in a particular direction direction z (z') at a time $t_a{}^*$ $(t_b{}^*) > t_0$ $(t_b{}^* > t_a{}^* > t_0)$ (Fig. 1). It is a distinctive feature of Griffiths' interpretation (see ref. 5 and 6) that under these and similar circumstances the component S_z^a of particle a spin along O_z is considered as objectively *having already had* at any time $t_a < t_a{}^*$ (with $t_a > t_0$) the value $+1$ or -1 (in units $\hbar/2$), which is registered at time $t_a{}^*$ by the instrument. This is expressed in Griffiths' notations by the formulas

$$\mathscr{P}(S_z^a = 1 \mid C^a_{+z}) = 1 \tag{2.1}$$

$$\mathscr{P}(S_z^a = -1 \mid C^a_{-z}) = 1 \tag{2.1'}$$

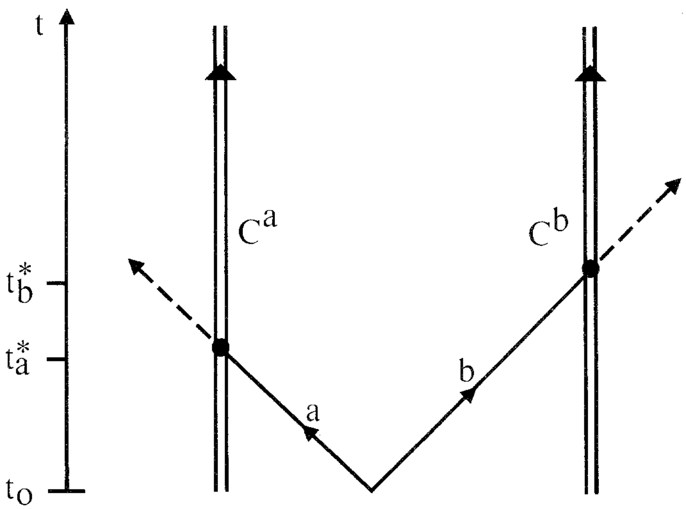

Fig. 1.

where C^a_{+z} means that the apparatus C_a has registered value $+1$, $S^a_z = 1$ means that S^a_z *has* value $+1$ at time t_a (similar conventions hold with -1 substituted for $+1$ at both places), and $\mathscr{P}(B \mid C)$ denotes the conditional probability of B given C. Similarly, we have (with obvious notations)

$$\mathscr{P}(S^b_{z'} = = 1 \mid C^b_{+z'}) = 1 \tag{2.2}$$

$$\mathscr{P}(S^b_{z'} = -1 \mid C^b_{-z'}) = 1 \tag{2.2'}$$

where now $S^b_{z'} = \pm 1$ refers to a time t_b, with $t_0 < t_b < t_b{}^*$. Moreover, assuming that $C^a(C^b)$ interacts only with a (b), we can write, following Griffiths, such formulas as

$$\mathscr{P}(S^b_{z'} = -1 \mid C^a_{+z} \wedge C^b_{-z'}) = 1 \tag{2.3}$$

where $\wedge$ means "and," and if we now take the spin correlation between particle a and b into consideration, we can also write, again following Griffiths, such formulas as

$$\mathscr{P}(S^b_z = -1 \mid C^a_{+z} \wedge C^b_{-z'}) = 1 \tag{2.4}$$

expressing that in a case in which C_a and C_b register values $+1$ and -1, respectively, S^b_z *has* at time t_b value -1.

The paradox, which, as Griffiths lucidly noted, is an unescapable one in this theory, is that, under the conditions stated, while (2.4) holds, simultaneously $S^b_{z'}$ must, according to (2.3), have a definite value, namely -1. So that at any time t_b *two* distinct components of $\mathbf{S}^b$ should possess definite values. This, however, is impossible, since in Griffiths' theory just as in ordinary quantum mechanics the fact that a physical quantity has a certain value is associated with a projector in the Hilbert space, whereas no projector can be associated with the conjunction of the two facts considered above.

Now, while Griffiths correctly acknowledges the existence of this difficulty (and even calls it a "paradox"), he nevertheless maintains both that (2.3) and (2.4) are true *and* that, however, given that C^a_{+z} and $C^b_{-z'}$ are true, the proposition $(S^b_z = -1 \wedge S^b_{z'} = -1)$ is meaningless. It must quite honestly be said that it seems difficult to follow him along these lines. To show why in detail, one may argue as follows.[9] First, observe that as soon as it is claimed that a proposition $\mathscr{D}$ *entails* a proposition $\mathscr{B}$, it logically follows (the logicians call this "*modus ponens*") that if a system S is considered on which $\mathscr{D}$ happens to be true, then $\mathscr{B}$ is necessarily true on S. Second, note that if two propositions $\mathscr{B}$ and $\mathscr{B}'$ are true (hence meaningful), it also follows, in any known logic, that the proposition $\mathscr{B} \wedge \mathscr{B}'$, which is the conjunction of both, is also meaningful and true.

