Jarrett's Locality Condition and Causal Paradox

Allen Stairs

University of Maryland

In this paper, I want to present a family of results that may seem to add up to a new proof of the impossibility of hidden variables. In fact, I very much doubt that that's really what really emerges, but I think the results are nonetheless interesting because they help to sharpen the discussion of Jon Jarrett's very useful decompostion theorem, in particular, of the condition he calls *locality*. Jarrett (1984) and Ballentine and Jarrett (1987) have suggested that the so-called condition of locality is the one that provides the conceptual link between hidden variable theories and relativity: if locality is violated, so is relativity. On the other hand, a theory may violate the condition Jarrett calls *completeness* without running afoul of relativity. Now I agree with Jarrett and Ballentine about completeness, but I strongly suspect that we have quite a way to go before we really understand what would be involved in a violation of locality.

Let us begin by stating Jarrett's conditions and his result. My notation will be a modification of Ballentine and Jarrett's (1987). We consider a typical Bell-type experiment with two wings, L and R, and with two possible outcomes + and - for each local measurement. Outcomes in the wings L and R will be denoted by x_L and x_R. If we need to say that the outcome in the L wing was +, we may write $x^L = +$, or just $x_L{}^+$. S_L and S_R represent the settings of the measuring devices in wings L and R. If more detail is needed, superscripts may be added, as in $S_L{}^A$. The symbols $S_L{}^O$, $S_R{}^O$ represent the cases in which no measurements are performed in the respective wings. Finally, λ, with or without additional decoration, represents a maximally informative state of whatever theory is under consideration.

Jarrett's conditions are stated in full generality: they characterize both determininstic and indeterministic theories. The condition Jarrett calls *locality* is given by

1a $p(x_L|S_L,S_R{}^O,\lambda) = p(x_L|S_L,S_R,\lambda)$

1b $p(x_R|S_L{}^O,S_R,\lambda) = p(x_R|S_L,S_R,\lambda)$

1a implies

1a' $p(x_L|S_L,S_R,\lambda) = p(x_L|S_L,S_R',\lambda) = p(x_L|S_L,S_R{}^O,\lambda)$

with a similar consequence for 1b. What this condition tells us is that the outcome in one

PSA 1988, Volume 1, pp. 318-325.

wing is independent of the setting in the other. Following Abner Shimony, we will refer to Jarrett locality by the more neutral name *Parameter Independence* (PI).

The condition Jarrett calls *completeness* is

2 $$p(x_L,x_R|S_L,S_R,\lambda) = p(x_L|S_L,S_R,\lambda) * p(x_R|S_L,S_R,\lambda).$$

This condition tells us that the *outcome* in one wing is independent of the *outcome* in the other. Again following Shimony's lead, we will refer to the "completeness" condition as *Outcome Independence* (OI).

Combining PI and OI yields

3 $$p(x_L,x_R|S_L,S_R,\lambda) = p(x_L|S_L,S_R{}^O,\lambda) * p(x_R|S_L{}^O,S_R,\lambda)$$

which is a way of stating a common pre-Jarrett "locality" condition. Since 3 allows the derivation of Bell-type inequalities, satisfaction of PI and OI entails satisfaction of such inequalities. This is Jarrett's "decomposition theorem".

Thus far, Jarrett's distinction between PI and OI has not been put to much systematic use in studying the earlier literature on foundations of quantum mechanics. In fact, I believe that it will prove to be a very helpful analytical tool. For one thing, there is a family of proposals often referred to as "contextual" that seem clearly to permit or even require violations of PI. For further discussion see, among others, Shimony (1984), Heywood and Redhead (1983), and van Fraassen (1973). To put the matter a little over-simply, we can think of contextual theories as assigning values to observables in a way that breaks up the connections on which the Gleason-Kochen-and-Specker argument depends. By way of illustration, think of quantities **XxY** corresponding to operators XxY as taking pairs of numbers as values. (I use "x" for the tensor product. Left and right in the expressions corresponds to left and right in the experiment.) Thus, in an EPR-type situation, the "hidden" state λ might determine that **AxC** takes the value-pair **<a,c>**, **AxD** takes value-pair **<a',d>**, and that **a=a'**. If the way to measure **AxC** on separated systems is with arrangement $(S_L{}^A,S_R{}^C)$ and the way to measure **AxD** is with arrangement $(S_L{}^A,S_R{}^D)$, then we could have

$$p(x_L|S_L{}^A,S_R{}^C,\lambda) = 1, \quad p(x_L|S_L{}^A,S_R{}^D,\lambda) = 0,$$

which violates 1a' above.

