nonlocality conclusion is provided by a rejection of those ideas. Nor is there any demand made in the proof for an *explanation* of quantum correlations. And no assumptions are made about the nature of physical reality, beyond the demands contained in the assumptions that the choices of the experimenters can be regarded, in this context, as free variables, and, for each of the four possible measurements under consideration, that *if* that measurement were to be performed, *then* nature would specify a single result.

If these two assumptions are accepted, together with *QT*, then the no-faster-than-light-influence idea expressed by *LOC* must fail. These three assumptions are completely in line with orthodox quantum thinking. The only thing outside normal quantum theory is the concept *LOC*, which is, however, formulated without introducing any new ideas, and which expresses in a pure form the requirement that there be no faster-than-light influence of any kind.

DO CORRELATIONS NEED TO BE EXPLAINED?

ARTHUR FINE

1. The experiment

Consider a generic correlation experiment of the sort proposed by EPR (Einstein, Podolsky, and Rosen 1935), later refined by Bohm and Aharonov (1957), and that is the subject of the Bell theorem (1964). In the experiment, pairs of objects are produced from an on-line source in the same, briefly interactive (joint) state. (Hopefully without prejudice, we refer to the objects as "particles.") After emission the pairs separate, each particle in a pair moving off to a different, spatially separated wing of the experimental apparatus, the A wing and the B wing. In these wings, for the simple ("2x2") experiment that we discuss, each particle is measured in one of two ways: A_1 or A_2 in the A wing, and B_1 or B_2 in the B wing. Using the language of quantum theory we sometimes say that what is being measured are "observables", i.e., physical parameters, such as components of position or momentum in the same direction (the original experiment of EPR), or spin components in different directions in the plane transverse to the "path" of the particle (the Bohm-Aharonov experiment).

Suppose that the observables are sufficiently coarse-grained to yield just two possible outcomes for each measurement, which we code as "positive" (+) or "negative" (−). Suppose also that the different observables measured in a wing are incompatible (as they are in the instances above), so that no two measurements in either wing can be performed at the same time. A run of the experiment consists of selecting one observable to be measured in each wing, and then measuring the particles in a pair accordingly, as successive pairs come off the production line and enter the wings. So far as possible, we try to make the measurements on the same pair at the same time. For simplicity assume that each run consists of the same number of successive measurements. In a *correlation experiment*, all four possible runs are made; that is, A_1 is measured with B_1, A_1 is measured with B_2, A_2 is measured with B_1, and A_2

is measured with B_2. Of course the runs need not be completed as a whole unit, back to back. We could shift between measurements, moving to and fro on line, and then separate out the distinct runs after the fact (Aspect et al. 1982).

In each wing the data from successive measurements of the same observable consist of a sequence of pluses and minuses that, in a typical case, might look like this:

+ − + + − − + − + + − + + + − + − − + − + − + − −

In any run there will be roughly the same number of positive as of negative outcomes. Moreover, the interspersing of the positive with the negative outcomes occurs at random. Thus for any one of the measurements, the probability for a particular outcome (whether positive or negative) is ½. A typical run for a sequence of 25 pairs might show outcomes like this:

+ − + + − − + − + + + − − − + − + + + − − − + − +
+ − + − − − + + + + − − + − − + + + − − − + − −

In this run, roughly half the outcomes of each measurement are positive and half negative. But in 40% of the pairs (i.e., in 10 out of the 25) a positive result of one measurement is accompanied by a positive result in the other. This is certainly a significant correlation between the positive outcomes, since if each particular outcome occurs at random (and so has probability ½), then one might expect that the odds for a positive outcome in both wings would be 1 in 4; that is, one might expect it to be just the product of the separate probabilities. This expectation, which amounts to looking for no correlation between the positive outcomes, would yield roughly 6 out of 25 positive pairs, not the 10 out of 25 displayed above. Typically, the runs in a correlation experiment yield significant correlations between positive outcomes. The correlations, just like the 50% chance of a positive for the individual outcomes, are stable and regular. Experiments show the same correlations in nearly all largish runs, when the same measurements are made. These stable correlations are grounded in the laws of the quantum theory, which predicts them accurately as functions of the common state in which the pairs are produced, and of the particular measurements in a run.

Below I will take up the question of whether one ought to be puzzled by these correlations in and of themselves, that is, whether the mere fact of correlations between sequences of randomly occurring events inherently calls for explanation. Not surprisingly, I will suggest that the answer depends on what attitude one takes toward explanation, and toward the reliable predictions of the quantum theory. Even those who may not find such correlations inherently puzzling, however, might still encounter some difficulty with the

relationships that can obtain between the correlations arising in different runs of the experiment. This is the subject of the Bell inequalities, to which we now turn.

2. *The Bell inequalities*

Consider the correlations that arise in certain quantum correlation experiments. (The numbers below are the joint probabilities, rounded to one decimal place, for a spin experiment where A_1 corresponds to measuring the vertical spin component and A_2 the horizontal; where B_1 splits the 90° angle between the As, and B_2 is the component 135° away from A_2, and perpendicular to B_1.) Suppose that the run in which A_1 is measured with B_1 yields a correlation of 4/10 between positive outcomes, as in the example in section 1 above. I will write it this way:

$$P(++/11) = 4/10.$$

Using the same conventions, suppose that the other runs yield:

$$P(++/12) = 4/10,$$
$$P(++/21) = 4/10,$$
$$P(++/22) = 1/10.$$

Thus all the runs, except for the one where A_2 is measured with B_2, regularly yield the same correlation of 4/10 between the plus outcomes in the two wings, and this last run reliably turns up a correlation of 1/10. Of course, the individual outcomes of any single measurement are still roughly half plus and half minus, as usual, which we could write (without having to pay attention to the particular run) as:

$$P(+) = ½.$$

As emphasized, the correlations that obtain between the positive outcomes for fixed measurements in the different wings are stable and reliable, depending only on the particular measurements performed. Although the two measurements in a single wing cannot both be done at the same time, we might still wonder whether that kind of stability of correlations between outcomes would continue to hold if we could conceive somehow of combining the outcomes of measurements in a single wing, at least hypothetically. We might, for example, pursue the following train of thought.