Under these conditions if it is the case that $\mathscr{D}$ implies $\mathscr{B}$ *and* that $\mathscr{D}$ also implies $\mathscr{B}'$, and if a system S is considered on which $\mathscr{D}$ is true, then, on S, $\mathscr{B} \wedge \mathscr{B}'$ is necessarily meaningful and true. In fact the opposite claim, which amounts to Griffiths' with the appropriate symbol identifications, could only be valid within some new and as yet unspecified logic, having nothing to do with either classical or quantum logic and of which it is not even known how it could be self-consistent.

3. THE PARADOX IN OMNÈS' THEORY

There are differences between the Omnès and the Griffiths theories. A noticeable one is that the first author introduces the probability concept only at a late stage of his developments. In the first stages he merely introduces "mathematical measures." However, just as Griffiths and others, he considers physical systems and physical quantities (observables) pertaining to these systems and he associates definite projectors to the facts that such and such observables *have* definite values on a system (or are in definite subsets of their spectrum). Moreover, by using "mathematical measures" in the same way as Griffiths uses probabilities (including conditional ones), he is able to define implications (symbol $\Rightarrow$) in such a way that, apart from very special cases that do not occur in the example used here, any relation written

$$\mathscr{P}(\mathscr{B} \mid C) = 1 \tag{3.1}$$

in Griffiths' notations and which is valid in Griffiths' theory is also valid in Omnès' theory, where it is written

$$C \Rightarrow \mathscr{B} \tag{3.2}$$

As a consequence, with propositions $\mathscr{B}$ and $\mathscr{B}'$ defined as

$$\mathscr{B} \underset{\text{def}}{=} \langle\!\langle S_z^b = -1 \text{ at time } t_b \rangle\!\rangle \tag{3.3}$$

and

$$\mathscr{B}' \underset{\text{def}}{=} \langle\!\langle S_{z'}^b = -1 \text{ at time } t_b \rangle\!\rangle \tag{3.4}$$

it seems at first sight that the paradox met with above reappears here in the form

$$C_{+z}^a \wedge C_{-z'}^b \Rightarrow \mathscr{B} \tag{3.5}$$

$$C_{+z}^a \wedge C_{-z'}^b \Rightarrow \mathscr{B}' \tag{3.6}$$

$$C_{+z}^a \wedge C_{-z'}^b \not\Rightarrow \mathscr{B} \wedge \mathscr{B}' \tag{3.7}$$

But does it? At the place in his articles[7] where Omnès investigates these kinds of problems (paper II, Section 4) he does not mention any paradox. On the other hand, he does not consider proposition $\mathscr{B}'$ explicitly. In fact, the only proposition concerning $S^b_{z'}$ that he considers is one taken at time t_b* (in our, i.e., Griffiths', notations). Since he mentions that at that place he has supplemented Griffiths' calculations by results from his own "measurement theory," we *a priori* could conjecture that in his views the wave packet reduction taking place at t_a* forbids implication (3.6). Concerning the central question of the present paper—whether statement A is falsified by theories such as Omnès'—readers who favors *this* interpretation of Omnès' views are referred to Section 8, from the content of which they will immediately infer that the interpretation in question does *not* falsify statement A.

But in fact this interpretation does not seem to be the right one for reasons of consistency internal to Omnès' developments. Indeed, in connection with the spatial aspects of the EPR problem, this author (see in particular his paper II, Section 5.4) discusses the case in which a spatial measurement—call it M—is made at a time t_2 on one of the two particles of the pair. He then considers that logical—and even causal—implication links do exist that, among others, associate two propositions bearing on the other, distant, particle, one of these, E_2' in his notation, also concerning time t_2 and another one, E_1', concerning some time $t_1 < t_2$ (with $t_1 > t_0$, t_0 being, as above, the time at which the pair is formed). When account is taken of the fact that if two spacelike separated events such as M and E_2' are simultaneous in one referential there always exist other referentials in which M takes place before E_2', the validity of the above implication links implies that the existence of measurement M at time t_2 does not vitiate implications concerning propositions associated with the particle other than the one on which M is made, even when these propositions bear on times one of which is anterior and the other one posterior to time t_2.

If we now, as above, introduce two measuring instruments C_a and C_b, we are thus led back, unescapably as it seems, to implications of the types (3.5) and (3.6). The reason why Omnès did not hit at that place on a paradox seems to be that in fact he carried out the just summarized detailed analysis only regarding *spatial* propositions (that is, propositions of the type "at time t particle $\mathscr{P}$ is in region J") and that, of course, propositions bearing only on one observable—here position—always commute. But as soon as we consider observables that do not commute, such as two different spin components of one particle, the problem reappears, very much as in Griffiths' work.

Now, should we really say that this problem constitutes a paradox also in Omnès' theory? To study the matter as it deserves, we must enter somewhat more into the detail of this theory. In fact, there are two features of it that might well, at first sight, be relevant. One of them (feature a) is the presence in it of special families of propositions called *consistent representations of logic* (a notion that, indeed, is central in the theory). In our example

$$L_1 \equiv \{C^a_{+z}, C^b_{-z'}, \mathscr{B}\} \tag{3.8}$$

and

$$L_2 \equiv \{C^a_{+z}, C^b_{-z'}, \mathscr{B}'\} \tag{3.9}$$

are consistent representations of logic, whereas

$$L_3 \equiv \{C^a_{+z}, C^b_{-z'}, \mathscr{B}, \mathscr{B}'\}; \qquad O_{z'} \neq O_z \tag{3.10}$$

is not (see the quoted articles for a formal definition and for details).

