

This article was downloaded by: [University of Wisconsin Madison]

On: 18 May 2010

Access details: Access Details: [subscription number 919988825]

Publisher Routledge

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Australasian Journal of Philosophy

Publication details, including instructions for authors and subscription information:

<http://www.informaworld.com/smpp/title~content=t713659165>

Conjunctive forks and temporally asymmetric inference

Elliott Sober ^a; Martin Barrett ^a

^a University of Wisconsin, Madison

To cite this Article Sober, Elliott and Barrett, Martin (1992) 'Conjunctive forks and temporally asymmetric inference', *Australasian Journal of Philosophy*, 70: 1, 1 – 23

To link to this Article: DOI: 10.1080/00048408112340013

URL: <http://dx.doi.org/10.1080/00048408112340013>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

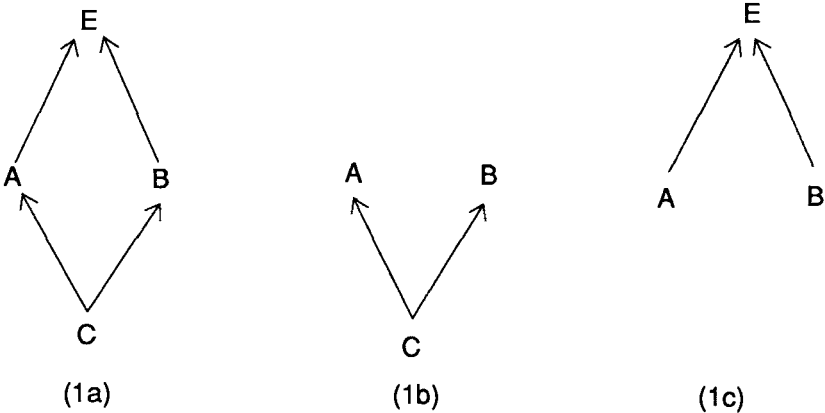
CONJUNCTIVE FORKS AND TEMPORALLY ASYMMETRIC INFERENCE

Elliott Sober and Martin Barrett

I. Introduction

Why do we know more about the past than about the future? In *The Direction of Time*, Hans Reichenbach argues that this epistemological asymmetry derives from a kind of causal asymmetry. Reichenbach [9, p. 159] discusses the three structures depicted in Figure 1. The first (1a) he terms a *double fork*; it involves a common cause, *C*, which produces two events, *A* and *B*, which, in turn, produce a common effect, *E*. The second structure (1b) Reichenbach calls a *fork open to the future*; it comprises a common cause *C* and two joint effects *A* and *B*. The third (1c), a *fork open to the past*, consists of two causes (*A* and *B*) and their common effect, *E*.

Figure 1



Reichenbach identifies a special statistical relationship that can obtain among the events in such forks. Forks are said to be *conjunctive* when the vertex *screens off* each of the tips from the other. Reichenbach's idea, and the meaning of the terms just italicized, can be clarified by considering one of his examples. Suppose that two actors in an acting troupe show a correlation in the days on which they experience gastro-intestinal distress. Each gets sick, let us say, about once every

hundred days, so the probability of sickness on a randomly selected day is $1/100$. If the two actors got sick independently, the probability that both would be ill on a randomly selected day would be $(1/100)(1/100) = 1/10000$. However, let us suppose that they get sick together far more often than this probability would suggest. Just for simplicity, let us suppose that when one gets sick, the other almost always does too. So the probability that both are ill on a given day is about $1/100$.

If A represents the proposition that the first actor gets sick on a given day and B represents the proposition that the second actor gets sick on that day, then we represent the fact that the sick days are correlated by the following inequality:

$$(1) \quad P(A \& B) > P(A)P(B).$$

Reichenbach suggests that this correlation be explained by postulating a common cause. We hypothesize that the two actors always take their meals together, so that on any given day, either both eat tainted food or neither does. Let C be the proposition that the actors eat tainted food at their common meals on a given day (and let $\neg C$ be the proposition that they do not eat tainted food at their common meals on that day). Now suppose that eating tainted food raises the probability of sickness for each actor:

$$(2) \quad P(A/C) > P(A).$$

$$(3) \quad P(B/C) > P(B).$$

Suppose, finally, that the actors' probabilities of sickness are independent of each other if they eat tainted food together, and that the same is true if they eat untainted food together:

$$(4) \quad P(A \& B/C) = P(A/C)P(B/C).$$

$$(5) \quad P(A \& B/\neg C) = P(A/\neg C)P(B/\neg C).$$

Reichenbach [10, pp. 157-163] proves that propositions (2)-(5) entail proposition (1). This deductive relation is intended to reflect an explanatory one: by postulating a common cause that obeys constraints (2)-(5), one thereby explains why the actors' sick days are correlated.

