

# Causal Inference in Quantum Mechanics: A Reassessment

Mauricio Suárez\*

Universidad Complutense de Madrid

## Abstract

There has been an intense discussion, albeit largely an implicit one, concerning the inference of causal hypotheses from statistical correlations in quantum mechanics ever since John Bell's first statement of his notorious theorem in 1966. As is well known, its focus has mainly been the so-called Einstein-Podolsky-Rosen ("EPR") thought experiment, and the ensuing observed correlations in real EPR like experiments. But although implicitly the discussion goes as far back as Bell's work, it is only in the last two decades that it has become recognizably and explicitly a debate about causal inference in the quantum realm. The bulk of this paper is devoted to a review of three influential arguments in the philosophical literature that aim to show that causal models for the EPR correlations are impossible, due to Bas Van Fraassen, Daniel Hausman and Huw Price. I contend that all these arguments are inconclusive since they contain premises or presuppositions that are false, unwarranted, or at least controversial. Five different causal models are outlined that seem perfectly viable for the EPR correlations. These models are then employed to illustrate various difficulties with the premises and presuppositions underlying Van Fraassen's, Hausman's and Price's arguments. In all cases it is argued that the difficulties cut deep against these authors' own theories of causation and causal inference. My conclusions are that causal models for the EPR correlations certainly remain viable, that philosophical work is still required to assess their relative virtues, and that in any case the mere theoretical conceivability and empirical possibility of these models sheds deep doubts over Van Fraassen's, Hausman's and (important elements in) Price's theories of causation and causal inference.

## Contents

### 1 The EPR Correlations

2

---

\*I want to thank the audience at my talk to the Causality and Probability in the Sciences conference at the University of Kent, as well as two anonymous referees for helpful questions and comments. Thanks also to Iñaki San Pedro for his feedback, and his help with the production of the paper. Research towards this paper has been funded by project HUM2005-07187-C03-01 of the Spanish Ministry of Education and Science.

<b>2</b>	<b>Reichenbach’s Principle of the Common Cause</b>	<b>6</b>
<b>3</b>	<b>The Arguments against Causal Models</b>	<b>12</b>
3.1	Van Fraassen’s Reichenbachian argument . . . . .	12
3.2	Hausman’s Independence Argument . . . . .	17
3.3	Huw Price’s Asymmetry Argument . . . . .	20
<b>4</b>	<b>Five Causal Models for the EPR Correlations</b>	<b>23</b>
<b>5</b>	<b>Replies to the Arguments</b>	<b>28</b>
5.1	Reply to Van Fraassen’s Reichenbachian arguments . . . . .	30
5.2	Reply to Hausman’s independence argument . . . . .	31
5.3	Reply to Price’s asymmetry argument . . . . .	34
<b>6</b>	<b>Conclusions and Prospects</b>	<b>38</b>

# 1 The EPR Correlations

In 1935 Einstein wrote and published jointly with two collaborators a notorious paper describing a thought experiment with correlated entangled pairs of particles.<sup>1</sup> The stated aim of the paper is to demonstrate that the quantum theory is incomplete since it does not describe fully all the “elements” of quantum reality, yet the paper is nowadays celebrated as the source of the burgeoning literature on what is known as “quantum non-locality”. On the one hand as Arthur Fine<sup>2</sup> and other have shown the real conclusion of the EPR argument is rather a dilemma between locality (in EPR’s own characteristic definition this entails that a disturbance of the state of the nearby particle can exert no change of any of the properties of the distant particle) and completeness. On the other hand Bell’s theorem is taken by many to demonstrate that any empirically adequate completion of the quantum theory is committed to the existence of EPR-like correlations between the measurement events of certain properties of distant particles that have interacted in the past. It follows that the conclusion that actually does follow from the EPR argument is the existence of distant correlations. Indeed these correlations have been tested experimentally on numerous occasions with the same positive result.

David Bohm’s version of the EPR thought experiment is most often discussed, and it is this version that provides the model for most of the real experiments that have actually been carried out.<sup>3</sup> In this Einstein-Podolsky-Rosen-Bohm (EPR-Bohm) experiment two particles (“1” and “2”) move in opposite directions, after either interacting in the past, or having been created simultaneously in some past decay event “E”. As a result of their

---

<sup>1</sup>(Einstein, Podolsky and Rosen, 1935). The exposition in this section borrows from a previous paper of mine, (Suárez, 2004).

<sup>2</sup>(Fine, 1987).

<sup>3</sup>(Bohm, 1951, chapter 22).

interacting history, quantum mechanics describes their composite state as an entangled singlet state. The initial angular momentum is zero, so their values of spin must be correlated throughout. Any particle's spin can be measured by means of a Stern-Gerlach apparatus. This is a magnetometer that impresses a force upon the particle proportional to its spin value, thereby correlating perfectly the particle's position with its spin value at the time the particle interacts with the magnetometer. A Stern-Gerlach apparatus can be rotated along 360 degrees, in order to measure the particle's spin value along any direction.

In a Minkowski space-time diagram, both particles describe symmetric paths along the time axis (see figure 1). The Stern-Gerlach apparatus that measure these particles' spin at each wing of the experiment are at rest in the laboratory frame so their world lines are represented by vertical lines "A<sub>1</sub>" and "A<sub>2</sub>" in that frame. Each time the experiment is repeated, laboratory technicians are at freedom to select a particular orientation of the measurement apparatus which will result in a measurement of spin along the corresponding direction. Such setting events are denoted by "a" and "b". The two arrows pointing towards such events represent the fact that each of those setting events is controllable by experimental means. (In the language of the causal inference literature: they are exogenous variables, and are moreover controllable by agents.) Each particle's spin is measured by means of a measurement interaction between the particle and the associated measuring device on the corresponding wing. The outcomes that are produced are denoted by "s<sub>1</sub>" and "s<sub>2</sub>" respectively, and are known as the "outcome-events".

An important feature of the EPR-Bohm experiments is that these outcome-events are spacelike connected, i.e. they lie outside each other's light-cone. Thus a signal from one event to the other must travel at speed greater than the speed of light, during a finite part of its trajectory at least. The implications of this fact regarding the special theory of relativity are both deep and complex, and have been the object of an intense debate.<sup>4</sup> Although this debate is not directly relevant to much of what I will say here, it is nonetheless important for the overall assessment of the prospects of causal inference in quantum mechanics. But the importance lies in the possibility of direct-cause models for the EPR correlations, as will be seen later, and I will center the bulk of my discussion upon the other type of causal models available, namely common cause models. It seems legitimate not to enter the debate in full here (other than by mentioning some options in the interpretation of relativity) since doing so can only strengthen the position defended in this paper, namely that there is a very large range of different causal models available for the EPR correlations.

---

<sup>4</sup>See e.g. (Maudlin, 1995).

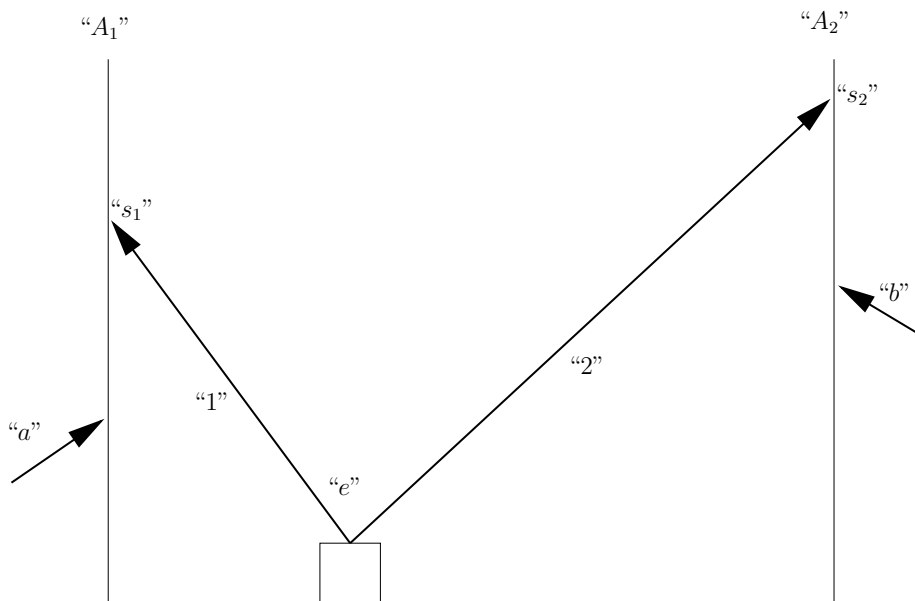


Figure 1.

Let us now return to the quantum mechanical description of the EPR-Bohm experiment. According to quantum mechanics, for the type of particles involved, there are only two possible values of spin in any direction of measurement ( $\theta$ ): positive spin ( $\uparrow_{\theta}$ ) and negative spin ( $\downarrow_{\theta}$ ). Quantum mechanics describes the spin states of the composite system of both particles at either the time of emission or measurement by means of what is known as the singlet state (here “1” and “2” refer to each particle):

$$\Psi = \frac{1}{\sqrt{2}}(|\uparrow_{1\theta}\rangle|\downarrow_{2\theta}\rangle - |\downarrow_{1\theta}\rangle|\uparrow_{2\theta}\rangle).$$

The theory offers two kinds of probabilistic predictions. First, it offers predictions about the outcomes of measurements performed on each particle. To calculate these, we must first apply what is known as the axiom of reduction, which allows us to derive the state of each particle, individually taken:<sup>5</sup>

$$W_1 = \frac{1}{2}|\uparrow_{1\theta}\rangle\langle\uparrow_{1\theta}| + \frac{1}{2}|\downarrow_{1\theta}\rangle\langle\downarrow_{1\theta}|,$$

$$W_2 = \frac{1}{2}|\uparrow_{2\theta}\rangle\langle\uparrow_{2\theta}| + \frac{1}{2}|\downarrow_{2\theta}\rangle\langle\downarrow_{2\theta}|.$$

We can apply the quantum statistical algorithm to  $W_1$  and  $W_2$  in order to calculate the probabilities of outcomes of measurements performed on

<sup>5</sup>See e.g. (Hughes, 1989, pp. 149-150).

each particle in any direction  $\theta$ :

$$\begin{aligned} \text{prob}(|\uparrow_{1\theta}\rangle) &= \langle\Psi|\uparrow_{1\theta}\rangle\langle\uparrow_{1\theta}|\Psi\rangle = \left(\frac{1}{\sqrt{2}}\right)^2 = \frac{1}{2}. \\ \text{prob}(|\downarrow_{1\theta}\rangle) &= \text{prob}(|\uparrow_{2\theta}\rangle) = \text{prob}(|\downarrow_{2\theta}\rangle) = \text{prob}(|\uparrow_{1\theta}\rangle) = \frac{1}{2}. \end{aligned}$$

It is important to emphasise that  $W_1$  and  $W_2$  are the states of each of the particles individually taken, *consistently* with the fact that the state of the composite is the singlet state. In other words the axiom of reduction allows us to derive uniquely  $W_1$  and  $W_2$  from  $\Psi$ . There is however no similar axiom in quantum mechanics that would allow us to derive  $\Psi$  uniquely from  $W_1$  and  $W_2$ . And this is at the heart of the notorious fact that the singlet state of the composite system is underdetermined by the states of its component subsystems, which often gives rise to the claims of quantum holism and quantum non-separability.<sup>6</sup> So our calculations of predictions for measurement results on each particle on the basis of  $W_1$  and  $W_2$  are perfectly consistent with the singlet state of the composite, and in fact required by it. But the predictions do not require the singlet state, and are instead consistent with all kinds of states of the composite that result from a mere phase difference from the singlet state (i.e. the minus sign in the singlet state could for all we care here be replaced by a plus sign).

In other words, the description offered by the singlet state  $\Psi$  of the composite system contains the greatest possible amount of information about both systems. By contrast, if we only consider the states of the systems individually taken,  $W_1$  and  $W_2$ , we can see that we have lost relevant information. Erwin Schrödinger was perhaps the first to note that “a portion of knowledge of the composite system” is found “squandered on conditional statements that operate between the subsystems” in the form of correlations between the measurement events that we can perform on each system.<sup>7</sup> Indeed by applying the quantum statistical algorithm to the entangled pair of particles in the singlet state  $\Psi$  we find the following conditional probabilities of outcomes of measurements on either particle, conditional on any

---

<sup>6</sup>I do not pursue these claims further in this paper. But a referee helpfully pointed out that both the arguments I describe in section 3, and the models I present in response in section 4 presuppose the assumption of separability, roughly: that in EPR-like situations it is perfectly legitimate to postulate the existence of two distinct physical systems at the wings of the experiment, however entangled their states. It might be harder to articulate some causal explanations for EPR correlations under a contrary assumption of non-separability, but then it would also be harder, I think, to articulate any arguments against such types of explanations. Moreover there is a sense in which any non-separable model of EPR is by definition causal: It merely postulates correlations between properties of one and the same entity.