Suppose we were to perform the B_1 measurement in the B wing, say with A_1 being done in the A wing, then we know we would get some definite sequence of positive and negative B_1 outcomes (indeed, ones that overlap

with the A_1 positives in 4 instances out of 10, on the average). But now suppose that, rather than B_1, we imagine measuring B_2 instead. Again, we would obtain some definite sequence of outcomes (whose positives would also overlap with those of A_1 an average of 4 instances out of 10). Surely, then, for the same, fixed population of pairs, we can at least imagine a string of hypothetical B_1 outcomes and a string of hypothetical B_2 outcomes. Of course, since the population of pairs is finite, whatever particular sequences of outcomes turns up and regardless of whether they are repeatable or not, there would be some definite overlap (or other) between the positive outcomes, pair by pair, had we measured B_1, and the positives had we measured B_2. Or so we must imagine. Indeed, we can say something quite definite about the possible overlap between the B positives.

We know that out of ten emitted pairs, on the average, we get five positives and five negatives in every measurement. Given the correlations assumed above we can see that there are constraints on the outcomes, as indicated below, where next to each measurement we list what an average run of ten would be like in a typical case.

$$\begin{array}{lllllllllll} A_1\text{:} & - & + & - & + & + & + & - & - & + & - \\ B_1\text{:} & . & X & . & X & X & X & . & . & X & . \\ B_2\text{:} & . & X & . & X & X & X & . & . & X & . \end{array}$$

Next to B_1, four of the Xs must be plus, and this is also true for four of the Xs next to B_2, in order that the correlations:

$$P(++/11) = P\,(++/12) = 4/10$$

be satisfied. *This means that in an average run of ten, there must be at least three overlaps among positive outcomes between* B_1 *and* B_2, since there is no way to distribute four objects among five places twice over without at least three duplications occurring.

Another constraint on the possible overlap between the B positives arises if we go through a similar line of reasoning for an average run of ten, where it is A_2 that is measured in the A wing. In this situation we get, for example:

$$\begin{array}{lllllllllll} A_2\text{:} & + & + & - & - & - & + & - & + & + & - \\ B_1\text{:} & X & X & . & . & . & X & . & X & X & . \\ B_2\text{:} & X & X & . & . & . & X & . & X & X & . \end{array}$$

Next to B_1, four of the Xs must be plus, in order to satisfy the correlation

$$P(++/21) = 4/10.$$

Next to B_2, exactly one of the Xs must be plus, in order that

$$P\,(++/22) = 1/10.$$

So, among the Xs at most one plus occurs in common between B_1 and B_2. Outside the Xs there remains only one other plus next to B_1. Even if it overlaps with a plus next to B_2, that would contribute at most one additional plus in common. *Thus in an average run of ten there cannot be more than two overlaps between the positive* B_1 *and* B_2 *outcomes.* We saw above, however, that there must be at least *three* overlaps in order to satisfy the other two correlations!

This contradiction shows that the quantum correlations do not fit together in a way that would allow a stable distribution of outcomes between hypothetical measurements in the same wing. This, finally, is where the Bell inequalities come in. For satisfaction of those inequalities by the correlations generated in the four runs of a 2x2 experiment are the necessary and sufficient conditions for the correlations to be consistent with such a stable, hypothetical distribution of outcomes of incompatible measurements (Fine 1982b). The Bell inequalities simply require that if we take the correlations for any three of the runs, add them up, and then subtract the correlation generated in the remaining fourth run, we get a nonnegative number that is no bigger than 1. If we apply this rule to the correlations above, then one possibility is

$$4/10 + 4/10 + 4/10 - 1/10 = 11/10 > 1.$$

Thus, as we have seen, these correlations do not fit together according to the rule. I might just add that to derive the Bell inequalities one simply carries out the counting argument that produced the contradiction above, under the stipulation that the overlap between B_1 and B_2 when measured with A_1 be consistent with that overlap when measured with A_2. This is possible just in case the Bell inequalities hold.

The reasoning that produced the contradiction depends on treating incompatible measurements differently from how they are treated by the quantum theory. It depends on thinking that even where one cannot carry out two measurement-operations at the same time, one can still imagine one having been done rather than the other, comparing the hypothetical results, and projecting that comparison in a lawlike and stable way into a variety of different circumstances. The trouble, I would suggest, comes not from the exercise of our imagination in a counterfactual way, but from the lawlike projection of the hypothetical results of the comparison. Contrary to this particular exercise in fancy, the quantum theory itself does not contain the resources for describing the overlapping results of such hypothetical measurements in a lawlike way. That is, it does not allow the joint distribution of observables to be defined as a function of the state of a system (and so projected lawfully), unless that state is one for which the observables are actually compatible (Fine 1982a). Thus the imaginative exercise can be viewed as providing a good and appropriate test for whether the boundaries of

the projectable, according to quantum theory, are properly drawn. Assuming that the quantum correlations, for instance those above, are correct for the experiments to which they relate, it seems that the contradiction in projecting the hypothetical overlap from one measurement context to another shows that the limits drawn by quantum theory as to what correlations are stable is quite appropriate.