Notice that the concepts of cause and explanation do not occur explicitly in (1)-(5). In principle, any triplet of events may satisfy these five requirements. Events that do so are said to form a *conjunctive fork* and the event corresponding to C is said to *screen-off* A from B . (1) expresses the idea that A and B are correlated (*nonindependent*); (4) and (5) express the idea that A and B are *not* correlated (are independent) once one conditionalizes on each of the states of the common cause. A correlation is explained by postulating an event that removes the correlation, so to speak. There is no paradox here, only the contrast between unconditional correlation (1) and conditional independence (4-5).

Reichenbach is a bit careless about the ontology of the common cause structure he describes. Is the 'common cause' the common meals the actors share, or is it the tainted food that they consume at some of their meals? This question is not an idle one, because inferring the existence of the common meals is a different problem from inferring whether the food on a given day was tainted, given the assumption that the actors dine together (Sober [15]). Reichenbach elides these inferential problems. In what follows we will distinguish the

common cause from the *states* that the common cause may occupy. The shared meal is the common cause. On any given day, the shared meal is in one of two states; either the food is tainted or it is not. A similar distinction will be deployed with respect to effects and their states. The terminology used to mark this distinction is somewhat stipulative, though the distinction itself is a real one.

The common cause and the states of the common cause can be viewed as explaining slightly different aspects of the phenomenon in Reichenbach's example. The fact that the actors dine together explains why their sick days are correlated. But when the actors are both sick on a given day, this may turn out to be due to the fact that they both ate tainted food. The first *explanandum* concerns a pattern that encompasses the longer time frame of the actors' enduring association with each other; the second concerns what happened on a single day. The common cause structure endures as long as the actors eat together; but on each day, the state of the common cause may change, from tainted to untainted food, or back again.

In Reichenbach's example, *A* and *B* are effects of a common cause *C*. But, as already noted, the probability relations defined by (1)-(5) do not explicitly say how the concepts of cause and effect apply to the three events. For example, *A* and *B* might be joint causes of some common effect. This raises the question of how the conjunctive fork idea applies to the three causal structures depicted in Figure 1.

Reichenbach [10, p. 161] claims that common effects 'usually' form conjunctive forks with their joint causes. And Reichenbach's better known *Principle of the Common Cause* says that if simultaneous events *A* and *B* are correlated, then they always have a screening off common cause. Reichenbach [10, p. 162] argues that conjunctive forks exhibiting the first two structures depicted in Figure 1 are quite common. He maintains, however, that conjunctive forks of the third kind are incompatible with the Second Law of Thermodynamics. According to Reichenbach, whenever one thinks one has found a conjunctive fork open to the past, it really is a double fork in disguise.

Although Reichenbach's views about these causal structures are clear enough, we do not entirely grasp what connection he sees between them and the temporal asymmetry in our knowledge. Even if conjunctive forks could exhibit patterns (1a) and (1b), but not (1c), why would that show that the past is better known than the future? This question splits into two: why should we focus on *conjunctive forks* in formulating this epistemological problem? And given that conjunctive forks are what we wish to examine, why should the existence of (1a) and (1b) but not (1c) engender an epistemological asymmetry?

We will not try to reconstruct Reichenbach's reasoning. Rather, we will argue against some of Reichenbach's claims about the forks just mentioned. We do not see why the Second Law of Thermodynamics rules out the existence of conjunctive forks open to the past. In addition, we will argue that a common effect rarely forms a conjunctive fork with its joint causes. That is, we will argue that (1c) is not impossible, but that (1a) and (1c) are both rare patterns for

conjunctive forks to exhibit.

Nevertheless, we think there is something to be said for Reichenbach's idea that forks of various kinds are relevant to explaining why we know more about the past than about the future. And entropy turns out to play a role in the explanation, though not the determining role that Reichenbach envisions.

II. Why an Effect Rarely Screens Off One of Its Causes From the Other

Reichenbach held that a common effect (E) 'usually' screens off one of its causes (A) from the other (B). When this is so, it must be because there is a common cause (C) that allows the structure to conform to pattern (1a). Without such a common cause (C), E cannot screen off A from B .

We doubt both these claims. In this section, we argue that common effects rarely form conjunctive forks with their joint causes. This arrangement will be rare, regardless of whether there is a common cause (C) in the background. The rarity of such arrangements derives from the fact that they are 'mathematical accidents,' so to speak. General facts about the structure of such forks suggest that effects rarely screen off one of their causes from the other.

