The Delicacy of Causal Ascription and Bell's Theorem

Erik Curiel

August 24, 2009

Contents

> The utility of a notion testifies not to its clarity but rather to the philosophical importance of clarifying it.
>
> Nelson Goodman
> *Fact, Fiction and Forecast*

1 Introduction

Quantum mechanics predicts startling behavior for pairs of certain types of particles (those described as having 'spin-$\frac{1}{2}$') in a certain joint state (the 'singlet' state) under certain experimental conditions. Such an experimental set-up is as follows. In the middle of a (very large) room, there is a device for producing pairs of spin-$\frac{1}{2}$ particles in the singlet state. The particles of each pair shoot off in opposite directions, one to the left, the other to the right, toward measuring apparatuses waiting at each end of the room to measure the spin of the particles in a particular direction. According to quantum mechanics, the measuring apparatuses will always record either a value of 'spin-up' or a value of 'spin-down' for each particle, each with a probability of 50%: quantum mechanics does not allow one to predict what value will be measured for any given particle, though it does demand that for any large enough ensemble of particles the two values will be measured with roughly the same frequency. If the apparatuses on opposite ends of the room are set up to measure spin in the same direction, moreover, then quantum mechanics also predicts that measurements on particles of the same pair will be strictly anti-correlated: if up is measured for the particle passing through the left measuring apparatus, for instance, then down will be measured for the other particle with probability 1. If the apparatuses on opposite ends of the room are poised to measure spin in directions different from each other, then quantum mechanics still predicts that the results of the measurements on particles of the same pair will be produced with particular correlations, though not with perfect anti-coorelation: when the directions are close, then the measurement results will be nearly perfectly anti-coorelated; as the directions get farther apart, the measurement results for particles in the same pair will become less strongly anti-correlated, until the directions become perpendicular, at which point there will be no correlation become the measurement results on opposite sides at all—the results will be statistically independent of each other; finally, as the difference in the directions of the spin measurements on the opposite apparatuses continue to increase beyond 90 degrees, the measurement results will begin to be correlated with each other, achieving perfect correlation when the spin measurements are made in exactly opposite directions.

This results by themselves are not terribly surprising. That the results of certain measurements on the particles should exhibit strong correlations, even when the measurements are made when the particles are far apart, can surely be explained by the fact that the particles have a common origin—they were produced in the same spot, in pairs, and so some aspect of their interaction with each other when they were created, or of their respective interactions with a common element of the production device, ought to be able to explain the information each particle appears to have about how its mate will behave under certain circumstances. In more evocative terms, the correlated behavior of the particles should be explicable by positing a common cause in their joint histories. Such a postulated common cause is usually referred to as a 'hidden-variable', since there is no explicit feature of quantum mechanics that plays the role this postulated common cause is supposed to play.

The truly surprising fact is that such a common cause cannot be found. That is the gist of Bell's Theorem. John Bell showed in 1964, roughly speaking, that the correlations in these sorts of experimental results could be explained by reference to such a common cause only if the measurement statistics satisfied a certain set of inequalities, the so-called Bell inequalities. Not only do the predictions of quantum mechanics not satisfy the Bell inequalities, but, more what is more striking, experiments designed to test them indicate that the real world does not either.[1] Because the measurements on the separated particles can be made as far apart as one likes, the upshot is that the real world appears to delight in the existence of striking correlations between arbitrarily separated phenomena, while refusing to countenance any explanation for these correlations in terms of the common causal past of the phenomena. One wonders mopre physicists are not conspiracy theorists in their politics.

In 1984 Jon Jarrett published an analysis of the "strong locality" condition used to derive Bell-type inequalities for stochastic, contextual local (models of) hidden-variable theories that reproduce the quantum mechanical spin statistics. It was known before this that Kochen-Specker takes care of all non-contextual theories (local or not), and the derivation of the Kochen-Specker theorem was

[1] *Cf.* Aspect, Dalibard, and Roger (1982).

uncontroversial. Likewise, there was no serious debate about the physical significance of the weaker locality condition needed to derive Bell-type inequalities for local deterministic theories. There was, however, no general agreement in the literature at the time on the physical significance (if any) of strong locality over and above its statistical requirements.[2] Since the results of experiments strongly suggest that the physical world violates Bell-type inequalities, by *modus tollens* one (at least) of the assumptions behind the inequalities had to go, and the faulty premise was almost universally adjudged to be strong locality—but, since no one was able to offer a convincing physical explication of strong locality, no one was able to provide a deeper physical analysis of what is involved in abandoning it.[3]

Jarrett offered a precise characterization of what may be involved in dispensing with strong locality by demonstrating that it is logically equivalent to the conjunction of two other (logically independent) conditions, "locality" and "completeness", each of which, he argued, has more perspicuous physical content than strong locality. Jarrett's analysis thus offered a refined set of alternatives: we can reject strong locality, and so (in accord with reality) avoid a physical theory satisfying the Bell inequalities, by dispensing either with locality or with completeness (or both). By arguing that locality is practically equivalent to a ban on the possibility of superluminal signaling in the context of Bell-type experimental arrangements, Jarrett concluded that any adequate physical theory should violate completeness, and not locality, since superluminal signaling[4] would *prima facie* constitute a serious conflict with relativity theory,[5] an infelicitous state of affairs.

In this paper, I shall neither attack nor defend Jarrett's conclusions concerning adequacy constraints appropriate to (models of) physical theories mooted in discussions of Bell's Theorem. Rather I shall attempt to clarify what sorts of arguments can and cannot coherently be made when Jarrett-type premises are accepted (or are accepted at least for the sake of argument). In particular I want to show that certain recent construals of Jarrett's 1984 argument that focus on the notion of causality not only are beside the point of Jarrett's argument but, more important, obfuscate what salutary can be gleaned from his argument. Martin Jones and Rob Clifton (1993) among recent commentators on Jarrett not only are notable for their clarity, precision and thoroughness of argument, but they also are typical in what, I suspect, is the main culprit behind confused arguments that are concerned with issues of causality, *viz.* a careless deployment of the notion of 'causality' itself. For this reason

[2] *Cf.* Hellman (1983), especially pp. 603–6, for a helpful survey of the issues as understood at the time.

[3] This paper will ignore the possibility of another of the assumptions behind Bell-type inequalities failing, as consideration of it would take us too far afield; for what it is worth, the prevailing philosophical winds are certainly against such an option.

[4] To avoid tedious repetition, it should be understood throughout this paper unless explicitly stated otherwise that by 'superluminal signaling' I always and only mean the possibility of superluminal signaling specifically in the context of Bell-type experimental arrangements.

[5] There is some controversy concerning the precise nature of the relationship between the possibility of superluminal signaling and the demands of relativity theory. For a thorough discussion of the issues involved see Friedman (1983, ch. 4, §§5–7). The arguments of this paper do not depend on the correctness of of any of the positions in this controversy. I thus should be understood to use 'superluminal' only as a convenient word to refer to any process that would connect events in spacelike relation to one another, with no express commitment as to whether there is 'really something there moving faster than light'.

I have chosen them as a foil, precisely because theirs seems to me the clearest, best case for what I take to be at bottom a confused way both of thinking of Jarrett's argument in particular, and of deploying causal arguments in general.

