entropy

MDPI

Article

Lapsing Quickly into Fatalism: Bell on Backward Causation

Travis Norsen [1,†] and Huw Price [2,*,†]

[1] Department of Physics, Smith College, Northampton, MA 01063, USA; tnorsen@smith.edu
[2] Trinity College, University of Cambridge, Cambridge CB2 1TQ, UK
* Correspondence: hp331@cam.ac.uk
† These authors contributed equally to this work.

Abstract: This is a dialogue between Huw Price and Travis Norsen, loosely inspired by a letter that Price received from J. S. Bell in 1988. The main topic of discussion is Bell's views about retrocausal approaches to quantum theory and their relevance to contemporary issues.

Keywords: quantum theory; retrocausality; superdeterminism; J S Bell

Citation: Norsen, T.; Price, H. Lapsing Quickly into Fatalism: Bell on Backward Causation. *Entropy* **2021**, *23*, 251. https://doi.org/10.3390/e23020251

Academic Editor: Lawrence Horwitz

Received: 27 January 2021
Accepted: 19 February 2021
Published: 22 February 2021

1. Introduction

This paper is a dialogue between Huw Price and Travis Norsen, loosely inspired by a letter that Price received from John Bell in 1988. The main topic of discussion is Bell's views about retrocausal approaches to quantum theory and their relevance to contemporary issues.

2. Price (I)

As far as I can recall, I first heard about Bell's Theorem at a workshop at Wolfson College, Oxford, in the Spring of 1977 (I was in Oxford that year as an MSc student in Mathematics). I remember little of the talk, except that the speaker noted in passing that Bell's argument required the assumption that the properties of the particles concerned did not depend on the future measurement settings. At any rate, that is what I took him to be saying. I certainly do not recall the exact words, and we will see in a moment that there is another possibility for what he might have meant—but that was the sense of the assumption I took away.

I remember even this much because I was puzzled at the time that this assumption seemed to be regarded as uncontroversial. I had read some philosophy of time by that point—enough to be convinced that past and future are equally real, and to be familiar with the idea that time-asymmetry in the physical world is a statistical matter. Yet here was a time-asymmetric assumption about what can affect what—on the face of it, not a statistical matter— playing a crucial role in Bell's argument. Everyone seemed to agree that the argument led to a highly counterintuitive conclusion. But the option of avoiding the conclusion by rejecting the assumption did not seem to be on the table.

More than 40 years later, I am still puzzled. I have returned to the issue at intervals over those years, always looking for a good reason for closing what seemed to me an open door—a door, among other things, to a potential resolution of the tension between quantum theory and special relativity. To me, nature seemed to be offering us a huge hint, a hint revealed in Bell's work, that our intuitions about what can depend on what are unreliable in the quantum world. Yet few of my growing circle of friends and colleagues who knew about these issues—who knew a great deal more than I did, in most cases—ever seemed able to hear the hint. I often wondered what I was missing.

After my year in Mathematics at Oxford, I shifted to Philosophy in Cambridge. There, as some sort of sideline to my main thesis project on probability, I spent some time on the puzzle I had brought with me from Oxford. In [1], a piece written in November 1978,

I discuss what is in effect the following assumption, central to what is now called the *ontological models framework* [2]:

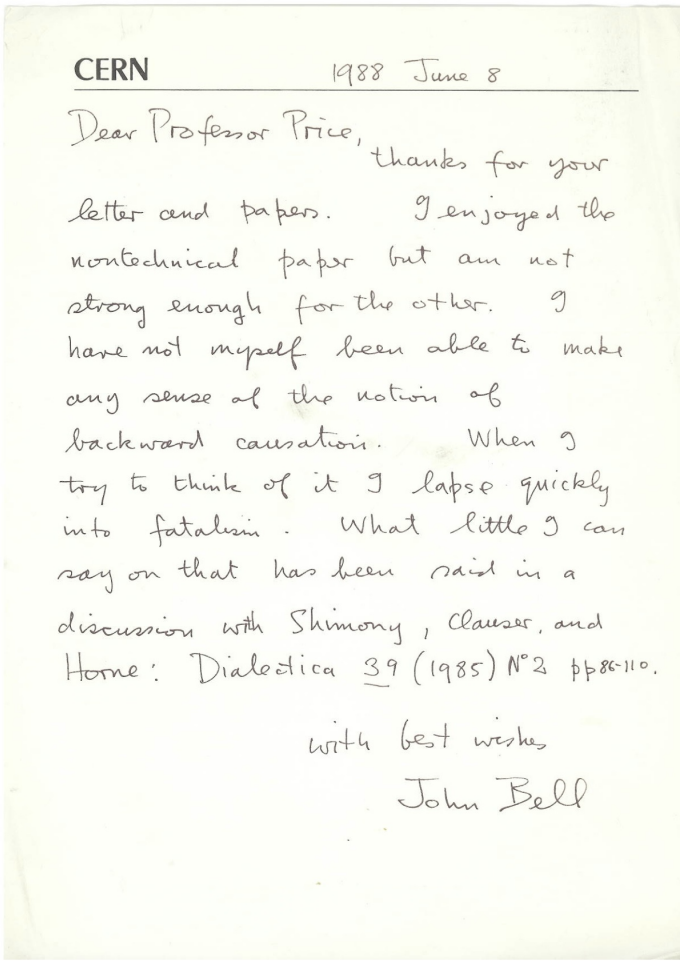
> In the ontological models framework, it is assumed that the probability measure representing a quantum state is independent of the choice of future measurement setting. (Leifer [3] (140))

My 1978 piece argues that this kind of assumption is 'very difficult to justify on metaphysical grounds', and notes that abandoning it has a very interesting potential payoff, given its crucial role in the no-go theorems of Bell and of Kochen & Specker. However, as the present ubiquity of the ontological models framework demonstrates, this has not become a common concern. Now, as in the 1970s, the assumption in question usually passes without comment—it is simply part of the model.

A couple of years later again, now a postdoc at ANU, Canberra, I worked on this little obsession some more. I focussed on the work of the Oxford philosopher Michael Dummett (who had been at the Wolfson workshop in 1977, I believe). Dummett had two well-known papers defending the coherence of retrocausality [4,5], and in a piece published in *Synthese* in 1984 [6] I offered some refinements to Dummett's arguments, and noted their potential application in the quantum case.

Back at ANU later in 1980s, I wrote the early drafts of a piece that eventually appeared in *Mind* in 1994 [7]. I think it was a draft of this piece, together with my *Synthese* piece [6] from 1984, that I sent to John Bell in 1988. His brief reply appears as Figure 1. On the idea of retrocausality, he says this:

> I have not myself been able to make any sense of the notion of backward causation. When I try to think of it I lapse quickly into fatalism.


CERN 1988 June 8

Dear Professor Price, thanks for your letter and papers. I enjoyed the nontechnical paper but am not strong enough for the other. I have not myself been able to make any sense of the notion of backward causation. When I try to think of it I lapse quickly into fatalism. What little I can say on that has been said in a discussion with Shimony, Clauser, and Horne: Dialectica 39 (1985) N°2 pp86-110.

with best wishes

John Bell

Figure 1. The letter from Bell.

For 'what little I can say', Bell then refers to a published discussion [8] with Shimony, Clauser, and Horne.

Bell's reference to fatalism in this letter certainly chimed with some of his published remarks, to the effect that to abandon the assumption in question, we would have to abandon the assumption that we are free to choose the measurement settings. This is from the piece referred to in his letter, for example:

> It has been assumed that the settings of instruments are in some sense free variables—say at the whim of the experimenters—or in any case not determined in the overlap of the backward lightcones. Indeed without such freedom I would not know how to formulate any idea of local causality, even the modest human one [8].

In the light of such remarks, the option of abandoning this assumption became known as the *free will* or *freedom of choice* loophole in Bell's Theorem.

As I will explain, Bell's letter did not do much to remove my sense of puzzlement. But it did help me to see that there are two very different models for what would be involved in abandoning the crucial assumption in Bell's argument—and that it was at least unclear whether Bell himself had properly distinguished them. Moreover, the model discussed in [8] was not the one that had interested me in the first place.

Accordingly, when I next wrote about these topics [9,10], I tried to distinguish the two models. Briefly, the difference is this. Both models involve correlations between measurement settings and properties of an incoming particle. In other words, both reject the assumption often referred to as *Statistical Independence.* But they propose to explain this correlation in very different ways.

It is a familiar idea that we need to distinguish correlation from causation, and that the same pattern on correlations may be compatible with more than one causal explanation. In susceptible folk, for example, eating chocolate is said to be correlated with the onset of a migraine, a short time later. If so, this might be because chocolate causes migraine, or because a migraine and a craving for chocolate are both effects of some underlying physiological cause (a common cause, as causal modellers say).

In the discussion in [8], it is assumed that the explanation of a correlation between measurement settings and underlying particle properties would have to be of the latter kind. In other words, it would have to be due to some common cause in the overlap of the past lightcones of the setting and the particle. Call this the Common Past Hypothesis (CPH). In Figure 2, adapted from a famous diagram due to Bell himself, the common cause is shown affecting both the measurement setting a on the left, and 'beables' λ on the right (which in turn can affect the measurement outcome B).

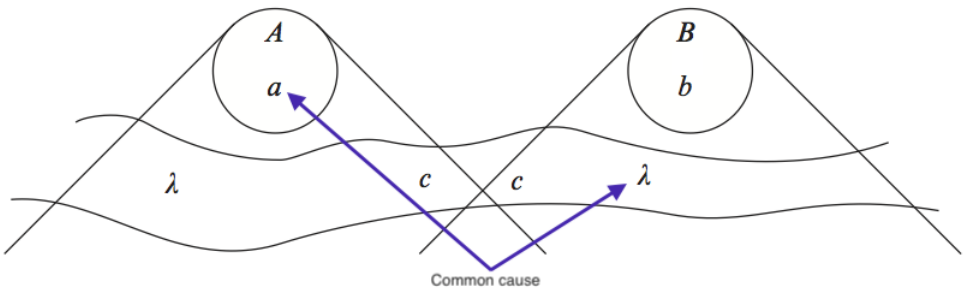


Figure 2. The Common Past Hypothesis (CPH).

In the case that I had in mind, however, the causal structure was different. As with chocolate-caused migraines, the correlation was to be explained by the hypothesis that the properties of the incoming particle were themselves causally influenced by the measurement settings. Of course, unlike in the chocolate and migraine case, this causal influence needed to work from future to past—it was retrocausality, as we now say. Setting aside for the moment the unfamiliarity of that idea—I am aware that some readers will see

it as the elephant in the ointment for this proposal—the difference between this explanation of correlations and the one involving a common cause is as stark as in the migraine case. They are completely different hypotheses.

Normally, of course, we take the measurement settings themselves to have causes in their own past, namely the choices of human experimenters, or the devices to which we humans delegate control. Adding this to the picture, we get the kind of causal model depicted in Figure 3. The green arrow from the left represents the experimenter's choice of the measurement setting a, and the black arrows represent the effect of this choice on λ and hence on B. A similar path would have to operate from right to left, of course. Note that nothing hangs on the fact that the green arrow is shown acting from outside the past lightcone (we will see that it is relevant that it comes from 'outside' in a different sense—it is an experimental intervention, of a kind that is universal in science).

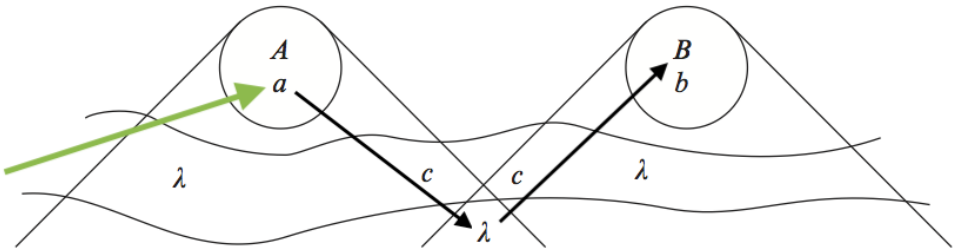


Figure 3. The Common Future Hypothesis (CFH).

In this second model, the correlation between measurement choices and earlier particle properties is explained in the future, by the fact that the particle encounters the measurement device there—a device whose setting, in turn, is determined by an earlier choice on the part of the experimenter. Accordingly, let us call this the Common Future Hypothesis (CFH).

It is easy to see how both CPH and CFH might lead someone to fatalism. However, it is crucial to see that they do so by very different routes. In the case of CPH, measurement settings are treated as effects of the postulated common cause. In the language of the causal modelling framework, the measurement settings are therefore being treated as endogenous variables (meaning, as Hitchcock puts it, 'that their values are determined by other variables in the model' [11]). Here, already, we have a stark contrast with the normal status of experimental settings in scientific models. Normally these are exogenous, 'meaning that their values are determined outside of the system' [11]. It is easy to see how this change of status might seem incompatible with a very down-to-earth sense of experimental freedom—and might seem, as Wiseman puts it, to 'undercut the core assumptions necessary to undertake scientific experiments' [12].

