Foundations of Physics, Vol. 11, Nos. 3/4, 1981

The Concepts of Influences and of Attributes as Seen in Connection with Bell's Theorem

B. d'Espagnat[1]

Received July 30, 1980

With regard to the notion of cause—or more generally of influence—the various methods of proof of Bell's theorem do not all have the same bearing. The differences between two of these methods are analyzed, with regard to both their conceptual basis and their conclusions. It is shown that both methods give valuable information but, not too surprisingly, the one that is based on the more detailed and specific definition of the concept of influences, and that makes use of the concept of attribute, leads to conclusions that are also to some extent more specific than those following from the other method.

1. INTRODUCTION

In 1935 Einstein, Podolsky, and Rosen (EPR) developed an argument purporting to show that, under very general locality assumptions, quantum mechanics must be an incomplete description of reality.[1] But in 1964 Bell[2] proved the impossibility of completing that theory in such a way as to restore locality. More generally, he and several authors[3–6] finally showed that some correlations between physical events taking place in different spacetime regions cannot be explained in terms of physical events in the overlap of the backward light cones of the two regions (unless the experimenter's freedom of choice is illusory, or some specific predictions of quantum mechanics are false: but these predictions have since been verified experimentally).

Then, if we *do* want to explain these correlations, it seems we must acknowledge the existence of some (direct or indirect) influences propagating

[1] Laboratoire de Physique Théorique et Particules Elementaires (Laboratoire associé au Centre National de la Recherche Scientifique), Université Paris-Sud, Orsay France.

0015-9018/81/0400-0205$03.00/0

faster than light. On the other hand, it was also pointed out by Bell[4] and others[7] that the influences in question cannot be used in order to transmit messages. They therefore lack the direct operational meaning conventionally associated with the concept of signals. This then raises some questions with regard to their nature, and so does also the observation (correlated to the one just made) that these influences do not, in general, propagate energy. In fact, in view of these two remarks some physicists are sometimes uncertain of the real significance of the Bell theorem just alluded to.

There are cogent reasons—which will become obvious below—for asserting that such uncertainties about the nature of the influences in question do not alter the importance and the meaningfulness of the Bell theorem. After all, the latter is *not* usually formulated in terms of influences, or at least not directly so. Nevertheless, it remains true that the theorem in question could not be expressed without some reference to the concept of an explanation going beyond mere description: and that it is difficult to think of a physical explanation that, in some way or other, would not call for such a general concept as that of physical influences. On the other hand, the status of the influences that are involved here has remained somewhat obscure. This is all the more true because, by reason of their velocity, there must exist referentials in which they propagate from the future into the past. But then the question arises: can any cause be posterior to its effects? Or, more precisely, (as Hume pointed out a long time ago!) if two events are in a mutual relationship of influence, how is it possible to define which one of the two influences the other—which one is the "cause" and which one the "effect"—if not by reference to their relative time ordering?

To be sure, these questions are largely matters of definitions. But that is far from making them trivial, for obviously problems of consistency are lurking around. First of all, the definitions must not be circular and they must be consistent with one another: this is what may be called "internal" consistency. Moreover, they should also bear at least some *remote* relationship with the entities to which we give the corresponding names outside physics. If no such relationship whatsoever could be pointed out, then statements to the effect that, for example, "influences" exist between arbitrarily distant (and even space-separated) events would in fact constitute *misleading* assertions about the world [unless it were made explicit that such words are but elements of a code; but then one could question the interest of introducing suggestive code names for entities that (a) would correspond to nothing in our experience and (b) lack direct operational meaning, as we pointed out a few lines above].

The present article aims at examining these problems in detail, not only as regards the notions of *influence*, *cause*, and *effect*, but also as regards that of *attribute* or *property* of a system. It turns out that the question of how to define the latter concept is of interest, not only in its own sake, but also because it plays a role, as will be shown, in at least one definition of the concept of influence.

Our purpose is essentially physical. In other words, we are not interested in the general philosophical problem of defining the concepts in question in their full range and for all cases. On the contrary, we purport to keep inside the smallest possible domain of generality compatible with the demand, that is, with the necessity of defining the said concepts in such a way that their use in connection with the Bell nonlocality theorem is made quite clear. In spite of that restriction, we shall be led to consider, for some concepts, not just *one* definition, but several possible ones. On the one hand, this corresponds to the existence in the literature of several nonequivalent proofs of the Bell theorem.[2] On the other hand, it corresponds to the existence of some free choice as regards the degree of affinity that we require our defined notions to bear with those already having the same names in the ordinary language.

The condition that such an affinity, however remote, should, in any case, be present was already formulated above. In view of the haziness of the concepts of *causes*, *influences*, and the like in the ordinary language it is by no means a trivial objective. At first sight it looks as if it would oblige us to indulge first into semantic analyses bearing on the general use of these key words: a task that falls outside the normal competence of physicists! Fortunately, however, this is not the case, for the simple reason that such analyses are already available. In fact, they are standard and they can be found in the appropriate textbooks. Conventionally, they are (unfortunately!) considered as being parts of the subject matter called philosophy. But, clearly, they have little or nothing to do with *speculative* philosophy, since their purpose is merely to make clear the miscellaneous and blurred notions that our everyday language covers with the words under discussion. In what follows our occasional references to philosophy are just references to these general classifications.

The article is arranged as follows. Sections 2 and 3 investigate the notions of cause (influence) and of attribute, respectively. Section 4 deals with the application of the considerations of Section 2 to a certain method of proof of the Bell theorem, with the purpose of specifying what exactly are the results that this method can establish. It is shown there that that method does establish a violation of a principle rightly called "local causality" but that

[2] A systematic survey of these proofs would be beyond the scope of this article. Here we consider only two of them, but of course we must remember that others also constitute interesting subjects of investigation. We think in particular of the method of proof developed by Stapp.[(8)]

nevertheless it does not allow us to speak meaningfully of "influences." Section 4 shows that by making use of some natural definitions of the concepts of influence and attribute reviewed in the previous section it is possible to take a step further and to really prove the existence of "faster-than-light" influences.

The article is constructed in such a way as to be very explicit. To be sure, this is at the expense of conciseness! Experience, however, shows that in such matters misunderstandings can easily emerge out of misplaced attempts at emphasizing formulas and at cutting off the "plain talk." The main results are described in Sections 4 and 5. Sections 2 and 3 are preparatory and could be skipped in a first perusal of the paper: in fact, their motivations would even appear more clearly in such a light.

2. CAUSES AND INFLUENCES

For the reasons described in Section 1 a short review of the key words and a classification of their various implicit meanings are necessary. These key words are those of cause, effect, influence, causal implication, influential relationship, attributes of a system, and the like. In this section, the notions of cause and influence and the related concepts are examined. The notion of attribute is investigated in Section 3.

Let us begin with the more elementary aspects of the question. Independently of what may be the ultimate meaning of the word "cause," everybody will readily agree that, as a rule, a phenomenon does not have just one cause determining its occurrence. Still, in ordinary language we often refer to causaity as implying some strict implication relationship between "the" cause and the effect. To avoid possible misunderstandings, it is therefore adequate in many cases to speak of *influences* rather than *causes*. To be sure, such a substitution of words does not by itself clarify the actual meaning of the concept, but at least it avoids surreptitiously introducing in it an elementary meaning that in many cases has no reason to be present: spontaneously, everybody thinks of influences as being, as a rule, multiple and nonbinding and this is the right attitude in general.

A second elementary point is that—whatever, again, the "deep" meaning of the involved concepts may be—we should distinguish ordering from mere relationship. When, in ordinary language, we say that event A influences event B in fact we make an assertion that can be decomposed into two steps: (a) we assert that there exists an "influential relationship" (or, loosely, a "causal relationship") between A and B and (b) what is more, we assert that it is A that influences B and not B that influences A. Clearly, it is conceivable that the analysis of the meaning of the concept of influence should be made easier if the analysis of the meaning of each of the two steps is carried out separately.