Let AEB be a fork made of a common effect and its (two) joint causes. When we talk about the *efficacy* of these causes, we mean the four conditional probabilities of the form $P(E/\pm A\pm B)$, which we call w , x , y , and z :

$$w = P(E/A\&B)$$

$$x = P(E/A\&-B)$$

$$y = P(E/-A\&B)$$

$$z = P(E/-A\&-B)$$

If AEB is conjunctive, there must be a nonzero (unconditional) correlation between A and B . This means that their *covariance* is not equal to 0. ($\text{Cov}(A,B) = P(A\&B) - P(A)P(B)$.) It will be useful to bear in mind that $\text{Cov}(A,B)$ can also be expressed as $P(A\&B)P(-A\&-B) - P(A\&-B)P(-A\&B)$. We will refer to the four probabilities of the form $P(\pm A\&\pm B)$ as the *frequencies of the causes*.

We now will establish the following theorem:

Theorem 1: Let AEB be a conjunctive fork made of two causes and their common effect. Unless the efficacies of the causes are tightly coupled with their frequencies, AEB will not remain conjunctive if the efficacies and/or the frequencies change.

The phrase 'tightly coupled' is imprecise, but our argument will clarify what we have in mind.

Bayes' theorem allows the three conditional probabilities we need to consider to be expressed as follows:

$$P(A\&B/E) = P(E/A\&B)P(A\&B)/P(E) = wP(A\&B)/P(E).$$

$$P(A/E) = P(E/A)P(A)/P(E)$$

$$= [P(E/A\&B)P(B/A) + P(E/A\&-B)P(-B/A)]P(A)/P(E)$$

$$= [wP(B/A) + xP(-B/A)]P(A)/P(E).$$

$$\begin{aligned} P(B/E) &= P(E/B)P(B)/P(E) \\ &= [P(E/A\&B)P(A/B) + P(E/-A\&B)P(-A/B)]P(B)/P(E) \\ &= [wP(A/B) + yP(-A/B)]P(B)/P(E). \end{aligned}$$

E screens off A from B precisely when

$$wP(A\&B) = [wP(A\&B) + xP(A\&-B)][wP(A\&B) + yP(-A\&B)]/P(E).$$

This simplifies to

$$(*) \quad wz/xy = P(A\&-B)P(-A\&B)/P(A\&B)P(-A\&-B).$$

Since A and B have a nonzero covariance, we know that:

$$P(A\&-B)P(-A\&B)/P(A\&B)P(-A\&-B) \neq 1.$$

Note that the left-hand side of (*) describes a relationship among the efficacies of the causes while the right-hand side describes a relationship among their frequencies.

Suppose that the system under consideration, at a given time, satisfies (*). Now let the frequencies and/or the efficacies *change*. Unless these changes are precisely 'coordinated,' the resulting values will fail to satisfy (*).

We believe that it is very often the case that w , x , y , and z are independent of how often the four combinations of causes are exemplified in the population of interest. For example, the probability of getting lung cancer, given that you smoke and are an asbestos worker, is independent of the distribution of smokers and asbestos workers in the population you inhabit. Many other similar examples could be cited.

We do not deny the existence of systems in which the efficacy of causes and their frequencies are connected. The phenomenon of frequency dependent natural selection is a case in point. For example, in a population of prey organisms, rarity may be an advantage, if the predators' search image is keyed to common appearance in the prey.¹ But the qualitative fact that the fitness of a character (its 'efficacy' with respect to survival and reproduction) declines as it becomes more common is not enough to ensure that (*) will be true. (*) demands an exact quantitative relationship between efficacies and frequencies, one that we believe is almost never exemplified, even in the 'best case' situation of frequency dependent selection.

In the real world, the frequencies of causes often wax and wane. It also is true that the efficacies of causes can change. In the space of values that these functions may assume, there is only a vanishingly small region in which an effect will screen off one of its causes from the other.

What are the corresponding facts about the mirror image case, in which a common cause (C) screens off one effect (A) from the other (B)? We believe that such forks often remain conjunctive when the probability of the cause or the conditional probability of an effect, given a cause, changes.

Consider Reichenbach's theatrical example. The two actors (A and B) in the theatre company exhibit a correlation in their sick days. The common cause explanation (C) of this correlation turns out to be that the actors take their meals together, so that on a given day either both eat tainted food or neither does.

¹ Some more detailed examples are discussed in Sober [12].

Supposing that this is a conjunctive fork, will it remain one if we manipulate the probability of the cause? That is, suppose we augment or reduce the frequency with which tainted food is placed on the actors' common table. In this case the fork will remain conjunctive, if it was conjunctive initially. Likewise, suppose we give one or both actors a drug that reduces the conditional probability of getting sick if one eats tainted food. Again, the numbers will be modified, but the fork will be conjunctive after the intervention if it was conjunctive before.