To be sure, contextual theories are strange, and their metaphysical peculiarities are not well-understood. However, it may be worth quoting Heywood and Redhead (some of whose ideas influenced my thinking on these matters) on the topic of theories that violate what they call "Ontological Locality".

> A sort of Holism is involved; physical magnitudes maximal in the product space are somehow prior to physical magnitudes maximal locally in the factor spaces. (p. 487)

The detailed relationship between PI and various proposals in the literature is a topic for another paper. Having had a glimpse of this, we turn to a different question: just why are violations of PI supposed to be more disturbing than violations of OI?

The reason most often cited for worrying less about OI than PI is that violations of PI amount to violations of special relativity. When PI is violated, it is apparently possible "in principle" to construct a "Bell telephone"–to send messages between space-like separated regions–and this is held to conflict with relativity. On the other hand, mere violations of OI, which are what quantum theory displays, do not permit signalling. If all this is correct, we can say that, peculiar though quantum theory may be, it needn't be seen as violating special relativity. However, a theory that violated PI *would* conflict with relativity.

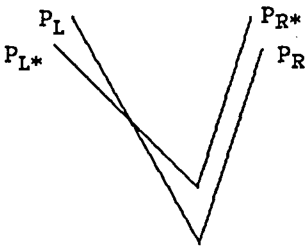
Prima facie this claim might seem too strong. For one thing, if PI is violated because an algebraic (or ontological) contextual theory is true, then the unseemly connections between regions L and R result not from the transport of mass-energy, but from a peculiar form of holism with respect to the properties of quantum systems. Since this is not even the sort of thing that relativity contemplates, it is by no means clear that it amounts to a violation of relativity. Perhaps it is best viewed as a strange sort of metaphysical "back door" that allows us to get around relativity by denying that the connections between L and R are space-time connections. However, there is a wholesale approach to violations of PI, implicit in the following remark by Ballentine and Jarrett.

> Even more serious is the possibility of using a tachyon relay to send a message to one's own past. This leads to very strange consequences: one could arrange to kill one's mother before one was born, and thus have a situation in which a tachyon message is sent if and only if it is not sent. For reasons such as these, it is agreed that the possibility of travelling or signalling at superluminal speeds is strictly incompatible with special relativity. (1987, p. 697)

Now this can not be quite right. A paradox is a paradox, and what forbids paradoxes is logic, not relativity. If the only way to resolve such paradoxes is to insist that a cause must unambiguously precede its effect, then superluminal signals are ruled out for reasons much more general than relativity. Furthermore, even if relativity is not the real issue, it may turn out that paradoxes can be generated *whether or not* superluminal connections are under our control. So the above passage suggests a general strategy: see if violations of PI generate causal paradoxes. If they do, this may provide a wholesale argument for insisting that any acceptable theory must incorporate PI, *not* because relativity *per se* demands it, but in order to avoid paradox. That is the possibility we will now explore.

Now it is not just obvious that violations of PI *in the context of QM* lead to causal paradoxes. Consider measurements on a pair of particles. On a contextual theory, my choice of parameter here may influence the result you observe there. Further, there is a *correlation* between the result you observe there and the result I observe here, but there is no obvious reason for a theory that merely aims to reproduce the predictions of quantum mechanics to posit any sort of feedback from your result to my setting. However, a twist in the experimental geometry gives us what we need.

Instead of considering one pair of particles, consider two. And allow an earlier local outcome to determine a later local setting. Now add one further assumption: we can arrange things so that we first emit a pair $P_L + P_R$, and then a pair $P_{L*} + P_{R*}$, but where P_L is slower than P_{L*}, and so enters the L-wing of the experiment *later* than P_{L*}, even though it was emitted earlier. A diagram may help.



https://doi.org/10.1086/psaprocbienmeetp.1988.1.192999 Published online by Cambridge University Press

Of course, a two-dimensional diagram is artificial. What we would really need is a more complicated geometry so that the particles P_L and P_{L*} never crossed paths. However, we will ignore this complication. Let S_L be the setting of the device that measures particle P_L, and S_{L*} be the setting of the device that makes the earlier measurement of P_{L*}. Outcomes will be similarly tagged. We will use similar labels in region **R**. We will suppose that S_{L*} is simply given, and corresponds to an **A** setting. Similarly, we suppose that S_R is simply given, and corresponds to a **C** setting. We suppose, however, that the setting of the device performing the later measurement is causally determined by the outcome of the earlier measurement. Specifically, we will suppose that the following conditionals are true.