After these two remarks let us take up the main subject of this section, that is, let us review the main different meanings that have hitherto been given to the word "cause" or to its substitutes. Essentially, they can be grouped under three headings, the *regularity theory*, the *entailment theory*, and the *energy theory*.[3] None of these three theories meets with universal approval, but most experts adhere to one of them.

2.1. The Regularity Theory

This is the theory that is most favored by empiricists, for it purports to define the concept of cause on a purely observational basis. Some authors also call it the extension theory. A first, but oversimplified, idea of it is given by the assertion: "A causes B if and only if event A is always or usually followed in time by event B." Mathematically, this can be expressed as follows. Let us denote by $(x|y, z,...)$ the (conditional) probability of event x if events y, z,..., are realized and let C be some general conditions under which events A and B are supposed to occur. Then, considering an event A occurring earlier than an event B, the above assertion would read

$$\text{“}A \text{ influences } B\text{”} \underset{\text{def}}{=} \text{“}(B \mid AC) \neq (B \mid C)\text{”} \tag{1}$$

where the probabilities are understood as frequency limits. Although such an attempt at a definition does reflect the general inspiration of the regularity theory, it is, however, quite insufficient, as it stands, to capture in a precise way the general ideas of causation, influences, and so on. This is immediately obvious if we remark that were we to accept such a definition literally, then we would be compelled to assert, for example, that in airports the lines forming from time to time at the arrival passport booths *cause* or *influence* the observed increases of the density of luggage in the luggage delivery area. And, on the other hand, we would *not* be able to assert that the 4K thermal equilibrium radiation in the Universe has its cause in the Big Bang. Both these consequences of the definition run counter to what we normally understand by the concept of causation. Clearly, the theory must therefore be refined before it can be considered as acceptable.

In fact, the two examples just given are qualitative illustrations of the two main difficulties that the regularity theory has to cope with. The first one is the fact that in many instances the correlation observed between two

[3] Here the word theory is not used in the sense in which we use it when we speak, for instance, of a physical theory. In fact, it is a shorthand expression for "coherent procedure for constructing definitions."

events A and B is obviously "due" not to the fact that one of them causes the other, but that they are both influenced by common causes (the arrivals of the planes in the example). The second one is that we feel the concept of cause should also be relevant as regards unreproducible events.

Let us begin by investigating the first of these difficulties. In nonrelativistic physics it can be considered as an experimental fact that if A is later than B and if C specifies the set of *all* the events that took place before B, then

$$(B \mid AC) = (B \mid C) \tag{2}$$

Similarly, in classical relativistic physics Eq. (2) is considered as being valid whenever A lies outside the light cone of B, provided that C stands for the set of *all* the events contained in the past light cone of B.[4] Could these facts be used in order to overcome the first difficulty just mentioned and could they therefore help in providing us with a definition of the words "cause" or "influence" that would remain purely within the realm of the regularity theory ?[5]

Intuitively that seems possible: after all, the easiest way of making Eq. (2) plausible to a non-epistemologist is to refer to "the well-known fact that the future cannot influence the past." For that reason it is common practice to say that any theory satisfying Eq. (2) (C being as specified above) is by definition a *causal theory* (or, in the relativistic case, a *locally causal theory*).

This definition of the expression "causal theory" is unambiguous and involves no internal contradiction, so that it is acceptable as such. However (and for that reason the choice of the expression "causal theory" could be misleading), by itself it does not open a possibility of reaching our aim, that is, of defining the concepts of cause or of influence. More precisely, it does not offer us a possibility to build up an operational definition of these concepts. For such a definition would have to take the form "A influences B if and only if X," where X stands for an empirically testable statement or at least for a statement that should be testable in principle. But even if we disregard the problem of the enumeration of the infinite sets of events constituting C,

[4] Classical physics is usually considered as deterministic, so that in fact the probabilities appearing in Eq. (2) are all either 1 or 0; but this is but a secondary point in the present context.

[5] Classical physics possesses time-reversal invariance, at least as regards those of its equations that are usually considered as basic. In other words, it also implies the validity of Eq. (2) with time ordering of events A, B, C reversed. Admittedly, both the common practice here referred to and the proposals that have been made for defining the concept of a locally causal theory (these proposals are briefly reviewed in Section 4) constitute a choice of words that breaks such a temporal symmetry (in a manner chosen so as to meet our intuitive notions concerning the cause–effect time ordering).

we must observe that, notwithstanding our intuition, we have no possibility of building up a consistent X that would satisfy our purpose. It is clear, for example, that replacing X by "$(B \mid AC) \neq (B \mid C)$, where C stands for *all* the events prior to B," would lead to inconsistency (it implies that *no* event A prior to B can influence B, since such an event is a member of class C, so that the two members are equal!): and the same is true if we replace "prior to B" by "contained in the backward light cone of B." Similarly, replacing X by "$(B \mid AC) \neq (B \mid C)$, where C stands for all the events that can possibly influence B" would obviously be incorrect, since then the definition of the word "influence" is circular. Of course, a corresponding difficulty occurs if we try to define the *noninfluence* of A on B.

These difficulties in defining the concepts of cause and of influence have an unpleasant consequence, at least if we keep to the definition of a locally causal theory given above. The consequence is that, if we are faced with a theory which is not "locally causal" in that sense, we cannot assert that that theory implies the existence of causes or influences that propagate faster than light. Or, at any rate, if we formulate such an assertion, it must remain somewhat hazy, since we have not been able to define what we mean by the words we use!

As regards the second difficulty of the regularity theory, it is possible to keep here to just one remark. It is that, as regards unreproducible events, there is no hope of making use of the concept of probability (defined as a frequency limit) in order to define causation on them unless we introduce the notion of counterfactuality, that is, if we accept to consider events that could have occurred but did not occur: such as (for all we know!) *other* Big Bangs, creating other universes (in the example given above).

2.2. The Entailment Theory

In view of the difficulties of rigorously defining the concept of cause and of influence entirely within the realm of the regularity theory, a number of thinkers (see, e.g., Ref. 9) have come to consider that such notions involve an element of *entailment* that is not present in the regularity theory. Obviously, there exists an affinity between that view and the view of the school of logicians who insist that we cannot do without the concepts of necessity, possibility, and so on, in spite of the fact that these concepts are foreign to conventional formal logic. On the other hand, a distinction has to be made between the concepts of *logical* necessity (or logical possibility) and the concepts of *causal* necessity (or causal possibility). Both varieties fall within the realm of the so-called *modal logic.* Unfortunately, that discipline has mainly studied *logical* necessity and possibility. Nevertheless, the importance of studying causal necessity and causal possibility has been stressed, even by thinkers, such as Hempel[(10)] and Carnap[(11)] in his later works, who, on the whole, were favorable to empiricism. Their main motivation was that it is only by explicitly considering such concepts that a logical distinction can be made between laws of nature and mere "accidental generalizations."

This is not a proper place for reviewing the various attempts that have been made to formulate clear-cut definitions of entailment, causal necessity, causal possibility, and related concepts. At the present stage we do not, in fact, need these definitions. We do not even need to make a choice between the thesis that these concepts can and should be defined and the thesis that they should be taken as primitive. We only need to know what *criterion* the supporters of the entailment theory have proposed in order to accompany their definition. For, obviously, it is in any case necessary to propose a workable criterion making it possible to assert, in the cases of interest, that event A causes or influences event B.

Since we are looking not for a definition but merely for some criterion, a complete generality is not required. Also, the criterion may well refer to human actions without necessarily implying that the notion itself is relative to human actions. In fact, the latter question can be left open, at least for the time being. As for the criterion, it is then simply the following (see, e.g., Ref. 9).

Criterion of Causation. Let A and B be two events of a physical nature[6] that is, let them be of such a type that there is no contradiction in considering that they can sometimes happen naturally, without anybody interfering. But let A nevertheless be such that by performing such and such an action we can make A happen at will. Then, we say that A causes B if it is the case that in any instance in which we make A happen, B happens also and in any instance in which we do *not* make A happen (and in which we verify that it did not happen by itself) B does not happen either.