The present argument goes contrary to Reichenbach's claim that common effects usually screen off their joint causes from each other. In addition, our argument is independent of Reichenbach's thesis about the ubiquity of double forks. As will be explained in the next section, lots of correlations fail to possess screening off common causes. When A and B are correlated though no common cause is at work, there is no special reason why A and B cannot have a common effect E . We do not see why it should be 'harder' for AEB to be a conjunctive fork in the absence of a common cause C .

We have compared forks made of *one* common cause and *two* effects with forks made of *one* common effect and *two* causes. However, this preoccupation with the numbers *one* and *two* is dispensable. Consider any number of causes $\pm C_1, \pm C_2, \dots, \pm C_n$ and any number of effects $\pm E_1, \pm E_2, \dots, \pm E_m$. Let the efficacies be the probabilities of combinations of causes given combinations of effects; these have the form

$$P(\pm E_1 \ \& \ \pm E_2 \ \& \ \dots \ \& \ \pm E_m \ / \ \pm C_1 \ \& \ \pm C_2 \ \& \ \dots \ \& \ \pm C_n).$$

The frequencies of causes are probabilities of the form $P(\pm C_1 \ \& \ \pm C_2 \ \& \ \dots \ \& \ \pm C_n)$. We have posed two questions:

If the causes screen off the effects from each other, will they continue to do so if the efficacies or frequencies of the causes are modified?

If the effects screen off the causes from each other, will they continue to do so if the efficacies or frequencies of the causes are modified?

We have argued that the answers are, respectively, *yes* and *almost always no*.

Our argument exploits no special explication of the concept of cause. We have talked about some events causing others, both before and after various constituent probabilities are modified. This imposes some modest constraints on what the causal relation can be; it cannot be so fragile that these changes turn causes into noncauses. However, much room is left open concerning how the causal relation should be understood.²

Our argument does not show that common causes and their joint effects often form conjunctive forks. We consider this important question in section IV. Rather, our claim is that once a fork open to the future is conjunctive, it usually

² One theory that is consistent with our claims is the probabilistic theory of causality, expounded in various forms by Suppes [17], Cartwright [2], Skyrms [11], Eells and Sober [6], Sober [12], and Eells [5]. If causes must raise the probability of effects in all background contexts, then the four efficacies in the AEB fork will be ordered as $P(E/A\&B) > P(E/A\&-B)$, $P(E/-A\&B) > P(E/-A\&-B)$. We emphasize that this theory suffices, but is not necessary, for the arguments propounded above.

will remain so when various constituent probabilities are modified.³ A conjunctive fork made of a common cause and its joint effects is *robust*, whereas a conjunctive fork made of a common effect and its joint causes is not.

III. Conjunctive Forks Open to the Past

In this section, we wish to describe a counterexample to Reichenbach's claim that any conjunctive fork between two causes (A and B) and their common effect (E) must be such that there is a screening off common cause (C) of A and B .

It is important that this thesis of Reichenbach's not be trivialized. Suppose A has a cause (C_a) and B has a cause (C_b). If these two causes have nothing much to do with each other, it will be false that A and B have a *common* cause. This point should not be obscured by the trick of constructing a composite event C_a -plus- C_b and calling this the common cause of A and B .

One of us has described a general kind of counterexample to Reichenbach's Principle of the Common Cause (Sober [13, 14]). If two causally independent processes both show a monotonic increase, then there will be a positive correlation between the state of one and the state of the other. For example, there is a positive correlation between bread prices in Britain and sea levels in Venice over the last 200 or so years, since each has increased with time. Yet no one thinks that there is a common cause explanation in this case. Each process developed on its own; the appropriate explanation is in terms of separate causes, not common ones.⁴

³ It should be clear that our talk of 'modifying' various probabilities accords no essential role to human intervention. The question is whether a fork can be expected to remain conjunctive when a constituent probability changes its value.

⁴ This is one respect in which our account diverges from the one Horwich [8, pp. 72-74] proposes for what he terms 'the fork asymmetry'. Horwich *assumes* that the Principle of the Common Cause is correct; he claims that its truth is explained by the hypothesis of 'initial micro chaos'. We think the principle is false; moreover, we don't entirely understand what Horwich means by initial micro chaos, nor how this hypothesis is supposed to explain Reichenbach's principle.

The derivation that Horwich provides does not show that correlated event types (each of which are rarely exemplified) probably have a common cause. Rather, Horwich assumes that this is true. For example, his argument does not show that the actors in Reichenbach's example, whose rare sick days are correlated, probably eat together. Rather, Horwich argues that if one assumes that they eat together, then on those days on which they get sick, one can infer that they probably shared tainted food rather than untainted food. Horwich's inference is to the *state* of the common cause, assuming that the common cause structure is already in place. This distinction between inferring the existence of a common cause and inferring the state of that common cause on the supposition that the common cause exists, is discussed in Sober [15].