To be sure, sometimes one can do better, for instance where the experimental correlations do satisfy the Bell inequalities; but not, it would seem, systematically. This is a nice and modest judgment. A theory that has proved itself extremely reliable experimentally turns out to set the boundaries for what joint probabilities can be projected from different finite runs just right. Thus those who may have been concerned with the connections among the correlations in the different runs of an experiment, and in particular with why they fail the Bell inequalities, may feel relieved. What these inequalities require is a lawlike projectibility of distributions for the outcomes of hypothetical and incompatible measurements, a projectability that well exceeds what the theory provides. Indeed the failure of the Bell inequalities is a nice demonstration that, with respect to joint distributions, in general the theory *cannot* provide, even in principle, what it *does not* provide. Moreover if the theoretical predictions are correct, nature does not provide for such projectability either.

Not everyone, however, is satisfied with these modest judgments. Instead many seem disposed to view the quantum correlations as inherently puzzling, and to see in their failure to satisfy the Bell inequalities a sign either of some new and quasi-mysterious physical process (maybe funny "influences" or odd "passions"), or an experimental demonstration of the end of the era of realist metaphysics. To be sure the idea of an experimental refutation of metaphysics is a charming (if oxymoronic) concept, and I for one would be happy to see realist metaphysics fade away. But I am afraid that the philosophical disposition in this case is grounded in the metaphysics of determinism, which seems to me no improvement, and that the puzzle over the correlations seems to require a similarly inflated, realist-style, essentialism in regard to explanations. We begin there.

3. *Explanations*

What surprises or puzzles is relative to context, which includes at least psychological set and background beliefs. Explanation is similar. Accordingly, what counts as an explanation can be expected to depend on particular features of the context of inquiry. Similar context dependency is to be expected in what calls for ("requires" or "needs") explanation. Moreover, there is no reason to presume any general or uniform concept of explanation

(or of what "needs" an explanation) that necessarily cuts across and unifies the various contexts. If one holds to such a nonessentialist attitude, then the general question of whether correlations require explanation is not a particularly useful question to pursue, for one anticipates that the answer will only be: "Sometimes yes, and sometimes no."

Instead it would be good to keep an open mind on the issue of what needs explanation as it arises in the case of particular correlations set in the context of particular beliefs and frames of reference. Thus, the "natural" or zero state of mind with regard to the correlations between outcomes in different wings of a correlation experiment should be no different from what it is for the randomness of the outcomes in each wing separately; namely, unless features are present that point to the correlations (but not the randomness) as in need of explanation, they are not.

I think it is fair to say that this is not the usual attitude toward correlations. General treatments set them in an essentialist framework, one that takes the mere existence of correlations as calling, all by itself, for an explanation. The explanatory resources, moreover, are usually pretty sparse. Either correlations are taken to be coincidental (or "spurious," i.e., not genuine in the sense of not indicative of a "real" connection between the correlates) or they are taken as signs of an underlying causal relation. In the latter case, the relation can either be direct, with one of the correlates causally connected to the other; or indirect, where the correlation is mediated by a network of common causal factors; or some combination of the two.

The framework of common causal factors involves two distinct requirements. One is that the initial correlation be derived by averaging over the contribution to the joint outcome made by each of the causal factors separately. The other is the requirement that no residual correlations remain when each factor is held fixed; that is, that relative to each causal factor, the outcomes are stochastically independent, and so their joint probability is just the product of their separate probabilities relative to the factor.[1] This last assumption embodies the essentialist conception that correlations inherently need to be explained, so that it would not be proper to use some correlations to explain others. An essentialist would hold that such an "explanation" could not be complete.[2]

When essentialists exhaust their explanatory resources (that is, treat

---

[1]Following Reichenbach (1956), this requirement of conditional stochastic independence was called "screening-off." See Salmon (1984) and Suppes (1984). In the Bell literature, the interpretation of this condition was prejudiced by calling it "locality." Fine (1981, 1982b) suggested the term 'factorizability' in order to free up the discussion.

[2]I believe this is one of the motivations for Jarrett (1984) calling this requirement "completeness." One should not confuse this terminology with Einstein's, where he charges the quantum theory with descriptive incompleteness. See Fine (1986) for the several senses in which Einstein used this term.

correlations that seem to be neither coincidental nor causal), they find themselves in the position of holding that the correlations are mysterious, for they need to be explained *and* they cannot be explained. Clearly something has to give. They might retreat from essentialism by expanding the conception of a common cause, admitting some basic residual correlations to be used in explaining others. Or they might withdraw the initial presumption that the correlations in question do require explanation. Thus without essentialism there would be several degrees of freedom.

The only essentialist alternatives, however, are to reexamine the possibility of coincidence or to propose new causal connections to do the explanatory job. For the quantum-correlations experiments, these options are rather constrained. The correlations arise between outcomes of measurements performed in spatially separated wings of the apparatus. We may suppose that the measurement-events are spacelike, that is, they are not separated by time enough to signal the results in one wing of the experiment to the other. Thus there can be no direct influence between the outcomes, unless that influence is conveyed between the wings with a speed faster than that of light. Neither is there a network of common background causal factors to produce the correlations. For the existence of such common causes would imply that the correlational data satisfy the system of Bell inequalities, and we may suppose that experiments have been chosen for which this is not the case.[3] The correlations obtained in these experiments are grounded in the quantum theory. They are stable, predictable and, by arranging for the appropriate measurements to be made, controllable in advance. This sort of correlation could hardly be called coincidental. Indeed, then, for an essentialist faced with the quantum correlations, something has to give.

