The Principle of the Common Cause faces the Bernstein Paradox

Jos Uffink *

September 20, 1999

Abstract

I consider the problem of extending Reichenbach's principle of the common cause to more than two events, vis-à-vis an example posed by Bernstein. It is argued that the only reasonable extension of Reichenbach's principle stands in conflict with a recent proposal due to Horwich. I also discuss prospects of the principle of the common cause in the light of these and other difficulties known in the literature and argue that a more viable version of the principle is the one provided by Penrose and Percival (1962).

1 Introduction

The principle of common cause (PCC) was proposed by Reichenbach in an attempt to characterize the asymmetry of time by exploiting a particular statistical distinction between cause and effect. The rule states, roughly speaking, that every occurrence of a pair of statistically correlated events is due to some cause operative in their common past. Reichenbach believed that this feature could be used to define the distinction between past and future.

This rule has been studied by many philosophers, and many of its shortcomings have already been pointed out. (See Van Fraassen (1980), Cartwright (1983, 1988), Sober (1984, 1988), Torretti (1987), Arntzenius (1990, 1993), Price (1996).) It emerges from this literature that although some authors sympathize with Reichenbach's project, there are only few willing to defend the principle in its original form. Accordingly, several variants of the principle have been proposed (Penrose & Percival, (1962), Salmon (1984), Horwich (1987)).

These studies have mostly been carried out in the context of the debate on time asymmetry and/or the nature of causation. The present paper starts by shelving these more difficult issues and studies the principle as a statement within probability theory. In particular, I discuss the problem of extending it to correlations between more than two events. In this way I hope to extricate some guidelines which shed light on the viability of proposed variants of the principle. In the final section, I argue that explicit introduction of spatiotemporal structure is needed to obtain a viable version of the principle.

*Institute for History and Foundations of Mathematics and Science, University of Utrecht, P.O.Box 80.000, 3508 TA Utrecht, The Netherlands (uffink@phys.uu.nl

2 The principle of the common cause

Reichenbach's original introduction of his principle was in the form

> "If an improbable coincidence has occurred, there must exist a common cause." (Reichenbach 1956, p. 157.)

This formulation is, of course, rather vague. However, he also provided a precise mathematical formulation. Consider two simultaneous random events A and B and suppose there is a positive correlation between them:

$$P(A\&B) > P(A)P(B). \tag{1}$$

Such a correlation would, according to Reichenbach, be in need of explanation; it is "a joint occurrence more frequent than can be expected for chance events" (ibid. p. 158).

The PCC states that in all such cases one can find an event C (the common cause) such that

$$\begin{aligned} P(A\&B|C) &= P(A|C)P(B|C), & (2)\\ P(A\&B|\neg C) &= P(A|\neg C)P(B|\neg C), & (3)\\ P(A|C) &> P(A|\neg C), & (4)\\ P(B|C) &> P(B|\neg C). & (5) \end{aligned}$$

A triple of events A, B, C obeying these conditions is called a conjunctive fork and often depicted as in Fig. 1.

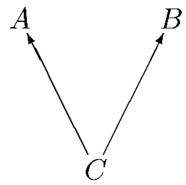

Figure 1: the conjunctive fork

A curious point is here that whereas Reichenbach's verbal formulation referred to 'improbable' coincidences, his mathematical version nowhere assumes that $P(A\&B) \ll 1$. Apparently, he intended the qualification 'improbable' for the coincidence of A and B to express that it occurred in unexpected or surprising frequency, and not in the literal sense that its probability is small.

The conditions (2, 3) can be regarded as the essential part of the principle. They express that C 'screens off' the correlation, that is to say, if we conditionalize the joint distribution of A and B on cases where C is present (absent), we recover their independence. By comparison, the conditions (4, 5) are more or less cosmetic. Presumably, they were added by Reichenbach because (2) and (3) by themselves are symmetrical with regard to the interchange of C and $\neg C$. This would conflict with every-day usage of the word 'cause', as referring to something favorable to the occurrence of its effects. Thus, while a heavy smoking habit is generally considered to be a common cause for developing cancer and heart diseases, General Surgeons do not issue the official warning that abstinence from cigarette smoking also causes serious health damages. However, one can easily circumvent this potential conflict by referring to C and $\neg C$ more neutrally as 'causal factors' or 'screening factors'.