This objection goes back at least to Bell's discussion with Shimony, Clauser, and Horne in [8] (an exchange originally published in 1976). As Shimony, Clauser, and Horne put it:

> In any scientific experiment in which two or more variables are supposed to be randomly selected, one can always conjecture that some factor in the overlap of the backwards light cones has controlled the presumably random choices. But, we maintain, skepticism of this sort will essentially dismiss all results of scientific experimentation. Unless we proceed under the assumption that hidden conspiracies of this sort do not occur, we have abandoned in advance the whole enterprise of discovering the laws of nature by experimentation.

This challenge to CPH seems to me to be entirely correct. As I say, however, what I had in mind was CFH. From the beginning at Wolfson College in 1977, I had taken the interesting and questionable assumption to be the claim that the properties of the particles could not depend on future measurement settings. CPH does not challenge that assumption at all. It simply makes both the measurement settings and the particle properties depend on some third thing in their common past.

This means that I have not been bothered by these objections to CPH. They do not trouble CFH, where there is no bar to treating measurement settings as exogenous variables (that is the sense in which the green arrow in Figure 3 'comes from outside'). In CFH, the threat of fatalism comes from the fact that the particle 'already knows' the choice of measurement setting, before it is made. In the case of an observation on a photon from a distant galaxy, for example, the model requires that the particle has 'known' for billions of years what the measurement setting would be. Is that not incompatible with the ordinary belief that we have a free choice in the matter?

At this point I felt there were two things to say. First, this argument is logically identical to an ancient argument for fatalism, an argument starting from the assumption that statements about the future have determinate truth values. Theologians once worried about problems in this vicinity. Does not God's knowledge of the future deprive us of free will, for example? Leibniz discusses such issues, noting that 'the sophism which ends in a decision to trouble oneself over nothing will haply be useful sometimes to induce certain people to face danger fearlessly' [13] (153). In the following instance the strategic fatalist is Henry V, admonishing Westmoreland for wishing for reinforcements for their impending battle at Agincourt:

> No, my fair cousin:
> If we are mark'd to die, we are enow
> To do our country loss; and if to live,
> The fewer men the greater share of honour. (Henry V, Act IV, Sc. III)

If the concern about fatalism in CFH amounts to no more than this, it is a very damp squib indeed. Physicists should not allow themselves to be mugged by medieval kings and theologians.

Second, and more interestingly, the concern about CFH might rest on issues about causal loops. Suppose we could find out the relevant property of the incoming particle, before it reaches the measuring device, and use the information to change the measurement setting. Contradiction? This is a reasonable concern, but as I noted already in [6], Dummett's work identifies the solution. Backward causation is safe from such concerns, so long as it is impossible to 'find out' about the effect in question, before the choice of the future setting on which it depends. The restrictions that quantum theory puts on measurements seemed to offer a prospect that it could exploit Dummett's loophole.

For these reasons, I felt that in the case of CFH, concerns about fatalism were unwarranted. This left me eventually with the following view about Bell's position, extrapolated from his letter. Either Bell had not sufficiently distinguished CPH and CFH in his own thinking, and was assuming that the (valid) concern about the former would also encompass that latter. Or he had distinguished them, and was relying on a much more questionable argument for fatalism in the latter case. This is what I meant when I said that Bell's letter did not resolve my sense of puzzlement. Perhaps there was some some third option that I was missing, but I could not see what it would be.

It would be nice to be able to report that these issues are more clearly understood these days, but confusion persists in some quarters. In particular, it still seems to be widely believed that ruling out CPH would be sufficient to close the loophole in Bell's Theorem associated with Statistical Independence. These quotations are from two recent pieces ([14,15]) by Anton Zeilinger and collaborators, for example:

> The freedom-of-choice loophole refers to the requirement, formulated by Bell, that the setting choices are "free or random" [16] (232). For instance, this would prohibit a possible interdependence between the choice of measurement settings and the properties of the system being measured.
>
> A ... major loophole, known variously as the freedom-of-choice, measurement-independence, or setting-independence loophole ..., concerns the choice of measurement settings. In particular, the derivation of Bell's inequality explicitly assumes that there is no statistical correlation between the choices of measure-

> ment settings and anything else that causally affects both measurement outcomes. Bell himself observed 40 years ago that, "It has been assumed that the settings of instruments are in some sense free variables—say at the whim of experimenters—or in any case not determined in the overlap of the backward light cones".

In both cases here, it is simply taken for granted that the only avenue for statistical dependence between measurement settings and 'anything else that causally affects both measurement outcomes' is one that involves some common cause, acting in the past (the second paper claims to be 'pushing back by ~600 years the most recent time by which any local-realist influences could have engineered the observed Bell violation').

I do not want to overstate this claim of confusion. Some recent writers are admirably clear that CFH, or retrocausality, is a distinct proposal for rejecting Statistical Independence (Leifer [3] links it to the option of rejecting the ontological models framework; see [17,18] for recent surveys). Moreover, we now have a better understanding of the issue of the relationship between time-symmetry and retrocausality, another factor in my interest in the case from the beginning. My own result [19], extended and generalised by Leifer and Pusey [20,21], shows how in quantum theory time-symmetry may require retrocausality, for a reason not present in the classical case. As Leifer and Pusey note, the argument depends on thinking about time-like analogues of EPR-Bell arguments: The EPR argument in my case, and Bell's extension in theirs.

For me, the effect of this recent work has been to increase my sense that nature is offering us a loud hint in Bell's results, a hint to which many people interested in these topics are curiously deaf (If anything, I now feel that there are two hints, one from Lorentz invariance and one from time-symmetry). As always, however, I am conscious that there may be objections of principle that I cannot see, and that Bell did see, perhaps. If there is any trace of such objections in Bell's letter from 1988, I am hoping that this discussion will unearth it. I am also conscious that I have not yet touched on the idea of so-called 'superdeterminism', or the suggestion that something objectionably 'conspiratorial' is required for a proposal such as CFH. Again, if there is a valid objection to CFH of this kind, I am hoping this discussion will highlight and clarify it.

Before I yield the floor to my colleague, let me mention again the (supposed) elephant in the ointment for CFH, the idea of backward causation itself. As noted above, my own early discussion of these ideas took a deliberate path via some classic philosophical work on the issue, that of Michael Dummett [4,5]. Since those early days, I have written extensively on the issue of the direction of causation [10,22]. In those investigations, I have not yet found anything that ought to count as a fly in the ointment for CFH, much less an elephant. On the contrary, Dummett's identification of a loophole in objections based on causal loops remains precisely the fly-free balm that CFH requires, in my view.

I know of one place where Bell himself discusses backward causation explicitly. In 'La Nouvelle Cuisine', seeking a characterisation of the sense of locality apparently implied by special relativity, he suggests that it might be defined in terms of cause and effect [16] (235):

> As far as I know, this was first argued by Einstein, in the context of special relativity theory. In 1907, he pointed out that if an effect followed its cause sooner than light could propagate from the one place to the other, then in some other inertial frames of reference the 'effect' would come before the 'cause'! He wrote:
>
> > … in my opinion, regarded as pure logic … it contains no contradictions; however it absolutely clashes with the character of our total experience, and in this way is proved the impossibility of the hypothesis … of a causal chain going faster than light.

Bell goes on to explain what Einstein had in mind—a case of causal loops, exploiting the fact that the effects precede their causes in some inertial frames. Perhaps this is evidence that Bell himself thought that backward causation is excluded by these causal loop arguments. If so, then the escape hatch he needed is close at hand. All such arguments

depend on the assumption that backward causation could be used to signal (to tell one's grandmother to avoid her unhappy marriage to one's grandfather, perhaps, in the classic paradox). However as Bell himself made clear, the Bell correlations imply causality without signalling. So long as retrocausality stays on the same side of the line, it is safe from paradox (once again, this is Dummett's loophole, in effect).

As for Einstein's own attitude to retrocausality, this little fragment suggests an admirable and characteristically empirical attitude. The objection is that 'it absolutely clashes with the character of our total experience'. Of course, Einstein did not know about the 'aspect' of our total experience that was to be revealed by Bell's work, and by the experiments it inspired. So it would be presumptuous, to say the least, to read into Einstein's remarks any general prohibition on retrocausal models.

3. Norsen (I)

Thanks to Huw for inviting me to join him in what promises to be an illuminating dialogue and for perfectly setting the context with that beautiful opening statement.

Since it does not involve any direct interaction with (or letters from) Bell, my own personal biography vis-a-vis Bell's theorem is far less interesting than Huw's. But in so far as I can reconstruct it, I can report that I first learned of Bell's theorem from David Albert's book 'Quantum Mechanics and Experience' [23], which I stumbled across at Orca Books in downtown Olympia, Washington (USA) during the winter vacation in the middle of my sophomore year of college, which would have been late 1994 or early 1995 (i.e., about four years after Bell's untimely death). Albert's book (and the many other books and articles on the foundations of QM (Quantum Mechanics) that I subsequently began to devour) made a significant impression on me and gave me a youthful confidence to raise skeptical questions about orthodox quantum mechanics as it was presented in my physics courses.

Given what I now know of the typical attitudes toward quantum foundations in the physics community, it is rather surprising that my undergraduate professors during this period were so incredibly supportive of my interests in unorthodox viewpoints. However they must have found my passion at least a little naive, and, in hindsight, I can now see that they were not completely wrong.

Huw remembers questioning a very specific and subtle assumption in Bell's theorem the first time he encountered it. For me the story is very different. I spent years knowing that Bell's theorem was crucially important, simply because everybody unanimously agreed that it was, but having basically no clear idea at all what to make of it, because different authors all seemed to present their own totally unique version of the theorem's logical structure. David Albert had claimed that Bell's theorem proved that nature was non-local, but other commentators told very different stories, usually along the lines of: Bell had refuted determinism, or the related and EPR-inspired hidden-variables program, and had therefore put the final nail in Einstein's coffin and proved once and for all that the orthodox interpretation of Bohr and Heisenberg was the only viable one.

Some semblance of clarity only emerged while I was in graduate school (nominally pursuing a PhD in theoretical nuclear astrophysics but really spending at least half of my time secretly studying quantum foundations), when it finally occurred to me that perhaps Bell himself might have an illuminating perspective on his own theorem. Reading Bell's collected papers [16] turned out to be very helpful indeed. Even his more technical papers were completely accessible and clear, and Bell's ability to explain his reasoning, crisply and cleverly, was truly masterful. I went from having no idea what Bell's theorem actually proved, to having no idea how any controversy could remain when Bell had laid everything out so perfectly.

Huw's opening statement focused mostly on the notion of *Statistical Independence*, which is the main focus of discussion. However to set the context and pre-empt any possible misunderstanding or miscommunication, I think it will be helpful to step back and lay out the overall structure of Bell's theorem (at least as Bell himself understood it and helped me to understand it).

There are two assumptions. One is the *Statistical Independence* that Huw has discussed, according to which (in the usual EPR-Bell sort of setup) the measurement settings a and b are "free" or "exogenous" and therefore (at least as long as we set aside the idea of retro-causation that Huw wants us to consider) uncorrelated with the variables λ, which characterize the physical state of the particle pair at some earlier time. There will be much more to say about this assumption as the discussion proceeds.

However I wanted to make sure to acknowledge explicitly that there is also another assumption—which Bell and I and most others would consider in some sense the more central and important assumption—namely, Local Causality. Bell's careful and important mathematical formulation of this notion is intended to capture the qualitative idea, motivated by relativity, that the causal influences on a given event are to be found exclusively in that event's past light cone, and the causal influences of a given event (on other events) are to be found exclusively in the future light cone. The idea, in short, is that causal influences propagate (from the past toward the future) always at the speed of light or slower. Readers unfamiliar with Bell's formulation of Local Causality are urged to read his final and clearest presentation in "La Nouvelle Cuisine" [16] (232) and/or my own dissection in "J.S. Bell's Concept of Local Causality" [24].

As a brief aside that will be relevant later, let me stress here that the core virtue of Bell's formulation of *Local Causality* is that it is totally and completely generic. The formulation is not in terms of the proprietary concepts (e.g., quantum mechanical wave functions) of some specific candidate theory, but is exclusively in terms of the un-sectarian notion that Bell invented for the purpose: "Beables", which simply means whatever some candidate theory posits to exist. And (unlike for example the conditions known as "Parameter Independence" and "Outcome Independence") Bell's *Local Causality* does not make reference to any specific type of process or situation and in particular does not imply or require any sub-classification of beables into distinct sub-types (e.g., those which are human-controllable "parameters" vs. those which are uncontrollable "outcomes").