This is in fact the criterion for strict causation. The criterion for the fact that A influences B can be formulated in a very similar way by requiring that B should happen more often (or less often) when we make A happen than when we refrain from doing so. Mathematically, this is expressed by an inequality very similar and formally equivalent to inequality (1), namely

$$(B \mid AC) \neq (B \mid \bar{A}C)$$

[6] An important but difficult question (and one that it is not the aim of the present article to investigate) is to know whether or not the supplementary condition that A be earlier than B should be inserted at this point. If it is *not* inserted, then admittedly the criterion implies, in some cases, that the word *cause* be given a meaning quite different from the one it has in common language. Some of the difficulties that this raises could be circumvented if the notion of "making A happen" could somehow be restricted to a kind of *direct* action from our part, involving no intermediate events.

where $\bar{A}$ means non-A. But the difference with the situation studied above is nevertheless considerable because of the conditions imposed on A by the foregoing criterion: $(B \mid AC)$ is now obtained by measuring the relative frequency of event B when we make A happen and similarly for $(B \mid \bar{A}C)$ (which we obtain by making A *not* happen). This bars the possibility that the correlation between A and B could be attributed to the existence of a common cause and therefore here it is no longer necessary to stipulate that C should stand for a *complete* description of the conditions under which these measurements are done.

To be sure, the criterion in question acquires its whole significance from the implicit assumption that our free will is something real [if, e.g., we were to measure $(B \mid AC)$ and $(B \mid \bar{A}C)$ under hypnosis, then again the objection based on the idea of "common causes" could not be discarded]. But this, after all, is more or less explicitly the case as regards any operational criterion or definition. And in this respect it should be underlined that the criterion here discussed *is* strictly operational. It remains of course to be checked that this operational criterion bears at least *some* relationship with our intuitive notions of causes and influences. That, however, raises no special difficulty. The only point that should be made in that connection is that, obviously, we are not exclusively interested in the cases in which *we* create events A: in fact we are often interested in the general case in which events A just occur or do not occur, possibly quite independently of human action. But then, whatever the source of our notion of entailment is (again, it would be inappropriate to enter here too far into that philosophical subject) the following must be acknowledged. When we assert that some event A "necessarily entails" event B we do not merely mean that *if* we do create A, then B will follow, but we mean also something more. In particular, we undoubtedly also mean the "counterfactual" assertion that if we did create A (although in fact we do not) then B *would* follow. It is in part through that channel that the hypothesis according to which some counterfactual assumptions make sense enters the entailment theory of causation; of course similar remarks also hold with regard to the (similar) theory of influences.

Once such an assumption is made, the second difficulty met with in the regularity theory also disappears for the very reason that was invoked at that place (with the proviso that we accept also the idealized view that we could always create A if we liked, a view that in the case in which A is the Big Bang is admittedly *extremely* idealized!).

2.3. The Energy Theory

When a physicist gets involved in a discussion bearing on the notions of cause and influence it is very likely that at some point or other he will remark that "anyhow there can be no real physical causation without some energy transfer." If we are not very strict about the use of the word "theory," we can say that this remark is perhaps the embryo of an "energy theory" of causes and influences. However, in spite of the great use theoretical physics makes of technical terms such as microcausality, primitive causality, and the rest, any such program of a theory of causality must be considered, at best, as a vague and partial expectation, since obviously the foregoing prescription is essentially negative. It stipulates that whenever there is no energy exchange between a system at a certain time and another system at another time we should not be allowed to speak of a causal (or "influential") relation between the two. But clearly, it cannot be the case that whenever one system loses and another one gains energy there exists a causal (or an influential) relationship between these systems: so that the prescription in no sense opens the way to real definitions of the words "cause" or "influence."

To summarize: both the regularity theory and the energy theory offer means for making apparently sensible assertions that presumably are somehow relevant to the concept of causation and influence. The former one even makes it possible to define a class of theories that it is, in some respects, appropriate to call "locally causal theories" and to which classical relativity theory belongs. It is only the entailment theory, however, that seems to be able to capture the essence of what we really mean when we speak of causes or influences. It is only within that theory that we could specify an operational criterion. To be sure, that criterion has no universal validity, since it applies only to "causes" (events A) that are in some cases at least at our disposal. Also it is not entirely free of circularity, since it treats as primitive the concept of human action (the act of creating event A), which after all is itself but one form of causation.

3. ATTRIBUTES

For reasons that will be explicated in due course, it is not possible to apply straightaway the results of the foregoing section to a dicussion of the epistemology of the Bell inequalities. Before that, it is necessary that the concept of attribute, or property, of a system should be examined and that the main theories concerning it should be discussed. There are essentially three of these. For convenience, let us call them the "partial definition" theory, the "preparation" theory, and the "counterfactual" theory.

3.1. The "Partial Definition" Theory

In this theory the attributes of physical systems (and in particular those of microphysical systems) should be defined essentially by the same method as that advocated by the bulk of modern epistemology for defining the so-called "dispositional terms," namely by the method of "partial definition" based on *bilateral reduction sentences* (considered in particular by Rudolf Carnap). Let it be recalled that the method in question is aimed at avoiding altogether the introduction of modal notions such as the notion of entailment; and that, to that effect, it defines the dispositional terms (e.g., the term "magnetic") *only* in the cases in which an "instrument" (e.g., a small iron object in the example) is present, the behavior of which is different according to whether the system under consideration has the attribute (is magnetic) or not.[(12)] Similarly, in full accordance with Bohr's emphasis on the importance of the notion of instrument, the "partial definition" theory of the attributes of microsystems defines the latter on a system S only in the cases in which a definite instrumental setup is present and is of one of the types that allows for a measurement of that attribute.

Let us agree that the value a of an observable physical quantity A on a system S *is* an attribute of S, and let us say that "the conditions P' are realized" when and only when an instrumental setup of the kind that allows for a measurement of observable A on S is present, together with S. Then, it should be clear that, according to the present theory, the assertion "A has the value a" (where a is one of the eigenvalues of the operator associated with A) *has no sense* whatsoever—is neither true nor false!—in the cases in which the conditions P' are *not* realized. In conventional quantum mechanics the necessity of avoiding the well-known difficulties connected with the notions of incompatible observables imposes that the foregoing remark should be taken particularly seriously. But it then has a significant consequence. It is that no attribute thus defined can be thought of as a being a property of S in the usual sense. By "a property in the usual sense" is here meant an attribute with some openness of meaning[(12)]: for example, we consider that some physical systems prepared in appropriate ways can meaningfully be said to be "magnetic" at a certain time even in the cases in which neither small pieces of iron nor any moving electric circuit (nor any other known "instrument" by means of which we can operationally test its magnetization) is then at hand. Within the realm of the "partial definition" theory such an attitude is possible only because we implicitly leave open—without any precise analysis! —the possibility that some other instrumental tests should then conceivably exist. In fact, that attitude is possible only because we consider it as legitimate to think, *anyhow*, as if such possibilities existed.[7] But in the case of quantum physics such an openness of meaning would, as is well known, lead to difficul-

[7] This is especially obvious in the case of certain dipositional terms, such as *soluble*, for in that case a sentence such as "this piece of sugar is soluble," if it has any sense at all, must clearly have a meaning even when the piece of sugar is not immersed in water!

ties or paradoxes: and, if we rely on Bohr's epistemology to save us from these paradoxes, then we must bow to the latter's instructions, which are akin, in a way, to a refusal of the openness in question. If a particle has been prepared in some accelerator and then sent into a room in which the measurement device that is present is not one that measures momentum but one that measures position, then—according to the most reasonable interpretation of Bohr's writings—we are allowed to think of that particle as having a definite *position* (presently unknown but that will be revealed by the measurement in question): and consequently—because of the uncertainty relation—we are in no case allowed to think of it as having a definite momentum, not even in the cases in which the particle has been prepared by an accelerator that, in classical language, "endows the particles with a definite momentum" (at any rate this is the only meaning we are able to give, in the present case, to Rosenfeld's interpretation of Bohr's epistomology as expressed in the sentence "it is now the indivisible whole formed by the system and the instruments of observation that define the phenomenon"[13]). The fact that in quantum physics we must thus give up, at least to some extent, the openness of meaning alluded to above is what differentiates most the "partial definition" theory as applied in quantum physics from that same theory as made use of in the conventional "Carnapian" epistemology. It is also what differentiates the notion of *attributes* as used in quantum physics from the notion of *properties* as defined in the epistemology just mentioned. In other words, as Stein[14] writes: "It would ⋯ be misleading to think of the quantum mechanical attributes as properties of a system in the ordinary, logical sense of 'property': on the domain of all possible conditions on the system, they are best conceived of as 'partial' or 'conditional' properties—functions into the set of truth values, defined on (*possibly proper*) subsets of that domain" (our italics).