Jones and Clifton argue against Jarrett in the following way: by producing a model of a theory that can accommodate the possibility of superluminal signaling and in which the signaling seems to be possible 'because of' completeness violations, Jones and Clifton argue that locality does not in fact have the privileged relation to superluminal signaling that Jarrett ascribes to it. I shall argue that Jones and Clifton's proposed model, while flawless technically, cannot serve as a counter-example to Jarrett's analysis. Nothing in Jarrett's paper rules out the model that Jones and Clifton propose as a counter-example—so far from it, Jones and Clifton's model is wholly consistent with, and can even be taken to support Jarrett's conclusions, in so far as locality violations necessarily accompany all superluminal signaling in Jones and Clifton's proposed model, even though it is a novel kind of model, one not explicitly considered by Jarrett in his arguments.

Their arguments fail because they misapprehend what it is that Jarrett argues the significant relation between locality and superluminal signaling to be. Jones and Clifton believe that determining which of the conditions, locality or completeness, can be causally implicated in superluminal signaling is the crux of the debate as Jarrett framed it, but I shall argue that in this they are mistaken.[6] Jarrett argues only for a *logical* connection between locality and superluminal signaling, completely sidestepping the nasty issue of what is 'responsible for' or 'causes' the signaling, or 'in virtue of' what is signaling possible, and indeed this logical connection is all he requires for the argument to work. Whatever else is the case about the real world, he argues, and no matter what mechanism, process or connection, causal or otherwise, one may wish to dream up to explain superluminal signaling, the possibility of superluminal signaling and of locality violations (in principle) imply each other, whereas the possibility of superluminal signaling and of completeness violations have no such logical connection. In any particular model that violates completeness superluminal signaling may be possible, or it may not—from the simple fact of completeness violations in a model, we can say nothing about the possibility of superluminal signaling in that model without further information, such information essentially, as I shall argue, consisting of the status of locality in the proposed model. From this it is easy to conclude that we should look for physical theories that violate completeness but respect locality, so long as we wish both to avoid possible relativistic headaches and to have a theory that violates Bell's inequality.

One need not (and indeed ought not) refer to causality in any way to make this argument. Jones and Clifton's introduction of causal considerations into Jarrett-type arguments not only obfuscates the fundamental issues but ultimately detracts from the cogency of their own analysis. Among other things, it is not clear at all what they mean by claiming that, in their model, there is superluminal signaling 'because of' completeness violations. Such carelessness derives, I think, in part from an overly formal approach, which obscures the important physical differences between completeness and locality as conditions on physical theories, and thus also obscures the respective power of each

[6] Kronz (1990), Butterfield (1992) and Maudlin (1994, pp. 94–6) misread Jarrett in a very similar way, and come to conclusions similar to those of Jones and Clifton.

to justify causal conclusions drawn from a set of simple correlations between distant physical events.

The problem of giving a general criterion for the presence of causality is famously murky, but there is no doubt that one always does stand in need of something else besides correlation to justify causal assertions. In specific cases, it may reasonably be hoped that one can specify what else is required besides the presence of correlations to make warranted causal judgments, even in the absence of a larger, more general theory of causality. Such a thought is at the heart of much work on Bell's Theorem: no matter what else, if anything, one thinks may mark the presence of a causal connection, surely the ability to send a signal bespeaks causality, at least in any everyday sense in which one uses the word.[7] Once so much has been said, though, there is still the problem of delineating precisely what one means by 'causality' in this context, and examining one's causal conclusions to ensure that they are consistent with the constraints on one's notion of causality.

There is a tendency in much of the contemporary philosophical literature dealing with causal issues, no matter the specific topic under discussion, to conflate different notions of causality that do not reside comfortably with each other, resulting in confused arguments and untenable positions. Though this confusion is endemic to many fields in contemporary philosophy, I shall focus on Bell's Theorem and Jarrett's argument because this seems to me an area where such confusions arise with peculiar rapidity and because there is the possibility here that some relatively simple and straightforward considerations can clear the air and allow larger lessons to be drawn easily. My overall intent is thus more broad than it may initally appear, for the considerations I present in this paper apply to all arguments that would move from correlation to causation.

2 Jones and Clifton's Argument Part 1: The Formal Situation

Jones and Clifton argue, on the basis of their proposed counter-example to Jarrett, that "there is nothing in the intrinsic nature of failures of COMP [completeness] which prevents their use in superluminal signaling, and so nothing intrinsically noncausal about such failures; it is not even true that violations of COMP are intrinsically less connected to superluminal signaling than violations of LOC [locality]."[8] They base this argument on a theorem they prove to the effect that "if the outcome at one end of an EPR-Bohm experiment is statistically dependent on the hidden microstate of the apparatus at that end, and if one other locality condition is satisfied, then violations of completeness

[7] I do not necessarily share the view that the presence of signaling implies the presence of a causal connection, at least not unless extreme care is taken with the phrase 'causal connection'; this assumption is merely the one that is used in such debates. One may, for instance, wish to deny a necessary link between signaling and causality because it is impossible to rule out entirely the possibility of conspiratorial mechanisms.

[8] Jones and Clifton (1993, pp. 3–4). The notion of 'intrinsic comparitive connectedness' that Jones and Clifton employ here is obscure to me. I believe what they intend is something like the following: they are going to argue that one needs precisely the same type of, if not the same, assumptions to demonstrate the possibility of superluminal signaling given locality violations as one would need to demonstrate this possibility given completeness violations, so locality cannot be privileged over completeness with respect to superluminal signaling.

can straightforwardly lead to superluminal signaling."[9] I shall first examine the logical structure of their arguments, then compare their arguments to Jarrett's original argument, and conclude with a more general discussion both of what it is legitimate to presuppose and of what it is legitimate to infer about the causal structure of such scenarios.

Jones and Clifton represent locality and completeness in the following way:
Locality:

$$\forall A, B, a, b, b', \alpha, \beta, \beta', \lambda \ (A|ab\alpha\beta\lambda) = (A|ab'\alpha\beta'\lambda)$$

and

$$\forall A, B, a, b, a', \alpha, \beta, \beta', \lambda \ (B|ab\alpha\beta\lambda) = (B|a'b\alpha'\beta\lambda)$$

Completeness:

$$\forall A, B, a, b, \alpha, \beta, \lambda \ (AB|ab\alpha\beta\lambda) = (A|ab\alpha\beta\lambda)(B|ab\alpha\beta\lambda).$$

$(X|Y)$ is the usual conditional probability of X given Y;[10] λ is the total joint state of the two-particle system at the point of measurement (they take this to be on a spacelike hyperplane that includes both experiments) including all hidden variables; A (B) is the measurement outcome on the left (right), +1 for spin-up and -1 for spin-down (assuming experiments performed on spin-$\frac{1}{2}$ particles); a (b) is the spin-direction measurement setting of the experimental apparatus on the left (right); α (β) is the total state of the left (right) apparatus including hidden variables, modulo the direction of spin-measurement. These parameters are also quantified over the no-measurement scenarios; we stipulate the convention in this case that the experimenter measures the value 0 with probability 1. Note that there are many possible distinct apparatus states that all correspond to 'no measurement made'. 'A', 'a', 'α', 'λ' and other terms of these types will be used to represent both values of the relevant parameters and the events of those parameters taking those values, with context indicating which is intended. 'L' ('R') will be used promiscuously to refer to the experimenter on the left (right) side, the entire left (right) side experimental apparatus arrangement, the left (right) side experiment, *etc.*, again with context indicating which is intended.