Bell's concept of *Local Causality*, that is, possesses the same virtues that Bell demanded of candidate theories when he complained that orthodox quantum theory, with its special ad hoc rules for how systems behave during measurements, was "unprofessionally vague and ambiguous" [16] (173). As he elaborated elsewhere, terms such as measurement, observable, system, and apparatus "...however legitimate and necessary [they might be] in application, have no place in a formulation with any pretension to physical precision" [16] (215). To avoid suffering from the sort of "measurement problem" that afflicts orthodox quantum theory, Bell thought, the ontological posits and dynamical laws of a proper candidate fundamental theory should be stated in precise mathematical way, without vague, anthropocentric terms or distinctions. Bell thus appreciated the professionalism of various unorthodox formulations of quantum theory such as the pilot-wave theory, spontaneous collapse theory, and (to a lesser extent) Everett's many-worlds theory and we should appreciate the professionalism of Bell's *Local Causality* on similar grounds.

Returning to Bell's theorem, the two assumptions, *Local Causality* and *Statistical Independence*, turn out to jointly entail something ("Bell's inequality") which I think Huw and I will agree is now known, from experiment, to be false. So at least one of the two assumptions—*Local Causality* and *Statistical Independence*—must be rejected.

Bell's view, and the view that I and many other commentators have tended to adopt, is that *Statistical Independence* is something like an unquestionable assumption of empirical science, the denial of which amounts to endorsing a kind of cosmic conspiracy theory. Indeed, in one important earlier article on Bell's theorem [25], my co-authors and I called the *Statistical Independence* assumption by the alternative name "No Conspiracies"—much, no doubt, to the annoyance of Huw and others of his ilk.

As part of this opening statement let me just stipulate for the record that Huw is entirely correct to insist that the idea of backwards-in-time causation—the Common Future Hypothesis, CFH—potentially provides (or at least appears to potentially provide) a non-conspiratorial way of violating the *Statistical Independence* assumption. So this very much

deserves to be peeled apart and examined carefully, hence my excitement to participate in this dialogue.

However I have a big-picture question that I think should be addressed here at the outset. The people who want to deny the *Statistical Independence* assumption—not on the basis of retrocausation and the CFH, but rather on the basis of the CPH and hence what Huw and I would agree is a scientifically-unacceptable kind of conspiracy—want to do so in order to save Local Causality. That is, the end-game of the conspiracy theorists is to find a way of reconciling the relativity-based idea that causal influences propagate (from the past toward the future) always at the speed of light or slower, with the empirical violation of Bell's inequality.

However, this cannot be your endgame, Huw, since (as I have tried to stress) the other premise of Bell's theorem, *Local Causality*, also has a (not merely statistical) arrow of time built into it. We could put the point like this. The kind of retro-causation that you want to use to provide a non-conspiratorial ground for rejecting *Statistical Independence*, also just blatantly and openly violates Bell's notion of *Local Causality*; it says, after all, that the causal influences on certain events are to be found in their future light cones. Thus, apparently, despite both focusing skeptical attention on the *Statistical Independence* assumption, you are not at all trying to achieve the same thing as the conspiracy theorists. So what exactly are you trying to achieve?

I think I have a sense of what your answer will be, but I am sure it will help focus the subsequent discussion to have this laid out explicitly.

4. Price (II)

Travis mentions a fateful Olympian encounter with Albert's excellent book, 'Quantum Mechanics and Experience', which led him into quantum foundations. This gives me an opportunity to recommend Travis's own recent text, 'Foundations of Quantum Mechanics'. Contemporary versions of the 1990s Travis, or the 1970s me, would be just as lucky to encounter this book as he was to encounter Albert's—in some ways, even more so. Among other things, Travis's book, unlike Albert's, offers a rich sense of engagement with the founders of quantum foundations—Einstein, Bell, and many others. Travis achieves this by working in well-chosen words from these greats, and I will draw on some of those in a moment.

Travis points out a contrast between retrocausalists such as me, who look for a violation of Statistical Independence via what I have called CFH, and those we are now labelling conspiracy theorists, who do so via CPH. As Travis says, proponents of CPH are trying to save Bell's principle Local Causality. That can not be my goal, because, as he puts it, Local Causality has an "arrow of time built into it". So what exactly am I trying to achieve? It is a good question. The main part of my answer is that I want to defend a more basic sense of Locality, and show how the world might violate Bell's Local Causality but respect the more basic notion.

What is the more basic notion? Here it is in the words of Einstein, quoted by Bell, in a passage reproduced in Travis's book:

> The following idea characterises the relative independence of objects far apart in space (A and B): External influence on A has no direct influence on B. [26] (109)

Let us call this *Einstein Locality.* Retrocausal models want to preserve Einstein Locality at the cost of Bell's Local Causality—at the same time explaining why this is not much of a cost at all, once we understand the limitations of Bell's version. The major limitation is the way in which *Local Causality* simply builds in a causal arrow of time. Einstein Locality says nothing about time—we could replace 'far apart in space' with 'far apart in spacetime, in any direction' and still have much the same idea.

Travis describes the ideas that Bell's *Local Causality* is intended to capture like this:

> Bell's ... mathematical formulation ... is intended to capture the qualitative idea, motivated by relativity, that the causal influences on a given event are to be found exclusively in that event's past light cone, and the causal influences of a given

> event . . . are to be found exclusively in the future light cone. The idea, in short, is that causal influences propagate (from the past toward the future) always at the speed of light or slower.

There are two parts to this qualitative idea, one the restriction to lightcones motivated by relativity, and other—much older, obviously—that causal influences propagate from past to future. Let us call the latter the Causal Arrow of Time, or CAT for short. Note that CAT actually combines two things, first a distinction between cause and effect, and second the claim that the cause–effect 'arrow' lines up with the earlier–later 'arrow' (unless the causal relation is itself asymmetric, it makes no sense to say that it points in a particular direction).

A generation before Schrödinger's famous feline, this CAT played a role in the most notorious rejection of causation in modern philosophy. In 1912, Bertrand Russell argued that modern physics had no use for causation, and that philosophy should therefore discard it too: "The law of causation," Russell said, "like much that passes muster among philosophers, is a relic of a bygone age, surviving, like the monarchy, only because it is erroneously supposed to do no harm" [27]. One of Russell's main arguments is that there is no asymmetric relation in physics that we could identify with causation. The computer scientist Judea Pearl, himself a leading contemporary writer on causation, sums up Russell's point like this: "[T]he laws of physics are all symmetrical, going both ways, while causal relations are unidirectional, going from cause to effect" [28].

Russell's argument had little practical effect, and indeed, as Patrick Suppes pointed out later [29], physicists themselves often use causal notions. Still, Russell had put his finger on a puzzle—'Russell's enigma', as Pearl calls it. As Pearl says:

> [V]ery few physicists paid attention to Russell's enigma. They continued to write equations in the office and talk cause-effect in the cafeteria; with astonishing success they smashed the atom, invented the transistor and the laser. [28]

The long debate about nonlocality shows that it is not just in the cafeteria that these things matter in physics, but this makes the enigma all the more urgent. What is this asymmetric relation doing in physics—or anywhere else, for that matter, in a world built on the symmetric laws of physics?

Simplifying a bit, we can distinguish three contemporary accounts of CAT (see [22] for discussion and references).

1. A matter of definition. Following Hume, we can treat CAT as a matter of definition. This view holds that the basic relations of dependence are among the symmetric relations identified by physics, and the terms 'cause' and 'effect' are just labels for the earlier and later of a pair of events related in this way;
2. Thermodynamics. We can try to explain CAT in terms of the thermodynamic arrow of time (and in particular the so-called Past Hypothesis, or low entropy initial boundary condition). This view has been defended in recent years by writers such as Kutach, Albert, and Loewer;
3. Interventionism. The third possibility seems to originate with Frank Ramsey [30]. Ramsey, one of the fathers of the subjectivist approach to probability, takes a similar line on causation. As he puts it, "from the situation when we are deliberating seems to . . . arise the difference of cause and effect". In effect, Ramsey proposes an explanation of the time-asymmetry of causation (indeed, causality itself) in terms of the epistemic perspective of agents like us. This approach has been influential in recent decades, thanks to the work of writers such as Jim Woodward, and Judea Pearl himself [28,31]. It is now called Interventionism, alluding to the central role of the idea of intervening on a system of interest—reaching in 'from the outside', to fix the value of an exogenous variable.

None of these accounts of CAT seem much use to Bell, seeking to build a fundamental causal arrow into a principle for quantum foundations. The first is empty, since it makes a matter of definition that effects are later than their causes. (It would have nothing to say about probabilistic dependence between hidden variables and future measurement settings,

though it would prohibit us from calling it 'causality'.) The second seems insufficiently fundamental. On the face of it, we want a theory of the quantum world that is independent of the thermodynamic environment in which a system happens to be embedded. And the third seems insufficiently fundamental for a different reason. Its dependence on the perspective of agents like us seems deeply in tension with Bell's desire, as Travis puts it, to avoid 'anthropocentric terms or distinctions'.

For Interventionism, anthropocentricity about CAT is only one part of a broader issue. As Pearl himself makes clear, the role of intervention threatens the idea that causation itself is fundamental:

> If you wish to include the entire universe in the model, causality disappears because interventions disappear—the manipulator and the manipulated lose their distinction. However, scientists rarely consider the entirety of the universe as an object of investigation. In most cases the scientist carves a piece from the universe and proclaims that piece in—namely, the focus of investigation. The rest of the universe is then considered out or background and is summarised by what we call boundary conditions. This choice of ins and outs creates asymmetry in the way we look at things, and it is this asymmetry that permits us to talk about "outside intervention" and hence about causality and cause-effect directionality. [28]

In case you feel tempted to respond 'so much the worse for Interventionism', look again at Einstein Locality. As Einstein says, "External influence on A has no direct influence on B". That looks very much like Interventionism. It is doubtful if we can formulate any notion of Locality, or indeed any causal notions at all, without implicitly relying on intervention. It is built into assumptions about what we treat as exogenous variables.

However perhaps we can at least do without an anthropocentric temporal arrow? After all, does physics not permit a time-symmetric notion of intervention? At least in a deterministic framework, we might interpret Einstein's 'external influence' in terms of an imagined change to properties in a small region of a Cauchy surface at an intermediate time. Such a change, propagated forwards and backwards in accordance with the relevant dynamical laws, will 'produce' changes elsewhere. (Here's a more homely example I once used elsewhere. Consider the perspective of someone planning to remake the entire series of Star Wars movies, with some tweaks to central characters. The prequels are required to be consistent with the original Episode IV, and hence tweaks made there will affect the plots of the remakes in both directions. If we substitute Harry Potter for Luke Skywalker the ramifications will spread backwards as well as forwards in the temporal dimension of the series.) Intuitively, Einstein Locality would be the requirement that such changes propagate only by continuous processes within the lightcones. (Spacelike influences would be allowed, but only indirectly, by indirect zig zags via the lightcones).

This may be a good way to capture the sense of causality that matters to relativity, where the idea that there is some sort of fundamental temporal asymmetry seems entirely gratuitous. It might seem an attractive approach for a retrocausalist, too, but I think it throws out far too much, in two senses. First, working science is simply not like this, either in the cafeteria or the laboratory. Our ordinary notions of causality, in science as in everyday life, are those of agents embedded in time with a particular temporal orientation. Throwing all that away would leave the view hostage to the objection rightly raised against CPH that it is incompatible with the assumptions we need to do science. As I explained above, CFH is immune from that objection.

Even more seriously, this symmetric option simply obscures the subtle idea at the core of the retrocausalist proposal. This idea is that even by the lights of the ordinary asymmetric perspective, it is possible that the world contains an indirect kind of retrocausality—hard-to-notice cases in which by intervening in the future, we can make a difference in the past.

Summing up, I want to make three points. First, nothing we know about CAT justifies taking it as fundamental, in the sense in which Bell's Local Causality differs from Einstein Locality. Second, it is doubtful whether Locality, or indeed any interesting notion of causality, can be captured qualitatively in wholly fundamental terms. The right response

to this is not to abandon talk of causality in physics. Instead we should keep a close eye on the role of the agent's perspective, in order to keep in mind a question like this: What sort of fundamental structure looks like this from here? (Travis, I think we are on the same page in wanting such a story).

Third, coming back to something I mentioned briefly above, we are going to need a distinction between direct and indirect spacelike influence. Einstein Locality rules out the direct kind, but not apparently the indirect kind. This was clear to the pioneer of the approach, Olivier Costa de Beauregard, who pointed out a decade before Bell's Theorem that zig zag causality, via the past lightcones, provided a potential loophole in the EPR argument [32]. It offered spacelike influence, without action at a distance. This distinction is still missed in some quarters. The following example is from an email I received from an experimentalist known for work in confirming the Bell correlations:

> For me, [the] Costa de Beauregard zig zag in space time, which you seem to consider equivalent to retrocausation, is nothing else than nonlocality. The addition of one time-like vector to the past and one time-like vector to the future, connecting the detections, results in a space-like vector, and a causal relation between both ends, spacelike separated, amounts to a non local relation.

