PRAMANA
— journal of
physics

Vol. 56, Nos 2 & 3
Feb. & Mar. 2001
pp. 199–209

Backward causation, hidden variables and the meaning of completeness

HUW PRICE
Department of Philosophy, Main Quad, A14, University of Sydney, New South Wales 2006, Australia

Abstract. Bell's theorem requires the assumption that hidden variables are independent of future measurement settings. This independence assumption rests on surprisingly shaky ground. In particular, it is puzzlingly time-asymmetric. The paper begins with a summary of the case for considering hidden variable models which, in abandoning this independence assumption, allow a degree of 'backward causation'. The remainder of the paper clarifies the physical significance of such models, in relation to the issue as to whether quantum mechanics provides a complete description of physical reality.

Keywords. Bell's theorem; backward causation; hidden variables; completeness.

PACS Nos 01.70.+w; 03.65.Bz

1. Bell's theorem and the independence assumption

Let us begin, as in figure 1, with a standard Bell-type experimental arrangement, involving particles postulated to possess 'hidden variables'. Here λ_x and λ_y are the supposed hidden variables of the particles x and y, and Set_x and Set_y are determinations of the settings of the measuring devices M_x and M_y.

The derivation of Bell's inequality relies on an assumption of the following kind:

Independence assumption (IA): The values of the hidden variables λ_x, λ_y are independent of the measurement settings Set_x, Set_y.

This assumption is rarely challenged. Even Bell [4] himself seems to have thought of giving it up as the 'most weird' of the four or so possibilities he could see for QM, in the light of his own famous results [3]. And yet it is not hard to find a *prima facie* reason for putting a question mark against IA (independently of the fact that it is a crucial assumption in what, as it stands, is perhaps the most puzzling single result in quantum theory).

What is this reason? Simply that the assumption is time-asymmetric: the states of the particles x and y between S and the measurements M_x and M_y are *not* assumed to be independent of what happens to them at S. In other words, we do not find it at all problematic that λ_x and λ_y should be correlated with what has happened to the particles x and y in the immediate past. Yet IA amounts to the assumption that there is no such correlation with what happens to them in the immediate future. Where does this time-asymmetry

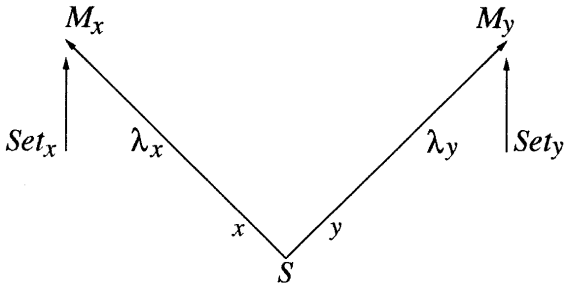

Figure 1. A Bell experiment.

come from? It is not given to us by observation. (The variables λ_x and λ_y are supposed to be 'hidden', after all!) By what right do we assume it so lightly?

Could it perhaps be associated with the macroscopic time-asymmetry of thermodynamics? Certainly, this asymmetry seems to have something to do with *our* 'knowledge asymmetry' — the fact that *we humans* know the past better than the future. However, we are many-particle systems, apt vehicles for embodiment of statistical laws. Not so the individual photon or electron, to which IA is supposed to apply.

A more promising suggestion might seem to be that the time-asymmetry assumed in IA is the familiar asymmetry of causation — the familiar fact that effects never precede their causes. But what sort of fact is this 'familiar' fact? If it is an observational matter, then we do not seem entitled to assume it prevails in regions of the microworld assumed to be unobserved. If on the other hand it is a logical matter, then it does not seem to be what underpins the temporal asymmetry of IA. There seems to be no *logical* barrier to HV models which violate IA. How then does our theorising about QM get to be constrained in this way?

More on causation in a moment. My immediate point is that IA embodies a time-asymmetry which is puzzling, at least *prima facie.* Despite this, models which give up IA have been considered very little (even, as I noted, by the unconventional Bell). Why is this? What makes IA so 'obvious' in the minds of people who care about the interpretation of QM? What makes giving up IA so 'counter intuitive', even by the notoriously relaxed standards of the discipline? I want to discuss what seem to me to be the two most significant objections to giving up IA, and to show that they are surprisingly weak.

