Retrocausality at no extra cost

P. W. Evans

Centre for Time, Department of Philosophy

The University of Sydney

May 19, 2010

Abstract

One obstacle faced by proposals of retrocausal influences in quantum mechanics is the perceived high conceptual cost of making such a proposal. I assemble here a metaphysical picture consistent with the possibility of retrocausality and not precluded by the known physical structure of our reality. I conclude that given the right mix of some reasonable metaphysical and epistemological ingredients there is no conceptual cost to such a picture.

Key words: Retrocausality, temporal symmetry, quantum mechanics

1 Introduction

It is no secret that by permitting retrocausal influences in quantum mechanics local hidden variables can be used to account for the violation of Bell's Inequality.[1] But can we rest easy knowing that such a retrocausal picture is *metaphysically tenable*? The metaphysics of retrocausality is often broached in the philosophical literature in and around discussions of time travel and causal paradoxes and there seems to be a general sentiment that there is nothing *manifestly self-contradictory* about the idea, strange though it may seem at first. There is, however, a significant challenge from the philosophy of physics literature: Maudlin (2002) claims that retrocausality is fundamentally at odds with the "metaphysical picture of the past generating the future" and thus cannot be entertained as a metaphysical possibility in a reality such as ours. The plausibility of Maudlin's metaphysical picture will not be of concern to us here.[2] The purpose of this paper is to counterbalance Maudlin's picture with a carefully considered metaphysical alternative that coheres with the possibility of retrocausality and that is not precluded by the known physical structure of our reality.

[1]See, for instance, Argaman (2008), Costa de Beauregard (1953; 1976; 1977), Cramer (1980; 1986), Hokkyo (1988), Miller (1996; 1997), Price (1984; 1994; 1996; 1997; 2001; 2008; 2010), Rietdijk (1978), Sutherland (1983; 1998; 2008) and Wharton (2007; 2010).

[2]Though see Evans, Price and Wharton (2010) for some discussion of this point.

This project does not introduce any explicitly new metaphysical ideas; on the contrary, the picture developed here is a conglomeration of developed material from various contexts that is merely collected together under the one roof. I begin by setting out in Section 2 two relatively uncontroversial positions that will serve as a solid conceptual foundation upon which to develop our metaphysical picture: the block universe model of time in Section 2.1 and the interventionist account of causation in Section 2.2. There are then two metaphysical intuitions that must be dismantled. The first is our ordinary asymmetric causal intuition: in Section 3.1 I describe an argument from temporal symmetry against the plausibility of extending our asymmetric causal intuitions to the microscopic realm. The second is our ordinary intuition about epistemic access to the past: in Section 3.2 I present an argument that clears a logical space for retrocausation at the expense of our intuition that our past is necessarily epistemically accessible independent of our own future actions. This then clears the way to build a symmetric picture of causation: Section 4 sets forth a model of agent deliberation that reconciles our notion of free will and unidirectional causation with a deterministic and causally symmetric metaphysical picture.

It is often claimed that the introduction of retrocausal influences into the interpretation of quantum theory is a higher conceptual cost to pay than the problems associated with the rejection of local hidden variables. The metaphysical picture assembled here is an attempt to show that there is no conceptual cost at all associated with retrocausality.

2 Foundations

2.1 The block universe model of time

We begin building our metaphysical picture with a temporal model popular among many physicists and philosophers: the block universe model of time. Rather than modelling reality as a three dimensional space evolving under the passage of time, reality is envisaged according to this view as a four dimensional block of which time is a mere passive ingredient. In the philosophical literature the block universe view is thought of as characterised by two claims: the past, present and future are equally real; and there is no privileged instant nor objective flow of time through the block. The spatial and temporal relations between all events in the four dimensional block are thus on an equal footing and exist atemporally. This view thus forges a strong analogy between the conception of time and our ordinary conception of space. Just as there is nothing objective about labelling a particular position in space 'here' nor claiming the contents of 'here' to be more real than the contents of 'there', there is nothing objective about labelling a particular time 'now' whose contents can be thought of as any more real than the contents of any other position within the block.

The block universe model of time is consistent with a deterministic model of reality. In a deterministic physical model, specifying data along a single hypersurface of spacetime is sufficient to determine the events of spacetime past and future of the hypersurface. Similarly, the block universe view is committed to the reality of all the events of four dimensional spacetime.