1) $(x_R = +) \rightarrow (S_{R*} = C)$

2) $(x_R = -) \rightarrow (S_{R*} = D)$

3) $(x_{L*} = +) \rightarrow (S_L = A)$

4) $(x_{L*} = -) \rightarrow (S_L = B)$

We now need only one more assumption to generate a causal paradox: the pairs $P_L + P_R$ and $P_{L*} + P_{R*}$ are in states λ, λ^* with the following partial specifications. (Values are in angle brackets, a, b, c and d unspecified.)

λ: $(S_L{}^A, S_R{}^C) = \langle a,- \rangle$ $(S_L{}^B, S_R{}^C) = \langle b,+ \rangle$

λ^*: $(S_{L*}{}^A, S_{R*}{}^C) = \langle +,c \rangle$ $(S_{L*}{}^A, S_{R*}{}^D) = \langle -,d \rangle$

(where a, b, c and d each can take either value + or -.) This assumption leads to the following four conditionals:

5) $(S_L = A) \rightarrow (x_R = -)$

6) $(S_L = B) \rightarrow (x_R = +)$

7) $(S_{R*} = C) \rightarrow (x_{L*} = +)$

8) $(S_{R*} = D) \rightarrow (x_{L*} = -)$

We now have the following two arguments: First,

1) $(x_R = +) \rightarrow (S_{R*} = C)$

7) $(S_{R*} = C) \rightarrow (x_{L*} = +)$

3) $(x_{L*} = +) \rightarrow (S_L = A)$

5) $(S_L = A) \rightarrow (x_R = -)$

Therefore,

9) $(x_R = +) \rightarrow (x_R = -)$.

Second,

2) $(x_R = -) \rightarrow (S_{R*} = D)$

8) $(S_{R*} = D) \rightarrow (x_{L*} = -)$

https://doi.org/10.1086/psaprocbienmeetp.1988.1.192999 Published online by Cambridge University Press

4) $(x_{L^*} = -) \rightarrow (S_L = B)$

6) $(S_L = B) \rightarrow (x_R = +)$

Therefore,

10) $(x_R = -) \rightarrow (x_R = +)$

and so

11) $(x_R = +) \leftrightarrow (x_R = -)$,

which gives us a paradox on the assumption that + and - are mutually exclusive, jointly exhaustive outcomes.

How "bad" are things here? Bad enough. Suppose we find a reliable way of producing pairs that exhibit the experimental geometry we have described. And suppose we find that in both wings, we can reliably use earlier results to determine later settings. And suppose we knew that pairs in states λ, λ^* are produced in succession with some frequency. Then we have the following very peculiar situation. Imagine that whether or not the earlier result will control the later setting depends on whether a certain switch is thrown. I, in region L, have agreed in advance that my switch will be on. So it's all up to you. *If* you turn your switch on, no λ - λ^* sequence of pairs will be produced. If you don't, then such a sequence might be produced. But suppose the particles are already in flight. You can ensure that they are not a λ - λ^* sequence with a mere flick of a switch. This seems to be almost a "logical" holism–a dependence among separated events that rests not just on physics but on logic itself. And this may seem to be too much to swallow.

The puzzle remains if we relax our assumptions–if we no longer insist on determinism. For now, I will simply present the results of the analysis. Details are in Stairs (1988).

There are three ways in which we might relax the assumption of determinism. First, we might suppose that the connection between a setting "here" and an outcome "there" is less than strict. In the example we have described, this would amount to assuming that, e.g., if $S_{R^*} = C$ then the probability that $x_{L^*} = +$ is greater than .5, but not necessarily one. It is not hard to show that if we retain the assumption that the earlier local outcomes determine the later local settings in accord with conditionals 1) - 4), even this very weak violation of PI is ruled out. A similar argument will dispose of the second case: suppose that settings "here" strictly determine outcomes "there" in accordance with 5) - 8), but that the connections between local outcomes and later local settings are not strict. (Again, we replace probabilities of one with inequalities of the form p(__) > .5) Finally, we might assume that none of the connections are strict, but all satisfy certain inequalities. (Specifically p(__) = 1)becomes p(__) > .5) and p(__) = 0) becomes p(__) < .5).) Then it can be shown that *some* violations of PI are consistent, but we can use a version of the Clauser-Horne inequality to place significant constraints on the extent of the violations.

Although the statistical cases aren't as striking as the deterministic case, they are nonetheless puzzling. A crude (and contentious) way of putting the point would be to say that by a flick of a switch, I can influence the probabilistic tendencies of particles already in flight.