1.1 *Initial randomness*

It might be thought that IA follows from the familiar assumption that all initial distributions of the relevant hidden states are equally likely — in other words, that failure of IA would require a very 'special' initial distribution of hidden states. For example, Lebowitz [9] suggests that a no-correlation principle of this kind follows from the 'reasonable minimalist assumption' that the 'initial *microstate* (of the universe) was *typical* with respect to some (at least vaguely defined) weight or measure on the different microstates compatible with the initial macrostate'.

However, this argument assumes that the failure of IA would be a *fact-like* matter. If it were a law-like matter, as it might well be in a HV model exploiting this loophole in Bell's theorem, then no special choice of initial distribution would be needed. The required correlations would obtain in all possible initial distributions, because 'possible' simply means 'allowed by the laws'. (By way of comparison, do we need a special choice of initial conditions to ensure that conservation of momentum holds in Newtonian mechanics? Obviously not.)

One corollary of this point is that we should not confuse the time-asymmetry of IA with the macroscopic time-asymmetry associated with thermodynamics, at least so long as the latter is held to be a fact-like matter, dependent on boundary conditions. IA should not be confused with Boltzmann's *stoßzahlansatz*, in other words. The latter is a fact-like matter, on most accounts, while the former might well be law-like.

1.2 *Fatalism and backward causation*

Another common objection to relinquishing IA is that it leads to fatalism. The thought is that if the incoming photon (say) already 'knows' how we are going to orient a polarizer which it is to encounter in the future, then we do not really have a choice in the matter. (This factor seems to have been influential in John Bell's thinking about these matters; see Price [11].)

This is a very puzzling objection, in several ways. For one thing, many people, including presumably many physicists, regard themselves as fatalists already, for a variety of reasons. From such a standpoint it could hardly be an objection to relinquishing IA that it confirms what they already believe. And even for non-fatalists, should not they acknowledge that physics might show that they are wrong, that fatalism holds after all?

But does relinquishing IA really imply fatalism in the first place? Why not say instead that in choosing the orientation of the future polarizer, we control the earlier state of the photon, rather than the other way around? It might be objected that would imply backward causation: our choice would be affecting the past. Indeed, but what is so bad about that?

At this point, the objection to relinquishing IA is no longer that leads to fatalism, but that it leads to backward causation. The usual objection to backward causation is that it would make possible 'paradoxical' or contradictory causal loops. In the philosophical literature this is known as the Bilking argument.

The basic structure of the Bilking argument is shown in figure 2. In essence, the claim is that there were backward causation (A causing E, in the terminology of the diagram), then it would be possible to set up a triangle of correlations, two of which are positive correlations, and the third of which (the diagonal link in the diagram) is negative. This supplies the required contradiction.

The Bilking argument was discussed by the Oxford philosopher Michael Dummett, in 1964 [7]. Dummett made the following observation. The argument depends on the assumption that it is possible to detect whether the event E occurs at time $(t-1)$, before A occurs (or does not occur) at time t. If this is not possible, then the Bilking apparatus is impossible to construct, even in principle. For the apparatus relies on such a detection, in order to guide our choice as to whether to bring about A. (Dummett's point is usefully amended in one respect: what is crucial is not merely that E should be detectable, but that it should be detectable *without disturbing the circumstances under which A is claimed to cause E*. Let us call this the detectability assumption.)

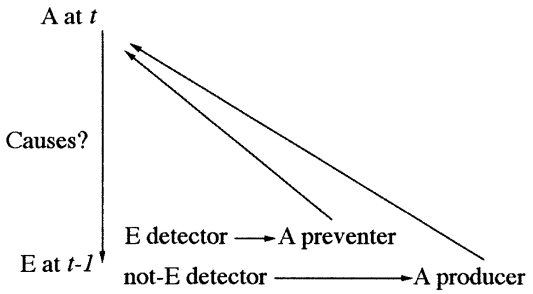

Figure 2. The Bilking argument.

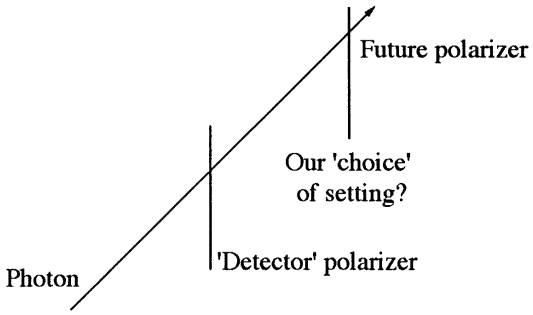

Figure 3. Dummett's loophole in quantum mechanics.