Another restriction in the principle which seems unimportant is that Reichenbach only considered *positive* correlations. It might thus seem as if anticorrelations, i.e. cases where

$$P(A\&B) < P(A)P(B)$$

do not stand in need of explanation. This point is inessential as long as one allows that A and B are arbitrary. It suffices to replace A by $\neg A$, or B by $\neg B$, to change a negative correlation into a positive one.

For the rest of the paper I propose to strip the principle from these cosmetic aspects and use the following formulation: whenever there are events A, B such that

$$P(A\&B) \neq P(A)P(B),$$

there exists an event C such that

$$P(A\&B) = P(A|C)P(B|C)P(C) + P(A|\neg C)P(B|\neg C)P(\neg C) \tag{6}$$

Obviously, this equation also implies that the probabilities of $A\&\neg B, \neg A\&B$ and $\neg A\&\neg B$ decompose in the same way.[1] The mathematical content of the principle is thus simply that any joint probability distribution which does not factorize itself is a convex combination of two factorizing probability distributions.

Pending a more serious discussion (section 5) a few comments are in order.

1. The motivation for the principle is the background intuition that independence of two random events is 'natural', or at least not in need of explanation. Any deviation of this, i.e. a correlation, ought to be explained, and the explanation is in fact provided by showing their conditional independence, thus recovering, in Reichenbach's line of thought, what we ought to expect.

Although one can cite many examples from physics and every-day life where this idea seems natural, its validity is sometimes put in doubt. A well-known example given by Sober (1988) is a correlation between bread prices in Britain and sea levels in Venice (both have been steadily rising over the last centuries). This provides a case of a correlation for which, intuitively, no common cause explanation is needed.

On the other hand, in cosmology a well-known problem arises of the remarkable uniformity of cosmic background radiation. Here, it is the absence of correlation (independence of intensity and direction of the radiation) which is by many scientists considered to be in need of explanation. This, no doubt, is due to the fact that the uniformity of cosmic radiation can also be presented as a striking correlation between the distant sources of this radiation. But then it appears that what we consider to be correlations or not depends on our way of posing the problem.

These and other problems show that the background intuition has only limited validity. The fact that Reichenbach's formulation allowed the events A, B to be arbitrary (except for the demand that they are simultaneous) makes it implausible that their independence is always the natural state of affairs. (There might simply exist logical relations between A and B, e.g. entailment or exclusion.)

[1] Thus, if one wishes, one can always choose one pair of events such that the inequalities (4) and (5) are recovered.

2. However, if we consider the mathematical form, one notes that the validity of (6) is trivial. That is to say, any 2×2 probability distribution can be decomposed into a convex sum of two independent distributions. (It suffices to take $C = A$ or $C = B$ or any event in 1-1 correlation to these to obtain (6).) The rather wide-spread opinion that the principle has been shown to be false in quantum mechanics, due to Bell's theorem, overlooks the fact that this theorem is based on assumptions of locality, something about which the PCC is silent.

This lack of empirical content is, of course, not an objection against the PCC. The principle was not intended to rule out particular joint distributions from occurring in nature. Its philosophical value lies rather in the claim (which I shelved) that the screening factors C and $\neg C$ always exist in the *past* of A and B, and in this way provide a criterion for the distinction of past and future.

3. A more serious problem for the PCC is that when some factor C succeeds in screening off the correlation, additional factors may very well succeed in restoring it, or even bring about an opposite correlation. To be precise, suppose C screens off a correlation between A and B:

$$P(A\&B|C) = P(A|C)P(B|C) \tag{7}$$

$$P(A\&B|\neg C) = P(A|\neg C)P(B|\neg C). \tag{8}$$

There may well exist a further event D (anywhere in the past or future of C) such that

$$P(A\&B|C) = P(A\&B|C\&D)P(D|C) + P(A\&B|C\&\neg D)P(\neg D|C) \tag{9}$$

$$P(A\&B|\neg C) = P(A\&B|\neg C\&D)P(D|\neg C) + P(A\&B|\neg C\&\neg D)P(\neg D|\neg C) \tag{10}$$