As I said, Costa de Beauregard originally thought of his idea as a challenge to EPR. The EPR argument assumes an intuitive notion of Locality, in arguing that measurement choices at A cannot affect measurement outcomes at a remote location B. Costa de Beauregard's point was that if we allow causal influence in both directions within the lightcones, and adapt our notion of Locality accordingly, this argument no longer works. There is now a zig zag path for local causal influence to reach from A to B. (I once met Costa de Beauregard, late in his life. I asked him when he had first had the idea for the zig zag, which he first published in 1953. He said in the late 1940s, when he had been a student of de Broglie; but that de Broglie would not let him publish it, until they saw Feynman's work treating positrons as electrons zig-zagging backwards in time.)

Thus in answer to your question, Travis, Costa de Beauregard's zig zag is still the endpoint that I have in mind. As you rightly point out, it involves rejecting Bell's version of Local Causality. But for the reasons I have sketched, the crucial thing that we need to drop—that is, CAT, the causal arrow of time—is on shaky grounds anyway, as a principle for fundamental physics. And Einstein's formulation of Locality from 1948 looks like the alternative that we need (I once met Costa de Beauregard, late in his life. I asked him when he had first had the idea for the zig zag, which he first published in 1953. He said in the late 1940s, when he had been a student of de Broglie; but that de Broglie would not let him publish it, until they saw Feynman's work treating positrons as electrons zig-zagging backwards in time).

5. Norsen (II)

There is a lot going on in that response, all of it very helpful in moving us toward what I see as a possible way of making sense of Bell's perhaps-puzzling dismissiveness about "fatalism". It may take some time to connect the various threads, though.

So, Huw, your vision involves rejecting both *Statistical Independence* and Bell's *Local Causality*. But you would hope to preserve a different, time-symmetric notion of locality in which the causal influences on a given event cannot be at space-like separation from it, but are equally allowed to be in either the past- or the future-light-cone. You suggest calling this alternative notion "Einstein Locality" and suggest that it is well-captured by a passage from Einstein's 1948 essay. For the record, and despite my appreciation of your praise of the book you quoted the passage from, I have serious reservations about basing the particular time-symmetric notion of locality that you have in mind on that particular (by the way, translated) passage, and thus attributing your proposed locality concept to Einstein. But a debate about exactly how to parse Einstein's words will be a pointless distraction here since you have made it abundantly clear what you have in mind. So, having noted

my reservation, I will simply follow you in describing your proposed time-symmetric alternative to Bell's *Local Causality* as "Einstein Locality".

My main concern (other than the terminology) about "Einstein Locality" is that, as you acknowledge, "we're going to need a distinction between direct and indirect space-like influences". From your point of view, the whole purpose of Einstein Locality is to endorse, as compatible with "the sense of causality that matters to relativity", the zig-zag sort of multi-step, "indirect" causal influence across space-like separation that Bell's Local Causality prohibits. Without this distinction between "direct" and "indirect", we would be stuck saying that Einstein Locality allows the very thing that it poses as prohibiting (namely, causal influence across space-like separation). In that case, Einstein Locality wouldn't actually prohibit anything, i.e., it would be rather empty and pointless.

As I think we agree, it thus seems that your proposal of replacing Bell's Local Causality with Einstein Locality swims or sinks with the project of situating this distinction between "direct" and "indirect" influences in the context of fundamental physics. And my initial gut reaction is that this project seems rather hopeless. When I look at extant theories that possess the appropriate sort of professionalism (e.g., Maxwellian electrodynamics, general relativity, some non-orthodox version of quantum theory) I see causal influences taking the form of continuous propagation. There is nothing "atomic", for example, about the sequence of steps whereby one charged particle affects the motion of another nearby charged particle via the intermediary electric and magnetic field. There are, if you like, a continuous infinity of infinitesimal intermediating steps, but in my opinion that description should be understood as a human theorist's perspective on something that is, in reality, a seamless whole.

So is one charged particle exerting force on another nearby charged particle a "direct" influence (because the process is a seamless whole) or an "indirect" influence (because it can be viewed as consisting of an infinite number of infinitesimal sub-steps)? I think the only good answer is to reject the question and whatever line of thinking motivated us to pose it.

Huw, I hope you will correct me if I am wrong, but I get the impression that instead of trying to find a sharp distinction between direct and indirect influences in fundamental physics, you perhaps want to ground that distinction in an appeal to the concept of causation itself, and in particular the notion of "intervention", without which, you said, we probably cannot formulate "any notion of Locality, or indeed any causal notions at all" (Incidentally, is not Bell's Local Causality a counterexample there? It does not explicitly involve, and does not appear to me to implicitly rely on, the idea of "intervention"). It may well be true that if we restrict our use of cause-and-effect terminology to processes involving agent-intervention, it might provide a clean way to say, for example, that the zig-zag influence depicted in Figure 3, from the setting a to the space-like separated outcome B (via the particle pair state λ), is unambiguously indirect. Both the setting a and the pair state λ are after all (at least in part) exogenous: Somebody sets the setting, and somebody sets up the equipment in a certain way to produce particle pairs in a certain state (or a certain distribution of possible states), and both of those interventions are inputs to (not subjects of) the model in question there. In short, we have, in this case, two distinct interventions, which (on this view) implies the two distinct causal influences symbolised, in Figure 3, with the two distinct black arrows. Maybe this could be said to render the effect of a on the space-like separated B unambiguously indirect.

However to me this kind of proposal for grounding the distinction between direct and indirect causal influences also seems highly suspicious and implausible. We want, at the end of the day, an account of fundamental physics that is not afflicted by anything like a "measurement problem", i.e., we want to avoid the use of, or need for, anthropocentric or otherwise "unprofessionally vague and ambiguous" concepts and distinctions at the fundamental level. We want, as you quote Pearl as saying, ultimately "to include the entire universe in the model". And, when we do that, I think it is exactly right that "interventions disappear".

Of course, Pearl goes on to suggest that, in practice, "scientists rarely consider the entirety of the universe as an object of investigation", that models always (or almost always) include some parts of the world only as boundary conditions, and that "this ... permits us to talk about 'outside intervention"'. That may well all be true, but I would nevertheless find a formulation of Einstein Locality which required explicit reference to "intervention" (to ground the distinction between direct and indirect influences) to be squarely in the "unprofessionally vague and ambiguous" category.

Incidentally, it might surprise some readers to know that Bell, who of course invented and applied the concept of Local Causality, was in some sense very sympathetic to at least part of the view that Huw quoted Pearl describing as "Russell's enigma", i.e., the idea that, at the level of fundamental physics, causality (at least described as such) is nowhere to be found. After presenting his formulation of Local Causality in "La Nouvelle Cuisine", for example, Bell remarks:

> Note, by the way, that our definition of locally causal theories, although motivated by talk of 'cause' and 'effect', does not in the end explicitly involve these rather vague notions. [16] (240)

This, however, does not mean that (appropriately "professional") theories do not describe processes which can legitimately be characterised in causal language. Bell once wrote that, in "pursuing [his] profession of theoretical physics" he was required to "insist ... on the distinction between analysing various physical theories, on the one hand, and philosophizing about the unique real world on the other hand" [16] (101) Continuing:

> In this matter of causality it is a great inconvenience that the real world is given to us once only. We cannot know what would have happened if something had been different. We cannot repeat an experiment changing just one variable; the hands of the clock will have moved, and the moons of Jupiter. Physical theories are more amenable in this respect. We can calculate the consequences of changing free element in a theory, be they only initial conditions, and so can explore the causal structure of the theory. [16] (101)

This comment occurred in the context of explaining that his concept of Local Causality (and, for example, the conditional probabilities which appear in its formulation) should be understood as referring to theories (which are, in turn, candidate descriptions of the unique real world) rather than to the unique real world directly. Physical theories (at least the serious ones that aspire to fundamentality) thus not only, in Bell's view, have "causal structures", they are our best and necessary tool for, in the long term, discovering the causal structure of the real world.

Anyway, le me summarise my concern about "Einstein Locality". I fear that, unlike Bell's Local Causality, this notion will never be formulatable in a meaningful and appropriately fundamental way that accomplishes, Huw, what you want it to accomplish. I fear, in particular, that it will be impossible to cleanly distinguish "direct" from "indirect" influences in an appropriately "professional" manner, and I fear that, without the terminological check provided by an appeal to explicit "interventions", an Einstein Local theory of the sort you claim to want will be riddled through with the sort of causal influences across space-like separation that (however you might want to classify them) are just prima facie contrary to "the sense of causality that matters to relativity." (On this last point, I am in complete agreement with your experimentalist correspondent.)

In the same discussion with Shimony, Clauser, and Horne that has been referenced several times already, Bell wrote, about the idea of saving Local Causality by rejecting Statistical Independence via the CPH:

> A theory may appear in which such conspiracies inevitably occur, and these conspiracies may then seem more digestible than the nonlocalities of other theories. When that theory is announced I will not refuse to listen, either on methodological or other grounds. But I will not myself try to make such a theory. [16] (103)

I feel this same way about the idea of rejecting both *Local Causality* and *Statistical Independence* but preserving "the sense of causality that matters to relativity" with some notion of Einstein Locality. I doubt this could be done, and so am not interested in spending my own time and effort on the project, but would be delighted to listen if and when somebody puts forward a precise formulation of Einstein Locality and/or a concrete example of a candidate fundamental theory which shows, if only in principle, how this project could work.

Am I correct, Huw, that neither exists at present? I have to admit that I was somewhat confused by your proposal to formulate Einstein Locality "in terms of an imagined change to properties in a small region of a Cauchy surface at an intermediate time". That actually sounded rather promising to me, so I was puzzled that in the end you seemed to reject it as "throw[ing] out far too much".

In particular, I did not understand the worry that this was somehow "incompatible with the assumptions we need to do science". It seems to me that this formulation is perfectly compatible with ordinary scientific practice. Indeed, do not several existing (serious candidate) theories, e.g., Maxwell's electrodynamics, respect this condition?

It seems to me that your rejection of this formulation—what I think you expressed when you said that it "obscures the subtle idea at the core of the retrocausalist proposal"—must be based on the recognition that such theories as Maxwellian electrodynamics do not appear to support the specific sort of zig-zag causality that would allow for a non-conspiratorial violation of Statistical Independence. However, it is not clear to me what would.

Let me try putting all my cards on the table here. In my response so far, I have not really addressed a core aspect of your proposal, namely, the tension between the apparently time-symmetric fundamental laws, and Bell's time-asymmetric Local Causality. You are clearly committed to the idea that time-symmetry is fundamental; as you explained from the very beginning, this is the motivation for the whole retrocausalist project. By contrast, I am more open to the possibility that some kind of Causal Arrow of Time (CAT) may remain in our fundamental physics. For one, I am not comfortable presupposing that the fundamental laws will turn out to be deterministic; maybe, as the founders of quantum mechanics seemed to believe, the world will turn out to be irreducibly stochastic. Although I never found the arguments of the founders convincing, I do not think this possibility has been ruled out, and as long as that remains true I think it is premature to insist that the fundamental laws are time-symmetric. (Of course, it is also not certain that irreducible stochasticity requires time-asymmetry. But to me at least it does not seem like time-symmetry and irreducible stochasticity play well together). But even if the fundamental laws do turn out to be time-symmetric, I am not convinced this means that there can not, or should not, be something like a fundamental causal arrow of time.

That said, though, such a fundamental CAT, like explicit notions of 'cause' and 'effect', may not appear as such, may be relatively invisible, in the formulation of a candidate fundamental theory—or at least a deterministic candidate fundamental theory. Indeed, in the context of such theories, my concern about the retrocausalist proposal is not so much that backward causation is impossible or unconscionable, but rather that the distinction between forward and backward causation seems to melt away. In Maxwellian electrodynamics, for example, the state of the particles and fields at one time determines the state of the particles and fields at a later time. So did the former cause the latter, or vice versa? The laws of the theory certainly do not answer that question; they just tell us that the states at the two times are necessarily connected. (This, I take it, was Russell's point.) So is Maxwellian electrodynamics a retrocausal theory? Maybe? I am honestly not even sure what the question means.

Of course, at the non-fundamental level, where we model only some narrow part of the universe and describe its surroundings as "exogenous" variables through which we might "intervene" on the narrow part under study, the distinction between forward and backward causation seems much clearer. If the system changes due to an intervention in its

past, that is forward causation, whereas if the system changes due to an intervention in its future, that is backward causation. But if we demand that the vague and anthropocentric classification of beables as "endogenous" or "exogenous"—if we demand that reference to "intervention"—should disappear at the more fundamental level, it just seems like the distinction between forward-in-time and backward-in-time influences will have to disappear too.