Dummett himself did not consider QM. (His discussion relies on imaginary examples, of an anthropological flavour.) But it turns out that QM is nicely placed to exploit the loophole in the Bilking argument identified by Dummett. For what does the assumption that it is possible to detect the claimed earlier effect amount to in the photon case? The claim is that the state of the photon is affected by the next polarizer it is going to encounter. To 'bilk' this claim, we would need to detect the polarization before it gets to the future polarizer.

However, the only way to detect the polarization is to put another polarizer in the path of the photon, before it gets to the future polarizer in question (see figure 3). And if we do that, the future polarizer is not the *next* polarizer any longer, and so the circumstances required for the claimed causal influence no longer obtain!

In other words, the restrictions that QM itself places on what it is possible to measure seem to ensure that the detectability assumption fails. If so, then the backward causation needed to make sense if QM does not lead to paradoxes.

1.3 *Conclusion*

The usual objections to relinquishing IA thus seem weak or misguided. On the other hand, as we saw, considerations of time-symmetry seems to count in favour of abandoning IA.

On this balance of pros and cons alone, then, HV models based on relinquishing IA seem to deserve some serious attention in the foundations of QM — no less attention, one might think, than other contextual models.

This assessment seriously understates the case, however. Unlike other contextual models, the backward causation approach offers the prospect of a resolution of the apparently space-like correlations implied by the failure of Bell's inequalities into a 'zig-zag' sequence of individually time-like correlations. In other words, this approach offers the prospect of a *local* explanation of the apparently non-local aspects of QM. This prospect might turn out to be illusory, of course. The required models might turn out to rule out on some other grounds, for example. But until these issues are addressed — until such models are investigated in depth — the whole issue of nonlocality in quantum theory is up in the air, and there can be no 'closure' for one of the great puzzles of twentieth century physics.

So much, at least in barest outline, for the case for investigating HV models which abandon IA. (I have developed this case at much greater length elsewhere [10].) However, much remains unclear, not only about the technical possibilities for such models, but also about their physical significance, and their relation to other approaches to the conceptual foundations of QM. In the remainder of the present paper I shall attempt to throw some light on these matters, by relating the HV approach in question to the issue of the completeness of the quantum-mechanical description of the world.

2. Two notions of completeness

At the heart of the debate between Einstein and Bohr in the early years of QM was the issue as to whether the wave function provides a *complete* description of a physical system. Does it, as Bohr claimed, give us all that is actually true in reality of the system in question? Or is it, as Einstein argued, merely a partial or incomplete description, telling us the truth but not the whole truth about the system?

In one sense, all HV approaches side with Einstein on this issue. By definition, a hidden variable is an aspect of physical reality not captured by the standard wave function. Often such views are associated with the claim that the wave function is not physically real, but this is quite inessential. After all, it is an incomplete description of Einstein to say that he was male, but the relevant aspect of his chromosomal structure (his possession of a Y-chromosome) is not therefore unreal. The crucial issue is that of completeness, not that of the 'reality' of the wave function.

It turns out, however, that although HV views agree with Einstein that the standard quantum mechanical description of reality is incomplete, they may need to disagree about what completeness requires. In fact, they may find themselves agreeing with Bohr that QM is complete in the sense of completeness which Einstein shares with Bohr; while maintaining that it is incomplete in another important sense, which neither Einstein nor Bohr seems to have entertained.

2.1 *Potentiality*

We humans are creatures with some knowledge of the past, but little knowledge of the future. Most of us care very much about what happens to our future selves. Unfortunately,

this often depends on unpredictable features of our, or 'their', future environment. The best we can do is to prepare our future selves to meet a wide range of possible futures.

As a result, it is useful for us — the best we can do, in fact, in many cases — to arm our future selves with a description of the world which is rich in 'potentiality'; in other words, a description which yields useful information about a wide range of possible futures. Potentiality is the best substitute for knowledge of the actual future itself. It gives us a kind of generic knowledge, useful in each of a wide range of future circumstances, many of which may be compatible with what presently we know.