Another one (feature b) is the presence in the theory of the expression "reliable statement" replacing, at most places, the expression "true proposition."

Let us here consider three questions in succession:

(i) Is feature a sufficient for removing the paradox?

(ii) If not, is the adjunction of feature b sufficient by itself to do the same?

(iii) If not, is it possible to remove the paradox by making feature b more precise and more specific?

(i) A hint as to the right answer to question (i) is provided by the fact that actually Omnès' notion of a *consistent representation of logic* (abbreviated as CRL in what follows) is very close to a notion central to Griffiths' theory, and which is known there under the name "consistent history." The fact that, as we saw, the *consistent history* notion is not sufficient to remove the paradox may already be considered as a first indication that the "consistent representation of logic" notion might well also be insufficient for that purpose.

Indeed, if we take a close look at the argumentation that was made explicit above (the argumentation concerning the occurrence of the paradox in Griffiths' theory), we observe that it can be transposed to Omnès' theory practically without change. To see this in detail, let us consider successively the two stages of the argument. The first one ("*modus ponens*") consists in saying that if we have two propositions $\mathscr{D}$ and $\mathscr{B}$ of which it is known that quite generally $\mathscr{D}$ entails $\mathscr{B}$ and if it is the case that in some particular instance we know $\mathscr{D}$ is (meaningful and) true, then we may quite confidently assert that, in that instance, $\mathscr{B}$ is meaningful, and that, moreover, it is true, without having to formulate any restrictive condition. In plain language, $\mathscr{B}$ is then "factually true, full stop," To be able to "water down" in some way this assertion while holding on to the premiss that $\mathscr{D}$ is "meaningful and moreover true, full stop" the only possibility seems to be to assert that $\mathscr{D}$ does not *generally* imply $\mathscr{B}$. When trying to proceed along these lines we could at first try saying that $\mathscr{D}$ implies $\mathscr{B}$ only if such and such conditions G are satisfied. But this would amount to replacing $\mathscr{D} \Rightarrow \mathscr{B}$ by $\mathscr{D} \wedge G \Rightarrow \mathscr{B}$. In our example this would work if we could replace (3.5) by

$$C^a_{+z} \wedge C^b_{-z'} \wedge G \Rightarrow \mathscr{B} \tag{3.11}$$

and (3.6) by

$$C^a_{+z} \wedge C^b_{-z'} \wedge G' \Rightarrow \mathscr{B}' \tag{3.12}$$

with G' describing some physical conditions differing from G. But in the example there are no such additional conditions G and G' differing from one another. Hence this first idea does not work. Nor is it the one that Omnès seems to have in mind. Indeed, his expression "consistent representation of logic" points to quite a different view, which essentially is that only some associations of propositions make sense. According to this view, we must say, for example, that (3.5) is meaningful and true in L_1 but not in L_2 since $\mathscr{B}$ cannot consistently enter L_2. There is, however, a trouble with this, which has to do with the "*modus ponens*" aspect of implication in logic. For example, if we say that (3.5) is meaningful and true in L_1 but not in L_2 and if we apply "*modus ponens*," we must say that—in a case in which $C^a_{+z} \wedge C^b_{-z'}$ is true—$\mathscr{B}$ is true in L_1 but is meaningless in L_2 and conversely as regards $\mathscr{B}'$, L_2, and L_1. So, finally, we end up with propositions that should be considered either as actually true or as meaningless, and that, not according to any factual differences in the systems themselves or in the instrumental setup, but just according to the way *we* choose to consider the matter at hand—more precisely, according to the way in which we choose to mentally associate these propositions with some other ones. Unquestionably this conclusion is at odds with the set of ideas that we normally have in mind when we speak of a factual truth, so that the use of the word "true" in this context is inappropriate and misleading.

(ii) Since relativizing the notion of factual truth to the extent of making it dependent on an arbitrary choice of ours as to how we associate propositions with one another is something quite hard to accept, we may now remember that Omnès in fact does not quite use this language. Instead of asserting that a proposition such as $\mathscr{B}$ [as given by (3.5)] is *true*, he merely says that it is a "reliable statement." Conceivably a way out of the difficulty may be searched for along this line. Omnès, however, does not actually specify what the difference *is* between a true proposition and what he calls a reliable statement. Clearly, it we had here to do with a mere renaming, nothing would be gained.

(iii) On the other hand, the analysis of paragraph (i) serves as a useful guide as soon as we try to take advantage of the fact of introducing a *new* notion, that of a "reliable statement," so as to remove the difficulty. The point is that, in physics, when we apply the qualificative "true" to a statement bearing on a property attached to a physical system we thereby make an assertion which is supposed to bear exclusively on that system and, correlatively, to be valid *per se*, quite independently of how we choose to associate the idea it expresses with other ideas. When, instead, we have to do with a *new* expression we obviously have more freedom in this respect. In particular, we are free to restrict ourselves to the search for a purely formal construction, of which we only request that it should be free from internal contradictions. Within such a limited purpose it is clear that the difficulty at hand is quite easily removed. For this, it is enough to speak, instead of just "reliable statements," of "statements reliable within such and such a CLR," and, instead of "implications," speak of "implications valid in such and such a CLR." Clearly the difficulty then dissolves. However, the price paid for that is heavy, for, instead of having statements actually bearing on the things as they are, what we actually now have is just a system of assertions that are valid merely as a result of some *conventions* we make and may change at whim.