2.2 The interventionist account of causation

The interventionist account of causation is introduced and defended in Woodward (2003). According to this account, we say that C is a cause of E just in case there is some possible intervention that can be carried out on C that will change E in some way or other, holding fixed all other properties of the system containing C and E. Woodward's account is explicitly counterfactual in the sense that there need be only some *possible* intervention that can be made on C to bring about a change in E. The advantage of this account is that it can be utilised in providing causal explanations without requiring that there exist a complete description of some spatiotemporal process connecting C and E. Let us consider an illustrative example.

Imagine the ignition system of a car. It seems that we would want to say that the turning of the key in the ignition (event K) is the cause of the starting of the car's engine (event E). According to the interventionist account we can say that K is indeed the cause of E since it is possible to carry out an intervention on K, by not turning the key say, that will change E in some way or other, in this case the engine would simply not start, provided all the other elements contributing to the system were held fixed. We can in fact claim a causal connection here without explicitly spelling out the mechanism by which the turning of the key brought about the starting of the engine. However, this does not mean that we cannot spell out such a mechanism if we wished.

Consider the mechanical chain of events connecting the turning of the key to the starting of the engine: turning the key (event K) completes the circuit between the car's battery and the starter motor (event C) which then starts the starter motor spinning (event S); the spinning starter motor then turns over the drive shaft of the engine (event D) which starts the pistons drawing in and then combusting the fuel (event P); the combusting fuel powers the engine to start running (event E). We have a chain of events, $K \rightarrow C \rightarrow S \rightarrow D \rightarrow P \rightarrow E$, with a mechanical account of how each event brings about the next. However, the content of any causal claim about any two of these events according to the interventionist account of causation is not that there exists a mechanical connection between the events. The key to the interventionist account is to imagine that each of these events is a handle or variable that can be manipulated and controlled. Accordingly, what makes each event the cause of the next is the fact that there exists a functional dependency between the variables; that is, some possible intervention on a particular variable will (over a range of conditions) bring about a consistent change in the values of the variables further down the chain. If we were to intervene on the above system by replacing the battery with an old or faulty battery, the starter motor would fail to spin, thus changing the value of the variable associated with event S (on or off, say) from what it would have been had we not made the intervention.

The chain of events may be more complex than the example above; events might have multiple causes or multiple effects. We can extend our example by imagining that event K, the turning of the key, also completes the circuit between the car battery and the dashboard, lighting up all the instruments inside the cabin (event I). We could also imagine that in order for the drive shaft to turn over (event D), the car's gearbox must be disengaged (event G). We can establish which correlations are causal by imagining possible interventions of these new variables while holding the rest of our variables fixed. If we sever the connection between the car battery and the dashboard, the battery would still connect to the starter motor bringing about the ignition of the car's engine. Thus, the dashboard lighting up is not a cause of the car's engine starting, even though these events are very often correlated. However, if we engage the gearbox with the drive shaft and attempt to start the car, we find that the car stalls. Whether the gearbox is engaged or not, i.e. whether the variable associated with event G is one value or another, has a functional relationship to whether the car starts or not and is thus a cause of event E.

There is one further issue which arises from this account of causation that will be crucial to our characterisation of retrocausality later in this chapter. It will be beneficial for our purposes here to view the interventionist account of causation as a kind of genealogical account of how we, as agents, come to acquire the concept of causation in relation to the world around us. To begin demonstrating how this might be the case, consider the way we might use causal concepts to describe a situation in which it is impossible for us as humans to intervene on the system. The gravitational pull of the moon is responsible for the tides of the ocean, and we would want to say that the moon *causes* the tides to come in and out as they do. Even though it is impossible for us to actually manipulate and control this system, we can attribute our causal intuitions in this sort of case to an ability to extend our causal intuitions from cases in which we can manipulate and control. Through our knowledge of the gravitational interaction between the moon and the tides, we can predict with confidence what the effect of some imagined (but perhaps physically impossible) intervention would be if we could in fact bring it about. It is this sort of knowledge which we usually gain by physical intervention and experimentation that allows us to make claims about the causal relations that exist within a system. Thus, it seems reasonable that we extend these causal notions to cases in which we do not in fact have the requisite ability to manipulate and control.