So I tend to think that you retrocausalists fool yourself into thinking there is a meaningful program to pursue here, by taking a much too interventionist perspective on causation, i.e., by thinking too exclusively about very narrow models of specific situations in which various things are treated explicitly as "exogenous" "interventions". And in particular I think that as soon as you try to imagine embedding one of these models, e.g., the one pictured in Figure 3, into a candidate fundamental theory, in which everything is treated on an equal footing, the very notion of retrocausation—and with it the distinction between the CPH and the CFH—will disappear like a mirage.

Could I be wrong about all of this? Absolutely. To me, the easiest and best way to find out would be to scrutinise a concrete example of a serious candidate fundamental theory that respects a time-symmetric Einstein Locality condition and which (unlike Maxwellian electrodynamics?) supports the needed kind of indirect, zig-zag causality in (but only in) the EPR-Bell type of setup where the retrocausalist needs Statistical Independence to be violated.

Unfortunately, I do not think any such theory exists at present, and everything I have said here should make pretty clear why this does not surprise me. But if (or when) I am wrong, and such a theory is presented, I will not refuse to listen.

In summary, these are what my cards look like. If it turns out I am (at least arguably) not wrong about all of this, there will be at least a bit more to say about how this relates to Bell's apparent conflation of the CPH and the CFH, i.e., his perhaps-puzzling dismissal of the retrocausality program on the grounds of "fatalism". However Huw, I think I should pause here and give you a chance to weigh in on what I have said.

6. Price (III)

In my opening Section I described my view that Bell's Theorem contains a loud hint from nature. Bell shows that if we assume Statistical Independence (SI), quantum theory implies nonlocality. The hint turns on the thought that we should read this as a reductio ad absurdum, and conclude that SI fails in the quantum realm. (I do not mean absurdity in the logical sense, of course, but rather what Newton had in mind, when he said of the the idea that 'one body may act upon another at a distance through a vacuum, without the mediation of anything else' that it was 'so great an absurdity, that . . . no man who has in philosophical matters a competent faculty of thinking, can ever fall into it.') This would be the obvious reading if we could already see on the shelf some plausible way in which SI might fail, but we do not. Looking further back on the shelf there are two possibilities, CPH and CFH—the former evidently much easier to see, from most vantage points. I suggested that deafness to the hint might rest on failing to distinguish them, and hence on the view that Bell's well-founded objections to CPH would apply to any attempt to abandon SI.

In response, Travis, you pointed out that Bell's own assumption Local Causality would still fail in CFH, and asked in what sense I could therefore claim to be defending locality (or avoiding nonlocality, as the reasoning just described requires). I offered Einstein Locality as a substitute for Bell's notion, and you have now said that you doubt whether the distinction between the two will be expressible in vocabulary permitted by fundamental theory.

It will not matter if this turns out to be the case, in my view, because the argument can simply fall back on Lorentz Invariance. If we treat a violation of Lorentz Invariance, and the need for a preferred frame, as the absurdum avoided by giving up SI, then the hint speaks to us just as before. To put this another way, suppose we concede to my famous 2015 correspondent that Costa de Beauregard's zig zag proposal still counts as nonlocality. No matter, so long as the zig zag offers us a path to an explanation of the Bell correlations

that avoids the tension Bell himself saw between quantum theory and special relativity. It has often been noted that there are two elements to the counterintuitive character of nonlocality in contemporary physics, the first linked to relativity and the avoidance of preferred frames, and the second, obviously much older (see Newton's remarks above), to the counterintuitiveness of action at a distance itself. A restriction to the language of fundamental physics might prevent us from expressing the latter, but presumably not the former. And this will do just fine, for the case for the hint.

As I said in §2, I now feel that there are two hints, one from Lorentz Invariance and one from Time Symmetry. Concerning the latter, I want to stress again that I think that quantum theory introduces a new reason, not present in the classical regime, for thinking that time symmetry requires retrocausality. A common objection to the retrocausal proposal is a challenge I could paraphrase like this: 'What about time-symmetric classical physics? Is that retrocausal?' This presents retrocausalists like me with a dilemma. If we say 'Yes', we are admitting that retrocausation is not novel or interesting, because it is common in classical physics; if we say 'No', we have conceded that time symmetry alone does not imply retrocausality, leaving it unclear what would do so.

Travis, I take you to be expressing the latter part of this challenge in remarks such as this: '[S]uch theories as Maxwellian electrodynamics do not appear to support the specific sort of zig-zag causality that would allow for a non-conspiratorial violation of Statistical Independence. But then it is not clear to me what would'. In response, I want to outline what I now think of as the best case for thinking that quantum theory is different.

The argument emerged from discussion in [33] of timelike versions of EPR-Bell experiments, such as the one depicted in Figure 4. The 'future' end of the experiment involves a polarising beam splitter, as used in many standard spacelike EPRB experiments. The 'past' end involves the same kind of device used in reverse, with a photon entering on one of two channels. In this timelike one-photon experiment the input–output correlations depend on the relation between the settings α and β of the two polarisers, just as in a regular spacelike two-photon EPRB experiment.

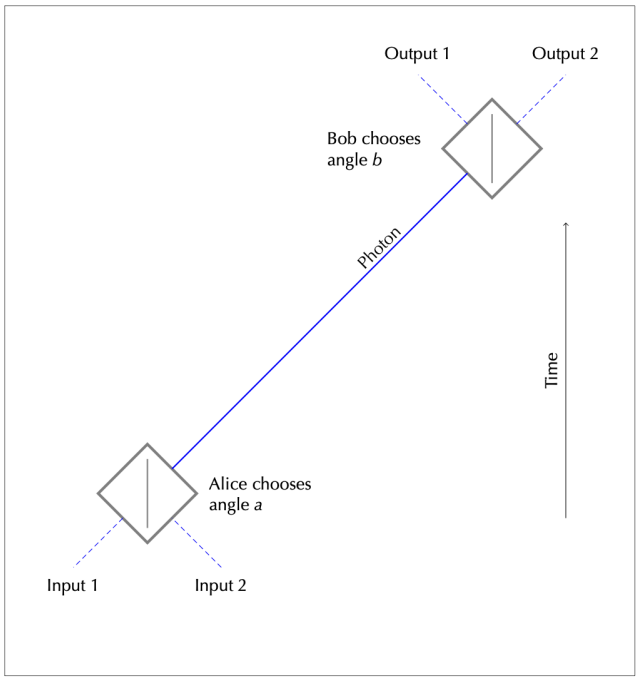


Figure 4. A timelike EPR-Bell experiment.

Normally, the earlier experimenter (Alice) would be able to control the inputs as well as the setting α, and would hence find it easy to signal to Bob, using photon polarisation to carry the required information. The key new idea in [19] is that we restrict Alice's control,

to make her situation analogous to Bob's. We do this by putting the input photon under the control of Demons, assumed to know the setting α. If the Demons are required to put the input photon on one input channel or the other—i.e., they are not allowed a superposition of the two—then their knowledge and options mirror those of nature at Bob's end of the experiment, in two senses. Nature knows Bob's setting β, and is required to produce a photon on one output channel or other (at least when a measurement is made).

Let us call the no superpositions rule the *Discreteness* condition. Without it, the Demons have complete control over the polarisation of the photon between the two devices. Knowing α, they can choose weights for an input superposition to produce any polarisation they want. With Discreteness, however, Alice retains very substantial control. She controls the polarisation completely, up to an additive factor of 0 or $\pi/2$. (This factor depends on which input channel the Demons choose.) Nevertheless, this degree of control does not guarantee that Alice can signal to Bob. If the Demons choose the input channel at random, and Alice does not know it when she chooses the setting, the control that results from Discreteness does not permit Alice to signal.

So in this artificial situation—interesting because of the way it mirrors normal circumstances at the other end of the experiment—Discreteness introduces what looks by ordinary interventionist lights to be a new element of forward causality. In effect, it gives Alice an extra degree of control over the probabilities at Bob's end of the experiment, compared to the case in which the Demon is not restricted in this way, (As [20,21] point out, this is effectively the EPR reasoning, transferred to the timelike case.) If this extra degree of control is reflected in an underlying ontology, and the ontology is time-symmetric, it will give Bob the same degree of control over the ontology at Alice's end of the experiment. This will amount to a violation of SI and to retrocausality, though not of a kind that would support signalling to the past, for the same reason as in Alice's case.

The full version of this argument comes with caveats I have not mentioned here. (Some of these caveats are removed in the generalisation by [20,21].) However I hope I have said enough to explain why I take it on the one hand that violation of SI and retrocausality are not an automatic consequence of time symmetry, but on the other hand that quantum theory has features that may make them so. Moreover, the subtlety of the new kind of forward causality revealed by this argument offers at least the beginnings of a response to a different challenge: If there were retrocausality, why would it not be obvious and everywhere? The response is that even its forward twin is hard to see. In effect, we had to move ordinary forward control out of the way first, and then focus on what remains. (Note that in familiar time-asymmetric models, in which the intermediate polarisation depends on Alice's setting but not Bob's setting, the forward causality explains the correlations all by itself—no retrocausality needed. This suggests that in time-symmetric models, with causal influence in both directions, the forward and backward components could be even more subtle because they share the explanatory work.)

Let me come back to the remark of yours I quoted above: '[T]heories as Maxwellian electrodynamics do not appear to support the specific sort of zig-zag causality that would allow for a non-conspiratorial violation of Statistical Independence. But then it is not clear to me what would'. The answer suggested by the reasoning just described is that under the constraint of time symmetry, non-conspiratorial violation of SI emerges from the same place as the Bell correlations themselves. We saw that if we consider Bell correlations in timelike settings, the EPR and Bell arguments reveal a distinctive kind of forward causality. From that point, modulo the assumption mentioned about an ontological basis for this causality, insisting on time-symmetry gets us to the violation of SI. (If we map this back to the space-like cases, we then have the zig-zag causal structure we wanted.)

Travis, I appreciate that you are less convinced than I am about both time symmetry and CAT. As you say, you want to hold open the possibility that one or both fail, perhaps independently. For me, both these options fall in the 'I would not refuse to listen' box. However what matters for present purposes is not our differing credences about these things. We seem to agree that—contrary to what many seem to think—a time-asymmetric

CAT does not have the status of a well-established piece of physical or metaphysical lore, something that can simply be invoked without argument, in order to dismiss the hint. On the contrary, work outside physics makes a significant case that the causal asymmetry is not fundamental—which means that we have at least some reason to be suspicious of proposals within physics that assume otherwise, implicitly or explicitly. Similarly, work inside physics gives us enough reason to take time symmetry seriously to imply that we certainly can not take its failure for granted. So again, there is no sign here of something sufficiently well-grounded to dismiss the hint.

The issue you raise of what goes into fundamental theory is very interesting, and in my view quite difficult. It is hard to make sure that we have eradicated the conceptual traces of our human perspective, especially our asymmetric temporal perspective. We are predictive creatures, always acting for the future on the basis of limited knowledge of the past. It is no surprise at all if we describe the world in terms appropriate for such a viewpoint. However we need to keep an open mind to the possibility that the fundamental level need not be described in these terms. (Again, I think we agree on the principle, even if we have different credences about where it might lead us.)

These lessons are as important for retrocausal approaches as for anyone else. For one thing, there is a risk that in pursuit of time symmetry, retrocausal approaches find themselves doubling-up ontology that would be better discarded. At the fundamental level we should try to prune away elements that reflect our time-asymmetric viewpoint, not balance them by adding elements reflecting a time-reversed viewpoint. In my view the so-called Two State Vector approach may be guilty of this mistake [34].

With these 'meta' issues about fundamental theory still open, and the path to CFH invisible to most in foundations of physics, let alone in physics more broadly, I think it is no surprise that we retrocausalists do not yet have anything that could claim to be a 'serious candidate fundamental theory', as you put it. I recommend [17,18] for recent surveys of various approaches that have been proposed.

In the spirit of putting cards on the table, I will take this opportunity to record a couple of preferences. First, I think that the de Broglie–Bohm theory (dBB) provides an attractive and under-explored framework for this approach. It has the advantage of an ontology both clear and sparse—I am thinking especially of the sense in which position is the only fundamental property, other properties being contextual and relational. This sparsity reduces the risk that we unwittingly build our own epistemic viewpoint into what is intended to be fundamental ontology. Another virtue is that in dBB, probabilities emerge much as in classical statistical mechanics, from a distribution over initial conditions. We do not need fundamental time-asymmetric chances, or anything of that kind. (I think there is no deep difficulty in the fact that we normally consider a distribution over initial conditions; final conditions would do the job just as well.) Finally, the dBB ontology offers an obvious place to hide some subtle retrocausality—in fact, two places, namely the particle positions and the pilot wave itself, if we give the latter an appropriate ontic status. There has already been some work exploring retrocausal versions of dBB: see, e.g., [35,36]). See also [21] for the sense in which orthodox dBB is time-asymmetric.