Thus conceived, the "partial definition" theory may be considered as being the cornerstone of what is often called the "orthodox" interpretation of quantum mechanics. Nevertheless, it has some unpleasant features. To mention but one of these, it has the consequence that a system S has a definite value of any observable A that can be defined on it (this is due to the fact that we can define A only in the cases in which an instrument is present that is capable of measuring A: but then, as mentioned above for the case of the *position* attribute, according to this theory A can always be thought of as having a definite value, known or unknown). In other words, here, just as in classical physics, the statement "A has a definite value" is either true or meaningless but never false. At any rate, this is what a strict requirement of internal consistency of the theory would force us to assert. For example, in that theory we should say for consistency that any system in which energy can be defined has a definite (possibly unknown) energy. This, however, is not, as a rule, what we assert when we describe quantum physics. On the contrary, we say, for instance, that a system on which some (possibly approximate) measurement of position has just been made has no strictly definite energy (and we write its wave function as a *superposition* of waves corresponding to various energies). This shows that the "partial definition" theory is not consistent with a practice of language which is quite common in quantum physics. This practice, to be sure, is not essential; however, it is so useful that its inconsistency of principle with the theory is a serious disadvantage of the latter.

Some other difficulties of principle of the partial theory have been pointed out by other authors.[(14,15)] It is not necessary to review them here, however: what has already been pointed out is sufficient to show that an inquiry into whether other theories are conceivable is justified.

3.2. The "Preparation" Theory

In this theory the attributes are defined by means of the mode of preparation of the system: if, upon measurement of an observable A on a system S a value a_k is found (one of the eigenvalues of the operator associated with A), then, immediately afterward, the system S *has* attribute a_k (at least if the measurement of A satisfies some ideal conditions that actual instruments can realize only within more or less reasonable approximations).

The interest of this approach is quite obvious. In fact, it is the one that matches most straightforwardly with the elementary formalism of quantum mechanics, based on the use of the wave function and of the projection theorem.[8] It thus corresponds quite closely with the world view according to which the wave function somehow constitutes a complete description of whatever may be called "physical reality." However, the approach in question also meets with difficulties when what is aimed at is a real definition of the word "attribute," a definition that would apply to all the cases in which we consider that a system S has such and such an attribute. The main problem is that, at least in the case of macroscopic systems, there exist many circumstances in which we must say that a system has some definite (possibly unknown) attribute, although it was not subject to any previous act of measurement. In fact, this also holds true for microscopic systems, at least as regards some observable such as energy, for we all agree that at low temperature the atoms of a gas are normally in their ground state, even though we have measured the energy of none of them. Even if we somehow manage

[8] Often called "postulate." But the "postulate" in question can be derived from the assumption that immediately after the measurement, S should be describable by *some* wave function: for that wave function can then only be an eigenfunction of A corresponding to eigenvalue a_k, since otherwise the probability that a second measurement of A would result in the same value would not be equal to 1.

to overcome the latter difficulty (e.g., by asserting that energy is "something special"), the problem remains acute as regards the macroscopic systems, for it seems that some other definitions of the notion of attribute should exist for macroscopic systems, and for macroscopic systems exclusively; and this in turn seems to imply the existence of qualitative and objective differences between microsystems and macrosystems (if the said differences were merely subjective, we should be able to extend *in principle* that other definition also to the attributes of microsystems, even if it could not be used there *in practice*)

3.3. The Counterfactual Theory

The counterfactual procedure for defining the attributes of microsystems is the method that resembles most closely the naive conception of what an operational definition of a property of a system should be. In that procedure the definition (for future reference let it be called "counterfactual definition") is as follows.

Counterfactual Definition: It is said that a system S has, at time t — 0, an attribute $A = a$ if and only if it is true that if a measurement of A were made at time t the value a would be obtained.

What is remarkable in that definition is the use of the subjunctive mood. As one knows, it is impossible to replace the latter by the indicative tense (...if a measurement of quantity A is made,..., then a is obtained), for the classical difficulty associated with some uses of the so-called "material" implication (if,..., then ...) would turn up: the definition would then mean, in particular, that A equals a on S in every case in which no measurement of A is made! It is equally impossible to replace the subjunctive by the future (when a measurement of A *will* be made,..., a *will* be obtained) without implicitly falling back on the "partial definition" theory. The subjunctive is thus necessary.

Within this method, the attribute of S consisting in the fact that the observable A has, on it, the value a is thus defined with the help of a proposition in the type "if,..., then ..." in which the premises can be "contrary to the facts." In other words, the proposition may have a meaning even in the cases in which no measurement of A is made: this is the reason why it is called "counterfactual."

The notion of counterfactual statements, or propositions, was already touched upon in Section 2. In particular, it was noted there that such a notion is closely linked with that of causal necessity: and that some use of the latter concept seems unavoidable to many leading epistemologists in order to establish a logical distinction between laws of nature and mere accidental generalizations. Clearly, under such conditions, there is no good reason for not considering the use of such concepts also in other domains of epistemology, and, in particular, in the present problem, provided that sufficient care be exercised.

Necessarily true propositions are sometimes called strict conditionals and, as noted above, there is in general a close similarity between counterfactual propositions and strict conditionals. In the particular case of the causal counterfactuals and strict conditionals that will be considered in this article (i.e., in the case in which the conditions are specified by the settings of instruments) the similarity even amounts to an identity (see footnote 10 below).

Let S be a system (microscopic or macroscopic). And let us formulate the working hypothesis *that S can be thought of separately from the rest of the world.* This hypothesis, which may be called the "splittability" hypothesis, means that, at least to some sufficient approximation, the various parameters (in the broadest sense of the word) that specify the state of the world at a given time are in some way partitionable (at least conceptually) into two classes: those that refer to system S itself and those that refer to the rest of the world. Our working hypothesis correlatively means also that, by modifying by thought the external parameters (those of the second class) but not the internal ones (those of the first class), it is possible to imagine state of affairs in which the *same* system S is subjected to *different* "external conditions."

To be sure, the precise nature of all these parameters as well as the detailed way in which they split into the two classes should, at some later stage, be specified. But, for the time being, we are merely postulating the principle that such a splitting is conceptually conceivable (our motivation being that, unless we are satisfied with such vague assertions as "everything is in everything," it would seem that any investigation must, explicitly or not, use a principle of such a kind). Then, let us, for short, call the set of values of the world parameters including those relative to S at a time t (in a given referential if relativity is considered) *a situation*[9] of S. And let us say that, as regards S, a situation is *accessible* from a situation i if the parameters that refer to S (the "internal" parameters in our language) are equal in both. With these conventions it is possible to define the precise meaning that we give here to the (causal) necessity (or certitude) operator N:

Definition. Let p be a proposition bearing on S and let i be a situation of S. Then, the proposition Np (it is certain that p) is true on S in situation i if and only if p is true in all possible situations j that are accessible from i as regards S.

[9] Here the word "situation" does not imply any particular reference to space. It is just a substitute for the word "state," which is usually given more specific meanings.

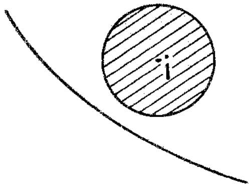

Fig. 1. The shaded area is the accessibility sphere of i.