I mention this feature of the interventionist account here to highlight the fact that a consequence of this view is that our role as agents in the world can be seen as at the root of our concept of causation. We will take this idea up again below where we will be in a position to expand on it in more depth. For now I simply wish tentatively to broach the outcome of this genealogical sketch that a being which interacted with the world differently to us as agents would have a very different concept of causation to the one that we have.

With these metaphysical foundations in mind, let us now move on to dismantling two of our ordinary temporal intuitions.

3 Dismantling intuitions

3.1 Macroscopic intuitions, microscopic symmetry

A familiar intuition, indeed one that seems almost trivial, is that the properties of interacting systems are independent before they interact. This is built upon the observation of many instances where this apparent principle holds true. In macroscopic systems we take this principle for granted. However, Price (1996; 1997) ask the question whether we are justified in extrapolating this familiar macroscopic principle to considerations of microscopic systems. Let us consider Price's analysis.

Firstly, it seems that the origin of this principle is related to the asymmetry of thermodynamics. When systems evolve from states of disequilibrium (lower entropy) to states of equilibrium (higher entropy) it is because the initial conditions are special; namely, the initial conditions are low entropy. Thus, if we were to consider a macroscopic system evolving in the reverse temporal direction, it would look strange because it would appear that highly correlated *incoming* influences were converging from disparate regions of space (imagine a pile of rubble 'un-collapsing' into a building) and these would be associated with a decrease in entropy. In such a case the violation of the principle that physical processes are uncorrelated before they interact would be a direct product of the violation of the second law of thermodynamics. Looking forwards in time again we can see that the temporal asymmetry which manifests itself in the correlations between *outgoing* influences is a result of special (low entropy) initial conditions and not a result of an inherent asymmetry within the laws of physics.

It appears to be assumed that this principle of outgoing correlations but incoming independence holds in the microscopic case just like in the macroscopic case. However, explaining this temporal asymmetry of microscopic systems in terms of boundary conditions simply does not work. The boundary conditions explanation is based upon the temporal asymmetry of entropy change. In a microscopic system, such as that of two particles which come together, interact and then separate, there is no entropy change to indicate a temporal orientation to the interaction. The temporal reverse of the interaction would look much the same as in the ordinary temporal direction; this is a function of the temporal symmetry of the dynamical laws of the system. Thus, there seems to be no reason to assume that outgoing correlations exist in one direction and not in the other. Furthermore, unlike in the macroscopic case, there is no *observed* asymmetry in microscopic systems that needs to be explained. We simply do not observe the independence of incoming particles nor the correlation of outgoing particles, yet we still assume that this principle holds for microscopic systems despite its incompatibility with the temporally symmetric nature of the dynamical laws of physical systems.

Therefore we are left with a dichotomy between two physical principles at the microscopic level: temporal symmetry in the dynamical laws of physical systems on the one hand and, on the other hand, the asymmetry of the independence of microscopic systems prior to interacting with each other. As such, one could make quite a persuasive argument against the independence of microscopic systems prior to interaction purely on symmetry grounds. Moreover, with no existing observational evidence in favour of the independence of incoming particles or the correlation of outgoing particles, it seems that such a principe may not deserve the status it currently enjoys in considerations of microscopic systems. If this principle is then abandoned, one is led to the conclusion that temporally symmetric causation in microscopic systems cannot be ruled out on analytic grounds. Thus if we take these considerations seriously then the physical structure of our reality does not preclude a metaphysical picture allowing such retrocausal influences.

3.2 The bilking argument

In our normal conception of causation, causes precede their effects. A causally symmetric viewpoint opens up the possibility that effects can precede their causes. This, however, immediately creates some conceptual difficulties. To demonstrate these difficulties, let us imagine a pair of events which we believe to be causally connected: a cause, C, and an effect, E. Let us further imagine that this connection is retrocausal; E occurs earlier in time than C. On first appearances it would then seem possible to devise an experiment which could confirm whether our belief in the causal connection is correct or not. Namely, once we had observed that E had occurred, we could then set about ensuring that C does not occur, thereby breaking any retrocausal connection that could have existed between them. If we were successful in doing this, then we would have *bilked* the effect of its cause. This is the bilking argument.