Horwich believes his derivation cannot be carried out for the common effect of the correlated events. The reason is that 'since the time reverse of the condition of initial chaos is false, we cannot suppose that the alternative to a joint effect, E , is some condition relative to which A and B would be independent of each other'. We find this argument puzzling. Even if the time reverse of the micro chaos hypothesis were not available, why would that show that no other principle could underwrite the forward-directed inference? In any event, we think that if a common effect exists, then its state may be inferable from the observed state of its joint causes.

This purported counterexample to Reichenbach's principle would be undermined if a composite event were defined and said to be the common cause of British bread prices and Venetian sea levels. If C_b causes the former and C_v causes the latter, then what prevents one from saying that C_b -plus- C_v is the common cause explanation of the correlation? The only thing that prevents this is that the contrast between common cause and separate cause explanations would lose all meaning. If Reichenbach's Principle of the *Common Cause* is to be meaningful, it must describe an alternative to the separate cause pattern of explanation.

Our present concern is not the Principle of the Common Cause, but Reichenbach's thesis that conjunctive forks open to the past are impossible. We wish to describe a case in which the event triplet AEB forms a conjunctive fork, but there is no event C that provides a screening off common cause of A and B . To be sure, our example will be one in which A has a cause; and B will have a cause as well. But we deny that these together form a *common* cause of A and B .

An example of a conjunctive fork open to the past is not far to seek. We have just noted that some correlations are not due to common causes. Let A and B be simultaneous events of this sort. In the previous section, we found that the condition for A and B to produce a screening off common effect (E) is stringent, but not impossible. It is set forth in proposition (*). So let A and B be positively correlated *because each develops autonomously*. And let the probabilities of A and B be related to the probabilities that E has conditional on different combinations of $\pm A$ and $\pm B$ as (*) requires. Here, then, is a conjunctive fork open to the past.

Many autonomous processes involve alternations of on/off states.⁵ Hotter and colder weather alternate seasonally on planets, no matter what the solar system is to which they belong. The same can be said of the pulsing of pulsars and of the dormant and active periods of many organisms. Select two quite dissimilar processes that develop on approximately the same time scale. Their states will be correlated, but not because of a screening off common cause. Let each process send a signal to some receiver (E), the probabilities of whose on/off states, conditional on the signals it receives (that is, w , x , y , and z), conform to (*).⁶ Such systems will form conjunctive forks open to the past.

The existence of structures of this sort does not contradict the Second Law of Thermodynamics. The easiest way to see this is to consider examples that are thermodynamically open; in this case, the Second Law says nothing about them.⁷

⁵ The idea of a conjunctive fork could be generalized to n -state or continuously varying characters, and arguments parallel to the ones given above would apply.

⁶ E can employ randomization to 'decide' whether to go into its on or off state, based on occurrences of $\pm A$ and $\pm B$.

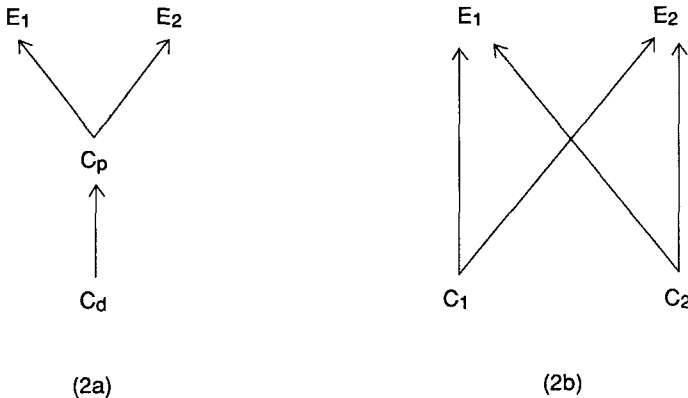
⁷ We also suspect that these structures can be thermodynamically closed, but will not argue the point here.

IV. Why Conjunctive Forks Open to the Future are Worth Considering

Reichenbach's Principle of the Common Cause does not say that every common cause screens off one effect from the other. Rather, it says that every pair of correlated effects has a screening off common cause.⁸ One of us has developed reasons for thinking that many common causes fail to screen off the effects from each other (Sober [13, 14]).