with

$$P(A\&B|C\&D) \neq P(A|C\&D)P(B|C\&D) \tag{11}$$

etc., so that the correlation between A and B is reintroduced. Indeed every factorizing distribution (except the extreme ones consisting of only zeroes and a one) can be decomposed in a sum of non-factorizing distributions, thus reversing the relationship in (6). The most obvious example for this is the factorizing 2×2 distribution

$$\begin{pmatrix} \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \frac{1}{2} + \begin{pmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{pmatrix} \frac{1}{2}$$

obtained from a mixture of perfectly correlated ones. More generally, a distribution showing any type of correlation (positive, negative or zero) can be decomposed into distributions with any other type of correlation. This scenario is known as Simpson's paradox (see e.g. Cartwright 1983, p. 24).

This situation is particularly bothersome, of course, if D is simultaneous with or in the past of C, so that the latter's screening-off effect is due merely to an incomplete specification of what has already occurred. It appears then, that we may be misled into thinking of C as a waterproof 'screen', when we are actually dealing with ignorance about which of two leaking umbrella's has been used.

One way to save the spirit of the principle is by adding further conditions, e.g. that C is, what one may call, a *sufficient* cause[2] for the correlation of A and B, i.e. the assumption

[2] I borrow the term 'sufficient' here from the analogous concept of sufficient estimators in mathematical statistics.

that

$$P(A\&B|C\&X) = P(A\&B|C) \tag{12}$$

for all events X in some class to be specified.

Obviously this raises the further problem of specifying such a class. Here, I only note that it seems reasonable to include at least all events in the past of C, but not A, B themselves, or any event directly causally dependent on these.

3 Extension of the PCC to multiple events

Let us now consider the question of extending the principle of the common cause to more than two events. Although previous discussions of the PCC have often dealt with triples or n-tuples of events, (e.g. Sober 1984, Cartwright 1988) I know of no attempts to give a general mathematical formulation.

Suppose that $A_1, \ldots A_n$ are correlated events, that is, at least one of the following 2^n equations is invalid:

$$P(a_1\& \ldots \&a_n) = P(a_1) \cdots P(a_n) \tag{13}$$

(Here I use lower case letters to denote event variables, i.e. $a_i \in \{A_i, \neg A_i\}$.) What would be a proper formulation of the PCC?

The most straightforward attempt to extend the PCC to this case would be to demand that there exists some screening factor C such that

$$P(a_1\& \ldots \&a_n) = P(a_1|C) \cdots P(a_n|C)P(C) + P(a_1|\neg C) \cdots P(a_n|\neg C)P(\neg C) \tag{14}$$

(cf. Fig. 2). However the fulfillment of this demand is not nearly so simple as in the case of

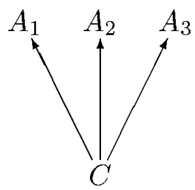

Figure 2: A conjunctive trident

$n = 2$. Indeed, the number of parameters needed to specify the distribution in the left-hand of (14) is $2^n - 1$, whereas the number of free parameters in the right-hand side is only $2n+1$. Thus, for $n \geq 4$, the relations (14) put non-trivial restrictions on the allowed correlations.

This would be a surprising result. If we adopt this version of the principle, certain correlations could not occur in nature because they would be inexplicable by any screening factor at all, whether in the past or future. There was nothing in Reichenbach's discussion suggesting his principle would have such drastic consequences. Thus, I regard (14) as contrary to the purpose of the principle, and look for a different formulation.

To facilitate this search, consider a simple example of the sort of correlations that one can encounter for three events. I call it

Bernstein's paradox Consider a universe with four random events: $\Omega = \{1, 2, 3, 4\}$ and suppose $P(\{i\}) = \frac{1}{4}$, $i = 1, \ldots, 4$. Now let $A_1 = \{1, 2\}, A_2 = \{1, 3\}, A_3 = \{1, 4\}$. Then

$$\begin{aligned} P(A_1 \& A_2) &= P(A_1)P(A_2) = \frac{1}{4}, \quad (15)\\ P(A_1 \& A_3) &= P(A_1)P(A_3) = \frac{1}{4}, \quad (16)\\ P(A_2 \& A_3) &= P(A_2)P(A_3) = \frac{1}{4}, \quad (17)\\ P(A_1 \& A_2 \& A_3) = \frac{1}{4} &\neq P(A_1)P(A_2)P(A_3) = \frac{1}{8}. \quad (18)\end{aligned}$$