The bilking argument seems to drive one towards the claim that any belief an agent might hold in the positive correlation between event C and event E is simply false. If this were the case then the agent would have to give up any belief in retrocausal influences between C and E. Dummett (1964) disputes that giving up this belief is the only solution to the bilking argument. In exploring the terms under which a belief in retrocausation can be maintained, Dummett suggests that what the bilking argument actually shows is that a set of three conditions concerning the two events, and the agent's relationship to them, is incoherent. In any incoherent set of conditions, all three conditions cannot hold simultaneously. Thus, depending on which of these three conditions fails to hold, there may be scope for an agent to maintain a belief that the later cause retrocausally influences the earlier event. To motivate these conditions, let us consider Dummett's own example.

Dummett imagines a tribe to exist with the custom of sending young men on a lion hunt to prove their bravery. The men travel for two days, hunt for two days and spend two days on their return journey. Observers travel with the young men and report back to the chief of the tribe whether the men acquitted themselves with bravery or not. While the young men are away, the chief performs dances intended to cause the young men to act bravely. Significantly, he performs these dances for the whole six days, i.e. for two days during which the events that the dancing is supposed to influence have already taken place. The chief notices that on occasions when he dances, he subsequently learns that the young men had hunted bravely and, on occasions when he does not dance, he subsequently learns that the young men had hunted in a cowardly fashion. The chief thus observes there to be a positive correlation between his dancing and the young men's bravery and therefore maintains a belief in retrocausation.

Imagine further that we are to convince the chief that this practice of his were absurd. We arrange that the observers who had accompanied the hunt return early and report to the chief whether or not the young men had acted bravely. We then set a bilking challenge to the chief to dance if and only if the young men had *not* acted bravely. There are two possible outcomes of this challenge. If the chief accepts this challenge and dances then he must concede that his dancing does not ensure the bravery of the young men. Alternatively, imagine that the chief accepts the challenge and then discovers he is inexplicably unable to dance, i.e. his limbs will simply not move. Then the chief would have to admit that dancing is not an action which is within his power to perform. If this were to occur, however, it would then be fair to say that it is not the chief's dancing that causes the young men to be brave, rather it is the young men's bravery that makes possible his dancing. Thus, regardless of whether the chief dances or not, it seems that the chief must give up his belief in retrocausation.

It appears then that there are two incompatible conditions here concerning the chief's dancing: (i) there is a positive correlation between the chiefs dancing and the bravery of the young men; and (ii) dancing is within the power of the chief to perform. If the first condition is to hold, then the second condition must fail, and vice versa, as we have just seen. Dummett, however, suggests that an implicit third condition can be violated which allows both of these conditions to hold simultaneously and thus allows the chief to maintain his belief in retrocausation. To see this, let us first consider an agent who believes a certain action is effective in bringing about a *subsequent* event. Such an agent would believe the action to be the cause of the later effect. Dummett recognises that there is a connection between the foreknowledge the agent possesses about the subsequent event and the intention the agent has to perform the action. The agent knows an event occurs in the future because they intend to bring it about by performing a certain action: the agent possesses knowledge in intention. This is in contrast to knowledge of the past which we can possess in more forms than merely in intention.

Let us then return to our example and imagine for the sake of argument that there is a parallel between the knowledge that the chief can possess concerning the bravery of the young men and the case of foreknowledge described here, i.e. the chief only knows that the young men are brave due to his intention to dance. This would then make our bilking challenge inconclusive. Since we can no longer arrange that the observers report the behaviour of the young men to the chief, we can no longer force the occurrence of a negative correlation. If we further rule out that there are no inexplicable incidents when the chief is unable to dance, then we are left with the original situation whereby the chief merely observes a positive correlation between his dancing and the young men's bravery and the chief can thus maintain his belief in retrocausation. To arrive at this result we have had to jettison the following condition: (iii) the chief has epistemic access to the behaviour of the young men independently of his intention to dance. These three conditions form a set which is shown to be inconsistent by the bilking argument.

Let us state these conditions in the more general terms we encountered at the beginning of this section.

(i) There exists a positive correlation between an event C and an event E.

(ii) Event C is within the power of an agent to perform.

(iii) The agent has epistemic access to the occurrence of event E independently of any intention to bring it about.

An interesting point to notice at this stage is that these conditions do not specify in which order events C and E occur. If we consider why it is not the case that it is possible to bilk *future* effects of their causes, this is because condition (iii) fails to hold for future events. If knowledge about future events could be obtained independently of an agent's intention to perform certain actions, then it would be possible to bilk those future events of their causes; this would amount, in a way, to changing the events we already know to occur in the future. Since this sort of foreknowledge is not possible, we can consistently believe our actions to *bring about* the future. Conversely, if it were the case that some past event was known only through our intention to perform a certain action, then it would be consistent to believe our actions to *bring about* the past.