4. IS GRIFFITHS' THEORY A COUNTEREXAMPLE TO STATEMENT A?

The answer is no, for several reasons (see, e.g., Section 6 below), the main one being the fact that this theory is vitiated by the logical paradox discussed in Section 2. It is true that, as Griffiths quite rightly point out, this paradox is of the same nature as one already taking place in this theory when successive measurements are carried out on just *one* spin-1/2 particle. But then what? The point is that in both cases we have to do, not merely with just counterintuitive features—features that "look paradoxical" from a classical perspective—but with a real *logical* paradox, as shown above; which means a state of affairs that is self-contradictory not only within classical logic, but also within any known other logic, quantum or otherwise. To this observation it cannot be validly replied (as Griffiths seems to suggest) that anyhow quantum physics is unavoidably deemed to labor under such logical paradoxes. In fact, there are interpretations of quantum mechanics that, unpalatable as they may be, at least do not involved logical paradoxes. Two extremal and opposite examples are Wigner's reduction-by-consciousnesses intepretation and de Broglie's (and Bohm's) pilot wave theory (it is consistent with statement A that both these theories turn out to be nonlocal).

5. IS OMNÈS THEORY A COUNTEREXAMPLE TO STATEMENT A?

For a theory to be a counterexample to statement A a necessary condition is that it be a realistically interpretable theory. This means in particular that it must contain realistically interpretable assertions concerning the physical systems it deals with and concerning their properties.

Now, the basic elements of Omnès' theory are propositions of the type (see his Definition 4) "On system S observable A has, at time t, its value in subset C of its spectrum." Such propositions explicitly bear on the systems themselves, as indeed they should within a realistically interpretable theory, rather than on our mental reconstructions of the systems. On the other hand, it was shown in Section 3 that most of these propositions cannot bluntly be said to be true (or false). The only correct assertions that in this theory can be formulated about them is that they are (or are not) "reliable statements within such and such a consistent representation of logic." Hence Omnès' theory can be said to be a realistically interpretable theory (of the considered systems) only if we grant that an assertion such as "$\mathscr{P}$ is a reliable statement within such and such a CRL and is not within such and such another CRL" (where $\mathscr{P}$ stands for a proposition to the type above) is itself a realistically interpretable assertion.

The question then is: can we grant that much? Formally, of course, this is quite possible. Subject to internal consistency, a set of definitions is always logically arbitrary. Since we have not yet defined the expression "realistically interpretable," we are logically quite free to choose for it a definition extensive enough to cover assertions such as the one above. On the other hand, this very freedom has the consequence that such a choice, once made, is hardly instructive or interesting. Clearly, for the sentence "assertion... is realistically interpretable" to have any interest it must be the case that the expression "realistically interpretable" corresponds, be it approximately, to some notion already present in our mental furniture. Otherwise it is just arbitrary or meaningless.

Seen from this angle, I am afraid the question above must be answered in the negative. In Western languages at least, the whole tradition converges toward an attribution of meaning such that the qualificative "real" or "physically real" conveys with it the notion that when applicable to some entity in some circumstances it remains applicable to it in these circumstances quite independently of how the thus formulated idea is mentally associated with other ideas. From this point on this kind of independence will be taken as being an inherent part of what we mean when we say that a statement or a theory is realistically interpretable. The Omnès theory is then not interpretable that way, hence it is not a counterexample to statement A.

6. COUNTERFACTUALITY

A property $\mathscr{P}$ is *counterfactually* defined on a system S when, for defining it, we refer to a measurement that *could* but perhaps *will not* be made on S. The alternative (partial definitions procedure) is to define $\mathscr{P}$ only regarding the systems S on which the measurement will *actually* be done. But in common practice we feel there are circumstances in which a system has a property even if nobody has prepared a device for testing whether or not this is the case.

Up to this point counterfactuality was not mentioned in this article. In other words, to show that the Omnès and Griffiths theories are no counterexample to statement A it did not prove necessary to refer to the links that, in ordinary ways of thinking, bind together counterfactuality and realism. This of course does not mean that these links are nonexistent or irrelevant. On the contrary they are very strong (see ref. 10 for details). The same argument as the one put forward in the last section thus shows that they must be taken into account; and if we do, then an additional argument is thereby provided to the effect of showing that neither the Griffiths nor the Omnès theory is realistically interpretable. This is due to the fact that obviously these theories do not allow for properties to be counterfactually defined.

Remark. The state of affairs just underlined, while it corroborates the fact that Omnès' theory is not realistically interpretable, is definitely not one that should induce us to reject the theory. Indeed, if it were, Bohr's interpretation of quantum mechanics would have been rejected long ago, for this interpretation *also* parts with counterfactuality. In fact, as Omnès noted, there are clear similarities between the approaches of these two authors. In his famous 1935 reply to the Einstein, Podolsky, Rosen paper, Bohr postulated that the "conditions which define the possible types of predictions regarding the future behavior of the system" constitute an "inherent element of the description of any phenomenon to which the term 'physical reality' may be attached."[11] Since the conditions in question have to do, as is well known, with the "whole experimental setup" used for making measurements on the system, they are for the most part external to the quantum system itself. Hence an implicit implication of Bohr's view is that no variable the very definition of which associates it exclusively to some specified physical system can be considered as having *all by itself* a definite value when the system is a quantum system. In other words, the corresponding system properties—position, velocity, spin component, etc.—viewed as attached to the system simply to not exist as such. If we nevertheless speak of such properties, as we often do, it can therefore only be as a convention, such a convention being both useful and harmless once the experimental setup is fixed.