This example was given by Kolmogorov (1933) (and attributed to S. Bernstein) to illustrate that pairwise independence between events does not entail mutual independence between all of them. In so far as this point is a source of trouble for the PCC, similar to the Simpson paradox, one may be justified in calling it a paradox. In fact the example can be used to show that the proposal (14) fails also for $n = 3$. (See the appendix for a proofsketch.)

However, the proposal (14) is obviously not the only way of generalizing the PCC. One may argue that for an n-tuple of correlated events more complicated tree-like causal structures can be constructed than the simple conjunctive fork, such as shown in Figure 3.

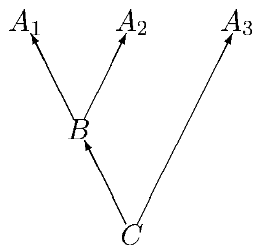

Figure 3: A conjunctive tree structure for three correlated events

The question is then what mathematical characterization should be given to such a structure. At first sight, one would propose the relations

$$\begin{aligned} P(a_1 \& a_2) &= P(a_1|B)P(a_2|B)P(B) + P(a_1|\neg B)P(a_2|\neg B)P(\neg B) \quad (19)\\ P(a_3 \& b) &= P(a_3|C)P(b|C)P(C) + P(a_3|\neg C)P(b|\neg C)P(\neg C) \quad (20)\end{aligned}$$

corresponding to the two conjunctive forks in the diagram. However this will not do to screen off correlation between A_1, A_2 and A_3. The reason is, as mentioned at the end of section 1, that while B may screen off the correlation between A_1 and A_2, the further conditionalization on C may restore their correlation.

Hence, to save the spirit of the principle, we should add the condition that B is a sufficient cause for the correlation of A_1 and A_2 with respect to C. But even this will not be enough in our case. Not only C, also A_3 might break the screening effect of B. To prevent this, let us replace (19) by the conditions

$$P(a_1 \& a_2 | b \& c \& a_3) = P(a_1 \& a_2 | b). \quad (21)$$

In that case we obtain from (20, 21)

$$\begin{aligned} P(a_1\&a_2\&a_3) &= \sum_{bc} P(a_1\&a_2|b\&c\&a_3)P(a_3|b\&c)p(bc) \\ &= \sum_{bc} P(a_1\&a_2|b)P(a_3|c)P(bc) \\ &= \sum_{bc} P(a_1|b)P(a_2|b)P(a_3|c)P(b|c)P(c), \end{aligned} \tag{22}$$

i.e. a decomposition into a convex sum of four factorizing distributions. There are now 9 free parameters in the right-hand side of (22), so one might hope that this will be enough freedom to reproduce all possible correlations for three events. Unfortunately, again, the Bernstein example does not allow a decomposition obeying (20, 21), as shown in the appendix.

Of course one could then investigate more complex tree-like structures, perhaps including even rejoining branches. But this does not seem a promising route to go. Instead, another option suggests itself almost immediately. In Bernstein's example one could argue that the correlation between A_1, A_2 and A_3 is best 'explained' simply by referring to the underlying elementary events $\{1, 2, 3, 4\}$. And indeed, when one conditionalizes on a value of i, the correlation between these events disappears. It thus seems one does obtain a satisfactory extension of the PCC in the form

$$P(a_1\&a_2\&a_3) = \sum_{i=1}^{4} P(a_1|i)P(a_2|i)P(a_3|i)P(i) \tag{23}$$

Thus, I conclude that the most natural way to extend the PCC is along the lines of (23), i.e. by demanding that there exist mutually exclusive screening factors $C_1, \ldots, C_m$ such that

$$P(a_1\&a_2\&\cdots\&a_n) = \sum_{j=1}^{m} P(a_1|C_j)\cdots P(a_n|C_j)P(C_j). \tag{24}$$

Note that, once again, the validity of this form is mathematically trivial. Any joint distribution can be decomposed into a finite number of factorizing ones. Thus, this formulation has the pleasing feature of not forbidding any type of correlation from occurring in nature.