The conditions under which it is possible to maintain a belief in retrocausation are especially relevant to quantum mechanics. In fact, once we make a suitable specification of how condition (iii) can be violated, we find that there exists a strong symmetry between the conditions which need to hold to justify a belief in bringing about the past and what we find to be the case in quantum mechanics. Following the prescription of Price (1996, p.174), let us not suppose that a violation of condition (iii) entails that the relevant agent has *no* epistemic access to the relevant past events independently of any intention to bring them about, rather let us suppose that the means by which knowledge of these past events is gathered breaks the claimed correlation between the agent's action and those past events. We can state our new condition as follows:

(iv) The agent can gain epistemic access to the occurrence of event E independently of any intention to bring it about and without altering event E from what it would have been had no epistemic access been gained.

In the dancing chief example a violation of this condition would entail that every time the chief attempted to discover the behaviour of the young men he subsequently affected their behaviour to be different from what it would have been had he not attempted his discovery. In those cases where the chief makes no attempt to discover the behaviour of the young men, we are back to our original violation of condition (iii).

The nature of this weakened violation of condition (iii) is just the sort of condition we would expect to hold if the system in question were a quantum system. As has been pointed out by Price (1996, p.174), according to Dummett's analysis of the bilking argument, quantum mechanics has exactly the sort of dynamics we would expect of a retrocausal physical theory. Thus we see again that the known physical structure of our reality does not preclude a metaphysical picture that allows retrocausal influences.

4 Keeping up appearances

Hopefully it is beginning to become clear what sort of limitations constrain the form of a metaphysical picture that allows retrocausal influences. We are now in a position to use these constraints, along with the causal and spatiotemporal structures we have taken to be most reasonable, to build a picture of what retrocausation actually involves. At the centre of this discussion will be the role that we play as agents as we interact with, and participate in, the world.

Let us start first and foremost with two conceptions of influence that are commonly conflated when talking about the future: the view that we *change* events and the view that we *affect* events. Consider a claim like the following: by deciding to catch the bus, I changed my day from one in which I was late for work, to one in which I was early. Regardless of one's model of time, there is an inconsistency in thinking that we change events through our actions. For an event to change, the event must have been a particular way in the first place. If we were partial to a dynamic view of time in which the future were *unreal*, it would make no sense to think of a future event as being any particular way before it is actual; there is simply no event that is my tardiness which can be changed before I am in fact late. However, we have explicitly signalled our intention to utilise the block universe model of time and in such a model we *can* speak of future events as being real and thus it might be possible for an event, one might say, to be a particular way *ab initio*. We might say that my tardiness was an event and that this event changed into my punctuality. But we must be careful here, because if a future event is real, it is in some sense already out there in the four dimensional block. If we *change* it at some point prior to it being a present event for us, we are left with the rather strange question: why was it as it was before we changed it? Why did the four dimensional block contain an event which was my tardiness, which then changed at some point into my punctuality? With respect to the block universe view, it is not even entirely clear that this question makes sense.

Before we get ourselves in a muddle, let us take a step back and see if we can clarify the above claim. We might do this by saying something like the following: when we say that we change a future event, we mean that we change it from being something that it *could* have been, say my tardiness, to something that it now actually *is*, say my punctuality. Expressing what we mean by change in counterfactual terms lets us sidestep the problems we encountered with the reality of the events under question. However, the notion we have ended up with by doing so has a significant causal ring to it (recall our characterisation of causation in terms of interventions); this is in fact just what we mean when we use the word 'affect'. I affect my day to be a day in which I am early for work, rather than a day in which I am late. I play a particular role in bringing about the future event and it is wrong to think that I change it from something that it already was. As long as we commit ourselves to the block universe view in which all events in the past, present and future are equally real, then we must think of influence in the 'affect' sense. Furthermore, we can now see that this argument is as much relevant to past events as it is relevant to future events. Under no circumstances does it make sense to change the past in any way, since one cannot change something that is already an actual event. Retrocausality is then not about changing the past, rather retrocausality is about affecting the past: playing a role in bringing about a past event.