<sup>7</sup>(Schrödinger, 1933, p. 161).

particular outcome of any measurement made on the other particle:

$$prob(\uparrow_{2\theta'} / \downarrow_{1\theta}) = \frac{prob(\uparrow_{2\theta'} \wedge \downarrow_{1\theta})}{prob(\uparrow_{1\theta})} = \frac{1}{2} \sin^2 \frac{1}{2} \theta \theta'.$$

In the specific case  $\theta = \theta'$  we obtain the following conditional probabilities, which we can immediately see imply a case of anticorrelation:

$$\begin{aligned} prob(\uparrow_{2\theta} / \downarrow_{1\theta}) &= 1 = prob(\downarrow_{2\theta} / \uparrow_{1\theta}), \\ prob(\downarrow_{2\theta} / \downarrow_{1\theta}) &= 0 = prob(\uparrow_{2\theta} / \uparrow_{1\theta}). \end{aligned}$$

This means that if we measure both particles' spin along the same direction, the singlet state predicts an anti-correlation between the spin values. If we measure the first particle's spin in the  $\theta$  direction, and we find the outcome corresponding to "positive" spin ( $\uparrow_{\theta}$ ), we can predict that the outcome of a later measurement of the second particle's spin in the same direction will be "negative" ( $\downarrow_{\theta}$ ) with certainty.

The kind of necessity expressed by these conditional statements, according to quantum mechanics, is merely nomological, since quantum mechanics does not describe any physical process capable of transmitting the information required from one system to the other. This of course is not to say that such a mechanism does not really exist. It is consistent to affirm both horns of the EPR argument's dilemma: i.e. that there are correlations between distant particles and that the theory is incomplete. What it means is that a causal explanation of these correlations would have to introduce some type of mechanism, or additional physical hypothesis, to explain these conditional statements. In a causal model the "additional portion of knowledge" would not be "squandered in conditional statements". For instance, in a model where the causes operate directly between the wings of the experiment, the "extra" portion of knowledge could be transmitted directly from one subsystem to the other by means of "mark-transmitters".<sup>8</sup> In a common cause model by contrast the causal influence might well follow the very same particle' trajectories. In any case, causal structure will need to be postulated that is not described by quantum theory, but will hopefully be consistent with it.

## 2 Reichenbach's Principle of the Common Cause

The EPR-Bohm experiment yields a typical case of statistical correlation, which is both predicted by a theory, and experimentally verified; it is somewhat surprising in retrospect that it took so long for it to be seen as fertile

---

<sup>8</sup> "Mark-transmitter" is the term employed in both Hans Reichenbach's (Reichenbach, 1956, p. 198) and Wesley Salmon's (Salmon, 1984, pp. 148-150) theories of causality. Nothing I say in this paper however hinges on such accounts.

ground for the application of techniques of causal inference from statistical data. Reichenbach’s principle of the common cause provides one of the earliest and most influential techniques, together with a complex theory of probabilistic causation and causal structure. In deriving his notorious theorem Bell employed essentially Reichenbach’s techniques, unbeknownst to him. Bell did not identify his statistical conditions as techniques of causal inference, but instead took them to be conditions on physical “locality”. It was the philosopher Bas Van Fraassen instead who, in a couple of influential and important papers in the 1980’s, first explicitly analysed Bell’s theorem in terms of Reichenbachian conditions of causal inference. Van Fraassen’s argument has been immensely influential in drawing philosophers of physics to sceptical conclusions regarding causation in the quantum realm. In this section I intend to describe Reichenbach’s conditions for causal inference, and Van Fraassen’s analysis of Bell’s theorem in terms of these conditions; in the next section I will analyse Van Fraassen’s full argument against causal models for the EPR correlations, together with related arguments to the same effect by Daniel Hausman and Huw Price. My conclusions will be critical, and I will defend that this scepticism is premature.

At its heart Reichenbach’s theory of causal inference is extremely simple. Its central principle (the “principle of the common cause”) asserts that “if an improbable coincidence has occurred, there must exist a common cause”.<sup>9</sup> Thus for Reichenbach the search for causes underlying correlation phenomena is a methodological maxim. Two comments however, regarding the principle, are in order. First, by “improbable” Reichenbach does not mean a coincidence between two token events with a low prior probability. Rather what he has in mind is a statistical correlation between two event-types  $A$  and  $B$  that is robust both theoretically and experimentally, i.e. (a) that it is predicted by some established theory and (b) that it has been verified empirically, or at least not refuted by experiment. But it must also be the case that the correlation between the event types can not be explained as a mere direct causal relation between those types. The methodological maxim to unearth common cause structure is not applicable for event types that have already been explained by means of a direct causal connection: those are not “*improbable*”.

Hence the first condition for causal inference, according to Reichenbach’s theory, is correlation between two event types  $A$  and  $B$  which are not directly causally related:

$$\text{prob}(A \wedge B) \neq \text{prob}(A)\text{prob}(B) \qquad \text{(Correlation)}$$

The second condition is the existence of a open fork, i.e. a third variable  $C$  representing an event type in the past of  $A$  and  $B$  that makes the

---

<sup>9</sup>(Reichenbach, 1956, pp. 157 ff).

correlation between  $A$  and  $B$  vanish:<sup>10</sup>

$$\text{prob}(A \wedge B|C) = \text{prob}(A|C)\text{prob}(B|C) \quad (\text{OpenFork})$$

This condition can be expressed in a mathematically equivalent way as a screening off condition. If  $C$ ,  $A$ ,  $B$  form an open fork as described above, then it follows that  $C$  screens off  $A$  from  $B$  and viceversa.<sup>11</sup>

$$\begin{aligned} \text{prob}(A|B \wedge C) &= \text{prob}(A|C) \\ \text{prob}(B|A \wedge C) &= \text{prob}(B|C) \end{aligned} \quad (\text{Screening – off})$$

Reichenbach was very aware that we must tread carefully at this point, since screening off is not a sufficient condition on common causes only a necessary one, and that only under a very strong presumption of completeness. Let me explain. To see that screening off is not sufficient for common causes it is enough to observe that for any correlation, there will always exist some variable  $D$  that is not common cause, but satisfies the screening off condition, i.e some variable  $D$  such that:

$$\text{prob}(A \wedge B|D) = \text{prob}(A|D)\text{prob}(B|D).$$

For example a common effect  $D$  lying in the future of  $A$  and  $B$  will screen off  $A$  from  $B$  and viceversa (figure 2). These cases can be dispensed with by means of the “in the common past of  $A$  and  $B$ ” qualification, which rules out a screener off in the future of either  $A$  or  $B$ . However, any common effect  $D$  lying to the future of  $C$  but to the past of  $A$  and  $B$  will also satisfy screening off (figure 3).

Hence screening off is not necessary for a common cause, unless we insist on a very strong assumption of completeness: i.e. unless we assure ourselves that  $C$  is the only causally relevant variable for  $A$  and  $B$ ; but to know this would be to know precisely what Reichenbach’s inferential techniques were meant to allow us to learn in the first place – i.e. that  $C$  is the common cause of  $A$  and  $B$ .

Thus causal inference for Reichenbach was to proceed negatively: by discovering violations of the screening off condition. In other words screening off was to be taken to be merely necessary for a common cause: Not all screener offs are common causes but all common causes screen off. Hence,

---

<sup>10</sup>The qualification “in the past of  $A$  and  $B$ ” is anachronistic, and very much my own. Reichenbach thought that open forks could be used to define the direction of time, so to say of an open fork that it is oriented towards the future (or, as I say above, that the screener off must lie in the common past of  $A$  and  $B$ ) would just amount, in Reichenbach’s theory, to the trivial truism that an open fork is oriented as an open fork.

<sup>11</sup>And, conversely, if  $C$  screens off  $A$  from  $B$ , and  $C$  lies in the past of  $A$  and  $B$ , then  $C$ ,  $A$ ,  $B$  form an open fork.



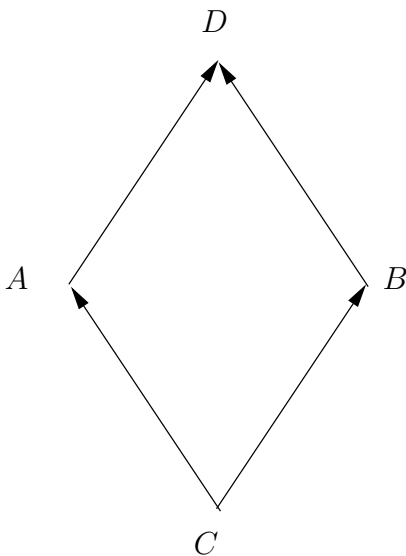


Figure 2.

roughly, if conditioning upon some variable  $C$  does not render  $A$  and  $B$  statistically independent then we can at least be sure that  $C$  is not the common cause. Yet, even this simple statement does not turn out to be generally true. To see why consider the following structure where two common causes  $C$  and  $D$  acting independently underlie the correlation between  $A$  and  $B$  (figure 4).

In this structure neither  $C$  nor  $D$  will screen off on their own. It is instead the conjunction of  $C$  and  $D$  that makes the correlation vanish:

$$prob(A \wedge B|C \wedge D) = prob(A|C \wedge D)prob(B|C \wedge D).$$

Hence a violation of this condition, in a structure with these four variables only, allows us to safely infer that either  $C$ , or  $D$ , or both fail to be common causes. A violation of the corresponding screening off conditions for  $C$  and  $D$  disjointly would allow us to infer nothing safely at all about  $C$  and  $D$  other than the very minimal conditional fact that *if  $C$  ( $D$ ) is a common cause of  $A$  and  $B$ , then  $C$  ( $D$ ) certainly is not the only cause.* And that again presupposes precisely some of the causal knowledge that Reichenbach's methods were supposed to allow us to discover.

Things actually get worse. The only piece of causal knowledge that we can possibly discover on the basis of statistical analysis by means of Reichenbach's screening off condition turns out to be conditional once again on a strong assumption of completeness. We concluded above that if the conjunction of  $C$  and  $D$  fails to screen off  $A$  from  $B$ , then we can be sure that

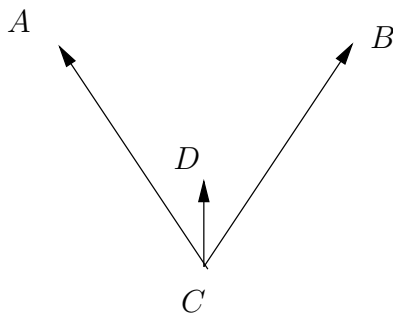


Figure 3.

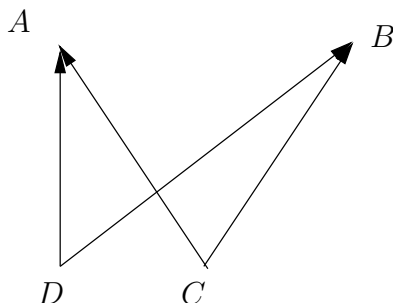


Figure 4.

$C \wedge D$  is not the common cause, *in a structure with four variables only*. But of course the same reasoning that led us to consider four instead of three variables might well lead us now to consider five. Consider the following structure (figure 5), with three putative common causes  $C$ ,  $D$  and  $E$  of the correlation between  $A$  and  $B$ . In this structure the conjunction of  $C$  and  $D$  will not necessarily screen off  $A$  from  $B$ . We can only expect the conjunction of  $D$ ,  $C$  and  $E$  to do so. So the only violation of screening off that would be informative about the actual causal structure would be:

$$prob(A \wedge B|C \wedge D \wedge E) \neq prob(A|C \wedge D \wedge E)prob(B|C \wedge D \wedge E).$$

But in turn this violation of the screening off condition will be informative about the actual causal structure, only *in a structure with only five variables  $A$ ,  $B$ ,  $C$ ,  $D$ ,  $E$* . Once again this implies a strong completeness condition is in place for causal structure. So it requires us to know a fair amount about the causal structure before we can apply Reichenbach's methods in order to discover any facts about the causal structure. And so on. By means of these simple examples, we can already easily appreciate why one of the often repeated main lessons of the recent literature on causal structure has been the insight that there is no causal discovery without background causal

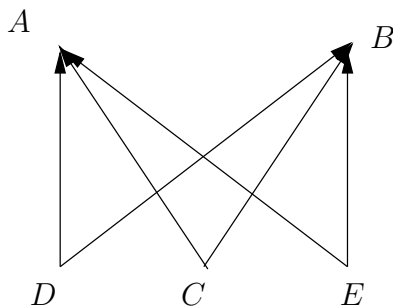


Figure 5.

knowledge: *no causes in, no causes out*. It is not possible to learn anything regarding causal structure from knowledge of statistical correlations alone. Additional causal knowledge is essential for informative inference from statistical correlations to causal hypotheses.