There are (at least) two general cases in which a common cause fails to screen off. Suppose a proximal cause screens off the effects from each other and that the proximal cause also screens off the distal cause from the effects (Figure 2a). In this case, the distal cause will *not* screen off the effects from each other. Secondly, suppose there are *two* common causes of the two effects (Figure 2b). If the total state of this pair of causes screens off each effect from the other, then neither common cause, taken by itself, will screen off.⁹

Figure 2



Common causes are often distal rather than proximal. And where there is one common cause, there often is another. For these two reasons, we believe that many common causes fail to screen off.¹⁰

⁸ We ignore here the fact that a correlation may be explained by saying that one of the correlates causes the other. That is, we will assume in the examples we'll consider that the correlation cannot be explained in this way; rather, the Reichenbachian approach will be to postulate a common cause.

⁹ Proofs of these claims are provided in the Appendix to Sober [13] and in Sober [14, pp.97-99].

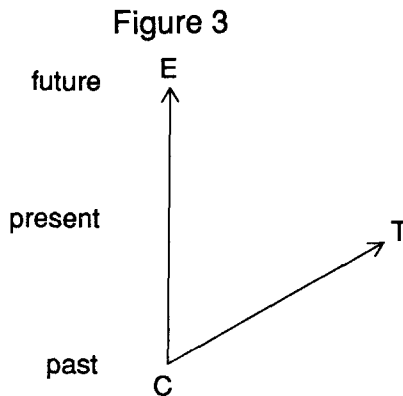
¹⁰ These results about nonconjunctive forks open to the future help clarify the asymmetry that should be associated with Theorem 1. A conjunctive fork open to the future can be rendered nonconjunctive, either by making the common cause distal, or by introducing a second common cause. The robustness claimed for conjunctive forks open to the future has to do with interventions in which it is the probability of the cause or the conditional probability of the effect given the cause that is modified. It is with respect to *these* interventions that forks open to the past differ from forks open to the future.

The somewhat modest point in Reichenbach's principle that is unaffected by these examples is that a *complete description of all the causal facts* will screen off the effects from each other. The only problem with this claim, as far as we know, is that it is inconsistent with ideas stemming from quantum mechanics.¹¹ But such considerations to one side, the principle seems quite plausible.

The principle, thus understood, has at least one methodological implication. It says that if one's description of the causal facts fails to induce a screening off relation, then one's description is incomplete. But the principle cannot be invoked to underwrite the claim that the common causes one has established in (as yet incomplete) science form conjunctive forks with their joint effects.

The problem is this: Reichenbach suggests that the existence of forks made of common causes and their joint effects helps explain why we know more about the past than about the future. Presumably, the forks that contribute to our ability to make better inferences about the past than about the future must be forks that we know about. But why should we think that these forks, which we know about as part of our incomplete knowledge of the world, are conjunctive? The Principle of the Common Cause does not provide an answer.

The forks we wish to consider involve what we will call a *main process* from cause (*C*) to effect (*E*) and a *branch process* that leads from the cause (*C*) to a trace (*T*). The trace is something we observe now. The effect term (*E*) in the main process will exist in the future. This structure is depicted in Figure 3.



Traces include such paradigm cases as fossils. An ancestral organism (*C*) has two effects. It produces a fossil (*T*), which we now observe. It also produces a lineage, which eventuates in a descendant organism (*E*) that exists in the future.

Written records provide another salient example. The Pharaoh's court in ancient Egypt (*C*) may have two effects. Scribes may write down a description of the court, which is transmitted through a series of scribes until we now have before us a derivative text (*T*). The Pharaoh's court also has an impact on

¹¹ See Van Fraassen [18] for discussion of why the violation of Bell's inequality conflicts with Reichenbach's Principle of the Common Cause.

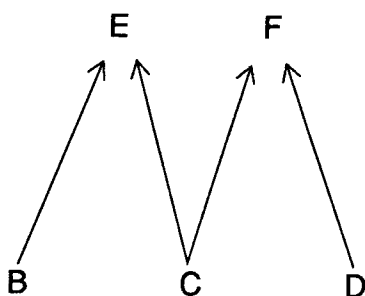
Egyptian society, which through a chain of interactions produces the Egyptian society that will exist in the future (E).

In both these examples, we eventually want to consider why (or in what circumstances) T provides more knowledge about C than it does about E .

What could entitle us to think that these examples involve conjunctive forks? The state of the ancestral organism is hardly the only cause of the shape of the fossil or the characteristics of the descendant. The state of the Pharaoh's court is hardly the only cause of what the derivative manuscript says or of the characteristics of future Egyptian society. For this reason, Reichenbach's principle (which formulates a *sufficient* condition for a fork's being conjunctive) does not entail that either fork is conjunctive. The suspicion therefore should arise that the conjunctive fork model is hopelessly idealized.

We believe that this suspicion is exaggerated. For there is a fairly general circumstance, depicted in Figure 4, in which a common cause does screen off the effects from each other, even though other causes also contribute to the effects. In such cases, the common cause does not constitute *all* the causal facts, but it screens off nonetheless.