Following Jarrett, we shall be concerned only with physical theories that

1. employ state-descriptions of measuring apparatuses and two-particle systems yielding unique values for all relevant probabilities, such values obeying all the usual requirements imposed upon probabilities;

2. postulate time-independent states, where 'time-independent' here need refer only to the character of the relevant probabilities under consideration

3. return the quantum mechanical spin statistics when averaged over all hidden variables.[11]

[9] *Ibid.* p. 3

[10] The nature of these probabilities, whether they result from deterministic or stochastic laws, is irrelevant for this section of the paper but will arise later. For a thorough discussion of the differences in the two cases *cf.* Hellman (1982a) and Hellman (1982b).

[11] For more on the second stipulation, see Jarrett (1984, pp. 571–2), and Jarrett (1986).

Again following Jarrett, who follows Skyrms (1982), "the assertions of the theory are to be regarded as subjunctive conditionals of the form: If measuring device A, in state $\alpha\ldots$, and measuring device B, in state $\beta\ldots$, were to measure respectively, the a-component of the spin of particle A and the b-component of the spin of particle B, for a pair of particles in the state $\lambda\ldots$, then the probability that the outcome of the L measurement is A and the R measurement is B would be $(AB|ab\alpha\beta\lambda)$."[12]

Locality stipulates that the probability of L's measuring any outcome A given a and α is independent of the performance of any particular measurement at R, *i.e.* is independent of the state of the measuring apparatus at R. Of course a similar condition holds for R; throughout this discussion anything said of L should, *mutatis mutandis*, be taken also to hold of R, and vice-versa. Locality thus denies L the power to alter the probabilities of measurement outcomes at R by manipulating his experimental apparatus, which *prima facie* appears to deny him the capability of transmitting information by performing such experiments. Completeness, on the other hand, stipulates that the L and R experimental outcomes be stochastically independent of each other: the value that the L measuring apparatus records is neither positively nor negatively correlated with that of R, but is statistically indepentent of it. Completeness does not exclude the possibility that, say, the probability of a certain L outcome may depend in some way on the performance of some particular experiment at R (on the state of the R apparatus).

Jones and Clifton begin arguing in earnest by introducing a few more conditions.
Measurement Contextualism:

$$\exists B, \beta, \beta', X (B|\beta X) \neq (B|\beta' X)$$

Constrained Locality:

$$\exists A, B, \beta, \beta' (\neq \beta), X (A|B\beta X) = (A|B\beta' X).$$

'X' here and in subsequent discussion replaces '$ab\alpha\lambda$', which is assumed to be fixed for the argument. Measurement contextualism represents the possibility that the experimental outcome at one end stochastically depend on some aspect of the state of the apparatus other than the angle of spin-measurement at that same end.[13] By postulating a model in which measurement contextualism and constrained locality hold but completeness is violated and in which the usual spin statistics of quantum mechanics can be represented, Jones and Clifton have an easy time showing that signaling is possible, given one more set of assumptions we shall come to in a moment. Here is how they do it.

[12] Jarrett (1984), p. 586, n. 6. I have altered Jarrett's notation so as to accord with that of this paper; the formulations are logically equivalent.

[13] That measurement contextualism must operate stochastically is necessitated by a recent argument due to Jarrett (1989, pp. 69–75), in which he concludes that violation of completeness entails the failure of "measurement determinism": if completeness fails, then there must be an element of randomness in the production of experimental outcomes, and Jones and Clifton are specifically considering models in which completeness fails. The fact that one of Jarrett's motives for making this argument is to show that completeness violations cannot be exploited to send signals, and thus is inherently noncausal, suggests that Jarrett and I do not share the same understanding of his argument and its consequences.

Formally, Jones and Clifton postulate a model that satisfies a condition they term 'ACTION':

$$\exists A, B, B', \beta, \beta', X\ [\{(B|\beta X) \neq (B|\beta' X)\} \wedge \{(A|B\beta X) \neq (A|B'\beta X)\} \wedge \{(A|B\beta X) = (A|B\beta' X)\} \wedge \{(A|B'\beta X) = (A|B'\beta' X)$$

The first conjunct represents satisfaction of measurement contextualism, the second violation of completeness, and the last two satisfaction of constrained locality. It is a straightforward matter to deduce a violation of locality from ACTION: "It then follows by Jarrett's original argument that superluminal signaling is possible in such theories."[14]

More informally, the scheme is as follows. First we assume that the experimenters can both have knowledge of the underlying state λ of the particles and produce particles in the appropriate states, and that the aspect of the experimental apparatus labelled by 'b' is (at least stochastically) manipulable by the experimenters.[15] Next we assume that the experimenters prepare a large ensemble of particles in the state λ associated with the locality violation and pass them through their experimental apparatuses in such a way that their respective measurements on the particles will be spacelike related to one another. Then the experimenter on the right chooses whether he will change the relative frequency of the outcome B at his end by manipulation of β and thus, because completeness is violated by hypothesis, alter the probability on the left that A occurs (of course, the same value of B must satisfy measurement contextualism and be involved in the completeness violation for this scheme to work). The experimenter on the left, by hypothesis knowing the probability of obtaining outcome A conditioned on the relevant values of B obtaining on the right, is then easily able to deduce whether or not the experimenter on the right manipulated β in a specific way, bracketing all the usual problems associated with verifying a probabilistic statement. The experimenter on the right has signaled the experimenter on the left his choice of β manipulation.

Jones and Clifton introduce constrained locality to ensure that changing the value of β on the right does not 'directly' affect the probability of A on the left, via some unknown mechanism, in such a way as to cancel out the effect that the change in the probability of B had on the probability of A. Jones and Clifton's explicit motivation for distinguishing a 'direct' from an 'indirect' effect of β on A is the "plausible reading" of completeness violations and measurement contextualism as both being causal in nature, since both are correlated with signaling. They feel justified in asserting that it can*not* be a 'direct causal link' between β and A that occasions the signaling, since in their postulated model manipulations of β alter the probability of the relevant values of A only when they also alter the probability of the relevant values of B, or, in fancier terms, alterations in B 'screen-off' the correlations between β and A. They conclude that the signaling must be propagated only via a causal link running from B to A, B having been influenced by β, thus demonstrating that completeness violations need not be noncausal. The situation they envisage is represented below in

[14] Jones and Clifton (1993, p. 13). It is worthwhile noting the *recherché* aspect of such a model, in that the conditions must collude in just the proper way so that the same set of parameter values satisfies measurement contextualism, satisfies constrained locality and violates completeness (or, more precisely, respectively satisfies and violates the schemata derived from measurement contextualism, constrained locality and completeness by removing the quantifiers via existential instantiation).

[15] I shall discuss the status of these assumptions concerning the knowability and manipulability of micro-states more thoroughly in the next section of the paper.

figure 1. It appears we must conclude with Jones and Clifton, therefore, that superluminal signaling is a distinct possibility in such models, and that, furthermore, it is the completeness violations that would "lead to" such signaling, at least for the parameter values in question.

To assume that L can manipulate β and monitor λ might at first blush appear unreasonable, since these are presumably hidden micro-states of macro-apparatuses and atomic particles, neither of which are, to the best of our knowledge, amenable to macroscopic measurement and manipulation. Jones and Clifton's answer to this charge is the final, and to them crucial, step in their argument against Jarrett. Since Jarrett has to make precisely the same sort of manipulability assumptions in his demonstration that violations of locality are practically equivalent to the possibility of superluminal signaling, Jones and Clifton conclude that the association of locality violations with superluminal signaling stands and falls with the like association of completeness violations they purport to have demonstrated: if one assumes manipulability of the relevant parameters, it seems one can show *both* locality and completeness to be causally implicated in superluminal signaling.[16] In short, Jarrett's argument that violations of completeness are to be preferred, as they are not necessarily implicated in superluminal signaling, whereas violations of locality are so implicated, has failed.