This is an insight that will turn out to be important for the rest of this paper. Let me here quickly recapitulate its implications towards Reichenbach’s Principle of the Common Cause. Two different commitments are often conflated under this rubric. On the one hand there is a commitment to the general maxim: “if an improbable coincidence has occurred, there must exist a common cause”. By itself this says nothing about whether common causes necessarily screen off. Reichenbach’s second commitment is that a common cause structure will satisfy screening off. This commitment conversely seems to be independent of the first, since there could be unexplained correlations even if all common causes that do in fact exist screen off.

Let us refer to these independent commitments as Reichenbach’s *principle of the common cause*, and Reichenbach’s *criterion for common causes*. The principle is then the assertion that every well established correlation must have causal explanation. This is a metaphysical statement regarding the nature of correlations, lacking any methodological implications in the absence of a more concrete algorithm for causal discovery. Reichenbach’s *criterion* on the other hand establishes that common causes necessarily screen off. The criterion is supposed to provide the principle with methodological bite, and to allow causal discovery to proceed on the basis of statistical analysis. But as we have seen the form of the criterion that seems sound already builds in causal knowledge from the start since it can only establish that *complete* common causes screen off. This is the main insight to be learnt from the preceding discussion in this section. However, we should be clear that this insight only compromises Reichenbach’s *criterion*. In and by itself the insight says nothing at all about Reichenbach’s

metaphysical *principle*, which might well be true in spite of our failure to find any grounds for causal discovery that would allow us to test it. In other words, Reichenbach’s principle of the common cause does not stand or fall with Reichenbach’s screening off criterion for common causes. The principle might be true even if the criterion turns out to be flawed.

### 3 The Arguments against Causal Models

In this section I review what I consider to be the three outstanding arguments against causal accounts of the EPR correlations, due to Bas Van Fraassen, Daniel Hausman and Huw Price. In the subsequent sections I shall attempt to rebut these arguments.

#### 3.1 Van Fraassen’s Reichenbachian argument

The most influential argument against causal models for the EPR correlations is due to Bas Van Fraassen. In a set of two overlapping papers in the 1980’s Van Fraassen argued that the EPR correlations can not receive a causal explanation in either the direct-cause or common cause varieties.<sup>12</sup> It would be difficult to overestimate the argument’s influence, even if it is not always explicitly acknowledged. I believe that this argument is historically the main source of many philosophers’ scepticism towards causal accounts of the EPR correlations. Van Fraassen’s papers have also deeply influenced the way philosophers of physics have come to analyse and understand the nature of quantum non-locality and its possible conflict with relativistic causation.<sup>13</sup>

Van Fraassen begins by establishing an analysis of the main statistical condition at the heart of Bell’s theorem (the notorious “factorizability” condition) in terms of three distinct and independent conditions called “causality”, “hidden locality” and “hidden autonomy”. Factorizability is a necessary condition for deriving Bell’s inequalities, which almost everyone agrees have been refuted by experiment.<sup>14</sup> Van Fraassen takes this to imply that it is an empirical fact that “factorizability” is false. It thus follows

---

<sup>12</sup>(van Fraassen, 1982, 1989).

<sup>13</sup>An instance is Jon Jarrett’s influential distinction between parameter and outcome independence (Jarrett, 1984), which tracks Van Fraassen’s “causality” and “hidden locality” conditions, and which has been widely adopted among philosophers of physics (for an acute dissenting criticism of Jarrett’s conditions see (Maudlin, 1995, chapter 4)).

<sup>14</sup>Not everyone agrees that Bell’s inequalities have been experimentally refuted. However, contrary to what some uninformed physicists seem to believe, philosophers have not been at all prominent among those disputing the experimental results – on the contrary philosophers of physics on the whole have shown at least as great, if not greater, a readiness to accept the standard understanding of the experimental results as any physicists. For some dissenting views among physicists see for instance Marshall, Santos and Selleri (1983), Foadí and Selleri (2000).

from his analysis that at least one of the three conditions that factorizability can be decomposed into must necessarily be empirically false. His argument purports to put the blame entirely on “causality”, which Van Fraassen then takes both to imply that no causal model is viable for the EPR correlations, and that Reichenbach’s principle of the common cause is false as a matter of fact: not all well established correlations admit of a causal model.

Let us now look into this argument in greater detail. Van Fraassen first aims to establish that the EPR correlations constitute an example of “improbable coincidence” in Reichenbach’s sense. So he aims to show that the measurement outcome event on each wing of the experiment can not be directly causing the outcome event on the other wing. Let us suppose that in the laboratory rest frame the measurement on particle “1” is carried out before the measurement on particle “2”; in other words measurement outcome  $s_1$  occurs before measurement outcome  $s_2$ . A direct cause model would be one in which the measurement outcome event in one wing,  $s_1$ , is a direct partial cause of the measurement outcome event in the other wing,  $s_2$  (figure 6). This is just the kind of model that Van Fraassen aims first to rule out:

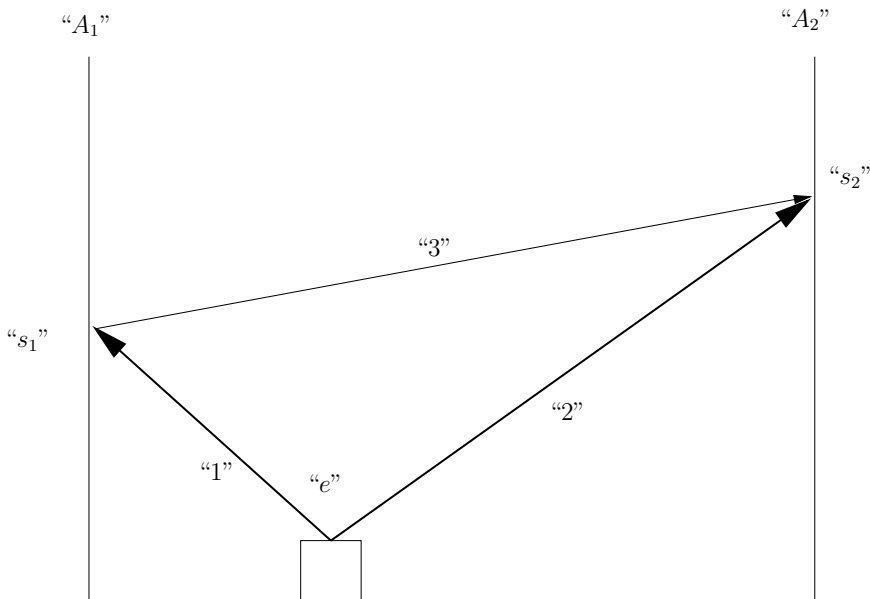


Figure 6.

In this spacetime representation of the EPR experiment  $A_1$  and  $A_2$  represent the worldlines of the measurement devices; “1” and “2” represent those of the particles; and the line comprised between  $s_1$  and  $s_2$  is the worldline of a direct causal process between the wings of the experiment.  $s_1$ ,  $s_2$ , and

$e$  denote event-types, where  $e$  is the particles' emission event, and  $s_1$  and  $s_2$  are the outcome-events that result from measurements on particle "1" by device  $A_1$ , and on particle "2" by device  $A_2$ , respectively.  $s_2$  is in addition the reception event by particle "2" of the causal influence emitted by particle "1".

Van Fraassen rules out direct cause models by appealing to relativity theory, since this theory implies that events  $s_1$  and  $s_2$ , which lie outside each others' light cone, are not absolutely oriented in time. This entails, according to Van Fraassen, that any direct causal model for the EPR correlations in these circumstances could only provide an "explanation by coordination".<sup>15</sup>

"By *coordination* I mean a correspondence effected by signals (in a wide sense): some energy or matter travelling from one location to another, and acting as a partial producing factor for the corresponding event. The situation need not be deterministic – there can be indeterministic signalling if the signal is not certain to arrive and/or not certain to have the required effect. But the word "travel" must be taken seriously. Hence this explanation cannot work for corresponding events with spacelike separation. To speak of instantaneous travel from X to Y is a mixed or incoherent metaphor, for the entity in question is implied to be simultaneously at X and at Y – in which case there is no need for travel, as it is at its destination already."

In other words special relativity entails that there exists some frame of reference, equally valid for the description of the physical facts, where the emission of the causal influence is simultaneous with its reception (figure 7).

In this frame of reference the material process that transmits the causal influence must travel at an infinite speed, which raises Van Fraassen's question regarding the inappropriate use of the word "travel" in this context. This part of Van Fraassen's argument is controversial (see section 4) but let us accept it here for the sake of argument. Let us then suppose for the sake of argument that a direct-cause model for the EPR correlations is impossible. This would indeed mean that the EPR correlations are precisely a case of Reichenbach's "improbable coincidence" and thus, if Reichenbach's *principle* is to hold, require a common cause explanation.

As has already been noted a common cause model for the EPR correlations is precluded, according to Van Fraassen, by the experimental violation of Bell's inequalities. His analysis starts from a consideration of the so-called "factorizability" condition that lies at the heart of the Bell inequalities:

$$\text{prob}(s_1 \wedge s_2 | a \wedge b \wedge \Psi) = \text{prob}(s_1 | a \wedge \Psi) \text{prob}(s_2 | b \wedge \Psi) \quad (\text{FACT})$$

<sup>15</sup>(van Fraassen, 1982, in 1989, p. 112).

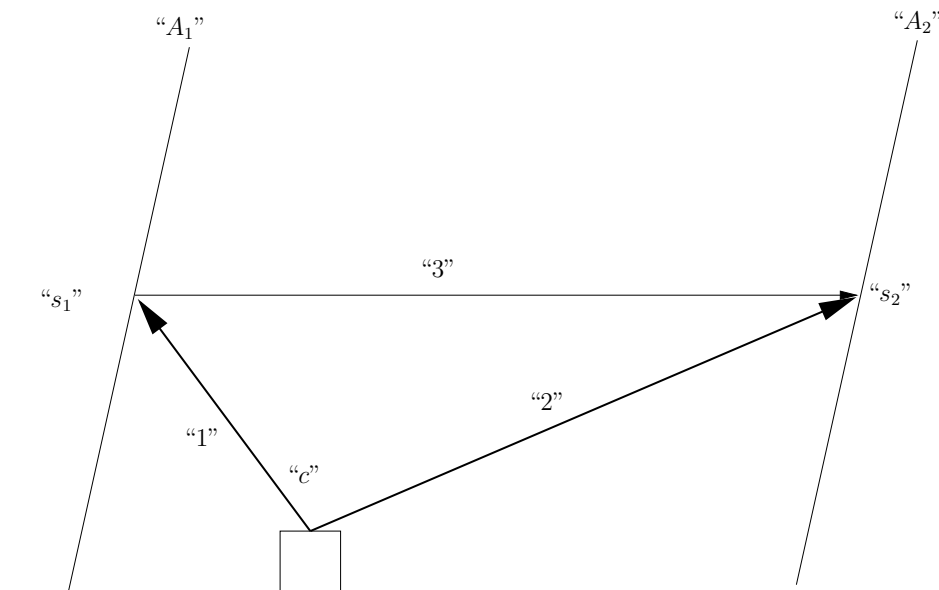


Figure 7.

Factorizability has sometimes been identified with a condition of physical locality, but this is nowadays considered a very contentious identification, so I will stick to the more neutral terminology in this paper.<sup>16</sup> Van Fraassen shows that this condition is the conjunction of three distinct statistical conditions, which he calls “causality”, “hidden locality” and “hidden autonomy”:

$$\begin{aligned} \text{prob}(s_1|s_2 \wedge a \wedge b \wedge \Psi) &= \text{prob}(s_1|a \wedge b \wedge \Psi) \\ \text{prob}(s_2|s_1 \wedge a \wedge b \wedge \Psi) &= \text{prob}(s_2|a \wedge b \wedge \Psi) \end{aligned} \quad (\text{Causality})$$

This condition is a straightforward application of Reichenbach’s screening off criterion for common causes, which Van Fraassen adopts. It states that the conjoint event  $(a \wedge b \wedge \Psi)$  makes event  $s_2$  statistically irrelevant to the probability of  $s_1$ , and viceversa. It has already been noted that according to Reichenbach’s criterion, screening off is a necessary condition on a common cause. Thus Reichenbach’s criterion implies the following conditional: if (Causality) is false in the EPR experiment then the conjunction  $(a \wedge b \wedge \Psi)$  can not be a common cause of the correlations.