In Omnès' theory (and also in Griffiths') the situation is in this respect quite similar, since in it the properties of a quantum system (the fact that such and such propositions are reliable statements on this system) are only meaningful when imbedded in a consistent representation of logic, which, as it seems, depends in the last resort of the choice of the instruments (C_a, C_b, and the orientation of each, in our example). But, as shown by the analysis in Section 3, Omnès' theory explicitly leads to a conclusion which was only implicit in Bohr's approach and which is that the convention by which we attribute values to variables of a system (or by which the propositions regarding a system are made *reliable statements*) can, in some cases, not be fully determined by the experimental setup. Instead, it can depend on a free choice of ours regarding the "representation of logic."

7. COMPLETELESS AND IDENTITY

Under what conditions can we speak of the physical state of a microsystem—that is, of a system of the type of those to which quantum physics is currently applied—and when can a specification of such a state be called complete?

These questions are by far not as elementary as it seems. Physicists usually shun them because the answers depend on a choice that most physicists hate to make, namely that of a philosophy. It is clear, though, that while a thoroughgoing operationalist may consistently answer these questions by merely referring to the experiments and measurements mankind can perform, a physicist who wants the theory to be realistically interpretable cannot be content with just this. He or she is bound to give answers that somehow refer, not only to what we can do on, or know about, a system, but to what it actually is. In the present paper, since its purpose is to discuss statement A, it is this "realist" standpoint that must be taken up as a basis of argument.

If we do, then it becomes clear that the Griffiths assertion that "measurements reveal properties that already existed,"[6] when made as general as this author makes it, is not compatible with the "orthodox" view according to which the "quantum state," as described, e.g., by a ket, fully specifies, at any given time, the physical state of a microsystem. To show this, almost any example will do. Using the one considered by Griffiths and introduced here in Section 2 (the apparatuses C_a and C_b being duly positioned) and taking into account Griffiths' consistency conditions, we observe that, within this author's theory and in view of the initial assumption and the evolution law, the system composed of the two particles a and b can be said to still be, at time t_a, in the spin-zero state in which it was at time t_0 (this "event," together with the "event" that the pair was already in that state at time t_0 and with the two measurement events taking place at times t_a^* and t_b^*, forms a "consistent history"). This fact should be compared with Griffiths' assertion[6] that "if the apparatus C_a is in the state C^a_{+z} at a later time we can be sure that the corresponding component of spin of particle a was positive at all time before measurement." Within Griffiths' theory as within the orthodox theory these two statements are definitely incompatible with one another, and the guiding line of the theory in question is to remove the corresponding inconsistency not by giving up one of them, but by considering that they describe "events" that belong to two distinct "consistent histories."

Thus, any one of these two "consistent histories" is good, but they are incompatible. Apparently in Griffiths' theory the question "which one is the true one" is dismissed as being meaningless, the truth being that we must choose one history and that we can do this at whim. But then this implies, in particular, that the statement "the pair can be said to be, at time t_a, in a spin-zero state" cannot be bluntly replaced by the statement "the pair *is* at time t_a in a spin-zero state," which would convey the erroneous impression that the pair is actually in this state independently of any option of ours (similarly, Griffiths' statement quoted above concerning the certainty that the spin component of a was positive should be watered down by adding "within a consistent history that we can choose at whim"). More generally, if, in Griffiths' theory, we were to formulate assertions such as "the quantum state, as described by a ket (or by a state operator) specifies, at any given time, the physical state of the system," we should have to add "within a consistent history that we are free to choose at whim among several possible ones." Similar observations hold regarding Omnès' theory, of course.

These remarks are useful, not only in that they corroborate the conclusions of Sections 4 and 5 (to the effect that the Griffiths and Omnès theories are not realistically interpretable), but also in that, by contrast, they reveal one of the conditions that we must assign to theories of which we claim that they are realistically interpretable, in order that this characterization should not be just a blank label. This condition is that proposed "complete specifications of physical states of systems" should, in such theories, refer exclusively to the systems themselves, without any implicit restriction that we may consider such a specification either as being or as not being valid according to the way we choose to associate ideas.

The assignment of this condition has an important consequence concerning the notion of two microsystems being in *identical* physical states. This consequence may seem trivial, but the mere existence of the Griffiths and Omnès theories shows it is not. It is referred to as follows as statement B.

Statement B. The assertion that such and such a mathematical description (ket, state operator, etc.) *completely* specifies the physical state of a system *implies* that two systems specified by one and the same such description are identical in every respect.

8. LOCALITY VERSUS "SIMPLE LOCALITY"

Above it has been shown that statement A is not falsified by the Griffiths and Omnès theories. The last item in our program is to show that, contrary to an impression one might get upon reading Ballentine and Jarrett,[8] the considerations on "simple locality" developed by these authors do not falsify statement A either.