One possibility would be to break with ordinary physics and to introduce the concept of superluminal influences. This would amount to inventing new physics, hopefully integrating it with the old. Such a move certainly cannot be ruled out *a priori,* but neither should it be accepted on that basis. That is, one should be suspicious of the argument that since the quantum correlations stand in need of explanation and since superluminal influences provide the best (and perhaps even the only) explanation for them, therefore one ought to accept the introduction of such influences. One does not need to have reservations about the strategy of inference to the best explanation in general, as I do, to be reserved about the soundness of this particular application. For we can break the argument at the very initial stage simply by stepping way from the essentialist conception of explanation and the general

[3]A common-cause explanation, as outlined above, is what the foundational literature calls a factorizable stochastic hidden-variables model. Fine (1982b) shows that there exists such a stochastic model for the data of a correlation experiment if and only if there is a deterministic model for the same data. For a 2x2 experiment, this is equivalent to the satisfaction of the Bell inequalities, as discussed in section 2.

suspiciousness of correlations on which it rests. We might also wonder, even were physical speculation allowed to be driven simply by one's essentialist cravings, whether the speculation has to be quite so conservative. In particular, why stick to "influences" propagating in space-time? The algebraic and topological structures of recent string and supersymmetry theories surely provide rich resources for reconceptualizing the experiments, resources that do not involve the prerelativistic worldview of things simply moving faster than light. Who knows, these other ways might even connect with progressive physics?

4. *Locality*

Since the explanationist case for superluminal influences seems entirely *ad hoc* and *a priori* (Latin sins than which not much could be worse), is there perhaps some other and better way to make a case? There is this: The Bell inequalities follow from several different sets of assumptions. If we could derive them from the *denial* of faster-than-light influences, then the violation of the inequalities in the correlation experiments would entail the existence of the influences that had been denied. Let us call the desired premise, whose denial entails the existence of superluminal influences, LOC (for "locality"). Whatever the precise formulation of LOC might be, it is clearly a principle that denies certain kinds of influence (or dependence) between the outcomes of measurements in one wing of the experiment and what happens in the other wing. It follows that a principle denying *any* influence between happenings in different wings would imply LOC. Hence the Bell inequalities could be derived from such a strong principle, if they could be derived from LOC itself. Call such a strong principle SLOC. Were SLOC consistent with the denial of the Bell inequalities, then the inference from LOC to those inequalities would fail, and so would this case for superluminal influences. But at least the *relative* consistency of SLOC with the failure of the Bell inequalities is well known. It is consistency relative to the quantum theory. For SLOC is actually built into the quantum theory, according to which there *is* no influence between the two wings of the experiment, that is, no physical interaction of any sort that is represented by terms in the Hamiltonian of the composite system at the time one or the other component is measured. As Bohr (1935) emphasized in his response to EPR, "There is in a case like [EPR] no question of a . . . disturbance of the system under investigation during the last critical stage of the measurement procedure."[4] Of course,

[4]It is true that Bohr (1935) goes on to talk about "an influence on the very conditions which define the possible types of predictions regarding the future behavior of the system." The word 'influence' here could be misleading. For Bohr is not referring to what is at issue above; namely,

correlations violating the Bell inequalities are also built into the quantum theory, as is well known; hence the case for superluminal influences fails here just as it did previously.

Despite Bohr's authority, not everybody will be persuaded by this argument; although to me Bohr's strategy here of calling on our best theories and respecting what they say seems pretty sensible. Let me, however, try again, concentrating this time on how one is to understand the idea of no-influence, or independence, in the case of the correlations. Suppose we do an experimental run where we measure A_1 in the A wing and either B_1 or B_2 in the B wing. The no-influence idea is that whatever one does in the B wing makes no difference to what happens in the A wing. Now what does happen in the A wing? I want to conceive of it this way: the A_1 measurement is carried out and *in a perfectly random way* an outcome (either plus or minus) occurs. I emphasize the randomness, for if we conceive of the outcomes as predetermined (in the sense that for every measurement there is an outcome such that if we were to perform the measurement that outcome would result), then the Bell inequalities automatically govern the statistics of the experiment (well, given some other reasonable assumptions; see Halpin 1986).

For definiteness, suppose that the measurement in the A wing turns up plus, and that in fact B_1 is measured in the B wing. The no-influence idea of SLOC is that what happened in the B wing (i.e., the performance there of the B_1 measurement) did not influence what happened in the A wing (i.e., the performance there of the A_1 measurement, yielding the plus result). Let us just assume that there is no wing-A-measurement to wing-B-measurement influence. (We might think to insure this by making the choice of what measurement is being performed depend on random selections made separately in each wing. But one can readily see that, in the context of this discussion, relying on this to rule out influences just begs the question!) The possible dependency left over would be between the B_1 measurement itself and the outcome of the A_1 measurement. However, we have supposed that the A_1 outcome is random. Nothing determines that particular outcome, and there are no factors on which it depends. So randomness implies that the outcome does not depend on which measurement is performed in the B wing, not even on *whether* a measurement is performed there.

Thus if we adhere to the idea that the measurement results are truly random, and we rule out influences affecting which particular measurements are made in the wings, we automatically have a framework in which the

an "influence" from measurements performed in one wing that affects the particular outcomes in the other wing. Rather, Bohr is pointing to his positivist conception of predication (what it means to attribute a property to an object), and its material presuppositions involving measurement preparations. That is a wholly different topic.

strong locality assumption, SLOC, is satisfied. It remains to show that in such a framework the Bell inequalities can fail. But that is easy. All we require are four pairs of random sequences of pluses and minuses corresponding to the four measurement runs in a correlation experiment, that carry the four quantum joint probabilities for the outcomes. Such random sequences are obtainable from the data of any actual correlation experiment by discarding all the pairs of results where one or the other of the detectors failed to register a result. For experiments in which the Bell inequalities fail (like the one discussed in the first section), the sequences show the consistency of SLOC with that failure. Hence the attempt to derive the Bell inequalities from even weaker locality assumptions, like LOC, breaks down.