The second condition, “hidden locality”, is also an application of the

<sup>16</sup>Some philosophers (as well as physicists) have contested that the experimental refutation of Bell’s inequalities entails the existence of physical “non-locality”. But on the whole philosophers accept that (FACT) has been violated by experiment. They just question that (FACT) can be identified in any straightforward way with a condition of physical locality (see most prominently, (Fine, 1982)).

screening off condition. It states, for each wing of the experiment, that the conjunction of the creation event at the source and the corresponding setting event at one wing together screen off the outcome event at that wing from the setting event at the distant wing:

$$\begin{aligned} \text{prob}(s_1|a \wedge b \wedge \Psi) &= \text{prob}(s_1|a \wedge \Psi) \\ \text{prob}(s_2|a \wedge b \wedge \Psi) &= \text{prob}(s_2|b \wedge \Psi) \end{aligned} \quad (\text{Hidden Locality})$$

The conjunction of (Causality) and (Hidden Locality) is sufficient for factorizability. The proof is trivial and begins with the observation that it is generally the case that:

$$\text{prob}(s_1 \wedge s_2|a \wedge b \wedge \Psi) = \text{prob}(s_1|s_2 \wedge a \wedge b \wedge \Psi)\text{prob}(s_2|a \wedge b \wedge \Psi).$$

By (Causality) it follows that:

$$\text{prob}(s_1 \wedge s_2|a \wedge b \wedge \Psi) = \text{prob}(s_1|a \wedge b \wedge \Psi)\text{prob}(s_2|a \wedge b \wedge \Psi)$$

And by (Hidden Locality) we obtain factorizability (FACT):

$$\text{prob}(s_1 \wedge s_2|a \wedge b \wedge \Psi) = \text{prob}(s_1|a \wedge \Psi)\text{prob}(s_2|b \wedge \Psi)$$

It then seems surprising that Van Fraassen invokes a third condition, namely “hidden autonomy”, which guarantees that the state at the source is statistically independent of the apparatus setting-events:

$$\text{prob}(\Psi|a \wedge b) = \text{prob}(\Psi) \quad (\text{HiddenAutonomy})$$

This condition establishes that the probability of the particles to be in a particular state  $\Psi$  at the time of their emission is independent of the selection of the setting-events in either wing. Note first that were a violation of this condition to entail a causal influence it would necessarily entail an influence backwards in time in the rest frame of the laboratory since in that frame the source event is prior to any of the setting events. (Hidden Autonomy) can be appealed to in this analysis for a couple of reasons. The first is simply that it is necessary for (FACT), just like the others. That is (Causality) and (Hidden Locality) are each entailed by (FACT) but so is (Hidden Autonomy). On the other hand it is clear that a violation of (Hidden autonomy) must in turn entail that either (Causality) or (Hidden Locality) is false since otherwise, given what we just proved above, (FACT) would hold – even if ex hypothesis (Hidden Autonomy) is false. The most natural culprit is (Hidden Locality): if the state statistically depends on the settings, then it seems that the outcomes – which in turn depend on the state – must statistically depend on the settings. This is the second and most important reason to include (Hidden Autonomy) explicitly among the conditions: There is an interesting case of failure of (Hidden Locality), and



consequently of (FACT), that turns on a violation of (Hidden Autonomy). And it is precisely this interesting case that will become relevant later on in assessing another argument against causality in quantum physics, namely the argument due to Huw Price.

Thus we have established, following Van Fraassen’s analysis, that the violation of the Bell inequalities requires that at least one among (Causality), (Hidden Locality) and (Hidden Autonomy) be false. Van Fraassen then goes to argue that the EPR correlations themselves show (Causality) to be false. This is because the state of the particles at the time of their emission does not screen off the outcome-events from each other. For, let us suppose that  $a = b = \theta$ , without loss of generality. Then (Causality) reduces to:

$$\begin{aligned} \text{prob}(s_1|s_2 \wedge \Psi) &= \text{prob}(s_1|\Psi) \\ \text{prob}(s_2|s_1 \wedge \Psi) &= \text{prob}(s_2|\Psi) \end{aligned}$$

And this condition is certainly false, since according to quantum mechanics:

$$\begin{aligned} \text{prob}(s_1|s_2 \wedge \Psi) &= 1 \neq \text{prob}(s_1|\Psi) = \frac{1}{2} \\ \text{prob}(s_2|s_1 \wedge \Psi) &= 1 \neq \text{prob}(s_2|\Psi) = \frac{1}{2} \end{aligned}$$

Van Fraassen thus concludes that a common cause model for EPR of the sort envisioned by Reichenbach is not viable: “The conclusion is surely inevitable: there are well attested phenomena which cannot be embedded in any common-cause model.” (van Fraassen, 1982, in 1989, p. 108).

### 3.2 Hausman’s Independence Argument

A different argument against causal accounts of the EPR correlations has been provided by Daniel Hausman, who aims to reproduce Van Fraassen’s negative conclusion by applying his own distinct theory of causation. At the heart of Hausman’s theory there are a couple of anti-Humean principles. The first one, partly inspired by Reichenbach, exerts the connection between causes and probabilities dependencies, and Hausman refers to it as the Necessary Connection Principle (N-Connection or NC Principle):<sup>17</sup>

N-Connection Principle (NC): Events  $a$  and  $b$  are n-connected if and only if they are distinct and (1)  $a$  causes  $b$  or  $b$  causes  $a$  or (2)  $a$  and  $b$  are effects of a common cause.

The relationship between (NC) and Reichenbach’s *principle* of the common cause is, according to Hausman, akin to the relation between tokens and

---

<sup>17</sup>(Hausman, 1999, p. 81). For Hausman’s own theory of causation see also (Hausman, 1998) where the (NC) principle appears as “The Connection Principle”.

types of causally related events. The (NC) principle applies to token events, and Hausman defends the view that this relation manifests itself as the kind of probabilistic dependence (correlation) among types that prompts Reichenbach's principle.<sup>18</sup> Another important point of clarification regarding this definition has to do with the notion of "distinctness". For Hausman, two events are *distinct* if they are neither logically related nor do they have any part in common. Thus Hausman's definition unpacks some of the commitments underlying Reichenbach's notion of "improbable coincidence" among event types. Finally it is important to note that to have wide application, the (NC) principle must be relativised to some causal "field", otherwise the principle would hold trivially for all events in relation to the big bang.<sup>19</sup>

The (NC) principle is in essence just an anti-Humean assertion of causation as an independent relation between token events which might give rise to and explain, but must not be confused with, probabilistic association among the corresponding types. The most distinct among Hausman's tenets, which identifies his theory, is another principle. The *independence* condition asserts that every effect must have a distinct and individual cause, unrelated to all its other causes (hence often represented by an exogenous variable), which in principle at least allows us to fix its value independently. I will in this paper call such distinct and individual causes a "handle" or "leverer", since at least in principle they allow us to control the presence of token effects and the corresponding probability of their associated types:

Independence Condition (I): If  $a$  causes  $b$  or  $a$  and  $b$  are n-connected only as effects of some common cause, then  $b$  has a cause that is distinct from  $a$  and not n-connected to  $a$ .

Hausman claims that (I) is a conceptual truth about causation: "a boundary condition on the possibility of causal attributions" (1999, p. 83); "a necessary condition for the possibility of causal attributions and causal explanations" (1998, p. 64); and "when all the same things have an n-connection to  $a$  and  $b$ , causal concepts are inapplicable" (1999, p. 83). The intuition is that causal concepts are inapplicable to cases that do not satisfy (I). Prominent among these are putative cases of singular proximate causes. Hence Hausman is ruling out the standard understanding of, for instance, radioactive decay – according to which the nuclear structure of the radioactive element is itself the sole and proximate cause of its decay with a certain probability. He states in response to this putative counterexample to (I) that "the phenomena that we identify as causes and effects are not at all like this" (1998, p. 69). The view that I will defend on the contrary is that the EPR correlations precisely show that we are quite prepared to entertain a causal relation that ascribes to an effect a sole and proximate cause.

---

<sup>18</sup>(Hausman, 1998, p. 59).

<sup>19</sup>(Hausman, 1999, p. 81; 1998, p. 40 and p. 60).

Together with an assumption of transitivity of causation (i.e. if  $c$  is caused by  $b$  and  $b$  is caused by  $a$  then  $c$  is caused by  $a$ ), (NC) and (I) jointly entail the following “theory of causation”:

Independence Theory of Causation (C):  $a$  causes  $b$  if and only if (i)  $a$  is n-connected to  $b$ , (ii) everything n-connected to  $a$  and distinct from  $b$  is n-connected to  $b$ , and (iii) something n-connected to  $b$  is causally independent of  $a$ .

The notion of causal independence that Hausman appeals to here is a straightforward application of the other terms already defined in his theory: “I shall say that events are causally independent if and only if they are both distinct and not causally connected [i.e. n-connected]” (1998, section 4.4). Thus substituting our prior definitions we obtain the following paraphrase of the (C)’s main implication: We can only meaningfully claim that some token event  $a$  is the cause of some other event  $b$  if 1)  $a$  and  $b$  are distinct events 2) causally connected to each other by either 3i) directly causing each other or 3ii) as effects of some common cause, and 4) such that all causes and effects of  $a$  are also causes or effects of  $b$ , while 5)  $b$  has at least one cause or effect that is neither cause nor effect of  $a$ . It is worth noting that each commitment in (C) essentially responds to either (NC), (I) or transitivity. Thus (NC) is essentially responsible for 1), 2) and 3); transitivity applied to (NC) yields 4); while condition 5) is essentially the result of applying the independence condition (I).

We may now return to Hausman’s analysis of the EPR correlations. An application of (C) to EPR yields the conclusion that the measurement wing on one wing can not be said to be the cause of the measurement event in the other wing (see figures 1 and 6).<sup>20</sup> The reason is that condition (5) is apparently violated: in other words the independence condition (I) fails. To check this claim in an EPR background, we must first translate condition (I) into that setting, as follows:

Independence Condition for EPR (I for EPR) : If  $s_1$  causes  $s_2$  or  $s_1$  and  $s_2$  are n-connected only as effects of some common cause, then  $s_2$  has a cause that is distinct from  $s_1$  and not n-connected to  $s_1$ .

---

<sup>20</sup>Hausman’s most detailed analysis appeals instead to the so-called GHZ experiment, which requires a smaller range of setting-events in order to generate the contradiction with a Bell-like inequality. But as far as I can see this is an unnecessary complication and detour since the reason why Hausman thinks that no causal model will apply to GHZ is exactly the same that leads him to the same scepticism in the EPR case, namely the failure of the independence (I) condition as described above (Hausman, 1999, p. 86). And my proposal in section 5 of the applicability of (I) to some possible causal models of EPR applies mutatis mutandis to causal models for GHZ.

A straightforward application of this condition to the simplest common-cause model – with causal influences travelling along the particles’ world-lines, and the setting events rendered irrelevant by setting them both in the same direction  $\theta$ , as in figure 1 – illustrates nicely the failure of (I) that Hausman alludes to. In that scenario there is no event that is a cause of  $s_2$  that is not also n-connected to  $s_1$ . In other words there is no independent “handle” on  $s_2$  that would allow us to control its value independently of the causal relation between  $s_1$  and  $s_2$ . The failure of (I for EPR) entails, according to Hausman, that causal concepts are inapplicable in this scenario: our putative common cause model for EPR is not a genuinely *causal* model.<sup>21</sup>

### 3.3 Huw Price’s Asymmetry Argument

Yet another argument against causal models for the EPR correlations is due to Huw Price.<sup>22</sup> Price’s views are close to Hausman’s in the following regard:<sup>23</sup> both seem to think that even if models for the EPR correlations may look causal, they nonetheless fail to be *genuinely* causal – since they do not employ fully articulate causal concepts. But they differ as to what they consider the key to a full articulation of the concept of causality. And they correspondingly differ in their analysis of what is lacking in putative causal models for the EPR correlations. For Hausman, as was noted in the previous section, the key is the failure of the Independence (I) condition, while for Price the key is the absence of a time-asymmetry in the relation of causal dependence.

It might seem odd to include Price in the list of critics of causal models, since he is well known for defending an explanation of the EPR correlations along the lines of the *zigzag* model of Costa de Beauregard, employing the notion of backwards in time influences; and this is often understood to be a causal model. In my view Price’s theses on causal asymmetry make it explicit that his conception of causation does not really fit a *causal* understanding of such backwards-in-time explanations of the EPR correlations. And his most recent defense of causal perspectivalism makes it plain that for him the notion of causation is unsuitable for microphysics altogether.<sup>24</sup> Hence the conclusion to be extracted is that the kind of explanation that Price is advocating for EPR correlations is not, on his own account, genuinely causal.

---

<sup>21</sup>Hausman then goes on to devote a fair amount of work to provide a revised version of the (NC) principle that accommodates the existence of nomological but non-causal correlations of the sort that he finds in EPR (Hausman, 1999, pp. 88ff.); but this is irrelevant to our purposes here, since the failure of (I) still impugns the application of his theory of causation (C) to the EPR correlations.

<sup>22</sup>(Price, 1996, chapters 7-9).

<sup>23</sup>Or rather “Hausman’s views are close to Price’s”, since Price’s work in this area precedes Hausman’s.

<sup>24</sup>(Price, 2005).