Secondly, and independently, I like proposals that seek to show how non-conspiratorial SI-violating correlations might emerge from global constraints—e.g., recent work by Wharton [37], Palmer [38], and Adlam [39,40]. The spirit of this approach is nicely captured by Adlam's remark that 'God does not play dice, he plays Sudoku' ([41]). In Adlam's version the approach is also linked to an interest in fundamental ontology in what seems to me a very interesting way.

Let me close by coming back to the idea that if we stick to the language of fundamental physics, it will be impossible to draw my distinction between CPH and CFH. If I understand you correctly, Travis, you think this may explain Bell's apparent wish to lump the two together. I am thinking of remarks such as this:

> Indeed, in the context of such [fundamental] theories, my concern about the retrocausalist proposal is not so much that backward causation is impossible or

> unconscionable, but rather that the distinction between forward and backward causation seems to melt away. In Maxwellian electrodynamics, for example, the state of the particles and fields at one time determines the state of the particles and fields at a later time. So did the former cause the latter, or vice versa? The laws of the theory certainly don't answer that question; they just tell us that the states at the two times are necessarily connected. (This, I take it, was Russell's point.) So is Maxwellian electrodynamics a retrocausal theory? Maybe? I am honestly not even sure what the question means.

In response, I want to distinguish two questions. First, does causal language appear in the fundamental theory? I think we agree in saying 'No' to this, at least if we set aside your lingering attachment to a fundamental CAT.

Second, does fundamental theory make distinctions which, when viewed from the ordinary perspective of agents like us, map onto the distinctions we make in causal language (including those involved in distinguishing between CPH and CFH, or between forward and retrocausal models)? Here I say 'Yes'. Why? Well, consider the corresponding questions about colour. Are red and green categories in fundamental theory? Obviously not. Do we expect fundamental theory to mark the difference between red things and green things? In some sense, obviously, yes—that is what it is to take colour to supervene on fundamental physics.

If there is a difference between the colour case and the causation case, it is that causal concepts are much more deeply embedded in scientific practice. Some philosophers would see in this an argument for thinking that causation itself needs to be fundamental, but set that aside. Here we are considering the possibility that causation is not fundamental, or at any rate not part of fundamental physics. My point is that this gives us no reason to reject the supervenience of causation on physics. We should expect causal distinctions to depend on lower-level differences, just as we expect for colour (This is not to deny that for both colour and causation, there is also an anthropcentric element to the story about the relation between the higher-level categories and underlying physics).

This means that I see no reason to think that fundamental physics will not continue to provide the distinctions we need to make causal judgements, including those needed to distinguish between CPH and CFH. However even if I turned out to be wrong, I think it would leave the case for retaining SI no better off. If we throw out the vocabulary we need to distinguish CPH and CFH, we also throw out the terminology we need to raise Bell's objections to rejecting SI. Even if we set aside such explicitly anthropocentric terms such as 'free will' and 'fatalism', the remaining objection turns on the idea that science treats measurement settings as exogenous variables. You are imagining that even this notion 'should disappear at the more fundamental level'. However if we do not have that notion in our vocabulary, we can not make Bell's objection, and if we do have it, then we can distinguish CPH from CFH. Even worse, we saw that the case for CFH (non-conspiratorial violation of SI) runs very close to the EPR and Bell arguments themselves. If it really were true that fundamental theory prevented us from making the former, I think it would follow that we could not discern the latter. So there is a danger that by blocking the hint we would deprive ourselves of the vocabulary to describe the problem itself.

7. Norsen (III)

In response to my skepticism about (so-called) Einstein Locality, you suggest that we "can simply fall back on Lorentz Invariance". However Lorentz Invariance is nearly as ambiguous and problematic as the various notions of locality we have been discussing. For example, theories with something like a "preferred frame" (an idea you mentioned, I think, as something obviously incompatible with the spirit of fundamental relativity) can be Lorentz Invariant and indeed can be argued to be fundamentally relativistic in a serious sense [42,43]. I also personally have questions about what compatibility with relativity could or should mean for theories (like virtually all extant, serious, and empirically viable quantum theories) which postulate non-local beables such as the wave function.

However even leaving that issue aside, I remain confused by the same big-picture point I raised initially. We know that it is possible to violate Bell's Local Causality in a Lorentz Invariant theory, in several distinct ways [35,42,44]. Thus if the goal is just to reconcile Bell's theorem with Lorentz Invariance, we can do this with (Lorentz Invariant) non-locality while still maintaining Statistical Independence. What is the motivation for instead rejecting both Local Causality and Statistical Independence, by introducing retro-causation? Of course, I cannot rule out the possibility that going down that road will yield a theory that, despite in some sense rejecting more of Bell's assumptions than is minimally necessary, is more natural or believable overall as a way of reconciling Bell's theorem and the associated experiments with fundamental relativity. But, as I have said, I would personally want to be more convinced that the idea of retro-causality was even coherent, in the context of a candidate fundamental theory without "measurement problem" issues, before investing my own time and energy on such a project.

On that issue of coherence, I must admit that I remain somewhat unsure about how to answer the question—"What about time-symmetric classical physics? Is that retrocausal?"—that you explained poses a dilemma for the retrocausalist. I think your answer must be no, because otherwise I would not understand why you would bother with the somewhat complicated setup and associated argument purporting to establish "that quantum theory is different"—different, I gather you mean, from straightforwardly time-symmetric theories (such as classical mechanics) in supposedly possessing the somewhat subtle and previously-unrecognised sort of causal (including specifically retro-causal) influences that you discuss.

Unfortunately, though, I find this argument totally unconvincing. It is based completely on the orthodox/textbook version of quantum mechanics including, in particular, the collapse postulate whose presence is the very core of the measurement problem. The crucial assumption you call Discreteness is just a kind of time-symmetrised collapse postulate, and, as you acknowledge in the more detailed discussion of [19], the argument fails to work for the two extant candidate quantum theories with time-symmetric dynamics, namely the de Broglie–Bohm (dBB) and Everett theories. (I will also note that the alternative notion of time-symmetry developed, for example, in [20], as a generalisation of your argument, is in my opinion revealed to be inappropriate/irrelevant by the fact that manifestly time-symmetric theories such as dBB and Everett do not respect it.) I would summarise the situation by saying that the argument fails to work precisely because these theories eliminate the need for a collapse postulate, i.e., because they do not suffer from the measurement problem.

Yes, as you point out there, this does not mean that such theories necessarily exclude retro-causality. Indeed, as I indicated before, the fundamentally time-symmetric character of the dynamical laws in those theories makes me perfectly open to admitting that they have retro-causality to exactly the same degree or extent, whatever that is exactly, that they have regular forward-causality. But I gather, from the fact that, if I am understanding correctly, you suggest somehow changing or supplementing dBB—"to hide some subtle retrocausality"—that you do not see any retro-causality, of the sort you want and need, in that theory's standard extant formaultion. However I do not understand that proposal at all. In [19] you suggest that, in order to deny that dBB is retro-causal in the needed sense, "it needs to be assumed that neither the wave function nor the initial positions of the particles are affected by later measurement choices". However both elements of the dBB ontology, the wave function and the particle positions, just obey deterministic evolution equations. Aside from "external fields", which would obviously not be included when the theory is applied to the world as a whole, there is simply no room for outside influences, by "measurement choices" or anything else.

To me, that is, it appears that the idea that dBB is somehow a promising candidate for retro-causality of the sort you want and need, arises only from a failure to appreciate that dBB can and should be thought of as a candidate fundamental theory, which does not need (and indeed does not even allow) special ad hoc exceptions to the basic dynamical postulates, associated with measurement or anything else. To me it instead seems clear that

the only sort of retro-causality one could plausibly attribute to the de Broglie–Bohm theory is just exactly the sort that one could, with equal plausibility, attribute to the Everettian quantum theory, Maxwellian electrodynamics, or classical particle mechanics. However, again, and as far as I can tell, this is a sort of retro-causation that you acknowledge is not what you need, so you do not even call it retro-causality.

To summarise that point, it seems to me that in order to find the sort of thing you are looking for, you retrocausalists need to set aside the best extant candidate fundamental time-symmetric quantum theories (dBB and Everett) and instead work with a time-symmetrised version of a theory that suffers from "measurement problem" issues (or, as in [20], implausibly re-define the meaning of "time-symmetric"). This just reinforces my previously-expressed sense that the very concept of retro-causation, of the sort you want and need, is a kind of mirage. It appears to be meaningful only when you put interventionist causality in by hand, by backing away from the fundamental level of description and instead treating certain things as "exogenous", i.e., outside of the quantum system under study and amenable to some kind of agent control. Doing this is, admittedly, part of the essential character of orthodox quantum theory. But that does not make it right.

You suggest that although causal language will likely not appear, as such, at the fundamental level, a candidate fundamental theory should probably still "make distinctions which, when viewed from the ordinary perspective of agents like us, map onto the distinctions we make in causal language". I completely agree. Causality should supervene on fundamental physics even if it does not appear there explicitly labelled as such. And as I have said I am completely open to the possibility that, for example, in the context of a candidate deterministic time-symmetric fundamental theory like dBB, there might well be some legitimate grounds for speaking in terms of retro-causal influences. I do not exactly see what those grounds might be, but I am happy to leave the door open. But at least as long as the fundamental physics remains deterministic, it seems to me that any correlations of the sort needed to violate Statistical Independence—even ones that we end up agreeing make sense to describe, "when viewed from the ordinary perspective of agents like us" who are part of the world described by the theory, in terms of retro-causality and the CFH—will imply correlated correlations, so to speak, in the physical state of the world at some much earlier time.

Let me try to explain more clearly what I am trying to get at here. There is some equipment in the lab that is arranged, in a certain way, to produce a sequence of particle pairs whose state we describe with the variable λ. And the experimenters set up some pieces of measuring equipment, including, say, some random number generators, which perform, on the incoming particle pairs, measurements of the particles' spins along directions a and b. We know, I take it, that λ is causally influenced by (even if not completely determined by) the state of the lab equipment: If something is not plugged in, no particle pairs will emerge at all; if some optical element is mis-aligned, the pairs may emerge in (what QM would describe as) a triplet state instead of the intended singlet state, etc. Similarly, the precise sequence of settings, a and b, is in fact determined by various details of the setup.

To make things really concrete, let us suppose that the particle source produces pairs with states coming in a certain order, starting from the moment the equipment was most recently powered up. And suppose that morning there was a windstorm in town, which resulted in a tree falling on a power line, briefly interrupting the supply of electricity to the lab, and thus causing the particle source to reboot and reinitiate its sequence of emitted pairs at that particular moment. Similarly, suppose that the seed for the pseudo-random number generating algorithm was chosen, on this occasion, by the number of blueberries in the lab assistant's pancake at the diner earlier that morning.

Now, in a fully deterministic theory like dBB, there does not seem to me to be any basis for claiming, nor does there appear to be any room for adding, causal influences from the settings, a and b, onto λ. The settings are just determined by various things in their past, as are the pair states. Thus to posit a Statistical-Independence-violating correlation between these is to posit a very special, "just so" type of correlation between the precise

physical details of the morning windstorm which resulted in that particular tree being blown down at that particular moment, and the precise physical details of the factors determining the subtle movements of the diner chef's hand which resulted in, say, 11, rather than 10 or 12, blueberries ending up in that one particular pancake. This is the sort of correlation needed to violate Statistical Independence with the CPH, and I think we agree that it seems unacceptably conspiratorial.

The new point I want to make here is that I do not see, exactly, how violating Statistical Independence instead with the CFH leads to a less conspiratorial picture. Suppose I keep an open mind and allow for the possibility that some hypothetical future theory might include, in some meaningful and compelling way, retro-causal influences from the settings a and b onto the pair states λ. My point here is that, presumably, that theory will also have to acknowledge the causal connection (and hence tight correlation) between the precise sequence of settings a and b and whatever complicated factors determined how many blueberries ended up in the lab assistant's pancakes. It will also, presumably, have to acknowledge the causal influence of the morning windstorm on λ. I do not know quite how the distinct causal influences on λ—one (backwards in time) from a and b, and one (forward in time) from the morning windstorm—would be reconciled in such a theory. Maybe the theory would say that a and b are, alone, sufficient to determine λ, and would thus be forced into saying that λ, instead of being influenced by the morning windstorm, retro-causally determines the relevant details of that morning windstorm. Or maybe there really would be two oppositely-direct causal influences on λ, neither of which alone would be sufficient to determine lambda, but which, together, are. In this scenario as well, we would (presumably) still have to end up with a very tight correlation between the morning windstorm and the settings, and therefore also between the morning windstorm and blueberries.