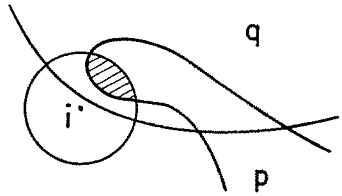

Fig. 2. The shaded area is the intersection mentioned in the text.

Consequence

$$N(p \supset q) \tag{3}$$

where $\supset$ is the usual symbol for material implication, is true on S in situation i if and only if q is true in all the situations that are accessible from i and in which p is true. $N(p \supset q)$ will be read as "p entails q on S."

In order to describe graphically this definition and its consequence, let us follow the method of Lewis[16] and to the pair constituted by the symbol N and the situation i let us associate a set S_i of situations, called the *accessibility sphere* of i, the elements of which (points) are the situations accessible from i. A proposition such as Np is then true for situation i of S (or more briefly, "true on i") if p is true within the whole accessibility sphere of i (see Fig. 1). Similarly, the proposition $N(p \supset q)$ is true on i if, as in Fig. 2, q is true for all situations within the intersection of the sphere of accessibility of i with the set of situations in which p is true.

As Fig. 2 shows, $N(p \supset q)$ can be true on i without p itself being true on i. $N(p \supset q)$ is what we shall call a counterfactual proposition for S in situation i. It explicates what we mean when we assert that "S being in situation i, if p were then true, q would be true also."[10] Instead of $N(p \supset q)$ we shall often use the simpler symbol

$$p > q \tag{4}$$

With the help of this symbol it is an easy matter to formalize and make precise the counterfactual definition given at the beginning of this subsection. Let p be the proposition "some instrument that is fit for measuring observable A is present at time t" and let q be the proposition "the value a is found." Then the definition in question takes the following form.

Counterfactual Definition Formalized

$$\text{“At time } t - 0,\ A = a \text{ on } S\text{”} \underset{\text{def}}{=} (p > q) \tag{5}$$

[10] For counterfactuals *in general* the situation is not quite so simple and, as shown, e.g., by Lewis, they cannot quite be identified with strict conditionals. But here we only deal with causal counterfactuals and then our assumption that the set of all the world parameters can be partitioned as specified above simplifies the problem.

This definition can be used freely, provided only that the splittability assumption formulated above is assumed to be valid. As one sees, it is meaningful even in the cases in which no instrument is actually present.

Two final remarks are appropriate.

Remark 1. This remark concerns the fact that an attribute—or property—of a system may exist (or "have a given value") at certain times and not at other times. As a rule the formulation just given, and which involves the two infinitesimally different times t and $t - 0$, is the most appropriate one. But in some cases it may be more convenient to consider noninfinitesimal time differences. For example, assuming for simplicity that we are at the equinox and standing at the equator, we may define the meaning of the proposition "the sun is now at the nadir" by saying "if, after waiting exactly twelve hours, we pointed a telescope toward the zenith, the image of the sun would form in it." Of course such definitions presuppose that between the times they involve neither a cataclysm nor, more generally, a perturbation of the system takes place.

Remark 2. In ordinary language the disjunction of several properties is also a property: for a human being, for example, the fact of having either year 1966 or year 1967 or etc. or year 1980 as the year of birth is in itself a property, independent of whether or not it has a name ("being now a child" in this example) and independent of whether or not the procedure for ascertaining the existence of the property necessitates a detailed check of the existence of one at least of the elements in the disjunction (in the example we may proceed in just one act, by looking at the individual in question; in other similar examples we would have to look successively at each birth register for the years of interest). Within a precise scheme, such as the one considered here, for defining the attributes of a system, it would be both difficult and arbitrary not to follow the same approach. In other words, it seems imperative that if several attributes (or properties) of a system S are defined by the counterfactual propositions.

$$p > q_1 , \qquad p' > q_2 ,\dots,$$

the disjunction

$$(p > q_1) \vee (p' > q_2) \vee \cdots$$

where the p, p',..., are equal or different, should also define an attribute (or property) that S can have. The only conceivable exception is the case in which the union in question would be "equal to 1," that is, would constitute a tautology.

4. LOCAL CAUSALITY

As mentioned in Section 1, the problem on which the present article is centered can be summarized as follows. If we want to *explain* the correlations involved in the Bell theorem, then, since we cannot account for them in terms of physical events in the overlap of the backward light cone of the two measurements, it seems that we have no other choice than to put forward the notion of influences propagating faster than light. On the other hand, as shown in Section 2, that very notion of influences is far from being as elementary as one might think. In fact, it is, especially in the present context, so ill-defined that, when engaging in a proof of Bell's theorem or analyzing its content, it looks safer to keep the notion in the background as much as possible. But then attention must be paid to the nontrivial problem of stating what the real content of the theorem precisely is.

To explain what we mean, let us consider somewhat in detail the epistemological content of what may, at least for the purpose of the present paper, be called the Bell, Clauser, Horne, Shimony (BCHS) proof. We give that name to a proof that, with some qualifications that it is not necessary to underline here, may be said to emerge from independent works of Clauser and Horne(3) and of Bell(4) together with the written exchange of ideas between Shimony and these authors.(5,6) Here, we are not particularly interested in the finer details that motivated the exchange of views in question, but rather by the central idea that is common to all these articles. This idea is called "local causality" in these papers but, for reasons of convenience that will appear shortly, we introduce also the alternative name "special noncorrelation assumption" to designate it. However, apart from that question of of semantics, we reproduce here literally the formulation that, following Bell the authors Clauser, Horne, and Shimony give of that assumption.

Special Noncorrelation Assumption ("local causality" assumption in BCHS terms). Let B be a variable beable[11] localized in some spacetime region, large or small. This region has a unique backward light cone; let C denote all the beables in this backward light cone. Then[12]

$$(B \mid A, C) = (B \mid C) \tag{6}$$

holds for any beable A localized in spacetime regions with a spacelike separation from the region of B.

[11] Bell calls *beable* any physical quantity, whether directly observable or not, that can be assumed to *be there*. Beables must include the settings of switches and knobs on experimental equipment and the readings of instruments, but they may include other things as well.

[12] The notation for probabilities and conditional probabilities is explained in Section 2.

In the quoted literature Eq. (6) is combined with a natural assumption on the independence of the beables of the source on the settings of the instruments used in the experiments actually done for testing the (generalized) Bell inequalities.[13] It is then shown that the latter inequalities follow from these two premises. In view of the violation (both observed and predicted by quantum mechanics) of these inequalities and in view of the plausibility of the second assumption just mentioned, it is concluded that the first one is presumably the one to blame or, in other words, that "local causality" is a faulty concept.

Here, this conclusion is fully accepted (with just the one small provisional reservation about the name "local causality") and an attempt at analyzing its content is made.

Let us first remark that, as mentioned in Section 2, the probabilities appearing in Eq. (6) are to be understood as limits of frequencies: this merely reflects the fact that ultimately the consequences of Eq. (6) and of the other assumption are tested by measuring the relative frequencies of various events. More precisely, the quantities that are measured are correlations. Hence, equality (6) applies in fact to correlations or absence of correlations. This is the reason why the assumption of which (6) is a part deserves a name emphasizing the role of that concept.[14]

Our next remark is that (as stressed, e.g., by Bell[(6)]) the events considered in Eq. (6) are of two types. Some of them are under human control: this is, for example, the case as regards the positioning of an instrument at a given time. Others are not: this is typically the case as regards the results of the measurements. In (6) the events B are all of this second type, whereas events C are of both type. The important fact in this respect is that the events labeled A in Eq. (6) also comprise events of both types. If we restricted condition (6) by requesting its validity only in the cases in which the events A are all of the first type (positioning of instruments and so on), then the Bell inequalities could not be derived [indeed, elementary quantum mechanics without hidden variables does satisfy such a "truncated condition (6)," since C is then just the wave function of the composite system and since the probability of a given result on one of the instruments is then independent of the posi-

[13] Also called the CHSH inequalities. As is well known, the Bell inequalities were generalized by Clauser, Horne, Shimony, and Holt[(17)] and it is only these CHSH inequalities that have been tested. As regards the "natural assumption" mentioned here, it is not necessary in this article to review its different formulations. Let it merely be stressed that, in some form or other, it cannot be dispensed with. Finally, the CH derivation is justified on the grounds of locality and reality assumptions similar rather than identical to those expressed by Eq. (6).