To that end, let it briefly be recalled that the quoted authors define simple locality as follows. Considering again the standard example of two correlated spin-1/2 particles, they denote by $\mathbf{d}_L$ ($\mathbf{d}_R$), S_L (S_R), and x_L (x_R) the unit vector that defines the component of spin measured by the device on the left (right) hand side, the other parameters specifying the states of these two devices (including the times at which measurements take place), and the corresponding results ($x = \pm 1$ in $\hbar/2$ units), respectively. The probability of obtaining the particular results x_L and x_R is then of the form

$$\mathscr{P}(x_L, x_R \mid \mathbf{d}_L, \mathbf{d}_R, S_L, S_R, \lambda) \tag{8.1}$$

where λ denotes the state of the pair. λ can be quite general. It can be a quantum state, a quantum state plus arbitrary variables, or some non-quantum-mechanical form of state description.

With the help of (8.1), several probabilities concerning x_L (or x_R) can be defined by applying the general formulas of probability theory. In particular, we may consider the probability

$$Q_L(x_L \mid \mathbf{d}_L, \mathbf{d}_R, S_L, S_R, \lambda) \tag{8.2}$$

that result x_L is obtained assuming $\mathbf{d}_L$, $\mathbf{d}_R$, S_L, S_R, and λ to be given, and also the (not considered in ref. 8) conditional probability

$$p_L(x_L \mid \mathbf{d}_L, \mathbf{d}_R, S_L, S_R, \lambda, x_R) \tag{8.3}$$

that x_L is obtained assuming $\mathbf{d}_L$, $\mathbf{d}_R$, S_L, S_R, λ, *and* x_R to be given. The corresponding formulas are

$$Q_L = \sum_{x'_R} \mathscr{P}(x_L, x'_R \mid \mathbf{d}_L, \mathbf{d}_R, S_L, S_R, \lambda) \tag{8.4}$$

(with a similar formula for Q_R, of course)

and

$$p_L = \mathscr{P}(x_L, x_R \mid \mathbf{d}_L, \mathbf{d}_R, S_L, S_R, \lambda)/Q_R \tag{8.5}$$

with a similar formula for p_R. "Simple locality" is then defined by the quoted authors as the hypothesis that, at least when the two measurements take place in spatially separated regions, Q_L does not depend on $\mathbf{d}_R$, nor Q_R on $\mathbf{d}_L$.

When λ is identified with the quantum state, the above-defined probabilities are obtained by applying the standard quantum rules and it can be shown that "simple locality" is obeyed. For this purpose what has to be done is not so much to check, as done in ref. 8, that (8.4) is satisfied since this formula *must* be valid in *any* theory whatsoever (with the only proviso that x_R is *really* measured). What must be shown is that Q_L does not depend on $\mathbf{d}_R$. This is not quite so obvious as it seems, because of the fact that a probability such as $\mathscr{P}$ may, in general, depend on the order in which the two measurements are made. As a rule, this is the case when the operators A_R and A_L describing the two measured observables do not commute, for then, if $|\psi_0\rangle\langle\psi_0|$ denotes the state of the pair, and D_i^α ($\alpha = L, R$) the projector onto the eigensubspace of A_α which corresponds to eigenvalue x_i^α of A_α, we have

$$\mathscr{P}(x_i^L, x_j^R) = \mathrm{Tr}(D_i^L D_j^R \, |\psi_0\rangle\langle\psi_0| \, D_j^R)$$

when the R measurement takes place before the L one and

$$\mathscr{P}(x_i^L, x_j^R) = \mathrm{Tr}(D_j^R D_i^L \, |\psi_0\rangle\langle\psi_0| \, D_i^L)$$

in the opposite case, so that [see Eq. (8.4)] the probability of obtaining x_i^L if A_R has been measured first but not registered is

$$q = \sum_j \mathscr{P}(x_i^L, x_j^R) = \sum_j \mathrm{Tr}(D_i^L D_j^R \, |\psi_0\rangle\langle\psi_0\rangle| \, D_j^R)$$

which in general depends on the choice of what observable A_R has been measured. However, when A_R and A_L commute, so do the projectors D_i^L and D_j^R, so that the foregoing formula reduces to

$$\sum_j \mathrm{Tr}(D_j^R D_i^L \, |\psi_0\rangle\langle\psi_0| \, D_j^R) \equiv \mathrm{Tr}(D_i^L \, |\psi_0 \times \psi_0|)$$

since $\sum_j D_j^R = 1$. This shows that indeed q is then independent of the choice of what observable A_R is measured on the right-hand side even when A_R is measured first, so that finally this independence holds independently of the order in which the two measurements are made. Since two observables pertaining to distinct, spacelike separated space-time regions always commute ("microcausality" principle), we have thereby proved that "simple locality" is obeyed by standard quantum mechanics whenever A_R and A_L commute, which is the case in the example considered here.