Have Jones and Clifton succeeded in their quest to demonstrate that locality is not privileged over completeness in its relations with superluminal signaling? It appears not, at least not in terms of logical relation, and by their own lights to boot. It is as if, as Howard Stein joked, one claimed to be able to fly to London by wristwatch and then boarded a London-bound 747 wearing the wristwatch to demonstrate the proposition. When the absurdity is pointed out, it will not do to reply, "Yes, you're quite right the 747 supplied the motive force, but, since I was in physical contact with both the wristwatch and the 747 in the relevant respective fashions, this shows that the wristwatch is intrinsically as related to my reaching London as the 747." If, in their model, violations of completeness can effect superluminal signaling *only in so far as* such violations are implicated in violations of locality, it is simply wrong, from a formal point of view, to claim that completeness bears the same relation to superluminal signaling as does locality.

This can most easily be seen in the following schematization of the upshot of their own arguments. They take as given

$$\neg(\text{locality}) \Rightarrow (\text{signaling}), \tag{2.1}$$

and they demonstrate

$$(\text{ACTION}) \Rightarrow \neg(\text{locality}). \tag{2.2}$$

From (2.1) and (2.2) they immediately deduce

$$(\text{ACTION}) \Rightarrow \neg(\text{signaling}). \tag{2.3}$$

Even though we also have

$$(\text{ACTION}) \Rightarrow \neg(\text{completeness}), \tag{2.4}$$

this plays no role in their argument, and

$$\neg(\text{completeness}) \Rightarrow (\text{ACTION}) \tag{2.5}$$

[16] *Cf. ibid.*, pp. 16–17.

is not even a sound inference. There is a manifest difference in the respective relations of completeness and locality to signaling. Jones and Clifton do not stress the fact that the completeness violation plays no logical role in their argument comparable to that of locality, but this fact raises an immediate problem for them: if they are right in assuming that it is the causal implications of the variables quantified over in a statement of a condition on a physical theory that determine how that condition is causally related to the possibility of signaling, then one must ask why Jones and Clifton could not make their argument directly from what they contend are in their model the causal implications of completeness violations to the possibility of signaling, without piggy-backing on locality violations. One certainly need make no reference to completeness violations in arguing that locality violations are related to signaling. We must therefore examine whether Jones and Clifton were in the right to introduce causal considerations as the salient feature in their analysis of the superluminal signaling possible in their proposed model. We shall begin by looking at Jarrett's original argument.

3 Jarrett's Argument

Jarrett's argument consists of his demonstration that strong locality as usually adduced for the deduction of Bell-type theorems,

$$(AB|ab\alpha\beta\lambda) = (A|a\alpha\lambda)(B|b\beta\lambda),$$

is equivalent to the conjunction of completeness and locality, together with his analysis of the respective relations of completeness and locality to superluminal signaling. Since the demonstration that strong locality is logically equivalent to the conjunction of completeness and locality is straightforward and uncontroversial, I shall take it as given, and concern myself with examining the arguments on signaling. I shall first quickly recapitulate Jarrett's arguments for the (in principle) equivalence of locality violations and the possibility of superluminal signaling before considering in detail Jones and Clifton's arguments against them. I do not completely follow Jarrett's methods nor do I endorse all his conclusions, as will be clear by the end of this section.

There are actually two slight problems in Jarrett's formulation of the argument it is necessary to address before I rehearse it. First, Jarrett assumes that the experimenters in the deterministic case know with certainty the hidden state of the two-particle system,[17] but this seems on the face of it unreasonable. It is one thing to assume that the hidden state determines the outcome of experiments, but it is something else entirely to assume that the experimenters *know* the determining hidden state. Because, to the best of our knowledge, any adequate physical theory must be essentially quantum mechanical, at least *vis-à-vis* our possible observations of the world, the most we can reasonably assume known by actual experimenters is a probability distribution over the space of hidden states.

A related problem is Jarrett's assumption that the experimenters can reliably prepare particles to be in a particular hidden state.[18] Preparability also seems an unreasonably strong assumption to burden a hidden-variables theory with, again at least in so far as actual experimenters are concerned,

[17] *Cf.* Jarrett (1984), *e.g.* p. 575, the parenthetical remark near the end of the first paragraph on the page.
[18] *Cf.* *ibid.*, pp. 574-5, n. 13.

for the same reasons that knowability seems burdensome. It might seem at first that knowability and preparability are really the same problem, but they are in fact distinct. For instance, it just may be the case that repeating the same experimental procedures would reliably produce particles in the same state, but the theory, for whatever reason, would not allow one to know this. Similarly, a theory may allow one always to discover the state of a particle and force one to destroy the particle in the process, in which case one could always know a particle's state and yet be unable to prepare particles in a particular state. Nothing in this paper hinges on the distinction between the two problems, and I shall ignore it; I point out the difference merely because the tendency in the literature is to conflate them.

The assumption of strong knowability and preparability stems from the (*prima facie* plausible) idea that to signal one must be able to control all the parameters relevant to the signaling. Jones and Clifton rely upon this tacitly when they argue that completeness and locality each bear the same relation to signaling since signaling 'via' completeness violations and signaling 'via' locality violations both depend on the same assumption, *viz.* that the appropriate parameters be controllable.[19] For discussion of signaling, however, we actually can make do with a weaker set of assumptions than that of controllability of all relevant parameters. We note first of all that, in Bell-type scenarios, we are concerned with large groups of pairs of particles, each pair of which is in the same quantum state, *viz.* the singlet state for spin-$\frac{1}{2}$ particles. We certainly ought not assume, though, that all these particles are in the same hidden state. We do assume that some subset of these particles are in a hidden state for which the theory in question predicts a locality violation when those particles interact with experimental apparatuses in one of a pair of given states (even though the ensemble as a whole will return the quantum mechanical spin predictions).[20] The experimenters then do not need to control nor even influence these underlying hidden states of the particles in any way; they need only to be able to *monitor* the underlying hidden states, *i.e.* to have the capacity to determine with some relatively high probability of success which particles are in which hidden states (or at least know which ones are in the hidden state associated with the locality violation), and to know something about the behavior of the particles in the various hidden states (or at least know something about the behavior of the particles in the hidden state associated with the locality violation).

As long as the ensemble of particles is large enough, the receiver of the signal will be able to winnow from it a statistically significant number of particles in the proper locality-violating hidden states, and so be able to interpret a superluminal signal. Of course, as noted above, this is an unreasonably strong assumption if we are talking about flesh-and-blood experimenters in skew laboratories subject to quantum mechanical fiat. The important point, though, is that as an idealization it does not seem unreasonable. If a physical theory predicts that there will be differences in correlations between certain experimental arrangements on one side and certain experimental

[19] *Cf.* Jones and Clifton (1993), pp. 16–17.