This is precisely the sort of correlation, between seemingly-random details of the states of seemingly-unrelated phenomena in the past, which we regard as unacceptably conspiratorial in the context of the CPH. Why should such correlations be considered any less conspiratorial in the context of the CFH? For this reason, to me, the distinction between violating Statistical Independence with the CPH, and violating it with the CFH, becomes blurry at best when we zoom out, include more of the world in our system, and refrain from treating various elements in a special way, as "exogenous interventions".

You claimed, at the end of your most recent contribution that to whatever extent we lose the ability to make the CPH/CFH distinction, we will also lose, to that same extent, the ability to argue for the reasonableness of Statistical Independence (or even to rehearse the EPR argument and Bell's theorem) in the first place. I do not understand this. As discussed in the exchange with Shimony, Clauser, and Horne that Bell referenced in his letter to you, the main case for Statistical Independence is a practical and empirical one. In particular, Statistical Independence is assumed in virtually every scientific experiment. Think, for example, of a randomised controlled drug trial whose standard interpretation requires assuming that the (say) coin flips, determining which patients got the drug and which the placebo, were uncorrelated with the previous health of the patients: If all of the patients who got the real drug survive and all the patients who got the placebo die, we would ordinarily infer that the drug works well to prevent death. However, of course, it is possible that the drug has no positive effect at all; instead, the coin came up tails for, and hence the placebo was given to, all and only those patients who had some other health condition, totally unrelated to that targeted by the drug, and were destined to die of that. As Shimony, Horne, and Clauser put it, in a passage you quoted earlier but which bears repeating here, denying Statistical Independence:

> ...will essentially dismiss all results of scientific experimentation. Unless we proceed under the assumption that hidden conspiracies of this sort do not occur, we have abandoned in advance the whole enterprise of discovering the laws of nature by experimentation. [8]

Science, that is, relies on the Statistical Independence assumption, and science unquestionably works.

This practical case for Statistical Independence does not appear to me to rely on any interventionist perspective on the applicability of causal terminology to fundamental theories. Indeed, it is easy enough to understand what Statistical Independence means in the context of such theories. For the randomised drug trial, it means that the specific facts, in the distant past, which determine the precise sequence of coin flips (used to assign placebo or drug to each patient) should not be conspiratorially correlated with the specific facts which determine whether those individual patients will or will not die of some unrelated malady. Similarly, in the EPR-Bell case, Statistical Independence means that the specific facts which determine the sequence of measurement settings a and b (e.g., the blueberries) should not be conspiratorially correlated with the specific facts which influence the sequence of particle pair states λ (e.g., the windstorm).

Of course, Huw, you will want to say I am begging the question here by inserting the word "conspiratorially". Your whole point is that, if the settings retro-causally influence the pair states, the needed correlations between (on the one hand) a and b and (on the other hand) λ would not need to be conspiratorial at all.

By way of concluding, let me try to summarise the several reasons I have explored, throughout this dialogue, for being sceptical of this project.

First, I am sceptical that, at the level of a fundamental candidate theory (i.e., without "interventions" and "exogenous" variables) the idea of retro-causation, of the rather special sort you want and need, even makes sense. (When I look at extant candidate fundamental theories, at least the time-symmetric deterministic ones we have available, the idea of assigning specific temporally-oriented causal arrows to specific sub-processes seems arbitrary and groundless. At the fundamental level, the theories just assert necessary connections between states at different times).

Second, I am sceptical of the claim that, if you did somehow produce a candidate fundamental theory that somehow made your desired sort of violation of SI compelling, the theory would respect some meaningful "Einstein Locality" or Lorentz Invariance condition. (It seems more likely to me that, without the terminological restraints introduced by interventionism, your "indirect" zig-zag causality would end up being ubiquitous and the theory would end up being wantonly non-local and blatantly incompatible with relativity).

The new, third, grounds for scepticism that I have been trying to raise here has to do with the fact that, even in a (deterministic) candidate fundamental theory in which it somehow made sense to speak of the settings a and b retro-causally influencing the pair states λ, something still determines the settings, and some other factors, besides a and b, will presumably still have to have some influence on λ. And these influences, traced further backwards in time, will imply some very particular and special correlations, in the earlier states, which, at least from this abstract perspective, do not seem different from—and so do not seem any less conspiratorial than—the correlations posited by the CPH.

A concrete candidate theory could, in principle, cleanly refute any or all of the grounds for scepticism I have expressed here. However, so far I have never seen a candidate retro-causal theory with the appropriate "pretension to physical precision". I have instead only seen various sorts of toy models that focus on some narrow system, postulate "interventions" by treating certain variables as "exogenous", and (hence) suffer from a kind of "measurement problem". I have not, that is, seen anything that refutes my scepticism, so I remain unconvinced that the retro-causal program generally, and the distinction between the CPH and CFH in particular, can survive the translation to a more serious, candidate fundamental theory.

Was it similar reasoning that led Bell, in the letter reproduced in Figure 1, to dismiss your retro-causal project? I am not sure. If anything, it seems more likely to me that he just did not appreciate that you wanted to violate Statistical Independence in a non-standard and purportedly non-conspiratorial way. But I think it is at least possible that he, as a consistent champion of the need for fundamental theories which avoid the

measurement problem, recognised that, from this fundamental point of view, the standard and non-standard ways of violating SI—that is, the CPH and CFH—are very difficult to even distinguish.

That, at any rate, is where I, having been profoundly influenced by Bell, end up. At least for the time being.

8. Price (IV)

Travis, it is a little frustrating to close at this point, when your latest comments raise several points for further discussion. However, wrap up we must, so I will confine these closing comments to just one issue, that of 'conspiracy'. Before that, I want to say a very warm thanks to you for taking this on. As you know, the idea for this dialogue originated in an exchange of Facebook comments, after I shared my letter from Bell there. Some fascinating discussion takes place on Facebook these days, but it has been a pleasure, as well as an education, to do this the old, slow way ('Too swiftly now the Hours take flight!/What's read at morn is dead at night', as Austin Dobson—a man well ahead of his time!—noted already in the 1880s [45] (233)).

Now to conspiracies. Helpfully, you suggest an answer to your own objection:

> [Y]ou will want to say I am begging the question here by inserting the word "conspiratorially". Your whole point is that, if the settings retro-causally influence the pair states, the needed correlations between (on the one hand) a and b and (on the other hand) λ would not need to be conspiratorial at all.

That would do, but I want to say a bit more. For one thing, I want to make sure it is clear why I take CFH to be much less vulnerable to the conspiracies charge in the first place.

I introduced the term 'conspiracy' into our discussion in §2. It occurs there in the quote from Shimony, Clauser, and Horne [8] that you quote again. But I did not use the term myself in distinguishing CPH and CFH. For me, the crucial difference was simply that unlike CPH, CFH treats measurement settings in the normal way, as exogenous variables. I noted that this explains how CFH escapes the objection that these authors and many since have raised for CPH, that it is incompatible with 'the whole enterprise of discovering the laws of nature by experimentation'. However, a few paragraphs later I added that I had not dealt with the suggestion that CFH also requires something 'conspiratorial'. Let me now come back to that, and explain why it is a very different issue from the one that confronts CPH.

In your account of the conspiracies objection, you note that physicists (notoriously playful folk) can easily put measurement settings under the control of such things as the number of blueberries in a pancake. And you point out—rightly, if you are talking about CPH—that this means that any theory trying to break Statistical Independence is going to have to concern itself with blueberries. Just look at the two (fortuitously) blue arrows in Figure 2. If there are blueberries on the causal chains that those arrows represent, the common cause needs to control the blueberries, along with everything else in the chain. Quite an ask.

However now look at Figure 3, depicting CFH. Here, the guts of the retrocausal proposal is some sort of lawlike constraint, correlating the measurement setting a with properties λ of the lefthand particle along the first black arrow. (The direction of the arrow just reflects the fact that a is an exogenous variable, on which interventions are possible.) Once we have these guts, the story about the blueberries follows for free. Blueberries are simply one among the endless ways that ingenious experimenters can devise to control the value of an exogenous variable (i.e., to provide the green arrow in Figure 3). We do not need anything novel or conspiratorial to control them. By way of comparison, imagine someone puzzled about how the buttons on a remote handset control a television. There is no additional mystery about how blueberries, too (and everything in the past on which they themselves depend), can control the television, if we hook them up to the handset.

To take this analogy a little further, we could imagine blueberries being used as a source of (effectively) random experimental interventions, in a project to test the hypothesis

that settings of the handset have a causal influence on the television. In that project it would be absurd, of course, to assume statistical independence between the handset settings and the state of the television—what we would be looking for would be, precisely, violations of such independence. This trivial example shows that science does not assume such Statistical Independence everywhere, and that it is indeed begging the question against CFH to assume it in the form of the principle we have been calling Statistical Independence.

So CFH is far less vulnerable to the charge of conspiracy than CPH. If there is something seemingly conspiratorial in CFH, it is the fact that the particles 'already know' the measurement settings, in some sense, before they arrive at the measurement devices. This certainly looks strange to ordinary intuitions, and words such as 'conspiratorial', or 'teleological', do something to capture this strangeness.

If someone objects to CFH on this basis I am inclined to accuse them of a temporal double standard—in other words, as I used the term in [10], applying different principles in one direction of time to the other, without offering any justification for the difference. This kind of 'conspiratorial' behaviour is exceedingly common in the direction from future to past, thanks to the thermodynamic asymmetry. Think of all the time-reversed videos you have ever seen of omelettes transforming themselves into unbroken eggs, and similar things. Many writers take that kind of behaviour to be explained by a lawlike constraint in the past, the so-called Past Hypothesis.

I am not suggesting that CFH involves a time-reversed version of the familiar thermodynamic asymmetry. However, if anyone objects to this remaining sense in which CFH looks conspiratorial, I will ask them whether they would still object to the time-reversed version of the same theory, in which a particle and a measuring device are correlated in virtue of a lawlike constraint on an interaction in their common past. If not, I will accuse them of a temporal double standard. That is close to the charge of begging the question, but a bit more specific in a useful way (I discuss these issues at length in [10] [Ch. 5]).

Obviously there is more to say here. For one thing, I have not said anything about your concern that CFH requires λ to be controlled from two directions. Indeed, reading through the whole thing again, it feels like we are just warming up. But for the moment, thanks again, Travis—and I hope this will not be our last opportunity to discuss these questions. Perhaps this dialogue will encourage others to weigh in on one side or other, and take up open issues.

9. Norsen (IV)

I share the sense that we are only now getting warmed up. However I think, rather than be frustrated to wrap up just when it feels like we are finally getting to the heart of the issues, we should appreciate the progress that getting to this point represents. And of course the end of this particular dialogue is not the end of dialogue as such. Our illuminating and highly enjoyable discussion here can hopefully generate and influence further constructive discussions in the future (and, not to rudely ignore the possibility of retro-causation, perhaps also in the past).

By way of wrapping up, I will just touch briefly on two points, making a special effort not to unfairly lob any new rhetorical grenades. In particular, I want to flag, for the purposes of future discussion, a point where I think we may have a substantive and unresolved disagreement. And then I will attempt to clarify my (admittedly potentially misleading) use of the word "conspiracy" in my previous contribution.

So, first, the flag. You said that my argument "that any theory trying to break Statistical Independence is going to have to concern itself with blueberries" would apply "if you are talking about CPH". I think you meant to imply that this argument does not apply in the context of CFH. But the whole point of my admittedly silly parable involving blueberries and windstorms, is precisely that it should still apply, either way.

You wrote that, for you, "the crucial difference [between CPH and CFH] was simply that unlike CPH, CFH treats measurement settings in the normal way, as exogenous variables". But perhaps the fundamental theme running through everything I have written here

is that everything we might want to say about the EPR-Bell scenario needs be compatible with the perspective of a candidate fundamental theory which is free of "measurement problem" issues and in which, in particular, there are no exogenous variables.

Therefore, if the reason you think that the CFH is not "going to have to concern itself with blueberries" is that the CFH by definition treats measurement settings as exogenous (and is hence freed from the responsibility of considering the factors that, in a real-world implementation of this setup, would in fact influence/determine the settings), then I would regard that as a fatal flaw in the CFH program.

As always, though, I suspect there may be some mutual misunderstanding that makes our differences appear greater than they really are. In particular, your reference back to the (indeed, fortuitously) blue arrows in Figure 2 gives me pause. You say: "If there are blueberries on the causal chains that those arrows represent, the common cause needs to control the blueberries, along with everything else in the chain. Quite an ask!"

But I believe the way you are thinking of Statistical Indepedence being violated here, in the context of the CPH, is not what I had in mind. It is of course true that Statistical Independence could easily be violated if, instead of the particle pair source spitting out pairs in a pre-defined order determined by the precise moment and location of the windstorm, the pair source instead spits out a sequence of particles in states that are, say, determined by the output of the very same random number generator (seeded by the blueberry count) that is also determining the settings. This sort of arrangement would seem to be the sort of thing you had in mind (in which the blueberries, the pair states λ, and the settings a and b are all on the same future-directed causal chain. But there would be nothing remotely conspiratorial about the failure of Statistical Independence in this kind of case. We would just blame the failure on an exceedingly stupid experimental design by the experimenters.