On the other hand, the conditional probability p_L that the L device obtains the result x_L *if* the device R has obtained the result x_R is in general different from the nonconditional probability Q_L and x_L is obtained. This remains true even if "simple locality" is obeyed. Does it entail the consequence that signals could be sent from the space-time region R (the one where the R device operates) to the space-time region L (or conversely)? As the quoted authors rightly point out (although not on the ground of the same argument), the answer is *no*. The main point here is (in my opinion) that while experimentalists operating at R can choose the value of $\mathbf{d}_R$, that is, the measurement they want to do, they cannot choose the value of x_R, that is, the result of that measurement. Hence, the fact that p_L depends on x_R is of no direct help for the purpose of sending a signal. The only way in which, *a priori*, an experimentalist could hope to make use of p_L for such a purpose would in fact be to sum over all the possibilities for x_R, with weights coefficients equal to the Q_R, of course. But the result of this operation is just Q_L and the fact that Q_L does not depend on $\mathbf{d}_R$ has therefore the effect that no such signalling is possible (see refs. and 12 and 13 for more detailed proofs). In particular, regarding the case in which the two measurements are performed in spacelike separated regions, the foregoing considerations bar any possibility that the dependence of p_L on x_R should make faster-than-light signalling possible between these two regions.

This argument adequately substantiates the old claim[14] that quantum mechanics does not violate the operationally formulated relativity law according to which "no signal can travel faster than light." It has led Shimony[15] to speak—extremely appropriately—of a "peaceful coexistence" between quantum physics and special relativity. A thoroughgoing operationalist such as the one already mentioned above could even go farther than that, and quite consistently consider that the results above suffice to remove *any* contradiction between quantum mechanics and special relativity. But can a supporter of the idea that physical theories should be realistically interpretable be satisfied with just this? Some reflection shows that the answer is no. The point is that the basic laws of a realistically interpretable theory cannot be formulated by merely referring to what mankind can or cannot do. The very principle that they should be realistically interpretable—whatever detailed meaning we ascribe to the expression—definitely rules this out. The only elements that should be allowed to appear in the basic laws of such theories are elements that somehow refer—or that at least can be considered as referring—to systems, events, and so on *as they really are.* Hence, in such theories the finite-velocity law must be expressed by statements that stand in conformity with this principle, such as the statement, "no event A can causally influence any event $\mathscr{B}$ that does not lie within the future light cone of event A."

This raises two questions in succession. The first one is: what is a measurement event? Obviously, a measurement consists in a certain instrument being positioned in a specific way at a given place at some given time. But it also consists in the instrument in question registering some definite result. In other words, in our example the measurement event consists in both $\mathbf{d}_R$ *and* x_R having given values, so that when formulated as above the finite-velocity law implies in particular that no event outside the future light-cone of the R measurement should in any way be influenced by the event of x_R taking up a definite value when the R measurement is made.

This then introduces the second question: is it possible to reconcile this condition with the fact that, in our example as in many other conceivable ones, a conditional probability such as p_L, bearing on an event occurring in region L, differs from the corresponding nonconditional probability Q_L in that it has some dependence on x_R, hence depends on a event that takes place in region R (which means that the two events are correlated)? In general, that is, when nothing is assumed concerning the completeness of the initial state description, the answer to this question is quite trivial. Two events A and B may quite well be correlated without one influencing the other. It suffices that they should proceed from a common cause. In our example, the explanation would be that the two particles composing one pair have some property in common, which some other similarly produced particle pairs are deprived of, or do not possess to the same degree. If events A and B are influenced to some extent by the property in question, a correlation between them normally results. Here, however, this commonsense explanation obviously fails, since we assumed that the quantum state $|\psi_0\rangle$ is a *complete* description of the physical state of the pair and since we showed (statement B, Section 7) that all the pairs produced in such a state must be quite strictly identical. Admittedly, we could still try to maintain that the difference between p_L and Q_L is just due, as in the "commonsense case," to the fact that the R measurement provides "additional information"; but then we would be at a loss to answer the question, "information on what?." It definitely cannot be on the state of the pair just before the measurement in question is made, since, by assumption, this state is completely known already, due to the fact that it is fully specified by $|\psi_0\rangle$. We could try to say, "it bears on the state of the L particle." But this would mean that there are some parameters attached to this particle and not described by the quantum state $|\psi_0\rangle$, which again contradicts the completeness hypothesis. Finally, therefore, we must come to the conclusion that what is at stake in such cases is not just simply information and that the mere observation that Q_L is independent of $\mathbf{d}_R$ is not in itself sufficient to remove the difficulty, a difficulty which can be pinpointed by stressing the fact that when some correlation takes place it must, in any realistically interpretable theory, receive a physical explanation.

Clearly, the hypothesis that the physical state λ of the pair is identical to its quantum state—or, otherwise said, that the ket $|\psi_0\rangle$ is a complete specification of the physical state of the system—is quite a vital assumption for the foregoing discussion to be valid. *A priori* it could therefore be conjectured that the difficulties met with are not real ones and that the appearance of their being there is just due to the fact that the hypothesis in question is not, in fact, true. Obviously, however, this new approach can only succeed if there is some possibility of defining the physical state λ in such a way that the conditional probability p_L becomes equal to the nonconditional one Q_L. Basically this is just how *locality* is defined in all the developments that have to do with the Bell theorem. We see, therefore, that, regarding the much looked-for realistically interpretable theories, the truly significant concept is *locality*, and not just "simple locality." On the other hand, the Bell theorem demonstrates, as is well known, that no realistically interpretable theory that obeys locality can exactly reproduce the observable predictions of quantum mechanics. The obvious conclusion is that statement A is corroborated.