Several objections can be raised to this line of argument:

Objection (1). The argument contains an undischarged assumption: namely, that performing a measurement in one wing does not influence what measurement is performed in the other wing.

To take this objection seriously would be to open up the possibility of what John Bell (Davies and Brown 1986, p. 47) calls "superdeterminism." It is a skeptical hypothesis that, once opened, could not easily be laid to rest. But like other skeptical hypotheses, it would require a lot of work, I believe, to get it going. Briefly, in the absence of a theory of "influences" between measurements performed, which shows how to integrate them with ordinary experimental practice, and in the absence of specific reasons to entertain suspicions about the presence of such influences, I think we can pass them by. Merely skeptical doubts ought not to stand in the way of judgments based on otherwise sensible arguments.

Objection (2). The argument seems to equivocate on the term "random." Sequences of numbers (or pluses and minuses) can be said to be random in the technical sense of a table of random numbers. This is the sense used at the end of the argument where the data-set from correlation experiments is cited as being random. But earlier that term is used to designate measurement-outcomes that are not determined at all, and hence independent and uninfluenced. This is a different sense of the word.

True. The idea precisely was to treat sequences of experimental data, which are random in the sense of random numbers, as representing outcomes of measurements where nothing at all determines or influences any particular outcome. The idea is to impose an indeterminist framework on the quantum experiments in order to demonstrate that, within such a framework, the strongest locality assumptions are perfectly compatible with experimental results that violate the Bell inequalities. If that demonstration holds up, then it will follow that those who see locality at issue in the violation of those inequalities do so only on the basis of additional determinist assumptions or presuppositions.

Objection (3). Strong locality requires that the particular outcome in one wing not depend on what is being measured in the other wing. This implies that the very same outcome would have occurred in a given wing regardless of whether one or another measurement were carried out in the opposite wing. So the sequence of A_1 outcomes would be the same whether B_1 or B_2 were measured with it, according to strong locality. Similarly, the outcomes of A_2 measurements would not vary between different B wing measurements. Suppose that we measure A_1 with B_1 and then A_2 with B_2. The sequence of B_2 outcomes in the second measurement might very well have been obtained had B_2 been measured instead with A_1, and in that case the A_1 outcomes, according to strong locality, would have been just whatever they originally were. Thus the A_1 outcomes fit together with those of the B_1 and the B_2 measurements to form a trio of sequences from which correlations for A_1 with B_1 outcomes as well as for A_1 with B_2 outcomes can be calculated. Similarly, the A_2 results fit together with the above outcomes from the two B wing measurements to form another trio of sequences from which correlations both for A_2 with B_2 and also for A_2 with B_1 outcomes can be calculated. But one of your very own theorems (says my knowledgeable interlocutor) shows that there are compatible trios from which correlations can be calculated just in case the correlations satisfy the Bell inequalities (Fine 1982a, 1982b). Hence the argument from strong locality to those inequalities goes through. No sort of determinism is involved, and randomness is of no avail in blocking it.

The preceding argument uses much more than strong locality. It infers from the requirement that the outcome in one wing not depend on what is being measured in the other wing (which strong locality does maintain) that the very same outcome would have occurred in a given wing regardless of whether one or another measurement were carried out in the other wing. Strong locality denies that there are any influences from the circumstances regarding measurements carried out in one wing to the actual outcome obtained in a measurement performed in the other wing. It says that the measurements carried out in the one wing make no difference whatsoever to the outcome obtained in the other. The outcome does not depend on them in any way at all.

What is the logic of this independence assertion? Does it imply that what did happen would have happened anyway (because nothing would have changed)? Surely it does, if the outcome in a wing is the result of stable local circumstances there. For then, according to strong locality, switching the measurement elsewhere would not affect those stable circumstances, which would therefore issue in the same result. But what if there are no such stable, local determinates? Then it would seem that although indeed nothing relevant would change had we switched measurements, that fact alone is entirely compatible with the occurrence of a different (undetermined) outcome.

To take a somewhat remote example. It is probably true that the color of my car does not influence my luck at poker. But it scarcely follows that had I a car of a different color, my luck would not turn out to be somewhat different from what it in fact is. Indeed there are many things not dependent on the color of my car, any number of which might in fact have been different had I a car of a different color. The inference from "no influence" to "things would have been just the same" requires supplementary assumptions in order to go through. To insist on it, as the above objection does, is to introduce the principle that where nothing relevant to an outcome changes, the outcome itself could not change. This is just a version of the idea that change requires a cause. Thus the objection relies on determinism, or something in that neighborhood. Indeterminism blocks it, hence it blocks this route from strong locality to the Bell inequalities.

Objection (4). One does not need the strong inference from "no cross-wing influence" to "the same outcome would have occurred." We can run the argument from the weaker principle that if there is no influence from the measurement performed in one wing to the outcome in the other, then had we switched measurements the same outcome *might* have (or *could* have) occurred. To take up the automobile example, if the color of my car were different, then although I am correct in noting that indeed my luck at poker might have been different, it might also have stayed just the same. So the principle proposed here is weaker than the one above. Moreover, this principle does follow from strong locality; for surely if stability of outcomes is not even possible, were the measurement to have been different, then the measurement does influence the outcome.