[14] The word "special" serves to emphasize that the validity of Eq. (6) is only postulated under the rather special conditions mentioned in the text.

tioning of the other one]. This remark is not of direct use here, but it will become essential in Section 5 and this place is the appropriate one for making it.

Now a question that can hardly be avoided is, "why do we consider that the consequences of the *special noncorrelation assumption* are worth investigating?" Of course, a theorist can always decline answering a question of this kind. He can, with good reason, uphold his right at considering any hypothesis he likes and investigating its consequence. But in the case that this should be his answer, it would be appropriate to inquire from him why he chooses to give to that assumption the name "local causality." If his answer to that second question were to be that this name is arbitrary, we would have to conclude that the experimental violation of the Bell inequalities merely informs us of the falsity of the assumption called "special noncorrelation assumption" above. As already stressed, such a piece of information bears—as it stands—exclusively on correlations. It is unquestionable, but it can hardly be considered as bringing new significant information, especially since the formalism of quantum mechanics has long since accustomed us to consider types of correlations that do violate the assumption in question. But, more likely, the answer of the theorist would be that the name "local causality" was not just chosen at random and that, on the contrary, it was chosen because the assumption it designates gives a precise content to some intuitive idea we all have on causality. This, at any rate, is the type of answer that seems to be favored by BCHS, as is manifest from several sentences in these papers.

Is it possible to make this kind of approach somewhat more precise? In view of the difficulties reviewed in Section 2 concerning the problem of defining the notion of cause, it seems that the best way of proceeding (perhaps even the only possible one!) is to reformulate the approach in question in the following way: "We do not really know how to define causality; at any rate we require no precise definition of it. But, whatever that definition might be, a restriction seems natural: the definition should be such that events A in Eq. (6) should not be causes of events B [with the definitions of events A and B given just before and just after Eq. (6)]."

We agree of course that such a restriction does in fact "seem natural." But why does it? And, what is more, why exactly does it lead to Eq. (6)? The answer to the second question lies obviously in the postulate that no systematically occurring statistical correlation can be purely accidental, and that whatever is not accidental "has a cause," so that the restriction (which bears on the possible causes) is relevant for the correlations considered here. The answer to the first question is somewhat less factual. Briefly, the restriction "seems natural" because it is known that neither energy nor utilizable signals travel faster than light and because it is a simple hypothesis to extrapolate this to any cause whatsoever, even at a stage at which we do not know what "a cause" exactly is.

Let us emphasize once again that we consider this whole conception as being a priori extremely reasonable, and that, as a consequence, the proof by BCHS (via the violation of the Bell inequalities) that what we called here the "special noncorrelation assumption" is violated seems to us significant also as regards the problem of causality. But its significance is essentially negative. It allows us to state that "local causality" is violated (both in quantum mechanics and in the real world) if by the words "local causality" an assumption is designated that bears, in fact, merely on correlations but that, however, is strongly motivated by the vague but persuasive conceptual links it has with known facts about signals and energy propagation. As it stands it does not allow us to formulate a positive statement to the effect that causes (or influences) can propagate faster than light: for, obviously it is not possible to assert that something exists if that something is not defined. The fact that BCHS do not express their conclusions in terms of influences has most certainly no other origin.

The problem that will be faced in the next section, and which is the central subject matter of this article, is the following one: is it possible to take a step further, and to render meaningful (and true) the statement just considered ?

5. INFLUENCES AND BELL'S THEOREM

Nowadays, whatever the philosophy a physicist adheres to, his methods are—quite appropriately!—heavily based on operationally defined concepts. Even though the epistemologists did show that *pure* operationalism is not quite sufficient to ground scientific research, still they did show also that scientists should, as much as possible, use operational definitions, operationally meaningful statements, and so on: the physicists did catch the lesson. In a way, the work of BCHS, as analyzed in the previous section, offers a good illustration of this. Obviously, this work is not compatible with the philosophy of the most radical operationalists, for the term "beable" (or equivalent notions) plays quite a central role in it: and it is a term that would be considered as meaningless by the radical operationalists. Nevertheless, the work in question is as faithful as possible to the *methods* of operationalism. This is reflected essentially in the fact that, in it, "local causality" is defined in terms of (absence of) correlations between observed events. This is also the general inspiration of the method of defining causality that was analyzed in Section 2 under the name "regularity theory."

The great advantage of this procedure is that it lays the emphasis on, and it requires a quantitative analysis of, only the entities that we really have a grip on. And correlatively, it succeeds in keeping so to speak in the background, and in mentioning, only qualitatively, the entities that lie, partly or totally, outside the realm of what is strictly operationally definable, such as beables, causes, and so on. This is sound scientific methodology. But, of course, such an advantage has unavoidably a counterpart, which, in the present case, is, as we said, the impossibility of meaningfully asserting the existence of faster-than-light influences. We are therefore in a situation which, though, properly speaking, it is not paradoxical, still is somewhat strange: for on the one hand, we consider as quite significant our assertion of the impossibility of *explaining* the observed correlations by referring to "common causes"[15] (which means that the verb "explain" has for us a meaning), but on the other hand, we consider we cannot meaningfully go as far as asserting in a positive way that they must then have some other explanation. And yet we reject the idea that these systematically observed correlations should be "purely accidental."

Clearly, if we want to make an attempt at asserting some positive explanation we must first have a way of defining the concept of physical cause or influence. And, as pointed out in Section 2, this is only possible if we adopt the procedure that was called there "entailment theory." In the present section it is that possibility that is explored.

At first sight, a direct application of the procedure in question to the BCHS method may seem possible and straightforward, for, as pointed out in Section 2, the entailment and the regularity theories both make use of the same relationships between probabilities, namely relation (2) and its opposite, which is just relation (6) of BCHS. A more careful examination of the problem shows, however, that the situation is not so simple. The difficulty stems from the remark made in Section 4 that within the BCHS method the events A appearing in Eq. (6) must include both events that are at human disposal (positioning of instruments and so on) *and also events that are not* (the readings on the instruments). The reasons why this remark reveals a real difficulty is that, as explained in Section 2, any application of the entailment theory must make use of *the criterion of causation*, and, in this criterion, it is stipulated that events of type A are such that, by performing such and such an action, we can make A happen at will. Since this is not the case as regards the readings of the instruments, the criterion of causation is not applicable to the BCHS method as it stands.

Admittedly, one of the great advantage of the BCHS approach is an advantage of simplicity, due to the fact that it considers only measurable events or, in other words, that it has no need to introduce (or even define)

[15] To common causes propagating with not faster-than-light velocities.

the concept of an attribute of a system. On the other hand, it seems that, when we try to introduce in a positive way the concept of (faster-than-light) influences, the difficulty encountered by the BCHS method and which has just been described is quite definitely with us. Does this mean that there is no way of applying the criterion of causation to the problem raised by the Bell inequalities ? Does it mean that we shall never be able to state meaningfully, as a conclusion of our investigations about these matters, that faster-than-light influences exist ? No, we claim it does not. But the only way of reaching these goals that we are aware of is to choose a procedure which is different (and less general) than the one of BCHS, a procedure that relies quite heavily on the notion of attributes, or properties, of microsystems, and on the analysis of these notions as carried out in Section 3.[16]

The intuitive idea of the method is elementary and it is by now quite well known. It consists in considering only the distant correlation experiments in which a *strict* correlation is predicted by the quantum rules. This is the case with pairs of spin-1/2 particles U and V created in a singlet state. After the component along a direction a of the spin of U has been measured at a time t, the component along a of the spin of V has a definite value. This, then, is an attribute of particle V. Particle V cannot be considered as having possessed that attribute already before time t since this assumption would entail the Bell inequalities. Hence, the attribute was imparted to V precisely at time t by the measurement carried out on U at that time. In other words, an influence has propagated instantaneously from U to V at time t.