The type of model advocated by Price for the EPR correlations can be represented in a spacetime diagram of the sort that we have been employing as follows:

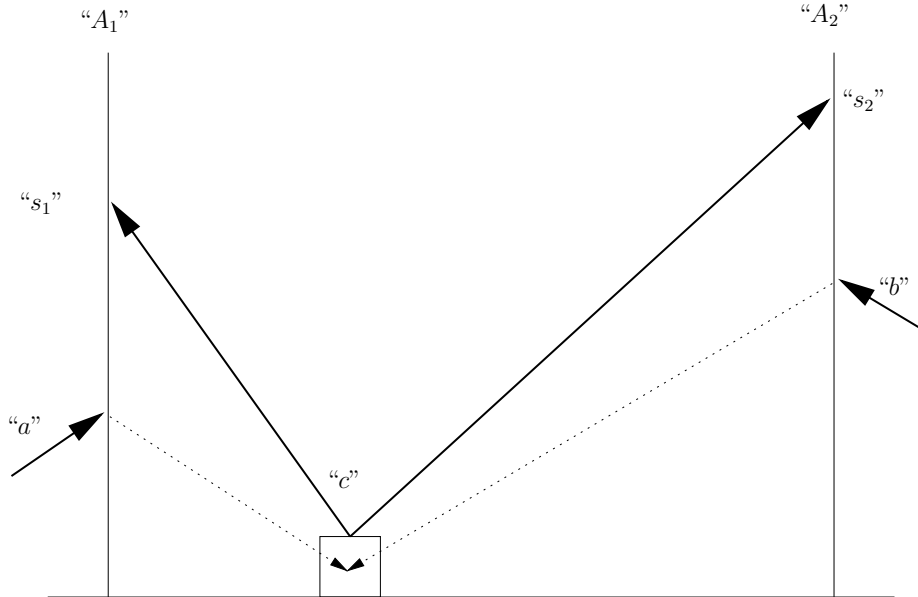


Figure 8.

In this model the setting events influence the creation event at the source. (The continuous arrows represent causal influences; the discontinuous arrows by contrast represent “influences” that need not be considered causal). They partly determine the state of the particle pair at the source. Hence the state of the particles by the time they reach their corresponding measurement devices is not the singlet state but the corresponding mixture over the possible values of spin in the direction selected by the settings. This explains causally the manifestation of both values of spin at the measurement outcome-events, and yields the appropriate statistical correlation over a large number of similarly prepared particles. And the model has the great advantage of being local, in the sense specified by both contiguity and relativity theory. All influences are transmitted by continuous worldlines in spacetime, and between events lying within the relativistic light cones (the arrows back from the setting events can be made to point as far back in the past of the creation event as we like – and thus can be pushed back into the shared part of the past lightcones of  $s_1$  and  $s_2$ ).

Yet the model has some apparent counterintuitive features from the point of view of our ordinary experience of the macroscopic world. Since in a typical EPR experiment the setting events  $a$  and  $b$  take place after the

creation event  $c$  in the rest frame of the laboratory, it follows that were this influence causal it would constitute an instance of backwards-in-time causation. But Price does not characterise it this way, preferring instead to refer to a backwards-in-time *influence*; and he defends this model as an instance of a time-symmetric explanation of physical phenomena that violates the *microscopic innocence* ( $\mu$ -*innocence*) principle.

$\mu$ -*innocence* establishes that the states of systems in microphysics record all their past interactions but none of their future ones.<sup>25</sup> So that the states of systems that have interacted in the past might well be entangled, but not so the states of systems that will interact in the future. In other words systems are innocent of any interactions that lie in their future, which can have no effect upon their present state. The corresponding principle applied to macroscopic phenomena (the macroscopic innocence principle) is of course intuitive, given the arrow of entropy defined by the second law of thermodynamics – so intuitive in fact that Price thinks that it is rarely made explicit. But it is not appropriate for microphysics, according to Price, since it is in open conflict with the time-symmetry of the equations of fundamental physics: “There really is a conflict in the intuitive picture of the world with which contemporary physics operates. [...] Our intuitive commitment to  $\mu$ -*innocence* is incompatible with T-symmetry” (Price, 1996, p. 123). Given the time-reversal invariance of the fundamental laws of microphysics, Price rejects  $\mu$ -*innocence*, and finds support for his rejection in the backwards-in-time influences model in figure 7.

But it is important to note that for Price  $\mu$ -*innocence* is not a principle of causation, and the arrow of time that it defines does not thereby provide the required asymmetry of causation. It is rather the other way round: the reason we find  $\mu$ -*innocence* intuitive in general is grounded upon our causal perspective upon the (macroscopic) world. Price adopts an agency based theory of causation, roughly:  $a$  causes  $b$  if and only if some agent can bring about  $b$  by producing  $a$ . Since, according to Price, agents are macroscopic creatures acting in the macroscopic world it follows from the time-oriented character of macroscopic phenomena that causes precede their effects, by definition.

So it is in a way misleading, according to Price, to refer to the model in figure 7 as a “backwards causation” model. For Price the idea of backwards causation only makes sense as a projection from our ordinary forward-oriented causal perspective of the macroworld. Price defends a model of this sort for the EPR correlations as the only type of model that preserves locality and a nearly fully classical understanding of the quantum world; but far from defending this as a *causal* model for the EPR correlations, he uses it precisely to show the limits of the causal perspective. The model shows

---

<sup>25</sup>(Price, 1996, pp. 120ff).

that the causal concepts that are properly employed to describe our experience of the macroscopic world, do not ultimately reflect any real properties of the physical world itself. These concepts are not properly applicable to the fundamental description of the processes underlying quantum correlation phenomena. Thus Price ends up embracing the same type of scepticism regarding causal accounts of the EPR correlations, and of quantum phenomena in general, as we have seen defended by Hausman and Van Fraassen.

## 4 Five Causal Models for the EPR Correlations

In this section I describe five different causal models, and I argue that they have not been refuted by any experimental results or theoretical considerations. (In the next section I will argue that several of these models are in no way compromised by the philosophical arguments previously reviewed). The first model is a more general version of the direct-cause model of figure 6.

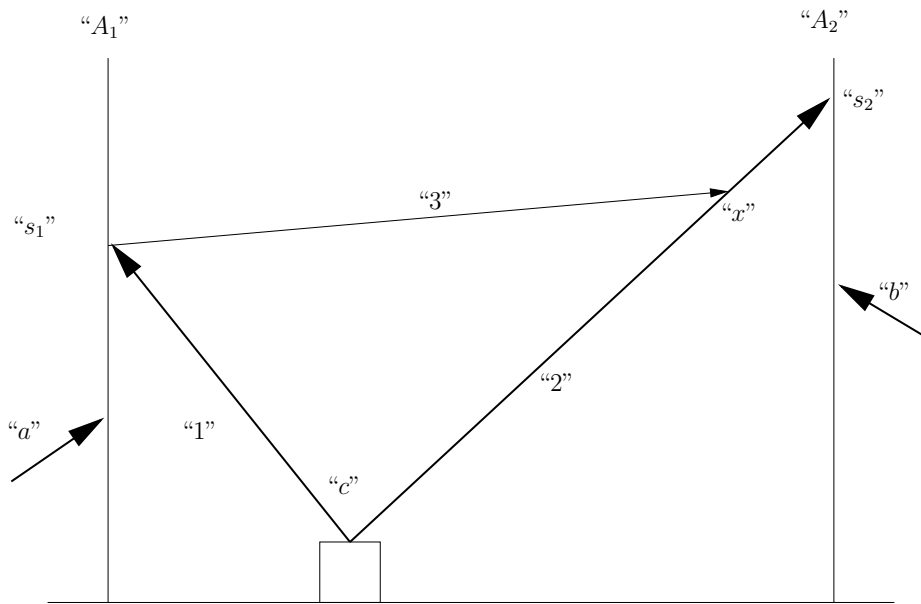


Figure 9: Causal Model I (Direct Cause)

In this model, unlike the model represented in figure 6, the causal influence does not “travel” directly from event  $s_1$  to event  $s_2$ , but it is assumed instead for the sake of generality that  $s_1$  is the cause of some change  $x$  in the state of particle “2” on its way towards the distant wing. This event  $x$  is in turn a partial cause of the subsequent spin measurement outcome event on that particle,  $s_2$ . I have argued elsewhere that this model is not actually ruled out by relativity theory, contrary to what some philosophers, including

Van Fraassen, have sometimes thought in the past. On the contrary there are at least three different readings of this model that allow it to successfully overcome any potential conflict with the theory of special relativity: an account of causal influence that does not require the transmission of energy or mass, such as a counterfactual account; an account in terms of hypothetical physical entities with superluminal velocities (tachyons); or an account that fixes a privileged frame of reference and abandons Lorentz invariance.<sup>26</sup> I will not rehearse these accounts here but will just reiterate that they remain experimentally and theoretically viable; so however controversial, Causal Model I remains a live option.

Hence it is not clear that, on Reichenbach’s Principle of the Common Cause, the EPR correlations are a case of improbable coincidence that requires explanation by means of common causes; for as we saw in section 2 correlations between directly causally implicated events are not “improbable coincidences” on Reichenbach’s definition. But let us go along with Van Fraassen in supposing so. We must then notice that Van Fraassen’s argument requires Reichenbach’s principle but also what in section 2 I referred to as Reichenbach’s criterion. Van Fraassen assumes that a Reichenbachian analysis of causation requires that correlations be explained by causal models and also that common causes necessarily screen off. But in the recent literature on causal inference the criterion has turned into an enormously controversial assumption, which is tightly related to the controversy regarding the causal Markov condition and its applicability to indeterministic phenomena.<sup>27</sup> I will not rehearse the debate here, but only mention that a straightforward non-screening off common cause model remains viable for the EPR correlations (figure 10).

In a common cause model of this sort the condition (Causality) need not generally hold:

$$\begin{aligned} \text{prob}(s_1|s_2 \wedge a \wedge b \wedge c) &\neq \text{prob}(s_1|a \wedge b \wedge c) \\ \text{prob}(s_2|s_1 \wedge a \wedge b \wedge c) &\neq \text{prob}(s_2|a \wedge b \wedge c) \end{aligned}$$

Nor does it need to hold in the particular case where the settings  $a$  and  $b$  are fixed to the same value and thus rendered irrelevant:

$$\begin{aligned} \text{prob}(s_1|s_2 \wedge c) &\neq \text{prob}(s_1|c) \\ \text{prob}(s_2|s_1 \wedge c) &\neq \text{prob}(s_2|c) \end{aligned}$$

Other common cause models are possible.<sup>28</sup> For instance, a model is possible where the common cause is not  $c$ , the emission event at the source,

<sup>26</sup>For the details see (Suárez, 2004). These issues are admirably treated in (Maudlin, 1995, chapter 5).

<sup>27</sup>(Hausman and Woodward, 1999), (Cartwright, 2000), (Hofer-Szabó, Rédei and Szabó, 1999).

<sup>28</sup>(Suárez, 2004).



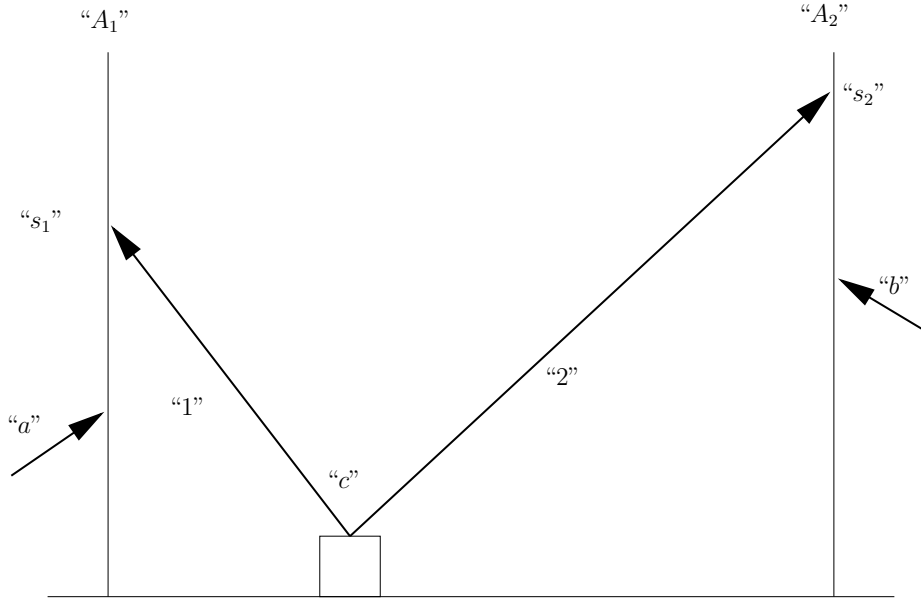


Figure 10: Causal Model II (Common Cause).

but some prior event,  $d$ . Indeed the common cause structure underlying the EPR correlations could be quite complex (figure 11).