Potentiality is a feature of many, if not all, commonplace descriptions of the world in which we live. Many properties are 'dispositional', or defined in terms of what *would happen* to the bearer of such a property in certain circumstances. Consider the property of fragility, for example, and the difference between saying merely that an object is fragile, and that it has or will *actually* break. Some philosophers argue that all properties are dispositional, and at least implicitly, such views are common in physics. Think of operationalism, for example, which insists that properties be characterised in terms of their 'disposition' or 'potential' to produce certain observable effects.

Whatever the truth about physical descriptions in general, the QM description is highly potential in nature. A familiar way to make this explicit is to use a propositional representation, in which the information carried by the wave function ψ is represented as a set of conditional probabilities for the outcomes of possible yes/no measurements.

$$\{P(Q_1|F_1), P(Q_2|F_2), \ldots, P(Q_j|F_j), \ldots\},$$

where $\{F_i\}$ is the set of possible future yes/no measurements, and Q_i is the proposition tested in case F_i.

Suppose that QM provides the best possible theory of this kind, in the sense that it encodes all the 'potentiality' that could be derived from a complete knowledge of a system's *past* interactions with the rest of the world (including us). In that case, let us say that QM is H_p*-complete*, where 'H' stands for 'human' and the 'p' for 'past', to remind us that this model of 'perfect knowledge' is tied to the kind of access to the world that we humans have: roughly, knowledge of the past but not the future. (Analogously, we could characterise H_f*-completeness* as perfect knowledge for creatures whose knowledge asymmetry is reversed.)

Now consider the same system from the perspective of an imaginary observer who is able to survey the future as well as the past; from an Archimedean point, so to speak. And let us suppose that the *actual* future measurement visible from this point is F_j. For the Archimedean observer, then, all the terms $\{P(Q_i|F_i)\}_{i\neq j}$ are *redundant*, in the sense that they provide information for futures known not to be actual.

From the Archimedean perspective, then, the standard quantum mechanical description looks both *incomplete*, in not providing the information that F_j is *actual;* and massively *redundant*, in providing a lot of information which is irrelevant — which has no application — in the light of this information about F_j.

Let us say that a description is A*-complete* if it includes all the information accessible from the Archimedean standpoint. Our conclusion is then that even if QM is H_p-complete, it is both A-incomplete and A-redundant.

One important note. The distinction between H_p-completeness and A-completeness is a matter of degree. A description which tells us that F_j is the actual future and gives us the value of $P(Q_j|F_j)$ (or equivalently $P(Q_j)$) is not entirely A-complete, because it does

not encode the information as to whether Q_j is actually true or false. But it is much more A-complete than the original quantum mechanical description.

2.2 *Hidden variables and kinds of completeness*

What is the relevance of the distinction between H-completeness and A-completeness to the possibility of HV models based on relinquishing IA? The crucial point is this. Once IA is abandoned, so that 'hidden' states are allowed to depend in part on what happens to the system in question in the future, then a specification of the hidden state provides information about the future. On the assumption that the standard quantum mechanical description is already H_p-complete, this information about the future must be information that could not be gleaned from a knowledge of the past interactions of the system. (In any case, such knowledge would open the door to bilking.) Hence it is information which could only be accessible from a more Archimedean standpoint.

Thus a HV model of this kind trades off some potentiality in return for a gain in A-Completeness. The values of the hidden variables λ_x and λ_y will not be predictive, in the fully counterfactual way that ψ is. To be precise, they lose predictivity with respect to the class of possible futures which they themselves exclude.

2.3 *The Bohr–Einstein debate revisited*

This discussion throws an interesting light on the Bohr–Einstein debate about the completeness of the quantum mechanical description of reality. The additional 'elements of reality' for which Einstein hoped *were* intended to be counterfactually-predictive. Recall Einstein's [8] 'criterion of reality': if we can predict with certainty what the result of a measurement *would be*, then there is an element of reality underlying that prediction, even if the measurement is not *actually* performed. The extra elements of reality provided by a hidden variable theory which violates IA would not in general be counterfactually-predictive in this way. If the state of a particle depends on the next measurement it is to encounter, then it will not be true in general that elements of reality which exist in the presence of later measurements would still exist if those measurements had not taken place.