Figure 4



We will establish the following theorem:

Theorem 2: Suppose that C is a common cause of E and F , that B causes E but not F , and that D causes F but not E . If (i) the total state of all the causes (B, C, D) screens off the effects from each other, (ii) the causes that contribute to an effect screen off that effect from the cause that does not contribute, and (iii) the causes are statistically independent of each other, then C screens off E from F .

We begin by defining some conditional probabilities:

$$\begin{array}{ll}
 h = P(E/B\&C) & w = P(F/C\&D) \\
 i = P(E/-B\&C) & x = P(F/-C\&D) \\
 j = P(E/B\&-C) & y = P(F/C\&-D) \\
 k = P(E/-B\&-C) & z = P(F/-C\&-D)
 \end{array}$$

The probabilities of one or both effects, given the common cause (C), are as follows:

$$\begin{aligned} P(E/C) &= \sum P(E/\pm B \& C)P(\pm B/C) \\ &= hP(B) + iP(-B). \end{aligned}$$

$$\begin{aligned} P(F/C) &= \sum P(F/\pm D \& C)P(\pm D/C) \\ &= wP(D) + yP(-D). \end{aligned}$$

$$\begin{aligned} P(E \& F/C) &= \sum P(E \& F/\pm B \& \pm D \& C)P(\pm B \& \pm D/C) \\ &= \sum P(E/\pm B \& C)P(F/\pm D \& C)P(\pm B \& \pm D) \\ &= hwP(B \& D) + hy(B \& -D) \\ &\quad + iwP(-B \& D) + iyP(-B \& -D). \end{aligned}$$

Note that $P(E \& F/C)$ is just the product of $P(E/C)$ and $P(F/C)$.¹²

So here is a case in which a common cause screens off its effects from each other even though other causes contribute to the two effects. Notice that the other contributing causes are not *common* causes; without this assumption, we would face the counterexample depicted in Figure 2b.

How general a circumstance does this theorem describe? We believe that it is realistic enough to be of some interest. Consider the two examples described before. In the first, an ancestral organism produces a fossil as well as a descendant organism. Admittedly, many other factors impinge on the shape of the fossil, and still other factors affect the characteristics of the descendant. But to a large extent, the further factors affecting the one will be quite different from the further factors affecting the other.

In the second example, the Pharaoh's court gives rise to a manuscript that is handed down to the present day. The Pharaoh's court also causally contributes to the characteristics of future Egyptian society. Admittedly, the present day manuscript and the future state of Egyptian society are both affected by factors additional to what went on in the Pharaoh's court. But to a considerable extent, the additional factors affecting the one will be disjoint from the additional factors affecting the other.

We do not advance this as an absolute claim, nor as one of total generality. It is easy to describe how the factors additional to C can include *common* causes. But we believe that in many cases, there will be no such further common causes, or their contributions will be modest enough that they may safely be ignored. This, at any rate, is our rationale for exploring the case in which a fork composed of a common cause and its effects is conjunctive.

V. Two Theorems about Conjunctive Forks

In a conjunctive fork TCE with vertex C , we observe the state that T occupies. From this observation, we infer the state that C is in; we also infer what state E occupies. We wish to investigate the circumstances under which observing T allows one to know more about C than about E .

T , C , and E may be in different states. For simplicity, we will suppose that each may be in state '0' or in state '1'. Let 'To' express the proposition that T occupies

¹² A parallel derivation establishes that $P(E \& F/C) = P(E/C)P(F/C)$.

state 0; similarly for the other nodes C and E and the other state the nodes may occupy. What information T provides about the other nodes depends on which state T is found to occupy. Learning that T_0 may provide more information about C (or about E) than learning that T_1 . So when we talk about ‘the information provided by observing T ,’ we have in mind an *expectation*; the quantity of interest will be an average over the information that would be provided by T_0 and by T_1 , the weighting being supplied by the probabilities $P(T_0)$ and $P(T_1)$.

We shall measure the knowledge we have about a given node — C , for example — in terms of the sharpness of the probability distribution we assign to its states. When we believe that $P(C_0) = P(C_1) = 0.5$, we are maximally uncertain about what state C occupies; when the probabilities are close to 0 and 1, we know a great deal.

It is important to keep clearly in mind three different measures of uncertainty (or information). There is the *prior information* we possess about C ; this reflects how sharp the probability distribution of C ’s states is *before* we observe what state T occupies. Second, there is *what we learn about C from observing T* ; this measures how much observing the state of T changes the probability distribution we assign to C . Information theorists (following R.A. Fisher) call this the mutual information ($M[C,T]$). Third, there is *the total information* we have about C , once we have observed T .¹³ This reflects what we knew prior to making the observation and what the observation teaches us.