[20] I am assuming that the locality violation is not so fine-grained as to require the experimental apparatuses to be in a particular hidden state, but that at least some locality violations in the theory are associated with apparatuses in particular macro-states. The same argument, in a more cumbersome form, would still go through even if one required locality violations to depend on apparatus micro-states. The experimenters would then simply need the ability to monitor the micro-states of the apparatuses.

outcomes on the other, then the theory in principle allows superluminal signals, even though we may never have the finesse to pull the trick off. For those who take signaling as a mark of causality, and who take relativity as denying the possibility of superluminal causality, this by itself is sufficient to conjure up clashes with relativity theory.

The obvious objection to this line of thought is to postulate hidden-variable theories that do not allow such fine access to the hidden states as needed for Jarrett-type arguments. There are several ways one may go about doing this. One may postulate as a fundamental axiom of one's model physical limitations upon state preparation in order to preclude such superluminal signaling, but this would not suffice, for as I argued above the experimenters need only to be able to monitor, not to prepare, the hidden states. One may then postulate as a fundamental axiom of one's model physical limitations upon the monitoring of hidden states. Taking a cue from Shimony,[21] I would argue that the anthropocentricity of such a maneuver leaves an unacceptable feeling of unease—it is surely inappropriate to mingle considerations concerning the possible experimental limitations of our ability to monitor hidden states with the fundamental axioms of a physical theory. Bell himself stressed a similar point repeatedly.[22] Finally, one may postulate a model that does not hold it as fundamental that we cannot monitor the hidden states, but rather one such that it is demonstrable within the model from its fundamental physical axioms that we will in the event be denied the ability to monitor the hidden states finely enough to signal. This, however, can*not* suffice as a rebuttal of Jarrett-type arguments, for any such demonstration would have to rely upon a particular modeling of our particular experimental practices in laboratories, a point that is not appreciated often enough. There is, after all, no canonical method of representing specific types of physical interactions, such as laboratory experiments, in any given physical theory. Such a demonstration could not, then, entirely rule out the physical possibility of monitoring the hidden states finely enough; at most, it could demonstrate only that such as we using techniques currently imaginable could not do it.

After so much is said, there certainly remain very general, and profound, philosophical questions concerning the modeling of experiments in physical theories. Such problems are important, and worthy of consideration, but they are not especially debilitating to Jarrett's arguments, no more so, at least, than to arguments concerning other types of *Gedankenexperimente*. In the end, even though Jarrett's argument can only show that superluminal signaling and locality violations are 'in principle' equivalent, this hedge is irrelevant to the use he puts his conclusion to.

First, then, a demonstration that if a (model of a) theory predicts locality violations, then we can expect, in principle, superluminal signaling to be possible.[23] I shall not discriminate between deterministic and indeterministic models in my treatment, presuming that the theory at issue is indeterministic. Deterministic theories can be understood in my treatment as the special case in which the relevant probabilities assume the values of 0 and 1. Since by hypothesis locality is violated, for some set of values of all the parameters we have:

$$(A|ab\alpha\beta\lambda) \neq (A|ab'\alpha\beta'\lambda).$$

[21] Shimony (1993a, p. 139) (in a post-paper comment).
[22] *Cf.*, *e.g.*, Bell (1987, pp. 117–118.).
[23] *Cf.* Jarrett (1984, pp. 573–6.).

For definiteness, let us assume that

$$(A|ab\alpha\beta\lambda) = A_b$$

and

$$(A|ab'\alpha\beta'\lambda) = A_{b'},$$

where $\{a,\ \alpha\}$ is a particular L apparatus state when spin is measured in the a direction, $\{b',\ \beta'\}$ is an R apparatus state when no measurement is made, and $\{b,\ \beta\}$ is a particular R apparatus state when a definite spin measurement is made in the b direction. Now suppose that L and R prepare a large ensemble of pairs of spin-$\frac{1}{2}$ particles in the singlet state. L and R agree to make spin measurements as follows: L will definitely measure spin direction a by preparing all of his apparatuses to be in $\{a,\ \alpha\}$; R, however, will have the choice of either measuring spin direction b by preparing his apparatuses to be in $\{b,\ \beta\}$, or else make no measurement at all by preparing his apparatuses in $\{b',\ \beta'\}$ (he can set them all instantly with one flip of a switch). R will, moreover, make the choice in such a way that the entire decision process and the setting of the apparatuses are both spacelike related to L's measurement—perhaps he plays rocks-scissors-paper with his lab assistant, and if R wins they make the b measurement, but if his assistant wins they make no measurement and have a cup of tea. Now, by monitoring the hidden states of the incoming particles, L will easily be able to determine whether R's assistant got his cup of tea. If, for that subensemble of particles in state λ, he measures A with a frequency of A_b, he knows R made a particular measurement; if, for the same subensemble, he measures A with a frequency of $A_{b'}$, he knows R made no measurement at all. R has signaled to L the results of his rocks-scissors-paper game 'more quickly than light could traverse the spatial distance between them'. Of course, we are assuming that the ensembles are large enough to make such a reliable determination. Nothing guarantees that L will get something approximating the correct result, and indeed L will not know with apodeictic certainty R's decision, but this is irrelevant to the case at hand. Whether or not we can ever determine probabilities with certainty, a physical state of affairs at L, *viz.* the probability that A assumes certain values, is affected by R's spacelike related decision in such a way that a signal could in principle be sent by exploiting the correlation.

A stronger conclusion than this is possible, though. By demonstrating that the possibility of superluminal signaling implies (in principle) a locality violation, we would have shown that locality violations and superluminal signaling are (in principle) logically equivalent. I shall now review the argument for this second implication. Again, I shall conflate the deterministic and indeterministic cases. Suppose that L and R prepare a large ensemble of pairs of spin-$\frac{1}{2}$ particles in the singlet state. Suppose further that by measuring, *e.g.*, the a spin-component of all the particles in the hidden state λ, L can determine whether or not R has performed a measurement at all. For this to be the case, there must be a suitable correspondence between the relative frequency of a certain outcome at L and something about the measurement process at R. For argument's sake, let us say that L experimentally finds a value of A with a relative frequency of A_b just in case R performs a measurement with his apparatuses in the state $\{b,\ \beta\}$, and L experimentally finds A with a relative frequency of $A_{b'}$ just in case R does not perform a measurement, with his apparatuses in state

$\{b', \beta'\}$. If 'p' represents some previously agreed upon relative frequency of A that, according to the theory in question, L 'ought' to measure in the subensemble of λ-particles if R makes a particular measurement, then we have

$$|p - A - b| \approx 0 \Leftrightarrow \text{particular measurement made at } R,$$

$$|p - A_{b'}| >> 0 \Leftrightarrow \text{no measurement made at } R.$$

Thus, for some values of A, a, α, b, β, b', β' and λ, we must have, if the theory is adequate, that it predicts

$$p = (A|a\alpha b\beta\lambda) = A_b$$

and

$$p \neq (A|a\alpha b'\beta'\lambda) = A_{b'}.$$

Otherwise, L's measuring A with anything approximating a relative frequency of A_b would be compatible with no measurement being made at R. But these two predictions constitute a violation of locality. We thus conclude that locality violations are (in principle) equivalent to the possibility of superluminal signaling.[24] Because completeness and locality are manifestly not equivalent, we can immediately conclude that completeness is not equivalent to the possibility of superluminal signaling.