In this causal structure:  $d$  is the partial common cause of  $c$ , but might also be a partial cause of  $a$ ,  $b$ ,  $s_1$ ,  $s_2$ ;  $c$  is the partial common cause of  $s_1$ ,  $s_2$ ;  $a$  is the partial cause of  $s_1$ ;  $b$  is the partial cause of  $s_2$ . The continuous lines represent causal influence, while the discontinuous lines represent possible causal influence. Thus the figure captures not just one, but a whole family of causal models. Individually taken, these “common causes” can not be expected to screen off their effects, on pain of a violation of factorizability. That is:

$$\begin{aligned} \text{prob}(s_1 \wedge s_2 | c) &\neq \text{prob}(s_1 | c) \text{prob}(s_2 | c) \\ \text{prob}(s_1 \wedge s_2 | d) &\neq \text{prob}(s_1 | d) \text{prob}(s_2 | d) \end{aligned}$$

However, the conjunction of both  $c$  and  $d$  might (although it need not) satisfy factorizability:

$$\text{prob}(s_1 \wedge s_2 | a \wedge b \wedge c \wedge d) = \text{prob}(s_1 | a \wedge c \wedge d) \text{prob}(s_2 | b \wedge c \wedge d).$$

In this case the failure of factorizability with respect to the initial state that yields the experimental violation of Bell’s inequalities might well be just the result of focusing our attention on a small part of the complex whole common cause structure.

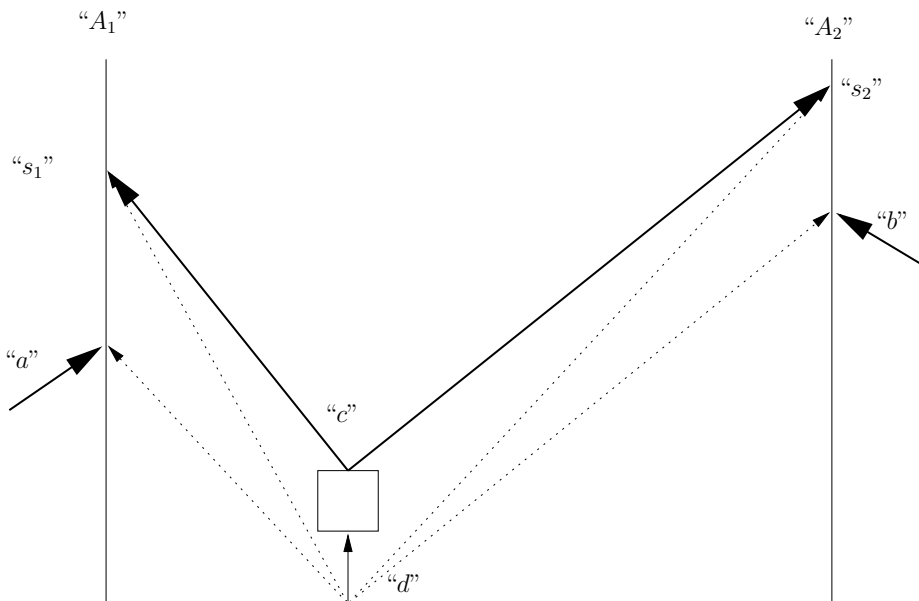


Figure 11: Causal Model III (Complex Common Cause).

Yet another possibility is a model where the common cause is not a discrete event but a whole part of a spacelike hypersurface with a value of the time parameter prior to the measurement events in both wings – we might for instance consider the hypersurface that includes the last setting event in the laboratory rest frame<sup>29</sup> (figure 12).

Other, more nuanced possibilities emerge once we realise that there is no real threat from relativity theory. For instance, a model becomes possible for the EPR correlations where the failure of factorizability is due to a failure of (Hidden Locality) rather than (Causality) as in (figure 13).

In this model a setting-event in one wing is a partial cause of the outcome-event in the opposite wing, so (Hidden Locality) is false:

$$\begin{aligned} \text{prob}(s_1|a \wedge b \wedge \Psi) &\neq \text{prob}(s_1|a \wedge \Psi) \\ \text{prob}(s_2|a \wedge b \wedge \Psi) &\neq \text{prob}(s_2|b \wedge \Psi) \end{aligned}$$

Yet the model is compatible with Reichenbach’s criterion for common causes, since (Causality) might well be satisfied, while keeping with the empirical predictions of quantum mechanics:

$$\begin{aligned} \text{prob}(s_1|s_2 \wedge a \wedge b \wedge c) &= \text{prob}(s_1|a \wedge b \wedge c) \\ \text{prob}(s_2|s_1 \wedge a \wedge b \wedge c) &= \text{prob}(s_2|a \wedge b \wedge c) \end{aligned}$$

<sup>29</sup>Some of these space-time options are described in (Butterfield, 1989).

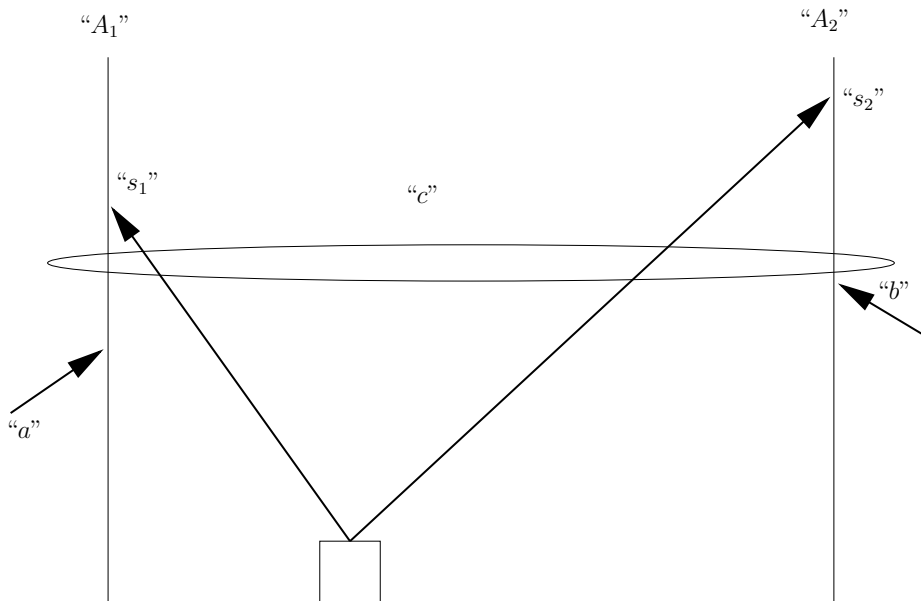


Figure 12: Causal Model IV (Hypersurface common Cause).

Finally, we must recall the type of model advocated by Huw Price, presented in section 3.3. In this model (regardless of whether we call the backwards influences causal or not) the (Hidden Autonomy) condition is violated, since the initial state of the particles at the source is statistically dependent upon the setting events, i.e.:

$$prob(\Psi|a \wedge b) \neq prob(\Psi)$$

In a thoroughly causal reading of this model (in contrast to Price's own non-causal reading), the setting-event in any of the wings is a partial cause of the state of the particles as they are emitted at the source – an earlier event in the laboratory frame:

The differences between this model and the one presented in section 3.3 (figure 8) are twofold. First, this model makes it the case that the influences from the settings to the creation event are causal in nature (therefore represented as continuous lines); so we are indeed assuming a case of genuine backwards causation. Second, we are assuming for the sake of generality that these events' influence upon  $c$ , the creation event at the source, is indirect and goes via a previous cause of this state  $d$ . This both provides the model with full generality and represents the fact, announced in section 3.3, that the model need in no way conflict with any intuitions from relativity theory, under any interpretation of the special theory, since all transmission of causal influence can be infraluminal, including the causal influence of the setting events upon the creation event at the source.

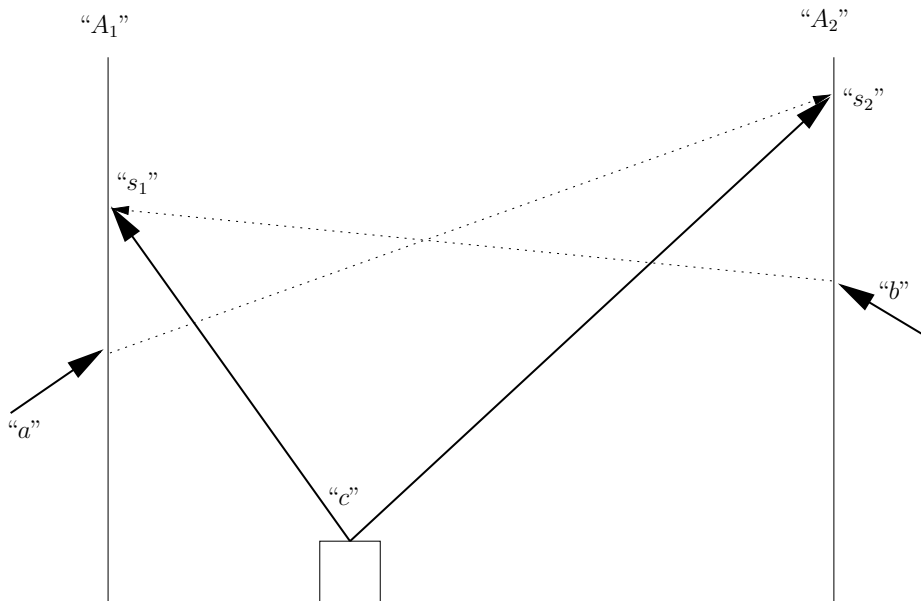


Figure 13: Causal Model V (Hidden Non-Locality).

## 5 Replies to the Arguments

I am now in a position to state and defend the main claim of this paper: The arguments so far advanced against causal models for the EPR correlations all include unwarranted assumptions and premises, and their scepticism about quantum causation is premature. But before showing this in detail, let me advance a disclaimer: I am not defending the view that causal concepts are necessary or compulsory in the description of empirical phenomena. I am not even defending the view that *there is* a causal model underlying the quantum correlations; on the issue of causation in quantum mechanics (and elsewhere in physics) the position adopted in this paper will be a thoroughly pragmatist one instead.<sup>30</sup>

The starting point is that Price’s pragmatism does not go far enough in

<sup>30</sup>In other words I oppose Nancy Cartwright’s causal fundamentalism just as much as I oppose Van Fraassen’s causal scepticism. Instead I adopt, here and elsewhere, the particularism and quietism of Arthur Fine’s NOA. The present paper is yet another NOA-driven exercise in my career: It presupposes that whether or not causality can be applied to the EPR correlations depends greatly on the details of the theory of causation employed, and on those of the particular model of the EPR correlations adopted. And there is no point attempting to answer the question in the abstract, independently of such details. For Fine’s criticisms of Cartwright’s causal fundamentalism see (Fine, 1991). My own criticism of Cartwright’s causal fundamentalism is (Suárez, 2002) – a paper that is regrettably still in press four years after it was written! I also share Price’s sceptical comments on Cartwright’s causal fundamentalism (Price, 2005, section 10).

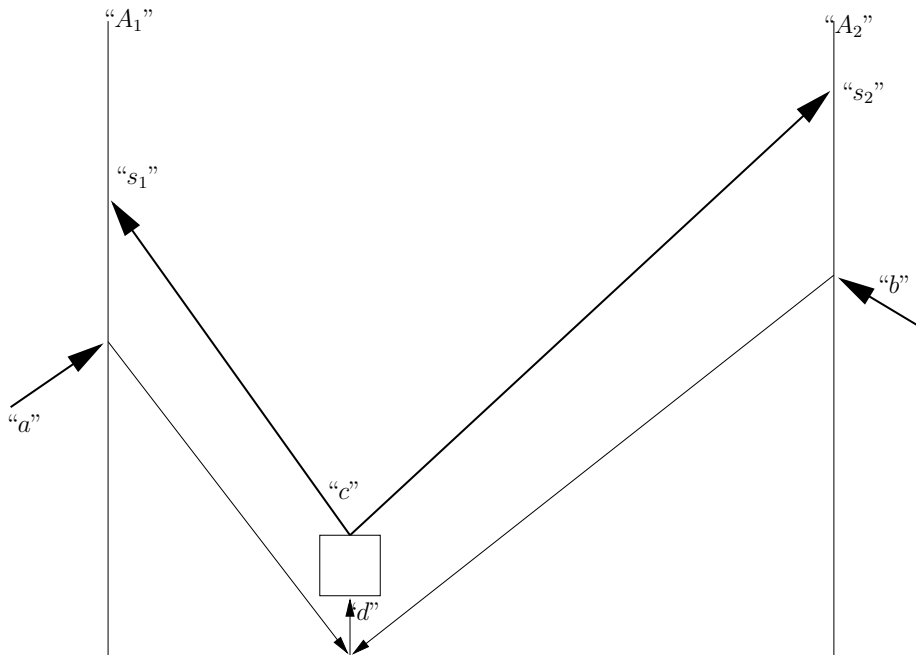


Figure 14: Causal Model VI (Hidden Non-Autonomy).

relation to causation: for the pragmatist the causal perspective ought to be a pragmatic option, which we might choose to adopt for particular purposes of explanation, prediction or coherence. So unlike Price, I will not take the causal perspective to be compulsory in macrophysics, but neither will I see it as incompatible with microphysics. Causal models can be made available for virtually any correlation phenomena. Provided that the completeness assumption discussed in section 2 is rejected (and provided enough imagination!) any causal structure can be suitably expanded to provide some causal model for any correlation phenomena, including the EPR correlations. We are not compelled to accept any of these models (and we might be forced to reject some of them on empirical grounds) but we are at freedom to adopt any of those not ruled out by the physics if we wish for pragmatic purposes.