There are various details of the argument I have offered and the assumptions I have made that should be discussed in a full-scale presentation of the results, but for now I leave them aside and plunge into the question "What do these results tell us?" In particular, should we conclude, as Ballentine and Jarrett would have us do, that we should reject PI? I am not convinced. To begin with, the most that has been shown is that violations of PI *of a certain strength*, coupled with outcome-setting connections *of a certain strength*, are ruled out. However, this may seem a weak reply. After all, if this is the most that can be said for violations of PI, the best guess may be that they really can't occur at all. Nonetheless, I am not content. One deep source of suspicion for me is that if this is the correct response, then we have drawn a strong substantive conclusion from an argument that is curiously close to *a priori*. It is of a piece with the general rejection of simultaneous and backward causation on the grounds that they lead to apparent paradoxes. But there is a coherent response to such arguments: it is not that temporally anomalous causal chains are ruled out *simpliciter*; it is rather that there are logical constraints on the sorts of causal chains that can exist.

There are two aspects of this point that I would like to explore. The first appeals to general considerations about probability. I suggest that in the stochastic cases, it is particularly unclear that we have grounds for ruling out violations of PI. The *experimental* consequence of our results is simply that no ensemble of experiments can display certain results with certain frequencies. But it is quite unclear what this tells us about individual cases, and the answer may depend on your interpretation of probability. Consider the case in which local settings don't strictly determine distant outcomes but merely "make" them more or less probable, but in which the earlier local outcomes strictly determine the later local settings. This means that certain conditional probabilties--conditional probability versions of **5 - 8**, as it were--are strictly between **0** and **1**. Call these the Group I probabilities. It also means that certain other conditional probabilities--conditional probability versions of **1 - 4**--are **0-1**. Call these the Group II probabilities. Suppose you think the Group I probabilities represent something like propensities or dispositions that inhere in the particles themselves. Then you will say: if a pair with one sort of propensity (or pair of propensities) is emitted, it will never be followed by a pair with another particular sort of propensity. But there is much to be challenged here, not least the very idea of propensities itself.

Even the friend of propensites may well object to the description. The idea that propensities are *local* properties of individual particles sits ill with the rejection of PI, not to mention with certain standard ways of talking about propensities. Why not say that the propensities are properties of the whole experimental arrangement? Granted, this means that we have to view "the experimental arrangement" as spread out in space *and* in time, but that is not a particularly surprising consequence of the rejection of PI. Indeed, it might be argued, if one were to imbibe this point of view, then what looked paradoxical before might come to seem "natural". (If this reminds the reader of some of Arthur Fine's recent reflections, that is no coincidence. But whether Fine would go this far is not so clear.)

There is another side to the point that all we have shown is the *joint impossibility* of certain conditions and not the impossibility of a breakdown of PI. Lewis (1976) has discussed the issues well. More recently, Paul Horwich (1987) has provided a very useful discussion under the heading of "bilking arguments". Here I will focus on Lewis.

Lewis notes that questions about what people can and can not do are at the heart of many worries about time travel and the like. In particular, if we could travel to the past, then it seems that we should be able to *change* the past. But if a time traveller can not change the past, his or her freedom and abilities would seem to be oddly restricted. Lewis argues that this is not so. Suppose that Tim is a time traveller who journeyed to 1921 to kill his Grandfather. To begin with, consistency requires that the adult Tim actually was present in 1921. The mere supposition that he was is not inconsistent, but it may be inconsistent with the facts. Suppose, however, that there really was a time-travelling Tim in '21. We know he did not kill his Grandfather, and so we know that if he tried to,

https://doi.org/10.1086/psaprocbienmeetp.1988.1.192999 Published online by Cambridge University Press

something went wrong. Does this mean that Tim was *unable* to kill Grandfather? Well, he certainly was unable both to have killed and not killed Grandfather, and so there is a restricted sense in which *relative* to the actual facts, Tim could not have killed Grandfather. But relative to the actual facts, I could not, at the moment of writing this sentence, be swimming at the YMCA. Does that mean I could not have gone to the "Y" this morning?

I raise these issues about freedom because they are part of a package that no doubt leads many to be suspicious of the very idea that PI could be violated. Recall our earlier example. Whether or not the connections will hold between earlier outcome and later setting in your wing–and hence, whether the strange loop at the heart of our puzzle exists depends on whether you throw a switch. We noted that if the particles are already in flight when you make the decision–with more precision, if their flight paths intersect your causal past–or even if they are in your absolute elsewhere when you decide, you can apparently ensure with a flick of a switch that they are not $\lambda - \lambda^*$ sequence, which seems to grant you more ability than you really have. But if we turn the case around, you may appear to have less ability than you really do. Suppose that the particles are a $\lambda - \lambda^*$ pair. Then you *will not* throw the switch, and this may suggest that, mysteriously, you can not. Since either of these cases may seem to be too much to swallow, you may argue that the potential violations of PI are the most dubious part of the story, and should therefore be rejected.