In terms of the distinction drawn above, it thus appears that Einstein took for granted the (potentiality-laden) view of physical properties associated with H_p-completeness. In these terms, it may well be the case that Bohr is right — that QM is H_p-complete. This is compatible with possibility that QM is also A-incomplete, and that there are 'hidden' properties in reality in addition to the properties ascribed by QM, though properties less counterfactually-predictive. Thus Einstein may have been right, too, in a sense — albeit a sense which he himself did not envisage.

3. Advice for hidden variable theorists

The moral of this discussion for HV approaches — especially for those involving backward causation, but also, I think, for other contextual approaches — is that the goal should be A-completeness, not H_p-completeness. (Arguably, QM is H_p-complete already.) However,

it should also be borne in mind that A-completeness is a matter of degree. A consistent model which extends QM without achieving full A-Completeness is already an important step forward, in that it shows that HV models are possible, once we loosen the demand for counterfactual predictivity.

Here is a simple way to produce such a model. Begin with the conditional probabilities of the propositional representation mentioned above:

$$\{P(Q_1|F_1), P(Q_2|F_2), \ldots, P(Q_j|F_j), \ldots\},$$

where $\{F_i\}$ is the set of possible future yes/no measurements, and Q_i is the proposition tested in case F_i.

Now let one of the items in this list be flagged as the one corresponding to the *actual* next measurement, replace it with the corresponding unconditional probability, and delete all the other items. What remains is a description of the system in question which is more A-complete than its QM predecessor. Is it consistent? Yes, surely, if QM itself is, for all we have added is the information that a particular F_i (F_j, say) is actual. If the result were inconsistent, then QM would already imply that F_j does not happen!

One might anticipate two kinds of objection to this simple model. First, some people will throw up their hands at the suggestion that IA might be false. I recommend that we ignore such people, until they offer us better arguments in defence of IA than the ones we considered above.

Second, people will object to details of the model. For example, they may object that the model does not tell us how the system knows about future measurements. To this kind of objection two replies are appropriate, I think. Firstly, we should acknowledge that as it stands, the model provides no mechanism for the influence of the future on the present. It simply has the status of a primitive law-like fact. But every physical theory relies eventually on facts of this kind, so the fact that our model does so is not a damning objection. Secondly, and perhaps more importantly, we should observe that this objection misses the point of the model. The real purpose of the model is to establish an existence claim, the claim that there are consistent HV models for QM which are based on relinquishing IA. We do not need an elegant model to prove such a claim — any model will do.

3.1 *Two-time approaches*

One possible source of more elegant HV models abandoning IA is in the so-called 'two-time' approach to QM, pioneered by Aharonov, Bergmann and Lebowitz [1] (see also [2]). The basic idea of such approaches is to take the state of a system between measurements to be constrained or described by two wave functions, one of them the standard wave function ψ (or ψ_p, as we usefully call it) 'coming from the past', and the other an analogous wave function ψ_f 'coming from the future'. ψ_f depends on what happens to the system in the future, just as ψ_p depends on what happens to it in the past. (If ψ_p is taken to be H_p-complete, as suggested above, then ψ_f might be the corresponding H_f-complete description.) Hence the resulting model violates IA, at least so long as a change in ψ_f corresponds to a physical change in the system concerned. (The qualification is important. We do not have backward causation if changing the future measurement merely changes our evidence about the nature of the system at earlier times.) It may seem odd to describe this as a HV model, since it is couched in terms of wave functions. But for our present

purposes, a HV model is simply one which takes the standard wave function to be an incomplete description. This will be true of the two-time approach, considered in the current framework, according to which the descriptions ψ_p and ψ_f are individually A-incomplete.

Interesting as these models are, I think they suffer from one serious design flaw. They move towards A-completeness by combining two descriptions which are not only A-incomplete (a flaw which might be remedied by the combination) but also massively A-redundant. If the goal is an Archimedean description, it seems unnecessarily circuitous to begin with two descriptions which, in their massive 'potentiality', reflect the great deficiencies of non-Archimedean perspectives. Why should we be trying to get to an Archimedean description by 'adding' something to the standard description, if (as I have suggested) the latter is designed for a different job all together, that of characterising reality from the H_p perspective? Why not start from scratch, or rather from the classical notions we used in the good old days when (thanks to determinism and the assumption that there were no principled restrictions on observation) A-completeness and H_p-completeness came to the same thing?