Before returning to Jones and Clifton's arguments, I want first to explicate an important subtlety of Jarrett's argument. We are thus far, from what we have argued, not licensed to infer just what about the act of measurement is causally implicated in the signaling and how it is so—if indeed we wish to discuss causal relations at all, for note that there was no need to mention causality in the demonstrations themselves. If we do wish to take the presence of signaling as a sufficient marker for a causal connection, then all we know for sure is that there is *some* difference between measuring and not-measuring that causally affects the relative frequency of the outcomes on the other side. It is tempting to think that the causal agent must be either the direction setting or the underlying

[24] Martin Jones (1989) has written a short note in which he claims that Jarrett's proof of the equivalence of a ban on superluminal signaling and locality is fundamentally flawed for a novel reason. Since Jarrett nowhere explicitly mentions the speed of transmission of the signal, Jones concludes that Jarrett has actually (albeit unintentionally) 'proved' that locality is equivalent to a ban on all signaling in a Bell-Bohm experimental scenario, not just a ban on superluminal signaling. This absurd conclusion, Jones claims, discredits Jarrett's argument. I confess to finding Jones' whole line of thought utterly baffling. Since Jarrett specifies that in the experiments he considers the relevant events are spacelike related, and since he does not discuss other types of scenarios, the most natural reading of his argument, as it seems to me, is to completely ignore the cases in which the relevant events are timelike related, as they have no bearing on the issues at hand. This amounts to reading Jarrett's quantifications over apparatus states, outcome events, *etc.*, as implicitly restricted to the particular set of them such that the relevant events are spacelike related, which Jarrett certainly implies should be done (*Cf.*, *e.g.*, Jarrett (1984), p. 587, n. 11). Jones's criticism would thus seem to be answered. If signaling were possible in the cases where the apparatus settings and the experimental outcomes are timelike or null related, moreover, this would still be a novel and striking phenomenon, one demanding investigation, since it is well known that quantum mechanics as presently understood does in fact prohibit signaling at any speed in Bell-Bohm type experiments (*Cf.*, *e.g.*, Eberhard (1978), Ghirardi, Rimini, and Weber (1980), and Shimony (1993a)). Such signaling would then be a contravention of quantum mechanics, however, but not of special relativity, a situation most philosophers would likely revel in.

apparatus state, but this is not in fact necessitated by the argument, for there are surely many more differences between not-measuring and measuring than what is explicitly encoded in these two parameters alone. The other extraneous factors that are associated with measuring but not with not-measuring, moreover, will be (if they are truly differences between measuring and not-measuring, and not merely artifacts of this or that particular instance of measurement) exactly correlated with those states of the measuring device that represent measurements being made, and, *mutatis mutandis*, the same for those extraneous factors solely associated with not-measuring. Direction setting and apparatus state, for various reasons, happen to be the two parameters we think most likely to be causally implicated in any such signaling, but nothing in Jarrett's argument demands that these be the only possible agents of causal relations, even though the explicit conditionalization over them might suggest this. Indeed, nothing in the argument even mentions causality, or even presupposes that anything is acting causally on anything else. No mention is made of possible causal paths, direct and indirect effects, *etc.*, nor is any needed. All we have are dependable correlations between the performance of an experiment on one side and the relative frequency of experimental outcomes on the other that can be exploited to transmit information, and it is precisely this that is supposed to license the *inference* that causality is present. Note, moreover, that not all the things one may want to implicate in the causal chain of the signal even need be controllable; all that is demanded is a certain amount of knowledge about them.

That the signal is associated with either measuring or not is a crucial difference between Jarrett's original argument and all recapitulations of it I have seen in print, which all assume that the signaler chooses between one of two different measurements, not between measuring and not-measuring. In the experiments where the choice is between two different spin measurements, if we take signaling as a sufficient criterion for causality, then we are forced to assume that it is either the direction setting or the apparatus state that is responsible for the signaling, since these are the only factors that reliably differ in different experiments. By choosing between one of two different measurements, one covertly assumes that only one of the two explicitly mentioned parameters could be causally implicated in the signaling, which is not only an unnecessary and obfuscating introduction of causal concerns into the argument, but also an illegitimate assumption to make at such an early stage in the game, when we are concerned with the presence or absence of signaling *simpliciter*, which itself may be used as a mark that causality is present (but need not be). Of course, if one detected locality violations by performing experiments wherein R either performs or does not perform a measurement, one could very well go on to test by performing various types of controlled experiments whether it was the direction-setting, some other manipulable aspect of the apparatus state, or something else correlated with the measurement process, that was specifically correlated with the aberrant relative frequency of outcomes on the other side. This, however, is beside the point of Jarrett's argument. I suspect Jarrett himself did not catch the subtle difference between the different versions of the argument, at least in part because he so scrupulously kept his original paper free of causal considerations.[25]

[25] *Cf.* Jarrett (1984, p. 587, n. 18).

4 Jones and Clifton's Argument Part 2: The Delicacy of Causal Ascription

It is time to return to Jones and Clifton's arguments against Jarrett and examine the treatment of causality in the model they present as a counter-example to their reading of Jarrett. Jones and Clifton declare, "Where there is signaling, it is hard to avoid the conclusion that there is causation."[26] This is all right as a first pass, but they then make a very puzzling claim: "Although locality fails in the class of theories we are considering, this is no reason to suppose that it is a direct causal link between the β's and the A's which makes the signaling possible; indeed, as we have claimed, the fact that constrained locality holds rules out such a link. . . . Thus there is nothing intrinsically non-causal about failures of COMP."[27] In their example, the signal is sent when one experimenter determines, by measuring the relative frequency of A, how the other experimenter manipulated β. It is thus relatively unproblematic to infer that something correlated with the manipulation of β is causally connected to changes in the relative frequency of A, in so far as the step from signaling to causality is unproblematic. After this first step, however, the thread of Jones and Clifton's argument quickly becomes tangled.

It is not clear first of all how 'causality', a hopelessly vague and contested concept in the most straightforward of cases, is to be understood when it appears in such unknown and forbidding territory as that between spacelike separated quantum mechanical events, and Jones and Clifton say nothing explicit about what they mean by it. Inspection of their arguments, however, does yield a sketch of what they intend. On the one hand, they need the notion of causality they employ to be rich enough to include such aspects as distinct possible paths of causality between the same two events, and a definite direction associated with each causal path, a 'causal arrow' so to speak. That Jones and Clifton require such a conception stems from their distinguishing an 'indirect' from a 'direct' effect of β on A. Presumably they mean by this something like the following. Since the correlations between β and A are screened off by B, we are justified in asserting that manipulations of β affect A, in so far as signaling is concerned, only 'via' B: manipulations of β are efficacious only in changing the relative frequency of B, and only changes in the relative frequency of B are efficacious on the relative frequency of A. Jones and Clifton thus seem to envisage causal efficacy as a fluid produced by certain manipulations of β, sloshing thence only to B and only then on to A (see figure 1 below). I shall refer to any notion of causality rich enough explicitly to sustain such ideas as differing paths and directions of causality as 'fluid causality'.

Merely giving a name to this concept will not get us very far, though, for our original question remains: what precisely do Jones and Clifton mean when they deny that β has "direct influence" on A here? In particular, why are we justified in concluding that, merely because constrained locality is satisfied, there is no direct causal connection between β and A (at least for the experimental values that satisfy ACTION), and that the collateral correlations that B exhibits with β and A are neither spurious nor simply 'outside the causal path', signaling detritus as it were? The only

[26] Jones and Clifton (1993, p. 14).