The models discussed in the previous section are all permitted by the physics, and we are therefore at freedom to adopt any of them on pragmatic grounds. These models are very helpful in displaying the unwarranted pre-suppositions and premises in the arguments, reviewed in section 3, against causal accounts for the EPR correlations. I will discuss each set of arguments against these models one by one.

## 5.1 Reply to Van Fraassen’s Reichenbachian arguments

Causal model I shows that Van Fraassen’s relativistic argument against direct-cause models is too harsh. The issues are threefold. First, it is unclear whether the concept of “travelling” is at all appropriate for this type of model: there are accounts of causal influence that do not require any material object travelling from cause to effect. Second, the existence of superluminal entities (tachyons), albeit controversial, has not been ruled out yet by the physics. Third, it is always possible that the EPR experiment fixes a privileged frame of reference (other interpretations of quantum mechanics do too, notably Bohm’s), in which case there is no issue of “infinite speed” travelling and the concept “travel” may apply fully. Since causal model I remains viable under any of these options, Van Fraassen’s brief argument against superluminal causation models is inconclusive.

Causal model II shows that Van Fraassen’s assumption of Reichenbach’s criterion, in addition to Reichenbach’s principle, is problematic too. Van Fraassen conflates principle and criterion, but twenty years on these two commitments seem clearly distinct. For there is now a large body of literature on the topic of causal inference that finds the criterion deeply controversial while agreeing fully with the principle.<sup>31</sup> Moreover model II is not ruled out by the physics. First the causal influences in the model are all infraluminal, so no relativistic issues arise. Second, since the common cause does not screen off, (Causality) will not hold in this causal model, and the failure of factorizability that gives rise to the violation of Bell’s inequalities can be accommodated in a straightforward and simple way.

Causal model III shows that the failure of (FACT) with respect to the creation event at the source (or more generally, the quantum state of the particle pair) can in principle be accommodated within a larger common cause structure. Moreover this larger structure might even satisfy Reichenbach’s criterion, not just his principle. Again the physics does not yet rule this out, although no credible candidates for the deeper common cause structure have been suggested to date.

Finally, models V and VI show that Van Fraassen is too rashly blaming the failure of factorizability upon (Causality) alone. (Hidden Locality) could be the culprit instead, and this would be so in a causal model in which the

---

<sup>31</sup>And since the criterion is controversial, the strongest defence today of the principle will abandon the criterion altogether. Thus the strongest defence of the principle of common cause takes it that common causes do not generally screen off. For a defence of this point of view see (Cartwright, 1988; Cartwright, 1989). Although I am adopting the principle for the purposes of this paper, I am aware of possible limitations – such as Sober’s well known argument against the application of the principle to arbitrary monotonically increasing time series. Sober’s argument is not relevant to the present discussion because the EPR correlations are not correlations between time-series, and his rejection of the principle in general is consistent with my adoption of it for the purposes of this paper.

setting events in one wing are causes, either directly of the measurement outcome events in the distant wing or indirectly via the common cause. The latter case is a particularly interesting case of failure of (Hidden Locality) since it also entails the failure of (Hidden Autonomy) and requires backwards causation. Model V by contrast does not entail backwards causation in the laboratory rest frame, but the relativistic considerations must once again be addressed. But neither case is precluded by the physics itself.

Hence Van Fraassen's argument contains three sets of presuppositions that require far more discussion than has been provided in the philosophy of quantum mechanics literature so far. First, it is in no way automatic that relativity rules out any of these causal models. Second, it seems at least necessary to distinguish between Reichenbach's principle and Reichenbach's criterion and to discuss their implications for each of these models. Third, it seems premature to put the blame for the failure of factorizability solely upon the (Causality) condition. I am suggesting here that there are definite avenues for defending that the models are not ruled out by relativity, that they might conflict with Reichenbach's criterion but not with his principle, and that it might well be the case that (Hidden Non-Locality) and not (Non-Causality) is to be blamed for the failure of factorizability in some of these causal models.

## 5.2 Reply to Hausman's independence argument

Let us now turn to Daniel Hausman's argument, reviewed in section 3.2. Hausman claims that the independence condition fails in general in quantum mechanics, and hence the quantum mechanical correlations are an exception to his theory of causation (C): they are established law-like correlations that can not be given a causal explanation. I believe the argument is mistaken on two grounds. First, I will argue that there are causal models for the EPR correlations that obey Hausman's independence (I) condition.<sup>32</sup> This might look like good news for Hausman, since it points out that some of these models are amenable to his theory; if so quantum mechanics is not the exception to his theory that he has taken it to be. But secondly, I will argue that at least some of the causal models for the EPR correlations are not amenable to this type of treatment. In these models the independence condition fails. However, the models are clearly causal, or at least they have been taken as such by virtually anyone in the literature. Hence, regrettably for Hausman's theory, causal concepts have been applied to situations that Hausman's theory rules out as non-causal. My conclusion will therefore be that Hausman's theory does not capture all conceptual truths about causation. If this is correct then, contrary to what he claims, the independence

---

<sup>32</sup>I conjecture that these models can be extended in a straightforward manner to cover the GHZ correlations.

condition (I) can not a conceptual truth about causation.

I maintain that model II satisfies Hausman’s theory of causality, including the independence assumption (I). The contrast is clearest with model IV which having a similar structure, nonetheless fails to satisfy Hausman’s independence condition (I). In model II the creation event at the source “ $c$ ” is the common cause of both measurement outcome events  $s_1$  and  $s_2$ . We already noticed in section 3.2 that in this scenario the independence condition becomes:

Independence Condition for EPR (I for EPR) : If  $s_1$  causes  $s_2$  or  $s_1$  and  $s_2$  are n-connected only as effects of some common cause, then  $s_2$  has a cause that is distinct from  $s_1$  and not n-connected to  $s_1$ .

This condition is satisfied because  $b$  is a cause of  $s_2$ , distinct from  $s_1$ , and not n-connected to  $s_1$ . The condition can be conversely applied to  $s_1$  in relation to  $s_2$ , with  $a$  as the independent cause of  $s_1$ . It is important to notice how the setting events  $a$ ,  $b$  are genuine independent causes, that can be used to affect the values of  $s_1$  and  $s_2$ . We can choose the setting events as late as we like. Suppose that event  $s_1$  takes place first, then we can choose at a later setting event  $b$  the orientation of the magnetometer that we desire, therefore fixing the decomposition of the singlet state for that particle, and the corresponding probabilities. Since the measurement event outcomes are just one of two (plus or minus) we effectively alter the probability of each possible outcome. On any probabilistic theory of causation  $c$  is a cause of  $e$  if  $c$ ’s occurrence changes the value of the probability of  $e$ . Thus each of the causally related events  $s_1$  and  $s_2$  have their own “leverers” or “handles” that can help to fix the probability for their values independently, namely the setting events  $a$ ,  $b$  in their respective wings.

There is a crucial difference between model II and the specific scenario that I described in section 3.2 as an illustration of Hausman’s theory. In that scenario all the settings are fixed in advance, so there is no “handle” *ex hypothesis*. But in the scenario described by model II there is a well defined handle on the effect’s wing for any setting and outcome event on the distant wing. True, an agent can make use of this “handle” to bring about differences in the probabilities as desired only if he or she has full information regarding the setting and outcome on the distant wing. However, the notion of “handle” presupposed by Hausman is ontological, and unaffected by the lack of knowledge of a particular situated agent, so this fact is irrelevant to the causal nature of model II.

The independence condition fails, by contrast, in model IV, even if this model has a very similar common cause structure. The main difference is that in model IV the common cause is a complete hypersurface of spacetime that contains  $b$  and (all the effects of)  $a$ . So it is impossible in this model to



control the value of  $a$  and  $b$  independently of the common cause  $c$  itself. Independence fails here because there is no cause of  $s_2$  that is not n-connected to  $s_1$ . So, according to Hausman's theory model IV is not a genuine causal model.

Then there are versions of model I that satisfy Hausman's conditions too, and are thus fully causal under his own theory. For instance, take the version of model I discussed by Van Fraassen (figure 6), where the event  $s_1$  is a direct partial cause of the event  $s_2$ . In this model the causal influence takes some time to reach  $s_2$  since this event occurs some time after its cause  $s_1$ . Now any setting event  $b$  on the measuring device in the wing of the experiment of particle "2" that takes place after event  $s_1$  will also be an effective leverer or handle for  $s_2$ , just as previously discussed. In the absence of any setting events, however (or if the settings have been preordained to be the same well before the experiment is run) the model will too fail to satisfy the independence condition. In the more general kind of model I described in figure 9, the causal relation between the intermediate event  $x$  and its partial effect  $s_2$  on that wing will satisfy independence, but not so the causal relation between  $s_1$  and  $x$ .

Similarly some versions of model III comply with independence, while others do not. A version of model III with continuous arrows between the deeper common cause  $d$  and the measurement outcome events  $s_1$  and  $s_2$ , and between  $d$  and the setting events  $a$  and  $b$  will not satisfy independence, as in this model every cause of  $s_1$  is n-connected to every cause of  $s_2$  and viceversa. However, a version of model III in which there are no arrows between  $d$  and  $a$  and  $b$  satisfies independence since there is some cause of  $s_1$ , namely  $a$ , which is not n-connected to  $s_2$ , and conversely there is some cause of  $s_2$ , namely  $b$ , which is also not n-connected to  $s_1$ . The details are crucial here to assess whether or not independence obtains.

On the other hand causal models V and VI unambiguously fail Hausman's test. For instance in model V there is no cause of  $s_2$  that is not n-connected to  $s_1$  *including* the setting event  $b$  which in this model is ex-hypothesis a partial direct cause of  $s_1$ . Or take model VI in its full causal reading: no cause of  $s_1$  ( $s_2$ ) fails in this model to be n-connected to  $s_2$  ( $s_1$ ), even if it is via a backwards in time causal influence upon the common cause  $c$ . Regardless of how we interpret the model as long as the influences are taken to be causal, the independence condition will fail, and Hausman's independence theory of causation (C) will be violated.

But notice what a bizarre consequence this seems to be. We have concluded that models IV, V, VI and some admissible versions of models I and III all fail the independence test. Hausman's analysis would unambiguously classify them as failed causal models, not because they conflict with the physics or the experiments, but simply because they fail his independence

test. And this is supposed to be as a result of some conceptual truths about causation, such as the independence condition. I find it then extremely implausible that these models would have been the subject of any debate regarding their possible incompatibility with special theory of relativity. For notice that the only reason why the threat of incompatibility has bothered physicists and philosophers alike is the causal reading of the special relativity theory, and particularly the light cone structure. If models I, III, IV, V and VI are not properly causal why would anyone have cared about their incompatibility with special relativity in the first place? Yet we intensely care about the incompatibility issue. What this shows, I believe, is that these models are eminently causal, and universally thought to be so among both philosophers and physicists. There seems to be nothing in our concept of causation to prevent us thinking that these models are causal (we might disagree on whether they are *plausible* models, but we do not disagree over the fact that they are causal in nature). Put in another way: if it fell naturally out of our concept of causation that these models for the EPR correlations are not causal, why would anyone like Van Fraassen (who presumably shares our concept) spend so much time and effort trying to rule them out as impossible or implausible?

To sum up, it seems to me that Hausman was wrong to rule out all models of the EPR correlations as non-causal, since at least three of these models satisfy his conditions. Instead of using his theory to rule out all causal models for the EPR correlations, the independence condition can be more helpfully applied to distinguish between these models. But then the fact that another four of these models are ruled out as non-causal by Hausman's theory is, I believe, just as much of an indictment of his theory – for it shows that our concept of causation, which we seem perfectly prepared to apply to microscopic physics, is neither exhausted by nor in agreement with Hausman's independence condition. These models seem to refute Hausman's claim that independence (the (I) condition) is a conceptual truth about causality.

### 5.3 Reply to Price's asymmetry argument

For the purposes of analysis it is helpful to explicitly list the premises in Huw Price's argument, reviewed in section 3.3. Price assumes that

- i) causation is agent based notion, roughly:  $a$  causes  $b$  if and only if an agent could bring about  $b$  by producing  $a$ .
- ii) this is explicitly a modal notion: what counts is what agents would be able to do in the appropriate circumstances, and not what a particular agent is capable of doing.