What I had in mind—what I thought we agreed would count as unacceptably conspiratorial—is instead the possibility that Statistical Independence could be violated (still leaving aside the possibility of retro-causation, i.e., working still in the framework of the CPH) with a more sensible experimental design, in which the causal chain leading up to the pair states λ and the chain leading up to the settings a and b, have no apparent connection. That was the point of the parable, with the windstorm affecting the one thing and the (causally disconnected) blueberries affecting the other. To me, the interesting question is whether Statistical Independence might still be violated in this kind of case. And of course it might be. But its being violated (with the non-stupid experimental design) would require certain subtle details in the kitchen of the diner to be correlated, just so, with certain subtle details of the weather pattern across town. The usual attitude, though, is that there is no reason those details should be correlated. So positing that they are—that is, rejecting Statistical Independence—amounts to asserting something that has a highly implausible, "conspiratorial" feel to it.

So much for the flag. Now to the clarification.

The point I was making previously is that, by considering the CFH (in which the settings a and b retro-causally influence the pair states λ), we in no way remove the correlation (between those subtle details in the diner and other subtle details in the weather pattern across town) that we would be committed to with the CPH. I think in the end you agreed with this; at least, that is how I understand your comment that "the story about the blueberries follows for free" (but you see, I am confused, because you also previously insisted that the CFH requires taking the settings as exogenous). Assuming this is correct, though, I think you then just want to object to my characterisation of the correlations as "conspiratorial". They would be (as I thought we agreed) in the CPH, because in that context there is no reason for them. However, I think you want to say, these same correlations are not at all "conspiratorial" in the CFH because the causal connection between the future end of the one chain (the settings) and the future end of the other chain (the pair states) connects the two otherwise-unconnected chains, and thus provides a perfectly good reason for the subtle details (on the past ends of the two, now-connected, chains) to be correlated. Indeed, as I think you want to say, once we allow retro-causation, those correlations are no

weirder, no more conspiratorial, than the future correlations which we already allow must exist between subtle details of things which have interacted in the past.

If you accept this as a fair summary of your view, Huw, I am happy to just concede all of it. I should not have continued to use the inflammatory word "conspiratorial" to describe the correlations (between subtle details of things like blueberries and windstorms in the past) that both CPH and CFH are, I think, committed to, when discussing those correlations from the point of view of the CFH. The correlations are the same, whether we adopt the CPH or the CFH, but their conspiratorial-ness is not the same.

However my intention was not to beg the question. Partly I developed the parable and pushed this point just because I think it is important to acknowledge that violating Statistical Independence implies these sorts of correlations, regardless of how exactly one does it. (Treating measurement settings as exogenous variables tends to obscure this fact).

In addition, just because these sorts of correlations (in so far as they arise via the CFH) should not be described (and subsequently dismissed) as "conspiratorial", does not mean they are beyond reproach. For example, as I think you hinted at, their existence raises questions about how to reconcile the posited retro-causation with the statistical/thermodynamic arrow of time. And I have to admit that I feel a vague additional sense of unease about such correlations, stretching, as they would clearly have to, all the way back to the big bang. I have a hard time putting my finger on the reason for this queasiness. Of course, one possibility is that, try as I might, I can not quite get myself to take retro-causation fully seriously, so, in my gut, I respond just as I would to the conspiratorial character that those same correlations would have in a theory without retro-causation.

I do think there is more to it than that, however, and what I meant to gesture toward before was the thought that maybe this could explain what Bell wrote in his letter to you. If "fatalism" is the uncomfortable idea that our apparent freedom (or even just apparent randomness) is illusory—that what we choose to (or just happen to) do has, in fact, been pre-written in a script going all the way back to the big bang—then is there not a kind of time-reversed fatalism inherent in your CFH? For me at least, and perhaps for Bell, the idea that my apparently free choices (for example, about how to seed the random number generator that controls the settings in a Bell experiment) end up causing incomprehensibly subtle correlations in the early universe, is no less uncomfortable than the idea that incomprehensibly subtle correlations in the early universe are the true causes of my apparently free choices.

However, is there more to this discomfort than a lingering bias against retro-causation? Answering that will, I think, require further reflection, discussion, and debate. I look forward to that, and I thank you again for inviting me to participate in this.

Author Contributions: The authors contributed equally to this piece. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: H.P. thanks Reinhold Bertlmann for advice concerning the publication of his letter from J. S. Bell, and Ken Wharton for comments on drafts.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Price, H. An Assumption in the Interpretation of Quantum Mechanics. 1978. Available online: philsci-archive.pitt.edu/id/eprint/18373 (accessed on 14 February 2021).
2. Harrigan, N.; Spekkens, R.W. Einstein, incompleteness, and the epistemic view of quantum states. *Found. Phys.* **2010**, *40*, 125–157. [CrossRef]
3. Leifer, M. Is the quantum state real? An extended review of ψ-ontology theorems. *Quanta* **2014**, *3*, 67–155. [CrossRef]

4. Dummett, M. Can an effect precede its cause? *Proc. Aristot. Soc. Suppl. Vol.* **1954**, *38*, 27–44. [CrossRef]
5. Dummett, M. Bringing about the past. *Philos. Rev.* **1964**, *73*, 338–59. [CrossRef]
6. Price, H. The philosophy and physics of affecting the past. *Synthese* **1984**, *16*, 299–323. [CrossRef]
7. Price, H. A neglected route to realism about quantum mechanics. *Mind* **1994**, *103*, 303–336. [CrossRef]
8. Bell, J.; Clauser, J.; Horne, M.; Shimony, A. An exchange on local beables. *Dialectica* **1985**, *39*, 85–96.
9. Price, H. Locality, independence and the pro-liberty Bell. *arXiv* **1995**, arXiv:quant-ph/9602020.
10. Price, H. *Time's Arrow and Archimedes' Point*; Oxford University Press: New York, NY, USA, 1996.
11. Hitchcock, C. Causal models. In *The Stanford Encyclopedia of Philosophy*, Summer 2020 ed.; Zalta, E., Ed.. Available online: plato.stanford.edu/archives/sum2020/entries/causal-models/ (accessed on 14 February 2021).
12. Wiseman, H. From Einstein's Theorem to Bell's Theorem: a history of quantum nonlocality. *Contemp. Phys.* **2006**, *47*, 79–88. [CrossRef]
13. Leibniz, G.W. *Theodicy: Essays on the Goodness of God, the Freedom of Man and the Origin of Evil;* Farrer, A., Ed.; Routledge & Kegan Paul: London, UK, 1951.
14. Giustina, M.; Versteegh, M.A.; Wengerowsky, S.; Handsteiner, J.; Hochrainer, A.; Phelan, K.; Steinlechner, F.; Kofler, J.; Larsson, J.-A.; Abellán, C.; et al. Significant-loophole-free test of Bell's Theorem with entangled photons. *PRL* **2015**, *115*, 250401. [CrossRef]
15. Handsteiner, J.; Friedman, A.S.; Rauch, D.; Gallicchio, J.; Liu, B.; Hosp, H.; Kofler, J.; Bricher, D.; Fink, M.; Leung, C.; et al. Cosmic Bell test: Measurement settings from Milky Way stars. *Phys. Rev. Lett.* **2017**, *118*, 060401. [CrossRef]
16. Bell, J. *Speakable and Unspeakable in Quantum Mechanics*, 2nd ed.; Cambridge University Press: Cambridge, UK, 2004.
17. Friederich, S.; Evans, P. Retrocausality in quantum mechanics. In *The Stanford Encyclopedia of Philosophy*, Summer 2019 ed.; Zalta, E., Ed.. Available online: plato.stanford.edu/archives/sum2019/entries/qm-retrocausality/ (accessed on 14 February 2021).
18. Wharton, K.; Argaman, N. Bell's Theorem and locally-mediated reformulations of quantum mechanics. *Rev. Mod. Phys.* **2020**, *92*, 21002. [CrossRef]
19. Price, H. Does time-symmetry imply retrocausality? How the quantum world says "maybe". *Stud. Hist. Philos. Mod. Phys.* **2012**, *43*, 75–83. [CrossRef]
20. Leifer, M.; Pusey, M. Is a time symmetric interpretation of quantum theory possible without retrocausality? *Proc. R. Soc. A* **2017**, *473*, 20160607. [CrossRef] [PubMed]
21. Leifer, M. Time Symmetric Quantum Theory Without Retrocausality? A Reply to Tim Maudlin. *arXiv* **2017**, arXiv:1708.04364.
22. Price, H.; Weslake, B. The time-asymmetry of causation. In *The Oxford Handbook of Causation*; Beebee, H., Hitchcock, C., Menzies, P., Eds.; Oxford University Press: Oxford, UK, 2010; pp. 414–443.
23. Albert, D.Z. *Quantum Mechanics and Experience*; Harvard University Press: Cambridge, MA, USA, 1992.
24. Norsen, T.; John, S. Bell's concept of local causality. *Am. J. Phys.* **2011**, *79*, 1261–1275. [CrossRef]
25. Goldstein, S.; Norsen, T.; Tausk, D.V.; Zanghi, N. Bell's Theorem. *Scholarpedia* **2011**, *6*, 8378. [CrossRef]
26. Norsen, T. *Foundations of Quantum Mechanics: An Exploration of the Physical Meaning of Quantum Theory*; Springer: Cham, Switzerland, 2017.
27. Russell, B. On the notion of cause. *Proc. Aristot. Soc. New Ser.* **1913**, *13*, 1–26. [CrossRef]
28. Pearl, J. *Causality*; Cambridge University Press: New York, NY, USA, 2009.
29. Suppes, P. *A Probabilistic Theory of Causality*; North-Holland: Amsterdam, The Netherlands, 1970.
30. Ramsey, F.P. General propositions and causality. In *Foundations: Essays in Philosophy, Logic, Mathematics and Economics*; Mellor, D.H., Ed.; Routledge and Kegan Paul: London, UK, 1978; pp. 133–151.
31. Woodward, J. *Making Things Happen: A Theory of Causal Explanation*; Oxford University Press: Oxford, UK, 2003.
32. de Beauregard, O.C. Méchanique quantique. *C. R. Acad. Sci.* **1953**, *236*, 1632–1634.
33. Evans, P.; Price, H.; Wharton, K. New slant on the EPR-Bell experiment. *Br. J. Philos. Sci.* **2013**, *64*, 297–324. [CrossRef]
34. Aharonov, Y.; Vaidman, L. The Two-State Vector Formalism of Quantum Mechanics: An Updated Review. In *Time in Quantum Mechanics*; Muga, J.G., Sala Mayato, R., Egusquiza, Í.L., Eds.; Springer: Berlin, Germany, 2008; pp. 399–447.
35. Goldstein, S.; Tumulka, R. Opposite arrows of time can reconcile relativity and nonlocality. *Class. Quantum Gravity* **2003**, *20*, 557–564. [CrossRef]
36. Sutherland, R. Causally symmetric Bohm model. *Stud. Hist. Philos. Mod. Phys.* **2008**, *39*, 782–805. [CrossRef]
37. Wharton, K. The universe is not a computer. In *Questioning the Foundations of Physics: Which of Our Fundamental Assumptions Are Wrong?* Aguirre, A., Foster, B., Merali, Z., Eds.; Springer: Cham, Switzerland, 2015; pp. 177–189.
38. Palmer, T. Bell's conspiracy, Schrödinger's black cat and global invariant sets. *Philos. Trans. A* **2015**, 373, 20140246. [CrossRef] [PubMed]
39. Adlam, E. Quantum Mechanics and Global Determinism. *Quanta* **2018**, *7*, 40–53. [CrossRef]
40. Adlam, E. The Operational Choi-Jamiolkowski Isomorphism. *Entropy* **2020**, *22*, 1063. [CrossRef] [PubMed]
41. Becker, A. Quantum time machine: How the future can change what happens now. *New Scientist*, 14 February 2018.
42. Dürr, D.; Goldstein, S.; Münch-Berndl, K.; Zanghi, N. Hypersurface Bohm-Dirac Models. *Phys. Rev.* **1999**, *60*, 4. [CrossRef]
43. Dürr, D.; Goldstein, S.; Norsen, T.; Struyve, W.; Zanghi, N. Can Bohmian Mechanics be made Relativistic? *Proc. R. Soc. A* **2014**, *470*, 20130699. [CrossRef]

44. Tumulka, R. A Relativistic Version of the Ghirardi-Rimini-Weber Model. *J. Stat. Phys.* **2006**, *125*, 821–840. [CrossRef]
45. Dobson, A. *At the Sign of the Lyre*; Kegan Paul, Trench & Co.: London, UK, 1889.