There is no doubt that the "might/could" principle is weaker than the "would" one. What is certainly doubtful is whether the restriction of a logical possibility should count as an "influence" of the sort intended by a physical locality principle. Fortunately, we need not get bogged down in a metaphysical squabble over what does or does not count as a real influence since, contrary to the objection, when the argument is weakened as suggested it no longer goes through.

To see this let us call different runs in the experiment *adjacent* (recall that a run consists of a sequence of simultaneous measurements of one observable in one wing and another in the other wing) if they have the measurement of some one observable in common, e.g., the A_1B_1 and the A_2B_1 runs are adjacent (having B_1 in common), as are the A_2B_1 and A_2B_2 runs, which have A_2 in common. We can adapt some terminology first introduced by Schrödinger (1935b) and say that a correlation experiment is *entangled* just in case there are adjacent runs in the experiment whose sequences of outcomes for the shared observable are different. (For present purposes, we still suppose that the total number of outcomes is the same for each run.) The point of the

third and fourth objections was to try to use strong locality to get unentangled experiments, for the correlations in such experiments satisfy the Bell inequalities (Fine 1982c). The determinist principle used in the third objection as a supplement to strong locality does entail that some experiments would be unentangled. But the weak "might/could" version of that principle in the fourth objection only implies that some experiments *might* be unentangled. Thus some experiments might yield statistics that do satisfy the Bell inequalities.

If we pick experimental arrangements whose statistics, according to quantum theory, do not satisfy those inequalities, then the argument in question only shows that one could in principle fail to verify the quantum statistics in such experiments, provided strong locality holds. I believe this to be a perfectly valid argument, but it does not show that strong locality implies the Bell inequalities, as it claimed to do, nor that strong locality conflicts in any way with the quantum theory. After all, the conclusion of an argument can be no stronger than the premises; so "might/could" *in* means "might/could" *out*. Nor can one see anything especially anomalous in the weak conclusion that the data might not be quantum mechanical. After all any experiment could fail to verify any set of predictions. We hardly need principles like strong locality to learn that. Moreover the jeopardy in quantum mechanics is double to begin with, since the predictions it yields are probabilities, and it is well understood that probabilistic predictions always have some actual likelihood of failure in any finite experiment. Nothing in this argument, however, suggests that the quantum experiments *do* fail or *would* fail; just that they might.

Objection (5). You have just identified the puzzle here that the previous objections were trying to get at. It is to understand why the quantum statistics, unlike others with which we are familiar, cannot be exhibited in unentangled experiments. For although the number of unentangled experiments is extremely small relative to the entangled ones (for a fixed largish number of outcomes in each run), we might suppose that among all the possibilities realized by various experiments sometimes we happen to find that as the data accumulate all the adjacent runs do show the very same outcomes, term by term, for the measurement they have in common, i.e., that the experiment is not entangled. But if the data are quantum-mechanical, this admittedly small possibility is entirely ruled out. How can that be? How can it happen that if we were to move among adjacent runs, by changing a measurement in one wing, the new sequence of outcomes in the opposite wing must differ somewhere along the line from the earlier sequence? Surely this is nonlocality, an influence from events in one wing generating changes in the other.

Is it? The argument shows this: Where the experimental outcomes are

governed by probabilities that fail to satisfy the Bell inequalities, the way the data accumulates involves a shift in some outcomes between at least one pair of adjacent runs. If the individual outcomes in a measurement sequence occur at random and independently of one another, then the odds of repeating the same sequence of n outcomes twice are 2^{-2n}. (For a pretty short run of length 50 (i.e., for $n = 50$), we already have that 2^{-100} is smaller than 10^{-31}, an impressively small number!) Thus the odds for there *not* being a shift in some outcomes between the several adjacent runs in a correlation experiment are negligible, regardless of the experimental data. If the correlational data do violate the Bell inequalities, these negligible odds vanish.

It is somewhat tedious to show that even this tiny difference is compatible with strong locality, but maybe it is worth the effort if it will help remove even tiny residual doubts. So consider two runs in such an experiment. First we measure A_1 with B_1 and then we measure A_2 with B_2. In accord with the drift of this objection we now imagine some alternative possibilities, which are sketched out below:

	FIRST RUN	SECOND RUN
ACTUAL	A_1/B_1	A_2/B_2
HYPOTHETICAL	$[A_2]$ / $[B_1]$	$[A_1]$ / $[B_2]$

POSSIBILITIES

[1] Actual A_2, 2nd run = Hypothetical $[A_2]$, 1st run.
[2] Actual A_1, 1st run = Hypothetical $[A_1]$, 2nd run.
[3] Actual B_1, 1st run = Hypothetical $[B_1]$, 1st run.
[4] Actual B_2, 2nd run = Hypothetical $[B_2]$, 2nd run.

Suppose in the first run we had measured A_2 instead of A_1, then it is entirely possible (although *extremely* unlikely) that the sequence of outcomes in the hypothetical A_2 measurement in this run would have matched exactly the actual sequence of outcomes of the A_2 measurement in the second run (cf. possibility [1] above), and that the sequence of B_1 outcomes would just have been a repeat of the original (cf. [3]). Similarly, if in the second run we had chosen to measure A_1 instead of A_2, then it is entirely possible (although again *extremely* unlikely) that the sequence of outcomes in the hypothetical A_1 measurement would have matched exactly the actual sequence of outcomes of the A_1 measurement in the first run (cf. [2]), and that here again the sequence of outcomes in the B wing, this time of B_2, would simply have been a repeat of the original (cf. [4]).