4 Application to a recent variation of the PCC

The previous discussion was very simple-minded. However, it can be used to create serious damage for recent variations of Reichenbach's principle.

In particular, Horwich (1987, pp. 73-74) has argued, following previous suggestions by Salmon (1984), for a weakened version of the PCC, called the 'principle of the fork asymmetry'. In this version, correlations between A and B are considered such that

$$P(A) \text{ and } P(B) \text{ small}, \tag{25}$$

$$P(A\&B) \gg P(A)P(B). \tag{26}$$

His principle claims that in such cases there will exist alternative events $C_1, \ldots, C_m$ in their past such that

$$P(A|C_i) \gg P(A), \quad P(B|C_i) \gg P(B), \tag{27}$$

$$P(A\&B|\neg C_i) = P(A|\neg C_i)P(B|\neg C_i). \tag{28}$$

This principle forms the basis of his theory of time asymmetry.

The discussion of the previous section sheds an unfavorable light on Horwich's version of the PCC. The crucial point here is that, in comparison with Reichenbach's version, the condition (2) is dropped altogether. But now, as soon as we admit a partition into several alternative causes $C_1, \ldots, C_m$ with $m > 2$, the negation $\neg C_i$ is no longer equivalent with the specification of an alternative cause. And indeed, it is easy to show that the corresponding analogue to (24),

$$P(A_1 \& \cdots \& A_n) = \sum_{j=1}^{m} P(A_1|\neg C_j) \cdots P(A_n|C_j) P(\neg C_j) \tag{29}$$

is impossible in Bernstein's example.

Perhaps, one might argue that this could be remedied by taking not the elementary events $C_i = \{i\}$ as causes, but rather their negations, so that (24) is recovered, in virtue of $\neg\neg C_i = C_i$. But this would overlook the fact that whereas C_i are alternatives (i.e. $C_i \cap C_j = \emptyset$), this does not hold for their negations. I conclude that any principle that fails so easily should have no place in a viable account of time asymmetry.

5 Prospects for common-cause-like principles

I finally like to come back to the relevance of our subject for the debate on time asymmetry. I want to argue that Reichenbach's PCC and its variants are crippled because they lack any explicit reference to space-time structure, and that another variant of the principle, provided by Penrose and Percival (1962) is in this respect superior to Reichenbach's.

The idea that one ought to assume some spatial separation between A and B was perhaps implicitly intended by Reichenbach in his drawing of fork diagrams such as Fig. 1. However, this is not as later commentators have understood his principle. Sober (1984) discusses applications of the PCC in evolution theory where these diagrams are interpreted as patterns in a genealogical tree, not in space-time. Let us see if an explicit restriction to localized events in space-time would help to overcome problems for the PCC.

We have noted that the value of the principle should come from the assumption that the putative screening factors C exist in the past of the correlated events. Now suppose that our universe is deterministic in the sense that every event is predictable with probability one from its state at some other instant of time. Arntzenius (1990) points out that then any correlation has screening factors both in the past and future. As he puts it, in such a universe "correlations are not born and do not die, they merely change variables". Thus the PCC will be trivially true.[3] In such a universe, neither Reichenbach's principle, nor the version proposed by Horwich, allows us to define a distinction between past and future.

A straightforward way to dodge this problem is by pointing out that these variables are often non-local. Take for example a collision of two molecules in a gas at time t. According to kinetic theory, this collision is determined by the state of the gas at any previous time t'. However, in order to specify the feature of this state that determines the collision of

[3] Arntzenius' own conclusion is the opposite (Arntzenius 1990, 1992), namely that the PCC is necessarily false in a deterministic universe. However, to reach this conclusion, he assumes that the PCC demands not only the presence of a common cause in the past of A and B, but also the absence of a screening-off common effect of A and B in the future. It is this latter condition which fails in a deterministic world. However, Reichenbach's discussion (op. cit. p.163) makes clear that he rejected this additional condition.

these two molecules, the positions and momenta of all other molecules have to be taken into account. Thus one might hope that by demanding that events are described by local variables, a more interesting version would result. (And then, of course, it remains to be seen if it would yield a successful criterion for distinguishing past and future.)