The problem is to investigate whether that intuitive argumentation can be turned into a rigorous one. This requires in particular that the key terms of the latter, namely "attribute" and "imparts" (or "influences") should be defined and applied to the situation under consideration. It also requires that the applicability of the criterion of causation should be checked. This, in turn, necessitates an appropriate specification of the notion of an "event." Let us investigate separately these various points, beginning with the last one.

Events. Of course it is very easy to give examples of events. The positioning of an instrument at a given time, and the measurement made by an instrument at a given time obviously are "events." But they are macroscopic ones. The decay of a particle at a given gime is a microscopic phenomenon that can presumably also be considered as an event. But, supposing we can meaningfully assert that a particle acquires at a given time an attribute, or property, that it had not, can we generally say that this constitutes an event ?

[16] The procedure in question was hinted at in the very first article of Bell.[2] In several articles[7,18,19] the present author has attempted to formalize it and to elucidate its real epistemological content.

The answer to this question is of course in part a matter of choice: we are free to define at whim the class of what we call "events." However, to the extent that we agree to call a radioactive decay an event, and to the extent that we can *meaningfully* define the acquisition in question (the word meaningfully is of course essential), there does not seem to exist any cogent reason for not also calling that acquisition an event.[17] The problem therefore is: can we meaningfully define such a process? Within the method of definition of attributes that was called the "counterfactual theory" in Section 3 the answer is yes. But, to show this we must first comment on the method in question and on its applicability to this problem.

Attributes. Let us consider once more the *counterfactual definition* of attributes of a system S given in Section 3. It is based on a counterfactual proposition, namely "if A were measured at time t, the value a would be found." Moreover, an examination of the nature of such causal counterfactuals has shown that, essentially, they are strict conditionals (see Section 3). So that, finally, we could formalize the counterfactual definition in question by means of Eq. (5).

Equation (5) has a very significant consequence. Again let p be the proposition "some instrument that is fit for measuring observable A is present at time t" and q be the proposition "the value a is found." Let $\bar{q}$ be the proposition "one of the eigenvalues of operator A that differ from a is found." Since the result of the measurement of A can only be one of the eigenvalues of the corresponding operator A (for which we assume a discrete spectrum for simplicity), we have (to be sure!)

$$(p \supset q) \vee (p \supset \bar{q}) = 1 \tag{7}$$

where the symbol $\vee$ means "or" and where the symbol 1 means a proposition that is alway true (tautology). But, and this is the point we want to make, in modal logic (e.g., Ref. 20) we have on the contrary

$$N(p \supset q) \vee N(p \supset \bar{q}) \neq 1 \tag{8}$$

In the case in which the symbol N stands for *logical* necessity this inequality corresponds to the fact that there exists *contingent* material implications and that if $p \supset q$ is one, then also $p \supset \bar{q}$ is one, so that neither $p \supset q$ nor $p \supset \bar{q}$ is necessary. In the case in which N stands for causal necessity (the case which is of interest here) the inequality has a similar interpretation and in terms of the definitions of $N(p \supset q)$ (see Section 3) it corresponds to the fact that there

[17] Events of such a type are called "transitions from potentiality to actuality" by Shimony.[21]

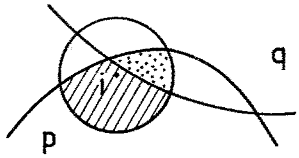

Fig. 3. Shaded area: situations j where p is true and q is false. Dotted area: situations j where p is true and $\bar{q}$ is false.

exists situations i such that some situations j accessible from i and in which p is true are such that q is false and some other are such that $\bar{q}$ is false (see Fig. 3 for an example). In the case considered here this simply means that the world (including the system S) may be in a situation i such that if the situation were the same as regards the system *and* if an instrument fit for measuring A were present, then the result of the measurement might be either a or $\bar{a}$ (depending on pure chance, or on hidden variables in the instrument or what not), so that neither result a nor result $\bar{a}$ is certain.

Since the proposition

$$Q \underset{\text{def}}{=} N(p \supset q) \vee N(p \supset \bar{q}) = (p > q) \vee (p > \bar{q}) \tag{9}$$

is not a tautology, it defines (see Remark 2 at the end of Section 3) an attribute that S can have under appropriate circumstances.

Applicability of Causation Criterion. Let us again consider the system of the two spin-1/2 particles $U + V$ in a singlet state at times at which they are distant from one another. Let us call t_0 the time at which they separated and let us assume that at a time t_1 we have the liberty of making U interact (or not interact) with an instrument I_a that is fit for carrying out a measurement of the spin component of U along a unit vector a, whereas V interacts at that time with no macroscopic system whatsoever. Let p be the proposition "I_a is set in position at time t_1." Moreover, let J_a be an instrument fit for measuring the spin component of V along a and p_v the proposition (which may, of course, be true or false) that J_a is set into position at a time $t_2 > t_1$. Finally, let $q_v(\bar{q}_v)$ be the proposition "upon measurement at time t_2 the spin component of V along a is found to be equal to $+1/2$ $(-1/2)$," and let us define in the lines of definition (9) the following disjunction of counterfactuals:

$$Q_V = (p_v > q_v) \vee (p_v > \bar{q}_v) \tag{10}$$

Before time t_1, Q_V is most certainly not true in general (that is, for any direction a), for if it were, then the Bell inequality would follow, according to a simple and well-known deduction.

Let us consider a direction a for which Q_V is not true before t. As already said, we are at liberty either to set or not to set instrument I_a in position at time t_1. Hence, the presence of I_a at time t_1 is an event of the type of the events A considered in the *criterion of causation* (see Section 2). Let us assume that no measurement at all is ever made on V. Nevertheless, if I_a is set into position at t_1, then at any time t_2 after t_1 proposition Q_V is true, whereas on the contrary it remains untrue in the opposite case. Hence, if (and only if) we set I_a into position at t_1 does V suddenly acquire at t_1 an attribute—the one defined by Q_V—that it did not previously have. According to what was pointed out above, this constitutes an event. Moreover, it is an event that is of the type of the events called B in the criterion of causation. In any instance in which I_a is set into position (event A), this latter event happens also and in the cases in which we do not set I_a into position it does not happen. Hence, applying the criterion of causation, we may say that setting I_a into position (at a time and place at which it is certain that it will interact with U) *causes* the event constituted by the acquisition of Q_V by V.

Since these two events are distant and simultaneous, the conclusion is that, within the entailment and counterfactual "theories," the existence of influences (and here, even, of strict causations) propagating instantaneously has been established, under, of course, the assumption that the correlations that would be measured by ideally selective and accurate instruments would be as predicted by the quantum rules.[18]

6. DISCUSSION

Two points seem to emerge in a particularly vivid way from the foregoing sections. One is the fact, established especially clearly by BCHS, that a certain special noncorrelation assumption is violated (both by the

[18] The introduction of counterfactuals in this problem has turned out to be a nontrivial operation. In order to stress in the simplest terms its distinctive features, a nonrelativistic description (in terms of absolute time) was found convenient. It seems, however, quite possible to take relativity into account. For that purpose we should start from the fact that Q_v is not *generally* true but that it is true when instrument I_a is present for *any* potential local observer (i.e., for a local observer at rest in any reference system). Then, if H is the hypersurface separating the region of spacetime where Q is true from the one where it is not, the fact just mentioned shows that H, which is just $t = t_1$ in the nonrelativistic version, should be changed to the past light cone of the U–I_a interaction event in the relativistic one. This changes nothing of the conclusions.

predictions of quantum mechanics and by the experimental results) and that, for reasons explicated in Section 4, the assumption in question rightly deserves the name of *local causality* given to it by these authors. Hence, we can assert that local causality is violated: moreover, we can ground such an assertion not only on our confidence in the verifiable predictions of ordinary quantum mechanics, but also on (highly reasonable interpretations of) actually existing experimental results.