Remark 1. A link quite obviously exists between this conclusion and the fact that in a realistically interpretable probabilistic theory a clear-cut distinction has to be made between objective and subjective probabilities. Contrary to the subjective ones, the objective—or *intrinsic*—probabilities must be viewed as real *properties* of systems. In a case such as the one considered here the probability of result x_L, for instance, can obviously be said to be an intrinsic one only when it is defined on a system the physical state of which is fully specified (otherwise it would, at least partly, be an ignorance one). But then, since it is a physical property, any change it suffers is a physical event. Within a realistic interpretation of the finite-velocity law such a change should obey a *principle of separability* (or *locality*) stipulating that this change cannot be induced by what takes place in the space-time region where the R measurement is made (if the two measurements are spacelike separated as here assumed). In other words, the intrinsic probability of result x_L should depend (in this case) neither on what measurement is made at R *nor on the state* x_R of the R *instrument pointer immediately after the R measurement has been made.* Requiring only "simple locality" amounts to giving up the second condition. Hence here again "simple locality" is shown not to be a sufficient condition for reconciling with the relativity requirements a theory of which it is required that it be realistically interpretable.

Remark 2. In spite of superficial appearances to the contrary, the thesis is *not* upheld in this article that a physical theory *must* be physically interpretable. Nor of course is the view that such a theory *must* be local. In fact the questions concerning what can reasonably be demanded now from a physical theory when account is taken of what is at present firmly known is a delicate and subtle one that lies very much beyond the scope of the article (my approach to these points can be found in ref. 10). Here the goal was a much more specific one. It was only to show that these two conditions cannot be imposed together.

9. CONCLUSION

The importance of the Bell theorem—or more precisely of statement A (see Section 1)—could hardly be overstressed. Quantum mechanics will presumably be superseded at one time or other by some more comprehensive future theory based on quite different axioms (at least this looks likely if the lessons of the past are to be believed). It may be hoped that many of the conceptual problems it raises will then automatically vanish. But since Bell's theorem is, as already noted, *not* based on these axioms, the validity of statement A must be expected to survive this momentous "revolution." In that sense it constitutes, so to speak, a fixed point in the moving sea of our conceptions of the world. It would be a pity if physicists derived the impression it does not hold true from an uncritical reading of some texts, which, indeed, if read that way, may well convey that impression.

The purpose of the present article was just to try to prevent this danger. It should therefore not be considered as actually constituting a *criticism* of the theories the content of which was discussed above. In fact, none of these theories or developments contains the expression "realistically interpretable." Strictly speaking, there is therefore no contradiction between their content and that of the present paper. At least, this is true if by "content" is meant the scientific content in the strict, and therefore restrictive, sense of the word (although a theory that, avowedly or not, involves a logical paradox may be viewed with reservation even from that angle). On the other hand, it must be granted that several of the interpretative comments the quoted authors make of their theories stand quite at odds with the main conclusions reached here. Indeed, while these authors do not actually *say* their theories are realistically interpretable, they somehow give at various places the impression that they *mean* just precisely that. Such a somewhat disquieting state of affairs seems to indicate that we physicists still have efforts to make before we succeed in imparting to the *words* we use (and especially to the nonoperationally defined ones) a strictness of meaning comparable with the strictness of our mathematical manipulations. This will presumably only be achieved when we have convinced ourselves that it is impossible to freely switch between an ontological and a purely operationalist usage of such words as "have," "is," "objective," and the rest.

REFERENCES

1. J. S. Bell, *Physics* **1**:195 (1964).
2. J. S. Bell, in *Foundations of Quantum Mechanics*, B. d'Espagnat, ed. (Academic Press, New York, 1971).
3. J. F. Clauser and M. A. Horne, *Phys. Rev. D* **10**:526 (1974); J. F. Clauser and A. Shimony, *Rep. Prog. Phys.* **41**:1881 (1978).
4. B. d'Espagnat, *Phys. Rep.* **110**:201 (1984).
5. R. Griffiths, *J. Stat. Phys.* **36**:219 (1984).
6. R. Griffiths, *Am. J. Phys.* **55**:11 (1987).
7. R. Omnès, *J. Stat. Phys.* **53**:893, 933, 957 (1988).
8. L. E. Ballentine and J. P. Jarrett, *Am. J. Phys.* **55**:696 (1987).
9. B. d'Espagnat, *Phys. Lett. A* **124**:204 (1987).
10. B. d'Espagnat, *Reality and the Physicist* (Cambridge University Press, Cambridge, 1989).
11. N. Bohr, *Phys. Rev.* **48**:696 (1935).
12. P. Eberhard, *Nuovo Cimento* **46B**:392 (1978).
13. G. C. Ghirardi, A. Rimini, and T. Weber, *Lett. Nuovo Cimento* **27**:263 (1980).
14. B. d'Espagnat, *Phys. Rev. D* **11**:1424 (1975).
15. A. Shimony, in *Proceedings of International Symposium on the Foundations of Quantum Mechanics* (Physical Society of Japan, Tokyo, 1983).