- iii) Since causation is agent-based, it is limited to the macroscopic world that agents operate in.
- iv) In particular the agent-based theory is not directly applicable to microphysics.
- v) The only sense of causation that would be applicable in microphysics is a projection from our macroscopic based concepts of causation.
- vi) Our macroscopic concept of causation is time-oriented, because agency in the macroscopic world always is: we act in order to bring about effects in the future, not in the past.
- vii) In other words, the macroscopic concept of causation is tied up to a notion of macroscopic innocence, according to which systems that interact keep a record of the interaction in their future history but not in their past history.
- viii) But the equivalent notion of innocence in microphysics ( $\mu$ -*innocence*) is false, since it conflicts with the fact that the fundamental laws of microphysics are invariant under time reversal, or T-symmetric.
- ix) Hence properly speaking there is no causation in microphysics.
- x) It is possible to build models of the EPR correlations that abandon  $\mu$ -*innocence*, such as model VI in section 4.
- xi) This model is not properly speaking causal, since it lacks the temporal asymmetry of our ordinary agent-based concept of causation.
- xii) But the model uniquely solves the problem of explaining the EPR correlations, without any appeal to non-locality.
- xiii) This in turn strengthens the claim (ix) that there is no genuine causation in microphysics.

Let us first consider Price's claim regarding model VI. Price assumes that this is not a properly speaking causal model. The arrows pointing back from the setting events  $a$  and  $b$  towards the common cause  $d$  are not to be thought of as causal influences (and should not be represented as continuous world lines). But it is very hard to see why, from the point of view of an agent-based theory of causation, these influences should not be counted as perfectly causal. After all if an agent was interested in bringing about a particular outcome of the spin measurement on particle "2", the best they could possibly do would be to orient the magnetometers in a particular direction. True, they would have to have some knowledge about the orientation of the distant magnetometer, in order to be able to change the probabilities of the possible outcomes in their own wing. And if they

knew both distant setting and distant outcome, then they would actually be able to bring about a plus or a minus outcome, with probability one, as the measurement outcome event in their own wing, simply by orienting the magnetometer appropriately. It is hard to see why this would fail to qualify as a proper intervention to bring about an effect. The fact that the causal influence must first go backwards in time to change the initial state of the particle pair before affecting the final outcome on the distant wing seems, from the strict point of view of an agency theory, neither here nor there. At least in this scenario the fact that some of the causing is backwards seems irrelevant to the possibility of the kind intervention demanded by an agency theory.

What I think this reflection shows is that Price conflates two different causal commitments (or axioms) into one. There is the commitment strictly embraced by an agency theory, which is expressed in our premisses i) and ii) above. Then there is the commitment deriving from the time-symmetry of causation in the macroscopic domain, namely that causes always precede effects. In the macroscopic world, given the second law of thermodynamics, these commitments coincide; and the agency theory then builds in an asymmetry in its concept of causation. This is in effect the implication of premisses iii) to vii) in Price's argument. But the two commitments come apart in the microscopic domain, where the second law no longer applies. Price extracts the conclusion that agency theories are not applicable there, and therefore causation overall fails (premise ix). But one could extract precisely the opposite implication, namely that agency theories no longer carry a commitment to temporal asymmetry in that domain. In other words, the microscopic domain forces us to distinguish between our two commitments regarding the nature of causation, and then choose which of them we will take as primary. If we take the temporal asymmetry commitment as primary then model VI is not a causal model, by definition, since the influences from the settings to the source are back-in-time. But if we take the agent-based commitment to be primary then model VI is a *causal* model, since agents set on bringing about particular outcome-values in the distant wing can affect the probabilities for those outcomes by means of their settings (as long as the settings in the other wing are known to them in advance).

It is unclear to me which of these two commitments is primary for Price. What is clear is that he has run them together in the microscopic domain in a way that I believe to be both mistaken and unhelpful. If instead we distinguish these commitments we can have a far more detailed and nuanced analysis of the status of agent based causality in the EPR experiment in particular, and in the domain of microphysics in general.

We can now approach all the other causal models in the light of this distinction. We will see that in some of these models we are in effect forced to make a similar choice between agency-based account and a temporal

asymmetry account. Models I, V and VI, and some versions of model III are causal in the agency-based sense but not in the temporal asymmetric sense if we apply relativity theory and consider time order in *all* frames of reference. Model IV is causal in the time-asymmetry sense but not in the agent-based sense (since clearly no agent can control a whole hypersurface of spacetime).

Model II is a curiously problematic model for Price's argument. This model is causal in both the agency-based sense and the temporal asymmetric sense. The cause  $c$  precedes the effects  $s_1$  and  $s_2$ , and the state can be altered so as to bring about different statistics (by e.g. replacing the singlet with a different kind of entangled state). Indeed in this model the conflation between the agency-based commitment and the time-asymmetric commitment that we saw was operating in the macroscopic domain is fully restored even though it is a model of phenomena in the microscopic domain. A perfectly legitimate causal model of microphysical correlations that obeys both of Price's causal axioms! Model II can hardly be said to be less explanatory than Price's own model, among other things because it does away with relativistic non-locality altogether, like Price's model. So it is difficult to see why anyone who is committed to both axioms regarding causality would opt for model VI at the expense of model II. The problem for Price is that he is one of those people apparently committed to both axioms, which suggests that on the grounds of his own preferred theory of causation he should abandon the model that he has been defending for the EPR correlations for the last twenty years, and adopt instead the simpler model II.<sup>33</sup>

In situations where the settings are predetermined, however, model II is not causal in the agency-based sense, since there is no handle for agents to operate to bring about effects independently. So the model becomes causal only in the temporal asymmetric sense. If Price preferred to choose model VI instead of model II he would be showing that his real commitment lies with the agency-based axiom, not the temporal asymmetric axiom. In any case, the choice between all these different models, in their different interpretations, opens up a route to test the real commitments underlying Price's theory of causation. My guess is that he would try to stick to both for as long as possible but if forced to choose might opt for temporal asymmetry. In either case, once the choice has been made between these two commitments, plenty of possibilities are still available to explain causally the EPR correlations: the causal perspective has been shown to be just as good for microphysics (and microphysicists!) as it is for ordinary cognition

---

<sup>33</sup>But of course the price to pay would be Price's argument for causal perspectivalism which requires us to accept that causation is not part of the ultimate furniture of the universe as described by fundamental physics. A thorough pragmatist should have no fear to pay that price, I contend, since a causal description of phenomena is a pragmatic option, available to physicists as well as anyone else.

and practice.

## 6 Conclusions and Prospects

I hope to have shown that all arguments so far to the effect that causation is ruled out in microphysics on account of the EPR correlations are unwarranted and premature. Much more philosophical work still needs to be done in order to establish the ways in which the different theories of causation can be applied to explain quantum correlation phenomena. It seems also clear that the question “are the EPR correlations causal?” in general has no informative answer. To answer this question we have to engage with the details both of the different theories of causation and the different possible models for the EPR correlations. Different combinations of causal theories and empirical models will yield different answers to this question.

In his brilliant 1998 book on quantum probability and non-locality, Michael Dickson showed how typical metaphysical questions regarding the notion of locality in quantum mechanics (such as “is the nature of the quantum world non-local?”) have no general or universal answers. These questions can only be informatively answered by considering how the different notions of locality fare with respect to each and every interpretation of quantum mechanics. And we should expect the answers to be different and even contrary in different cases.<sup>34</sup> The kind of non-locality built into Bohm’s theory is not the same as is built into Ghirardi-Rimini-Weber theories, or the modal interpretation. The same lesson I think applies to the nature of causation in quantum mechanics. We are unlikely to learn much from attempts to completely rule causation out, or completely rule it in. We will probably learn much more if we proceed in an unprejudiced and cautious way to a detailed and piecemeal study of all the different concrete possibilities instead.

## References

- Bohm, D. (1951) *Quantum Theory*. Prentice Hall.
- Bohm, D. (1952) “A Suggested Interpretation of the Quantum Theory in Terms of Hidden Variables, I and II”. *Physical Review*, **85**: 369–396.
- Bovens, L. and Hartmann, S. (Eds.) (2006) *Nancy Cartwright’s Philosophy of Science*. Routledge.

---

<sup>34</sup>(Dickson, 1998). In my review of this book (Suárez, 2000) I embraced the position, which I referred to as a deflationism, and argued that Dickson actually did not go far enough in its defence!

- Butterfield, J. (1989) “A Space-Time Approach to the Bell Inequality”. In J. Cushing and E. McMullin (Eds.), *Philosophical Consequences of Quantum Theory*, pp. 114–144. University of Notre Dame Press.
- Cartwright, N. (1988) “How To Tell a Common Cause: Generalizations of the Conjunctive Fork Criterion”. In J. H. Fetzer (Ed.), *Probability and Causality*, pp. 181–188. Reidel Pub. Co.
- Cartwright, N. (1989) *Nature’s Capacities and their Measurement*. Oxford University Press.
- Cartwright, N. (2000) “Against Modularity, the Causal Markov Condition, and Any Link Between the Two: Comments on Hausman and Woodward”. *The British Journal for the Philosophy of Science*, **53**: 411–453.
- Cushing, J. and McMullin, E. (Eds.) (1989) *The Philosophical Consequences of Quantum Theory*. Notre Dame University Press.
- de Beaugrand, O. C. (1977) “Time Symmetry and the Einstein Paradox”. *Il Nuovo Cimento*, **42B**: 41–64.
- Dickson, M. (1998) *Quantum Non-Locality and Probability*. Cambridge University Press.
- Dummett, M. (1954) “Can an Effect Precede Its Cause?” *Proceedings of the Aristotelian Society, Supp. Volume*, **38**: 27–44.
- Einstein, A., Podolsky, B. and Rosen, N. (1935) “Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?” *Physical Review*, **47**: 777–780. Reprinted in Wheeler and Zurek (1985), pp. 138–141.
- Fine, A. (1982) “Hidden Variables, Joint Probability and the Bell Inequalities”. *Physics Review Letters*, **48**: 291–295.
- Fine, A. (1987) *The Shaky Game: Einstein Realism and the Quantum Theory*. Chicago University Press.
- Fine, A. (1989) “Do Correlations Need to Be Explained”. In J. Cushing and E. McMullin (Eds.), *Philosophical Consequences of Quantum Theory*, pp. 175–194. University of Notre Dame Press.
- Fine, A. (1991) “Piecemeal Realism”. *Philosophical Studies*, **61**: 79–96.
- Hausman, D. (1998) *Causal Asymmetries*. Cambridge University Press.
- Hausman, D. (1999) “Lessons from Quantum Mechanics”. *Synthese*, **121**: 79–92.

- Hausman, D. M. and Woodward, J. (1999) “Independence, Invariance and the Causal Markov Condition”. *The British Journal for the Philosophy of Science*, **50**: 521–583.
- Hofer-Szabó, G., Rédei, M. and Szabó, L. (1999) “On Reichenbach’s Common Cause Principle and Reichenbach’s Notion of Common Cause”. *The British Journal for the Philosophy of Science*, **50**: 377–99.
- Hughes, R. I. G. (1989) *The Structure and Interpretation of Quantum Mechanics*. Harvard University Press.
- Jarrett, J. (1984) “On the physical significance of the locality conditions in the Bell arguments”. *Nous*, **18**: 569–589.
- Maudlin, T. (1995) *Quantum Non-Localilty and Relativity*. Oxford Blackwells.
- Price, H. (1996) *Time’s Arrow and Archimedes’ Point*. Oxford University Press.
- Price, H. (2005) “Causal Perspectivalism”. In H. Price and R. Corry (Eds.), *Causation, Physics and the Constitution of Reality: Russell’s Republic Revisited*. Oxford University Press.
- Price, H. and Corry, R. (Eds.) (2005) *Causation, Physics and the Constitution of Reality: Russell’s Republic Revisited*. Oxford University Press.
- Reichenbach, H. (1956) *The Direction of Time*. University of California Press.
- Salmon, W. (1984) *Scientific Explanation and the Causal Structure of the World*. Princeton University Press.
- Schrödinger, E. (1933) “The Present Situation in Quantum Mechanics: A Translation of Schrödinger’s ‘Cat Paradox’ Paper”. *Proceedings of the American Philosophical Society*, **124**: 323–38. Reprinted in Wheeler and Zurek (1985).
- Suárez, M. (2000) “The Many Faces of Non-Localilty: Dickson on the Quantum Correlations”. *British Journal for the Philosophy of Science*, **51**: 882–892.
- Suárez, M. (2002) “Experimental Realism Defended: How Inference to the Most Likely Cause Might be Sound”. Delivered at the Konstanz conference in honour of Nancy Cartwright, December 2002, and forthcoming in Bovens and Hartmann (eds.).
- Suárez, M. (2004) “Causal Processes and Propensities in Quantum Mechanics”. *Theoria*, **19**: 271–300.



van Fraassen, B. C. (1982) “The Charybdis of Realism: Epistemological Implications of Bell’s Inequality”. *Synthese*, pp. 25–38. Reprinted with corrections in J. Cushing and E. McMullin, eds. (1989).

Wheeler, J. and Zurek, W. (Eds.) (1985) *Quantum Theory and Measurement*. Princeton University Press.