[27] *Ad loc.*

thing like an argument they give for the claim that there is no "direct influence" of β on A for the parameter-values in question is that it is precluded by constrained locality. It is thus vital to note that constrained locality is a wholly statistical criterion, and as such only conditions possible correlations between events. On the other hand, then, Jones and Clifton offer only purely statistical considerations when they propose conditions on the possible causal relations between manipulations of β and changes in the relative frequency of A. I shall refer to any notion of causality that employs only statistical considerations to characterize and condition possible causal relations between events as 'statistical causality'.

We have, now, two different conceptions of causality. One, fluid causality, is characterized by the richness of the ideas of causal relation it sustains. The other, statistical causality, is characterized by the types of criteria, *viz.* purely statistical, one may use to define and distinguish causal relations. One must ask whether the conclusions of Jones and Clifton involving the first notion can be supported by evidence involving only the second notion. We can give an immediate and unqualified answer to this question. Correlations alone cannot tell us much about the possible paths of a fluid, and thus the conclusion that β does not 'directly' act on A is unjustified, if one considers, as Jones and Clifton do, only the correlations predicted by Jones and Clifton's model. For example, let us imagine a causal influence of β that operates 'directly' on both A and B in the following way: when β is manipulated in the proper way, a 'wavefront' of causal efficacy propagates out to affect both A and B. This wavefront is created stochastically, *i.e.* one cannot predict with certainty just which of all possible wavefronts will actually propagate when β is manipulated, but every possible wavefront acts on both A and B deterministically in such a way that there is only a non-zero probability that manipulations of β alter the probability of A when those manipulations definitely alter the probability of B. Measurement determinism fails in this scenario, since the values of the variables a, b, α, β and λ do not suffice to determine the relative frequency of either A or B, and so the door is open for completeness to be violated. We have stipulated that the relative frequency of A is unaffected by manipulations of β so long as the relative frequency of B remains the same, and so our model satisfies constrained locality. Finally, by hypothesis there are values of β that alter the relative frequency of B, and so measurement contextualism is satisfied. Thus this mechanism can exactly replicate the statistics Jones and Clifton employ in making their argument.

[The wine example]

If we could conceive of such a possible mechanism, and we were faced only with the correlations given by Jones and Clifton in their model, would not the simplest explanation be that the signaling was effected by this 'direct' influence of β on A? How could we even experimentally distinguish in this case a 'direct' from an 'indirect' influence of β on A?[28]

[28]The obvious examples of experiments that come to mind to distinguish the two will not work. For instance, one may think that blocking the R particles at the source, or deflecting them, or even just moving the R experimental apparatus out of the way, while allowing the L particles to pass unmolested to the L experimental apparatus, and still having the R experimenter manipulate β or not, would suffice to test between the two different suggested paths of causality. There is an implicit assumption in such an experiment, however, that the states of the two particles are separable in the strong sense that doing something different to the R particle will leave the L particle in the same state as it was in the original set of experiments. Whatever is done to alter the 'state' of the R particles, though,

It appears that the statistical resources Jones and Clifton have mustered so far are inadequate for supporting the notion of distinct causal paths between the same events. It will also be instructive to examine how they fare with respect to the second aspect they hope their causal connection to manifest, that of directionality. Instead of asking now why we cannot infer that β directly influences A, I shall ask why we cannot infer that A influences B, or even why we cannot infer that A directly influences β. Since Jones and Clifton's diagram of the causal nexus instantiated by their proposed experimental arrangement (reproduced below as figure 1) indicates an asymmetry in the causal relationship, in that the arrow of causality moves from β to B to A, and not vice-versa, we must ask why they are justifed in assuming it so. One may want to ask what difference it makes which way the arrow of causality points, so long as the completeness violation is shown to be causal in character; but a different direction of causality would undermine a crucial step in Jones and Clifton's own argument, *viz.* the justification for assuming the presence of causality in the first place.

The notion of signaling crucially involves the thought that the message is sent at the instigation *of the signaler*, irrespective of what we think about what 'really happens' on the physical micro-level and where we come down on free-will versus determinism.[29] Appeal to signaling as a marker for causality *assumes* that β influences A. Appeal to signaling cannot tell us anything about possible causal routes from A to β, and it certainly cannot tell us anything about causal paths between β and B, or between A and B, since neither the relation of β to B nor the relation of A to B has anything to do with the presence of signaling in this case. So long, then, as signaling is taken as the criterion of superluminal causation, it does not make much sense to have the outcome on the left cause the experimenter's choice on the right. Otherwise any strong spacelike correlations, such as those in the classic EPR experiment, should by themselves suffice to demonstrate the presence of superluminal causality, and we would not need to go mucking about with turbid notions like 'signaling' in the first place. But now Jones and Clifton cannot even make their argument, for constrained locality, the only criterion they have allowed themselves to ensure not only that causality is present, but also that the path of causality is as they require, *viz.* going from β to B and only thence to A, tells us nothing about the direction of possible causal paths between the events.

The most such arguments as Jones and Clifton's can do is to show that *imagining* completeness violations to be causal in nature, by describing how one might interpret a set of correlations to make them so, is neither internally inconsistent nor necessarily incoherent. But presumably no one ever thought that was the case anyway.

Where then does all this leave Jones and Clifton's final argument against Jarrett, that since both violations of locality and of completeness require a controllability assumption to be implicated in signaling, neither can be privileged over the other with respect to their relation to superluminal signaling? We see now that this consideration is a red herring. The only feature that matters in any

cannot simply be assumed to have no effect on the L particles, since it is possible paths of causal propagation we are investigating. 'Paths of causality' as envisaged by Jones and Clifton are simply not amenable to easy experimental testing in the absence of background knowledge.

[29] Note that I mean 'at the instigation of the signaler' most loosely. I do not, for instance, think that every signaler must intentionally, or even consciously, encode the signal.

situation where one would consider the possibility of signaling is whether or not reliable correlations exist between the performing of experiments on one side and the outcomes of experiments on the other. But this feature is precisely captured by the formulation of locality. Herein lies the connection between locality and the possibility of signaling that is at the heart of Jarrett's argument. Questions of causality, if they arise at all, ought only do so after so much is settled.

AlthoughJones and Clifton's general broadside on Jarrett misses its target, I do think one particular claim of theirs is right. In light of their proposed model, Jones and Clifton criticize the penchant of many recent writers (including Jarrett himself) to argue, largely on the basis of the alleged inability of completeness violations to effect superluminal signaling, that such violations are inherently 'non-causal'. To recount a notable sample of some (by and large) tentative 'metaphysical' conclusions drawn from the alleged noncausality of completeness violations: a new type of physical interaction, "passion-at-a-distance";[30] some sort of physical holism;[31] reworking the notions of 'event' and 'process' in terms of actuality and potentiality;[32] relativized identity of physical individuals.[33] Jones and Clifton are absolutely in the right to criticize this trend toward a characterization of completeness violations as noncausal, though not for reasons I endorse. While I have argued that Jarrett was essentially correct in his conclusions concerning the respective relations of completeness and locality to the possibility of superluminal signaling, I must also conclude that those (including Jarrett himself) who would try to draw 'metaphysical' conclusions, even experimental ones, concerning the nature of completeness violations can obtain no warrant at all for doing so from Jarrett-type arguments concerning the possibility of superluminal signaling. So long as we know nothing more about the completeness violations, we can say nothing more about their character that is not pure speculation. In the end, the considerations I marshall suggest that arguing about possible causal connections between spacelike related experimental outcomes, when the criterion for causality is signaling, is misguided. Additional arguments are needed to say *anything whatsoever* about the nature of completeness violations, that they are causal or noncausal, and this does indeed portend grimly for much of recent 'experimental metaphysics'.