The other point is related to the fact that, strictly speaking, the developments just referred to do not really allow us to speak in a positive way of causes—or influences—propagating faster than light, since these developments are apparently constructed with the very purpose of short-circuiting the delicate epistemological problem of the definition of causes (or influences). To be more precise, this point is that there does exist a proof of Bell's theorem that makes it possible to speak of causes or influences propagating faster than light.[19] However, that proof is conceptually rather different from the one worked out by BCHS and it is both more involved (although intuitively even more obvious) and less general.

The reason why that other proof is conceptually different and more involved than the one of BCHS has been explicated above. It stems from the facts, acknowledged in Section 2, (a) that the only consistent way of defining causes or influences implies the use of the *causality criterion* (as formulated there) and (b) that the only cases in which that criterion can be applied are those in which the events that are to be defined as causes can be *created at will*. A somewhat unfortunate consequence of this second fact is that the causality criterion cannot be applied directly to the BCHS method. In fact, we have found no way of applying it to any of the methods that have been developed for proving the *generalized* Bell inequalities. Since the experiments have only checked the violation of *these* inequalities, this, in turn, implies a severe restriction in the generality of the proof. Strictly speaking, it seems that, at least for the time being, our positive claim that "there exists influences that propagate faster than light" can only be based on our confidence that the predictions of quantum mechanics are correct also in the cases in which they predict strict correlations at a distance. Such an observation might motivate a search for more refined experiments, which would test the nongeneralized 1964 Bell inequalities. However, before actually engaging in such experiments it would be advisable to pursue the epistemological investigation initiated here somewhat further, so as to ascertain that really there is no exact way of introducing the concept of influences also in the general case.

[19] Provided that the working hypothesis called "splittability" above is adhered to. An alternative formulation of the same conclusion would be that that hypothesis is wrong.

As we saw, the "restriction to the applicability of the causality criterion" called (b) above (the restriction that the "causes" can be created at will) does not prevent *our* proof from going through. This is due to the fact that we chose to define the concept of an attribute of a system according to the "counterfactual theory": To see this we must remember that in the proof in question there is an assertion that plays a decisive role as regards the applicability of the causality criterion, namely the assertion that the emergence, in a physical system, of an attribute is an *event*. And *that* assertion could not even be formulated if the attributes were not defined according to the "counterfactual theory" or according to some equivalent method.

This use of counterfactual propositions of course does not imply that any proof that makes use of such propositions should be satisfactory as regards our standards. One point that should be kept in mind in that respect is that there are several types of counterfactual propositions. If, for example, we have to do with an assembly of systems on every element of which an observable is measured, if we assume that the results are not strictly determined by the past histories of the systems and if we consider an event *e* occurring in a region of spacetime that is spatially separated from that in which the measurements are contained, we may like to consider the proposition according to which *if* event *e* did *not* take place, all the results of the measurements in question would still be what they actually are. Such a counterfactual proposition has obviously nothing to do with the propositions called strict conditionals in modal logic (see Section 3). Now our present proof does not depend for its validity on whether a counterfactual proposition such as the one just described is considered as true or not. In fact, our proof only involves *causal* counterfactual propositions, that is (see Section 3 again) propositions that are in fact identical with some *strict conditionals* of (causal) modal logic, namely with those that are used to define properties of systems. As the analysis of Section 3 indicates, it seems extremely doubtful that any science at all could develop if *such* strict conditionals were meaningless.

Some final remarks are in order concerning induction and reality.

As regards induction, our point is as follows. The procedure consisting in defining attributes exclusively by means of the partial-definition method obviously constitutes a limitation on the free use of induction. At least this is true if, as was found necessary above, for an application of the method to the conventional quantum mechanical definition of attributes, no loose "openness of meaning" is considered as acceptable: for, then, a system prepared in such and such a way has a definite attribute *only if* an instrument fit for measuring the said attribute is "present" (whatever the precise meaning of such a statement may be); so that, even if that attribute is found on an arbitrarily large unbiased sample taken from a population of systems thus prepared, we cannot infer from that observation that the elements of *another* sample of the same population also have that same attribute if, for *them*, no instruments are present. On the contrary, the procedure consisting in defining the attributes by means of the counterfactual method imposes no such limits on induction. And conversely, it seems that a free use of induction, such as the use we make of it both in daily life and in most sciences, is only possible if the attributes are defined by that method or by some equivalent one.

Finally, as regards the concept of reality, it must first be noticed that neither the word "reality" nor any synonym to it has been used in this article: so that, strictly speaking, the conclusions of the article do not depend on any philosophical opinion we may entertain on such problems as "the reality of the outside word" and such like. However, the fact must be acknowledged also that the counterfactual method for defining attributes obviously has some *indirect* connections with our intuitive conception of an independent reality. And that it is perhaps made attractive and plausible only just through that conception. Conversely, the method in question (taken in association with the entailment theory of causality) might well constitute the most definite way that we have at our disposal for approaching in a scientific manner such an elusive concept as that of "an independent reality."

Some sort of general consensus is now developing that the Bell theorem constitutes a very significant new piece of information. But information on what ? On physics as it is actually done now, within the existing centers for physical research, over the world ? Certainly not. On a kind of physics that in fact ought to be done there ? There are arguments for that view, but we cannot, at present, be quite sure that they are not based on overambitious aspirations to realism. On epistemology ? The conclusion of this article might be that this is perhaps the most appropriate answer. Epistemology would then no longer be based on the mere abstracting power of professional epistemologists. For a long period, epistemology has striven, and with good success, to provide physics with reliable foundations. It is perhaps time that the converse movement should develop and that physics should pay its debts.

ACKNOWLEDGMENTS

It is a pleasure to thank A. Shimony and H. Stapp for very useful comments and criticisms.

REFERENCES

1. A. Einstein B. Podolsky, and N. Rosen, *Phys. Rev.* **47**, 777 (1935).
2. J. S. Bell, *Physics* **1**, 195 (1964).

3. J. F. Clauser and M. A. Horne, *Phys. Rev. D* **10**, 526 (1974).
4. J. S. Bell, *Epistemological Lett.* **9**, March 1976 (17.0).
5. A. Shimony, M. A. Horne, and J. F. Clauser, *Epistemological Lett.* **13**, October 1976 (17.1).
6. J. S. Bell, *Epistemological Lett.* **15**, February 1977 (17.3); A. Shimony, *Epistemological Lett.* **18**, January 1978 (17.4).
7. B. d'Espagnat, *Phys. Rev. D* **11**, 1424 (1975).
8. H. P. Stapp, *Phys. Rev. D* **3**, 1303 (1971); *Nuovo Cimento* **40B**, 191 (1977).
9. G. H. von Wright, *Causality and Determinism* (Columbia University Press, New York, 1974).
10. C. G. Hempel, *Philosophy of Natural Science* (Prentice Hall, Englewood Cliffs, New Jersey, 1966).
11. R. Carnap, *Philosophical Foundations of Physics* (Basic Books, 1966).
12. C. G. Hempel, in *The Validation of Scientific Theories*, P. Frank, ed. (Beacon Press, Boston, 1956).
13. L. Rosenfeld, in *Louis de Broglie physicien et penseur* (Albin Michel, Paris, 1953).
14. H. Stein, On the Conceptual Structures of Quantum Mechanics, in *Paradigm and Paradoxes*, Vol. 5 (University of Pittsburgh, 1972).
15. B. d'Espagnat, *Epistemological Lett.* **22**, March 1979 (40.0).
16. D. Lewis, *Counterfactuals* (Blackwell, Oxford, 1973).
17. J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, *Phys. Rev. Lett.* **23**, 880 (1969).
18. B. d'Espagnat, Quantum Logic and Nonseparability, in *The Physicist's Conception of Nature*, J. Mehra, ed. (Reidel, Dordrecht, 1973); *Phys. Rev. D* **18**, 349 (1978).
19. B. d'Espagnat, *Sci. Am.* **241**, 158 (1979).
20. G. E. Hugues and M. J. Cresswell, *An Introduction to Modal Logic* (Methuen, London, 1968).
21. A. Shimony, *Epistemological Lett.* **26**, June 1980 (47.0).