3.2 *A natural alternative?*

Here is another possibility. Let us think of the 'hidden' reality in terms of Feynman paths, between an initial state (e.g., an electron being emitted by a source) and a final state (e.g, detection of that electron at a particular point on the screen in a two-slit experiment). In Feynman's path integral approach, calculation of the probability of the outcome in question depends on an integration over the possible individual paths between the given initial state and the given final state, each weighted by a complex number. The fact that the weights associated with individual paths are complex makes it impossible to interpret them as real-valued probabilities, associated with a classical statistical distribution of possibilities.

However, there is no such difficulty at the level of the entire 'bundle' of paths which comprise the path integral. If we think of the hidden reality as the instantiation not of one path rather than another but of one entire bundle rather than another, then the quantum mechanical probabilities can be thought of as classical probability distributions over such elements of reality. (For example, suppose we specify the boundary conditions in terms of the electron source, the fact that two slits are open, and the fact that a detector screen is present at a certain distance on the opposite side of the central screen. We then partition the detector screen, so as to define possible outcomes for the experiment. For each element O_i of this partition, there is a bundle B_i of Feynman paths, constituting the path integral used in calculating the probability of outcome O_i. We have a classical probability distribution over the set of such B_i.

Of course, this conception of the hidden reality violates IA. The range of possible bundles depends on all the boundary conditions, including those in the future. However, this is the kind of model we were looking for.

It might be objected that the model does not fully restore a classical picture of reality, because it does not assign an individual classical trajectory, but only membership of a bundle of trajectories. However, this misses the point of the exercise. We noted that A-completeness is a matter of degree, and that it is legitimate and interesting to enquire into the existence of HV models which are *more* A-complete than the standard QM description,

without insisting that they be entirely A-complete. The present model seems to provide a positive answer to the existence question, in a way which is much more 'naturally' motivated than our previous proposals. It is a further question whether it is possible to do better, and to find a HV model which is even more A-complete; but a question for another occasion.

4. Conclusion

I have argued that HV models violating IA deserve far more attention than they have yet received. On the one hand, the usual arguments in favour of IA are very poor. On the other hand, HV models which give up IA have two big potential advantages. First, they avoid the time-asymmetry inherent in IA itself. Second, they offer the prospect of a time-like resolution of the Bell correlations, and thus the elimination of the apparent tension between QM and special relativity.

I have also argued that in order to understand the significance of such models, it is important to realise that they do not offer further 'completeness of description' *in the sense taken for granted both by Bohr and by Einstein.* Instead, they offer further completeness in a different sense. The issue of which sense of completeness is more important for physics warrants further discussion. But for the moment, I think, we can afford to be pluralists. We should simply recognise that they are different, and get on with the business of exploring HV models with this improved understanding of what their goals may be.

Acknowledgement

I would like to thank David Atkinson, Dipankar Home and Jason Grossman for many helpful discussions of this material.

References

[1] Y Aharonov, P G Bergmann and J L Lebowitz, Time symmetry in the quantum process of measurement, *Phys. Rev.* **B134**, 1410–16 (1964)

[2] F J Belinfante, *Measurements and time reversal in objective quantum theory* (Oxford, Pergamon Press, 1975)

[3] J S Bell, On the Einstein-Podolsky-Rosen Paradox, *Physics* **1**, 195–200 (1964); reprinted in Bell (1987)

[4] J S Bell, Bertlmann's socks and the nature of reality, *J. Phys.* **42**, C2-41–C2-62 (1981); reprinted in Bell (1987)

[5] J S Bell, *Speakable and unspeakable in quantum mechanics: Collected papers on quantum philosophy* (Cambridge University Press, 1987)

[6] J S Bell, J Clauser, M Horne and A Shimony, An exchange on local beables, *Dialectica* **39**, 85–110 (1985)

[7] M A E Dummett, Can an effect precede its cause? *Proc. Aristotelian Soc. Suppl.* **38**, 27–44 (1954)

[8] A Einstein, B Podolsky and N Rosen, Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.* **47**, 777–80 (1935)

[9] J Lebowitz, Review of H Price, *Time's arrow and Archimedes' point* (Oxford University Press, 1996); *Physics Today* **1**, 68–69 (1997)
[10] H Price, *Time's arrow and Archimedes' point: New directions for the physics of time* (University Press, New York, Oxford, 1996a)
[11] H Price, Locality, independence and the pro-liberty bell (1996b), preprint archived at http://xxx.lanl.gov/abs/quant-ph/9602020