This analysis is beginning to push us into a position about determinism and the nature of the block universe that may seem highly undesirable; namely, that we have no freedom in choosing our own actions. If we cannot change the future in just the same way that we cannot change the past, and if affectation is merely bringing about an event that in some sense already exists, then it would seem that we are mere spectators of our reality in a rather uninteresting sense. Fortunately, we are not pushed into this position by adopting typically block universe notions as above. Moreover, coming to grips with why this is the case will tie together many of the issues with which we have so far dealt and it will give us our first glimpse at the metaphysical picture of reality that allows for retrocausal influences. The solution to this seeming incompatibility between the conception of reality as a block universe and our ability as agents to control and manipulate our surroundings lies in thinking of causation as a perspectival notion. According to Price (2007), evidence suggests that causation is in fact a perspectival notion; indeed, we have already been introduced to the idea when we were considering the interventionist account of causation above. The tentative outcome that I flagged of what we called a kind of genealogical account of causation in terms of intervention was that a being who interacted with the world differently to how we interact with the world as agents would have a different concept of causation to the one that we have. Let us consider how we can use this to help us find some sort of compatibility between the block universe view and our causal intuitions.

The essential point to solving this problem is to realise that considering the block universe 'from the outside' is availing oneself of a very different perspective of the world to the one which we have while we are inhabiting some spatiotemporal region. The important difference between the two viewpoints is that there is a discrepancy between the parts of the spacetime block that are epistemically accessible from each perspective. The spatiotemporally constrained perspective by which we are bound permits us only limited epistemic accessibility to other spatiotemporal regions. This is significant because it is as spatiotemporally bound agents that we have evolved and it seems reasonable to suggest that we are in possession of a concept of causation that reflects this very fact. Once we imagine ourselves to be omniscient beings that have epistemic access to the whole spatiotemporal block, as we have done in the above analysis of change and affect, it should not come as a surprise that our causal intuitions get confused when we attempt to consider how a spatiotemporally bound agent can deliberate about whether or not to affect a particular event that is already determined from our imagined omniscient perspective. The solution that I am pushing towards here is that it is because we do not *know* which events are determined to occur that we can deliberate, and therefore be causal agents, at all. The relationship between deliberation and epistemic accessibility, and the role this plays in our concept of causation, has been explored in Price's (2007) and it will be important for our characterisation of retrocausality that we set out those details here.

Price (2007, p.20) sets out "an abstract characterisation of the structural, or functional architecture, of deliberation" with a view to separating out the intrinsic features of deliberation itself from those aspects of deliberation that are a function of our perspective as spatiotemporally bound agents. To begin with, a deliberator must be deliberating over whether to bring about some particular occurrence out of a range of possible occurrences. Following Price, we will call the set of events of which this range consists the OPTIONS that the deliberator is considering. The set OPTIONS can be thought of as consisting of two subsets: all those occurrences over which the deliberator has immediate control, the DIRECT OPTIONS, and all those occurrences that can be brought about indirectly via the DIRECT OPTIONS, the INDIRECT OPTIONS. All other events that are not under consideration during the deliberation we will call the FIXTURES. An integral subset of the FIXTURES is the set of events that the deliberator already knows, or are in principle knowable, at the time of deliberation which we will call the KNOWABLES. The KNOWABLES must be a subset of the FIXTURES since if these events are knowable to the deliberator at the time of deliberation, then they cannot be under consideration to be brought about and thus cannot be part of the set OPTIONS. For this reason, all the events in OPTIONS must fall into the set we will call UNKNOWABLES. Thus a deliberator makes two dichotomous distinctions: the distinction between FIXTURES and OPTIONS; and the distinction between KNOWABLES and UNKNOWABLES. The set KNOWABLES is a subset of FIXTURES and the set OPTIONS is a subset of UNKNOWABLES. Let us now consider how spatiotemporally bound deliberators such as ourselves might map these distinctions onto the past and the future.