I demur. First, even granting the point that there are puzzles here about human ability and freedom, it seems to me to be bad procedure to draw conclusions for physics from our sense of how these puzzles should be resolved. It seems to me that at *best* for the defender of PI, the situation is this: the assumption that PI fails is very far from our usual set of beliefs about how things in the world are connected. Consequently, we are *unsure* how our usual concepts of human freedom and ability, let alone causation, apply in cases in which potential failures of PI are at issue. And this makes the inference that PI cannot fail particularly dubious.

It is also worth noting that if the pairs in the case we have described are a $\lambda - \lambda^*$ sequence, no logical invigilator will wrest your hand from the switch. This may simply be one of the times when you decide not to throw the switch. Or you may, indeed, throw the switch, but this may be one of those rare occasions on which the mechanism fails. Or the particles may really behave as a $\lambda - \lambda^*$ sequence *if measured*, but this may be one of the cases in which, for whatever reason, one or more of the particles proves to be a "drop-out" and doesn't yield a result. (Perhaps it is diverted from its course by a bit of stray matter.) Or the particles may be emitted with the wrong geometry, even though the violations of PI that they potentially incorporate have the structure we have described. What we know is that a rather large constellation of events will not all happen "at once", though various subsets of them may occur. But this seems to me very far from showing that there can be no violations of PI.

Another point: if PI is violated, then it is clear from our discussion that the setting of the instruments is not independent (or, at least, not always independent) of the states of the particles. But this is not the mark of conspiracy. It is part and parcel of the peculiar assumption we are considering.

I conclude, then, that if PI is violated, it is indeed possible to generate some *apparently* paradoxical results, but that it would be incautious to conclude that there can be no violations of PI. On the other hand, positing violations of PI is optional at best, and in fact, not clearly motivated. Taking quantum mechanics at face value, it is a theory that gets its job done by respecting PI and violating OI. This is surprising, as the last twenty years of debate makes clear, but it may be no more than surprising. And while violations of PI might not be any more surprising, they are surely no *less* surprising. However, there is a point that needs stressing. The conventional wisdom has been that there is an important conceptual difference between OI and PI. What I have argued suggests that there are differences, but that they may not be as important. or as well-understood as we thought.

There is, in any case, one aspect of all this that is still puzzling. Suppose one accepts the idea that there *could* be violations of PI, but that the simultaneous truth of conditionals like **1) - 8)** can be ruled out on logical grounds. Then, various soothing remarks of Lewis's notwithstanding, this may leave one with the feeling that somehow logic is getting very close to physics; that the sorts of constraints it imposes have a degree of substantiality rather like actual physical laws. That is what was hinted at earlier by the phrase "logical holism". What stands between the possible and the paradoxical is mere consistency. But as our discussion of ability, freedom and the like may suggest, even if there is no flesh and blood logical invigilator, his ghost seems to be throwing substantial weight around. Perhaps this is mere illusion; perhaps not. But if not, it is a prospect that I find intriguing rather than disturbing.

References

Ballentine, L.E. and Jarrett, Jon P. (1987). "Bell's Theorem: Does Quantum Mechanics Contradict Relativity?", *American Journal of Physics* 55: 696-707.

Heywood, P. and Redhead, M.L.G. (1983). "Nonlocality and the Kochen and Specker Paradox", *Foundations of Physics* 13: 481-498.

Horwich, P. (1987) *Asymmetries in Time* Cambridge: MIT Press.

Jarrett, J.P. (1984). "On the Physical Significance of the Locality Condition in the Bell Argument", *Nous* 18: 569-589.

Lewis, D. (1976). "The Paradoxes of Time Travel", *American Philosophical Quarterly* 13: 145-152.

Shimony, A. (1984). "Contextual Hidden Variables and Bell's Inequalities", *British Journal for the Philosophy of Science* 35: 25-45.

Stairs, A. (1988). "Locality and Causal Paradoxes", (forthcoming).

Van Fraassen, B.C. (1973). "Semantic Analysis of Quantum Logic". In C.A. Hooker (ed.), *Contemporary Research in the Foundations of Quantum Mechanics.* Dordrecht: Reidel, pp. 80-113.

https://doi.org/10.1086/psaprocbienmeetp.1988.1.192999 Published online by Cambridge University Press