Arntzenius also shows that the PCC can fail in indeterminist universes, by an example of time-homogeneous Markov processes. While it would take us too far afield to go into this example, I note that, here too an injection of explicit locality might help.

Note that by wedding the PCC to an assumption of locality, the principle will gain empirical content. This, of course, is not surprising because the assumption that events are exhaustively characterized by local variables is itself non-trivial. Thus, the quantum mechanical violation of the Bell-inequalities in Einstein-Podolski-Rosen experiments is only a problem for the PCC when one insists that only localized events are acceptable causes. In Bohm's interpretation of quantum mechanics as a deterministic but non-local theory, the PCC is not violated.

Our suggestion can also help to see why the uniformity of the cosmic background radiation is considered to be a problem. When we describe this uniformity as an absence of correlation between intensity and angle, we are considering probabilities in a phase space, defined by the variables intensity and angle. This is not the space in which we live and breathe. What is striking in the uniformity is that it learns us that physical conditions at the sources of this radiation must have been very similar, in spite of their immense spatial separation. Thus, if we suppose that the natural habitat for the PCC is an application for localized events in space-time, rather than in formal phase spaces, we can understand why the observed uniformity is considered problematic.[4]

Also, in order to evade the Simpson paradox, it seems that one can save the principle by specifying that the cause C is a sufficient causal factor with respect to a class of events. It would be reasonable to take this class to include at least all events in the past of C, perhaps also those outside of C's causal future. However, this means one needs to introduce concepts from the space-time background in the principle.

This point that one needs a restriction of the principle to localized events is also clear in a recent discussion of the PCC in the context of relativistic universes (Earman, 1995, ch. 5.). Unfortunately, Earman's attempt at providing a translation of the spirit of the principle in this space-time setting was, by his own judgement, unsatisfactory (op. cit. p. 138).

Remarkably, a variant of the principle of the common cause taking explicit account of relativistic space-time has been around for a long time, although it is seldom discussed in the philosophical literature. It is Penrose and Percival's (1962) principle of conditional independence.

These authors consider two spacelike separated bounded regions A and B in spacetime, and let C be any region which dissects the union of the past light-cones of A and B into two parts, one containing A and the other containing B. (See Fig. 4.) Then

$$P(\mathcal{A}\&\mathcal{B}|\mathcal{C}) = P(\mathcal{A}|\mathcal{C})P(\mathcal{B}|\mathcal{C}) \tag{30}$$

where $\mathcal{A}, \mathcal{B}, \mathcal{C}$ are the histories of the regions A, B and C, i.e. complete specifications of all events in those regions.

For our discussion, the salient points in which this formulation differs from other formulations are, first, in this version only non-local correlations are to be explained; secondly,

[4] I leave here aside the question of how to interpret the required probabilities in this problem.

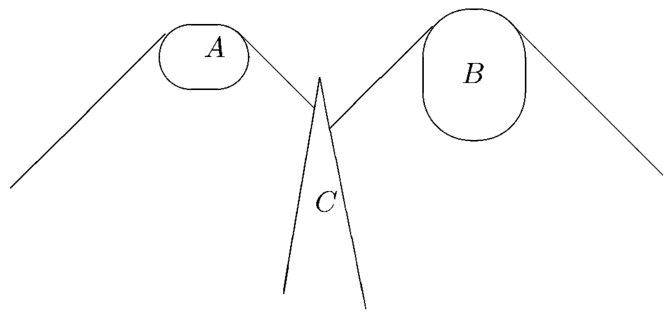

Figure 4: Region C cuts the union of the past light-cones of A and B

the approach is just the opposite of Horwich's: here it is (3) that is is dropped while (2) is retained. Thirdly, conditional independence is demanded only upon conditionalizing upon the *entire* history of a region C. This entails that the problems such as Simpson's paradox connected with incomplete specification of the factors cannot appear. Note also that, whereas Reichenbach's principle claims only the existence of a screening factor, the present principle demands –and also explicitly specifies!– infinitely many, namely, the history of *any* region C cutting the past of A and B. Further, Penrose and Percival's formulation can be generalized straightforwardly to multiple events.[5]

I conclude that this principle is a viable formulation of Reichenbach's intuition. Of course, it is false in quantum mechanics, and probably also in the case of the cosmic background radiation. But even this has a merit: putting the genuine problems in these examples more sharply into focus than in any other of the formulations mentioned above.