If, however, we had chosen *both* to measure A_2 instead of A_1 in the first run *and* to measure A_1 instead of A_2 in the second run, then, assuming that the correlational data do not satisfy the Bell inequalities, at least one of these four

possibilities would have to go. Either the hypothetical sequence of outcomes of the A_2 measurement in the first run would differ from the actual A_2 sequence in the second run [1], or the hypothetical sequence of outcomes of the A_1 measurement in the second run would differ from the actual A_1 sequence in the first run [2], or the actual sequences of outcomes of the B_1 measurement in the first run (where A_1 was measured with B_1) would differ from the hypothetical sequence of outcomes (where A_2 would have been measured with B_1) [3], or the actual sequence of outcomes of the B_2 measurement in the second run (where A_2 was measured with B_2) would differ from the hypothetical sequence of outcomes (where A_1 would have been measured with B_2) [4]. Only the latter two alternatives, corresponding to the failure of [3] and [4] above, could possibly suggest a challenge to strong locality, for only they involve any possible dependence between measurements made in one wing and outcomes in the other.[5]

Consequently, we can derive the failure of strong locality from the failure of the Bell inequalities only if we make some additional assumptions. We have to assume something that necessitates that were we to shift the measurement in the A wing in both runs, the new sequences of outcomes there would match, across the runs, both of the old sequences of outcomes, i.e., that both possibilities [1] and [2] are realized. Clearly, nothing can guarantee the occurrence of both of these highly unlikely possibilities short of determinism; that is, the requirement that if the local conditions do not change, then a repeat of the same measurement would produce exactly the same outcome each and every time. Hence here once again we see that it is only the combination of strong locality with determinism from which the satisfaction of the Bell inequalities follows. Locality alone is not enough. To put it differently, in answer to the objection, it is true that correlation experiments whose statistics fail to satisfy the Bell inequalities are entangled. It is not true that such entanglement implies the existence of any nonlocal influences or dependencies.

When all these objections are considered together, they seem to add up to no more than the original cry, which was that if nonlocal influences (or the like) are not invoked to explain the tangled statistics of the correlation experiments (i.e., statistical correlations that can only arise in tangled experiments), then how are we to explain them at all? I have tried to suggest the way out above. Let me try to reinforce the suggestion below.

[5]It is important to stress the *possibility* here, for it is by no means clear that the association between changing measurements in one wing and changing some outcome or other in the second wing represents a real influence or dependency from one wing to the other. Whether and when such associations count as real influences is an issue that we have only just begun to investigate. The answer in this case is not yet in.

5. *Indeterminism*

If we adopt an indeterminist attitude to the outcomes of a single, repeated measurement, we see each outcome as undetermined by any factors whatsoever. Nevertheless we are comfortable with the idea that, as the measurements go on, the outcomes will satisfy a strict probabilistic law. For instance, they may be half positive and half negative. How does this happen? What makes a long run of positives, for example, get balanced off by the accumulation of nearly the *very same number* of negatives? If each outcome is really undetermined, how can we get *any* strict probabilistic order? Such questions can seem acute, deriving their urgency from the apparent necessity to provide an explanation for the strict order of the pattern, and the background indeterminist premise according to which there seems to be nothing available on which to base an explanation. If one accepts the explanationist challenge, then one might be inclined to talk of a "hidden hand" that guides the outcome pattern, or its modern reincarnation as objective, probability-fixing "propensities."

This talk lets us off the hook, and it is instructive to note just how easily this is accomplished. For if propensities were regular explanatory entities, we would be inclined not just to investigate their formal features and conceptual links, but we would make them the object of physical theorizing and experimental investigation as well. However, even among the devoteés of propensities, few have been willing to go that far. The reason, I would suggest, is this. Once we accept the premise of indeterminism, we open up the idea that sequences of individually undetermined events *can* nevertheless display strict probabilistic patterns. When we go on to wed indeterminism to a rich probabilistic theory, like the quantum theory, we expect the theory to fill in the details of under what circumstances particular probabilistic patterns will arise. The state/observable formalism of the quantum theory, as is well known, discharges this expectation admirably. Thus indeterminism opens up a space of possibilities. It makes room for the quantum theory to work. The theory specifies the circumstances under which patterns of outcomes will arise and which particular ones to expect. It simply bypasses the question of how any patterns *could* arise out of undetermined events, in effect presupposing that this possibility just is among the natural order of things. In this regard, the quantum theory functions exactly like any other, embodying and taking for granted what Stephen Toulmin (1961) has nicely called "ideals of natural order." What then of correlations?

Correlations are just probabilistic patterns between two sequences of events. If we treat the individual events as undetermined and withdraw the burden of explaining why a pattern arises for each of the two sequences, why not adopt the same attitude toward the emerging pattern between the pairs of

outcomes, the pattern that constitutes the correlation? Why, from an indeterminist perspective, should the fact that there is a pattern *between* random sequences require any more explaining than the fact that there is a pattern internal to the sequences themselves?

We have learned that it is not necessary to see a connection linking the random events in a sequence, some influence from one event to another that sustains the overall pattern. Why require a connection linking the pairs of events between the sequences, perhaps some influence that travels from one event in a pair to another (maybe even faster than the speed of light) and sustains the correlation? We have explored part of the answer above. Our experience with correlations that arise in a context in which there generally are outcome-fixing circumstances has led us to expect that where correlations are not coincidental, we will be able to understand how they were generated either via causal influences from one variable to another or by means of a network of common background causal factors. The tangled correlations of the quantum theory, however, cannot be so explained.