5 Conclusion

I want to conclude with a few remarks of a more general nature. I have concentrated on Bell's Theorem for various reasons, no least of them that the issues I wanted to discuss are laid out with particular clarity in some of the disputes surrounding it, but my considerations have a much broader import. My arguments in this paper have turned on the idea that 'causality' is not a univocal notion, and that its particular meaning in any given situation is largely determined both by the context in which it is put to use and by the collateral notions used in that context to condition it and flesh it out, as the notion of 'signaling' functioned in this paper. There are several lessons I want to draw

[30] Shimony (1993a); Redhead (1986), Redhead (1989).
[31] Howard (1985), Howard (1989) Teller (1986), Teller (1989).
[32] Shimony (1993b).
[33] Howard (1989).

from this point.

First among them is the need to guard against putting the cart before the horse when causality is discussed. There is too much a tendency to begin arguing immediately about the presence or lack of causality, about the possible causal nature of certain correlations and interactions, before any attempt is made to clarify what precisely one means by 'causality' in the context, what is at stake in declaring that causality is or is not present, how one could in practice determine whether causal relations are present or not, and, if so, *which* ones are.

Second, on a more fundamental level, it behooves us to ask what reason we have to think that the (generally unclear) notion of causality may be profitably, or even coherently, applied to any novel, bizarre situation one can dream up. There are, in general, many clear and profitable questions that can be asked of a situation before one goes muddying the waters with causality. This is one of the lessons to be learned straight from Jarrett: the question, 'If x obtains, will y obtain?' is always clearer than 'Does x cause y?' when not much is known.

Third, once one has determined that causal issues are appropriate to introduce into an argument, one must ensure that one has given oneself a rich enough structure to support all the work one hopes to extract from the causal notions one is employing. There are (at least) two general, distinct notions of causality employed in many discussion of science and philosophy today, what I referred to above as 'fluid causality' and 'statistical causality'. I also believe that Jones and Clifton, far from being alone in this regard, exemplify a pandemic tendency to rest the ideas of the one illegitimately upon the scaffolding of the other, which cannot support it, all the while unaware that there is any internal tension. Statistics by itself cannot support any notion of causality rich enough for use with ideas such as signaling, which are too anthropocentric to find their home in a purely formal analysis of an event or process. I do think that statistical analysis of data to investigate possible causal relations can be a valuable tool, but only so long as it is not asked to do what it cannot by itself. And even when this is recognized, and good-faith efforts are made to harmonize the requirements and dictates of these two notions of causality, it must be kept in mind that there is no guarantee of success. The common, intuitive ideas of causal paths and directions of causality arose long before statistics was codified as a rigorous science, in many and variegated walks of quotidian life, and we have no *prima facie* reason to believe that these archaic, nebulous notions can be cleanly or even coherently recast in the stern and abstract stuff of modern physics and mathematics.[34]

References

Aspect, A., J. Dalibard, and G. Roger (1982). Experimental tests of bell's inequalities using time-varying analyzers. *Physics Review Letters 49*, 1804–1807.

[34]This paper owes a great debt to the conversation and guidance of David Malament and Howard Stein—most, if not all, of what is good in it can be attributed to their influence. Thanks are also due to Rob Clifton, Arthur Fine and Martin Jones for extensive comments on an earlier draft of this paper. It goes without saying that all deformity and fallacy here is of my own device. This work was supported in part by a National Science Foundation Fellowship.

Bell, J. (1987). Quantum mechanics for cosmologists. In *Speakable and Unspeakable in Quantum Mechanics*, Chapter 15, pp. 117–138. Cambridge: Cambridge University Press.

Butterfield, J. (1992). Bell's theorem: What it takes. *British Journal for the Philosophy of Science 43*, 41–83.

Cushing, J. and E. McMullin (Eds.) (1989). *Philosophical Consequences of Quantum Theory: Reflections on Bell's Theorem*. Notre Dame, IN: University of Notre Dame Press.

Eberhard, P. (1978). Bell's theorem and the different concepts of locality. *Il Nuovo Cimento 46B*(2), 392–419.

Friedman, M. (1983). *Foundations of Space-Time Theories*. Princeton: Princeton University Press.

Ghirardi, G., A. Rimini, and T. Weber (1980). A general argument against superluminal transmission through the quantum mechanical measurement process. *Lett. Nuovo Cimiento 27*, 293–298.

Hellman, G. (1982a). Einstein and bell: Strengthening the case for microphysical randomness. *Synthese 53*, 445–460.

Hellman, G. (1982b). Stochastic einstein-locality and the bell theorems. *Synthese 53*, 461–504.

Hellman, G. (1983). Stochastic locality and the bell theorem. In *PSA 1982*, Volume 2, pp. 601–615. Philosophy of Science Association.

Howard, D. (1985). Einstein on locality and separability. *Studies in the History and Philosophy of Science 16*, 171–201.

Howard, D. (1989). Holism, separability and the metaphysical implications of the bell experiments. See Cushing and McMullin (1989), pp. 224–253.

Jarrett, J. (1984). On the physical significance of the locality conditions in the bell arguments. *Noũs 18*, 569–589.

Jarrett, J. (1986). Does bell's theorem apply to theories that admit time-dependent states? *Annals of the New York Academy of Sciences 480*, 428–437.

Jarrett, J. (1989). Bell's theorem: A guide to the implications. See Cushing and McMullin (1989), pp. 60–79.

Jones, M. (1989). What locality isn't: A response to jarrett. In M. Kafatos (Ed.), *Bell's Theorem, Quantum Theory and Conceptions of the Universe*, pp. 77–79. Kluwer Academic Publishing.

Jones, M. and R. Clifton (1993). Against experimental metaphysics. In T. U. P. French and H. Wettstein (Eds.), *The Philosophy of Science*, Volume XVIII of *Midwest Studies in Philosophy*, pp. 295–316. Notre Dame, IN: University of Notre Dame Press.

Kronz, F. (1990). Jarrett completeness and superluminal signals. In *PSA 1990*, Volume 1, pp. 227–239. Philosophy of Science Association.

Maudlin, T. (1994). *Quantum Non-Locality and Relativity: Metaphysical Intimations of Modern Physics*, Volume 13 of *Aristotelian Society Series*. Oxford: Basil Blackwell.

Redhead, M. (1986). Relativity and quantum mechanics—conflict or peaceful coexistence? *Annals of the New York Academy of Sciences 480*, 14–20.

Redhead, M. (1989). Nonfactorizability, stochastic causality, and passion-at-a-distance. See Cushing and McMullin (1989), pp. 145–153.

Shimony, A. (1993a). Controllable and uncontrollable non-locality. See Shimony (1993c), pp. 130–139.

Shimony, A. (1993b). Events and processes in the quantum world. See Shimony (1993c), pp. 140–162.

Shimony, A. (1993c). *Search for a Naturalistic World View: Natural Science and Metaphysics*, Volume 2. Cambridge: Cambridge University Press.

Skyrms, B. (1982). Counterfactual definiteness and local causation. *Philosophy of Science 49*, 43–50.

Teller, P. (1986). Relational holism and quantum mechanics. *British Journal for the Philosophy of Science 37*, 71–81.

Teller, P. (1989). Relativity, relational holism and the bell inequalities. See Cushing and McMullin (1989), pp. 208–223.