Appendix

Here I give a proof-sketch of the mathematical claims in the text connected with Bernstein's paradox.

To see that (14) cannot be fulfilled, note first that since $P(a_1\&a_2\&a_3)$ does not factorize, one must have $0 < P(C) < 1$ in this equation. Secondly, note that $P(A_1\&A_2\&\neg A_3) = 0$. Thus, by (14) one obtains

$$P(A_1|C)P(A_2|C)P\neg A_3|C)P(C) + P(A_1|\neg C)P(A_2|\neg C)P\neg A_3|\neg C)P(\neg C) = 0 \qquad (31)$$

This entails that some pair of factors from each term must vanish. Suppose, e.g. that $P(A_2|C)$ and $P(\neg A_3|C)$ vanish. Then, by (14), $P(\neg A_1\&A_2\neg A_3)$ would have to vanish too, which it does not. It is straightforward to repeat this argument for any other possbile choice of a pair from each term in (31).

To see that (21) cannot hold in our example, observe that it would lead to

$$\begin{aligned} 0 = P(A_1\&A_2|\neg A_3) &= \sum_{bc} P(A_1\&A_2|b\&c\&\neg A_3) = \sum_{bc} P(A_1\&A_2|b) \\ &= \sum_{bc} P(A_1\&A_2|b\&c\&A_3) = P(A_1\&A_2|A_3) = \frac{1}{2}. \end{aligned}$$

[5] One demands that for any finite set of spacelike separated regions $A_1, \ldots, A_n$ and any region C which cuts the union of the pasts of $A_1, \ldots, A_n$ into n disjoint subsets such that each A_i is part of only one of them, the analogue of (30) holds: $P(\mathcal{A}_1\& \ldots \&\mathcal{A}_n|\mathcal{C}) = P(\mathcal{A}_1|\mathcal{C}) \cdots P(\mathcal{A}_n|\mathcal{C})$.

References

[1] Arntzenius, F. (1990), "Physics and common causes", *Synthese*, 82: 77-96.

[2] Arntzenius, F.(1993), "The common cause principle", *PSA 1992*, Vol. 2, D. Hull, M. Forbes and K. Okrulik, (eds.) East Lansing: Philosophy of Science Association, pp. 227–237.

[3] Cartwright, N. (1983) *How the Laws of Physics lie.* Oxford: Clarendon Press.

[4] Cartwright, N. (1988), "How to tell a common cause: generalizations of the conjunctive fork criterion" in *Probability and Causality.* J.H. Fetzer, ed., Dordrecht: D. Reidel Publishing Company, pp. 211-228.

[5] Earman, J. (1995), *Bangs, Crunches, Whimpers, and Shrieks.* Oxford: Oxford University Press.

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

[7] Kolmogorov, A.N. (1933), *Grundbegriffe der Wahrscheinlichkeitsrechnung.* Berlin: Springer.

[8] Penrose, O. and Percival, I.C., (1962), "The direction of time", *Proceedings of the Physical Society*, 79: 605-616.

[9] Price, H. (1996), *Time's Arrow and Archimedes' Point.* Oxford: Oxford University Press.

[10] Reichenbach, H. (1956). *The Direction of Time.* Berkeley: University of California Press, reissued 1991).

[11] Salmon, W.C. (1984), *Scientific Explanation and the Causal Structure of the world*, Princeton: Princeton University Press.

[12] Sober, E. (1984) "Common cause explanation" *Philosophy of Science* 51: 212-241.

[13] Sober, E. (1988) 'The principle of the common cause', in *Probability and Causality.* J.H. Fetzer, ed., Dordrecht: D. Reidel Publishing Company, 211-228.

[14] Torretti, R. (1987), "Do conjunctive forks always point to common causes?", *British Journal for Philosophy of Science* 38: 384-387.

[15] Van Fraassen, B.C. (1980), *The Scientific Image.* Oxford: Oxford University Press.