Considering the future first, we are going to want to say that much of the future belongs to the set FIXTURES. This is largely due to the finite nature of deliberation: since we do not deliberate about bringing about the whole future all at once, there are then many future occurrences that we take as part of the fixed background during the deliberative process. It also seems as given that the set DIRECT OPTIONS must also be comprised of future events. We can attribute this to the fact that we are temporally constrained agents of a particular sort; the set DIRECT OPTIONS consists if our immediate actions and we simply cannot deliberate about whether to bring about our past actions, only our future actions. Further to this, we might want to say that the set INDIRECT OPTIONS also is comprised exclusively of future events, but this would be so only if we were committed to classifying all past events as belonging to the set FIXTURES. Ordinarily, this is exactly how we consider past events: as fixed. This is for the most part a function of the fact that we consider the past as knowable in principle, and as we have seen above, the set KNOWABLES is a subset of the set FIXTURES. But is it the case that our spatiotemporally bound perspective commits us to the past being fixed?

If such a commitment is indeed a function of the fact that we consider the past as knowable in principle, then it would seem that the possibility of the past being unknowable in principle would purge us of this commitment. Recall that this is exactly the condition we found to be suitable to avoid the bilking argument in the above analysis of Dummett: an agent is immune to having a belief in a particular retrocausal correlation bilked if the past effect in question is epistemically inaccessible to the agent at the time of the causal action. In the language of our current analysis, if some past event belongs to the set UNKNOWABLES then it does not *necessarily* belong to the set FIXTURES, an agent may believe it to belong to the set INDIRECT OPTIONS. As we noted above, the very nature of quantum mechanics ensures that it *is* immune to the bilking argument. Thus, in the right circumstances, there is information about the past of some quantum systems that is epistemically inaccessible *in principle*! If this is the case then it is a live possibility that the set INDIRECT OPTIONS contains some events which are past; or rather, the architecture of deliberation does not rule out the possibility of bringing about the past on analytic grounds.

This schematic of where retrocausality fits in to the structure of deliberation highlights an important feature of a metaphysical picture that allows retrocausal influences: that agents within such a reality will always deliberate towards the future, i.e. the set DIRECT OPTIONS will always be comprised of future events. Thus retrocausality is not deliberation towards the past, or in other words, it is not our normally directed causation in the reverse temporal direction.

As a final comment on the perspectival nature of causation, the way that any particular agent divides the set of all events into FIXTURES and OPTIONS, KNOWABLE and UNKNOWABLE and past and future will depend completely upon the agents spatiotemporal perspective. For spatiotemporally constrained agents such as ourselves, there is a specific recipe for how these distinctions are made which is a function of the way we have evolved from within the spacetime block. If we imagine ourselves as omniscient beings who are observing the events in the spacetime block 'from the outside', there will be no past or future (though there may be past and future directions along the temporal axis of the block) and all the events will be in the set KNOWABLE and thus in the set FIXTURES. This is how we can imagine the spacetime block to be entirely determined without having this intuition be in conflict with our usual sense of free choice in the deliberative process; these are vastly different perspectives and causality is perspectival. It is the extent of our ignorance, of both the future and of the complete set of prior causes of our actions, that creates the illusion, so to speak, of free choice.

5 Conclusion

This then is the package of metaphysical ideas that combine to give a picture that is consistent with the possibility of retrocausality. We begin with two uncontroversial metaphysical foundations in the block universe model of time and the interventionist account of causation. We then remove two potential obstacles originating in our ordinary temporal intuitions: we realise that we have no evidence to suggest our macroscopic asymmetric causal intuitions can be extrapolated to the microscopic realm and we realise that we do not necessarily have epistemic access to the past independent of our own future actions. With these obstacles gone, the emerging metaphysical picture of a temporally and causally symmetric reality viewed from an epistemically limited vantage point concords well with the possibility of retrocausality. A significant aspect of this assembly of ideas is that none of the included elements are precluded by the known physical structure of our reality. Indeed, if anything, these elements are supported by the structure of at least one of our best physical theories: quantum mechanics.

Thus while Maudlin is clearly correct in noticing that retrocausality is fundamentally at odds with the metaphysical picture of the past generating the future, this by no means renders retrocausality metaphysically untenable. Given the right mix of some reasonable metaphysical and epistemological ingredients, an alternative metaphysical picture arises that is consistent with the possibility of retrocausality. Moreover, the conceptual cost of these ingredients cannot be a higher price to pay than the interpretational problems associated with the rejection of local hidden variables, simply for the fact that we were given all these ingredients for free by the metaphysical structure of our reality and the epistemological structure of our existence.

References

Argaman, N. (2008). On Bell's Theorem and the Causal Arrow of Time, arXiv:0807.2041v1 [quant-ph].