The search for "influences" or for common causes is an enterprise external to the quantum theory. It is a project that stands on the outside and asks whether we can supplement the theory in such a way as to satisfy certain *a priori* demands on explanatory adequacy. Among these demands is that stable correlations require explaining, that there must be some detailed account for how they are built up, or sustained, over time and space. In the face of this demand, the tangled correlations of the quantum theory can seem anomalous, even mysterious. But this demand represents an explanatory ideal rooted outside the quantum theory, one learned and taught in the context of a different kind of physical thinking. It is like the ideal that was passed on in the dynamical tradition from Aristotle to Newton, that motion *as such* requires explanation. As in the passing of that ideal, we can learn from successful practice that progress in physical thinking may occur precisely when we give up the demand for explanation, and shift to a new conception of the natural order. This is never an easy operation, and it is always accompanied by resistance and some sense of a lost paradise of reason. If we are to be serious about the science that we now have, however, we should step inside and see what ideals *it* embodies.

The quantum theory takes for granted not only that sequences of individually undetermined events may show strict overall patterns, it also takes for granted that such patterns may arise between the matched events in two such sequences. From the perspective of the quantum theory, this is neither surprising nor puzzling. It is the normal and ordinary state of affairs. This ideal is integral to the indeterminism that one accepts, if one accepts the theory. There was a time when we did not know this, when the question of whether the theory was truly indeterminist at all was alive and subject to real

conjecture. Foundational work over the past fifty years, however, has pretty much settled that issue (although, of course, never beyond *any* doubt). The more recent work related to EPR and the Bell theorem has shown, specifically (although again, not beyond *all* doubt), that the correlations too are fundamental and irreducible, so that the indeterminist ideal extends to them as well. It is time, I think, to accept the ideals of order required by the theory. It is time to see patterns *between* sequences as part of the same natural order as patterns *internal* to the sequences themselves.

A nonessentialist attitude toward explanation can help us make this transition, for it leads us to accept that what requires explanation is a function of the context of inquiry. So when we take quantum theory and its practice as our context, then we expect to look to *it* to see what must be explained. This leads us to the indeterminist ideal discussed above, and to the "naturalness" of (even distant) correlations. There is a small bonus to reap if we shift our thinking in this direction. For the shift amounts to taking the correlations of the theory as givens not in further need of explanation and using them as the background resources for doing other scientific work. One thing they can do is to help us understand why the theory has correlational gaps. From the very beginning, one wondered about the incompatible observables and why one could not even in principle imagine joint measurements for them. After all, as Schrödinger (I believe) first pointed out, in the EPR situation, one could measure position in one wing and momentum in the other and, via the conservation laws, attribute simultaneous position and momentum in both wings.

The conventional response here has been to point out that only the direct measurements yield values that are predictively useful. (See note 4) Not everyone has been happy with the positivism that seems built into this response. But if we recall the discussion in section 1, then we see that (at least in part) there is a better response at hand. For we have seen how the correlations that the theory does provide actually exclude the possibility that there could be any stable joint distributions for incompatible observables in those states where the correlations are tangled. This shows us that there is no way of augmenting the theory with values for incompatible observables, and distributions for those, that would follow the same lawlike patterns as do the distributions of the theory itself. To put it dramatically, the shadow of the given correlations for compatible observables makes it impossible to grow stable correlations for the incompatibles. There is a sense, then, in which there would be no point in trying to introduce more for incompatible observables than what the theory already provides.

This way of thinking turns the Bell theorem around. Instead of aiming to demonstrate some limitation or anomaly about the theory, this way proceeds in the other direction and helps us understand why the probability structure of the theory is what it is. That understanding comes about when we

take a nonessentialist attitude toward explanation, letting the indeterminist ideals of the theory set the explanatory agenda. Such an attitude means taking the theoretical givens seriously, and trusting that they will do good explanatory work. Thus, in the Bell situation, we shift our perspective and use the given quantum correlations (and the simple sort of counting argument rehearsed in section 1) to explain why, even in principle, correlations forbidden by the theory cannot arise. Nonessentialism leads us to engage with our theories seriously, and in detail. In the end, that is how better understanding comes about.

What then of nonlocality, influences, dependencies, passions, and the like, all diagnosed from correlational data? As one good statistician remarked about the similar move from linear regression to causal connection, and as we have seen demonstrated above, "*Much less is true.*"[6]

[6] "It is easy to think that when we . . . find a linear regression of y on x (a statistically significant regression), we have evidence that increasing x causes y to increase. *Much less is true*" (Moses 1986, 294).

BELL'S THEOREM, IDEOLOGY, AND STRUCTURAL EXPLANATION

R. I. G. HUGHES

1. Ideology and explanation

Duhem regarded with skepticism the suggestion that the aim of scientific theory was to furnish explanations.[1] Explanations, on his view, are attempts to account for phenomenal laws in terms of prior metaphysical assumptions; it follows that,

> If an appeal is made, in the course of the explanation of a physical phenomenon to some law which that metaphysics is powerless to justify, then no explanation will be forthcoming, and physical theory will have failed in its aim.[2]

Further,

> no metaphysics gives instruction exact enough, or detailed enough to make it possible to derive all the elements of a physical theory from it.[3]

The conclusion he draws is not that physical theory is doomed to fail in its aim, but rather that to view this aim as the production of explanations is a mistake.

At first sight, the problems arising from Bell's theorem seem a striking corroboration of Duhem's views. On any sane account, quantum mechanics is a remarkably successful theory. Yet some of the events it successfully predicts not only resist explanation, they actually undermine a cluster of metaphysical beliefs. Specifically, putative explanations of EPR-type correlations

[1] P. Duhem, *The Aim and Structure of Physical Theory*, trans. P. P. Wiener (New York: Athenaeum, 1962), chap. 1.

[2] Ibid., p. 16.

[3] Ibid.