Costa de Beauregard, O. (1953). Méchanique Quantique, *Comptes Rendus de l'Académie des Sciences* **T236**: 1632–1634.

Costa de Beauregard, O. (1976). Time Symmetry and Interpretation of Quantum Mechanics, *Foundations of Physics* **6**: 539–559. DOI: 10.1007/BF00715107.

Costa de Beauregard, O. (1977). Time symmetry and the Einstein paradox, *Il Nuovo Cimento* **42**: 41–63. DOI: 10.1007/BF02906749.

Cramer, J. G. (1980). Generalized absorber theory and the Einstein-Podolsky-Rosen paradox, *Physical Review D* **22**: 362–676. DOI: 10.1103/PhysRevD.22.362.

Cramer, J. G. (1986). The transactional interpretation of quantum mechanics, *Reviews of Modern Physics* **58**: 647–687. DOI: 10.1103/RevModPhys.58.647.

Dummett, M. (1964). Bringing About the Past, *The Philosophical Review* **73**(3): 338–359.

Evans, P. W., Price, H. and Wharton, K. B. (2010). New Slant on the EPR-Bell Experiment, arXiv:1001.5057v2 [quant-ph].

Hokkyo, N. (1988). Variational formulation of transactional and related interpretations of quantum mechanics, *Foundations of Physics Letters* **1**: 293–299. DOI: 10.1007/BF00690070.

Maudlin, T. (2002). *Quantum Non-Locality and Relativity*, Blackwell Publishing, Oxford.

Miller, D. J. (1996). Realism and time symmetry in quantum mechanics, *Physics Letters A* **222**: 31–36. DOI: 10.1016/0375-9601(96)00620-2.

Miller, D. J. (1997). Conditional probabilities in quantum mechanics from time-symmtric formulation, *Il Nuovo Cimento* **112B**: 1577–1592.

Price, H. (1984). The philosophy and physics of affecting the past, *Synthese* **61**: 299–324. DOI: 10.1007/BF00485056.

Price, H. (1994). A neglected route to realism about quantum mechanics, *Mind* **103**: 303–336. DOI: 10.1093/mind/103.411.303.

Price, H. (1996). *Time's Arrow and Archimedes' Point*, Oxford University Press, New York.

Price, H. (1997). Time Symmetry in Microphysics, *Philosophy of Science* **64**: S235–S244. arXiv:quant-ph/9610036v1.

Price, H. (2001). Backwards caustion, hidden variables, and the meaning of completeness, *Pramana Journal of Physics* **56**: 199–209. DOI: 10.1007/s12043-001-0117-6.

Price, H. (2007). Causal Perspectivalism, *in* H. Price and R. Corry (eds), *Causation, Physics, and the Constitution of Reality: Russell's Republic Revisited*, Oxford University Press, Oxford, chapter 10, pp. 250–292.

Price, H. (2008). Toy models for retorcausality, *Studies in History and Philosophy of Modern Physics* **39**: 752–776. DOI: 10.1016/j.shpsb.2008.05.006.

Price, H. (2010). Time symmetry without retrocausality: how the quantum can withold the solace, arXiv:1002.0906v1 [quant-ph].

Rietdijk, C. W. (1978). Proof of a retroactive influence, *Foundations of Physics* **8**: 615–628. DOI: 10.1007/BF00717585.

Sutherland, R. I. (1983). Bell's theorem and backwards-in-time causality, *International Journal of Theoretical Physics* **22**: 377–384. DOI: 10.1007/BF02082904.

Sutherland, R. I. (1998). Density Formalism for Quantum Theory, *Foundations of Physics* **28**: 1157–1190. DOI: 0.1023/A:1018850120826.

Sutherland, R. I. (2008). Causally symmetric Bohm model, *Studies in History and Philosophy of Modern Physics* **39**: 782–805. DOI: 10.1016/j.shpsb.2008.04.004.

Wharton, K. B. (2007). Time-Symmetric Quantum Mechanics, *Foundations of Physics* **37**: 159–168. DOI: 10.1007/s10701-006-9089-1.

Wharton, K. B. (2010). A Novel Interpretation of the Klein-Gordon Equation, *Foundations of Physics* **40**: 313–332. DOI: 10.1007/s10701-009-9398-2.

Woodward, J. (2003). *Making Things Happen: A Theory of Causal Explanation*, Oxford University Press, New York.