Intuitively, we might expect this third quantity to be the sum of the first two:

$$\text{Total}(C/T) = \text{Prior}(C) + \text{Mutual}(C,T).$$

This intuition turns out to be correct, once these three quantities are defined in the way we propose.¹⁴

Prior(C) is the neg-entropy of C . Total(C/T) is the expected neg-entropy of C conditional on T , where the average is over the different states that T might occupy. And Mutual(C,T) is the mutual information that each of C and T provides about the other. These are defined as follows:

$$\text{Prior}(C) = \sum_i P(C_i)[\log P(C_i)].$$

$$\text{Total}(C/T) = \sum_{i,j} P(C_i/T_j) \log [P(C_i/T_j)] P(T_j).$$

$$\begin{aligned} \text{Mutual}(C,T) &= \sum_{i,j} P(C_i \& T_j) [\log P(C_i/T_j)] - P(C_i) [\log P(C_i)] \\ &= \sum_{i,j} P(C_i \& T_j) \log [P(C_i \& T_j) / P(C_i) P(T_j)]. \end{aligned}$$

Mutual information is a symmetrical measure of the relationship between the two random variables. It is zero when observing the state of T does not modify the probability assigned to C .

The summations in the above expressions are needed since we wish to compute expectations over the possible states of the random variables involved. This complication aside, the relationship between these three quantities is rather sim-

¹³ Properly speaking, mutual information and total information are *expectations*, since they are averages over the states of T . This will be explained presently.

¹⁴ A fuller defence of these choices of measures is provided in Barrett and Sober [1]; see Khinchin [9] for discussion of the relevant mathematical ideas.

ple. If $P(H)$ is the probability one assigns to the hypothesis H before determining whether O is true, and if $P(H/O)$ is the probability that H has, once one learns that O is true, then $P(H/O) - P(H)$ measures the degree to which O confirms H . It may be positive (if O confirms H), negative (if O disconfirms H), or zero (if O fails to say anything about H). The posterior probability is thus the sum of the prior probability and the degree of confirmation.

Note that it is possible for an observation to strongly confirm H although H remains highly dubious after the observation is made. Suppose that $P(H) = 0.0001$ and $P(H/O) = 0.5$. H has gained in credibility. But this does not mean that we are entitled to be very confident that H is true.

Similarly, an observation can have little or no impact on the plausibility of H and yet H may be quite plausible nonetheless. Suppose that $P(H) = 0.95$ and $P(H/O) = 0.95$. In this case, H is plausible in the light of the observation O even though O does not confirm H .

We do not wish to argue that ordinary usage of words like ‘confirms’ precisely coincides with the distinctions we have just drawn. Suffice it to say that these distinctions are important. The *diachronic* measure of how much difference an observation makes to the plausibility of a hypothesis should not be confused with the *synchronic* measure of how plausible the hypothesis is, either before or after the observation is made.

The first theorem about conjunctive forks was independently established by Van Rijsbergen [19] and Forster [7]. It says that T provides at least as much information about C as it does about E :

Theorem 3: If TCE is a conjunctive fork with vertex C , then $M(C,T) \geq M(E,T)$.

A proof is provided in the Appendix.

What does this theorem say about the conditions under which observing a trace provides more knowledge about the trace’s cause than about some other effect (possibly in the future) of that cause? We wish to emphasize that mutual information describes *how much one learns*, not how much one knows in the light of the observation. In the vocabulary introduced above, it is a diachronic measure of the (average) change in opinion brought about by the observation. This is quite different from the synchronic measure — provided by what we have called $Total(-/-)$ — of how sharp the distribution can be expected to be in the light of the observation.

So in terms of *learning*, it is quite reasonable to think that conjunctive forks open to the future give rise to a temporal asymmetry. We observe a trace *now*. This may affect our degree of confidence in the state of the *past* event C ; it also may affect our degree of confidence in the state of the *future* event E . When the fork is conjunctive, T ’s impact on C cannot be less than T ’s impact on E .

As noted above, how much you learn from T about C (or E) should not be confused with how much you know about C or (E) in the light of T . It is not an unconditional fact about conjunctive forks that $Total(C/T) \geq Total(E/T)$. This is easy to see once two simple sums are placed side-by-side:

$$(6) \quad Total(C/T) = Prior(C) + Mutual(C,T).$$

$$(7) \quad Total(E/T) = Prior(E) + Mutual(E,T).$$

Theorem 3 asserts an unconditional relationship between the mutual information terms in each equation. Simple arithmetic then provides a sufficient condition for $\text{Total}(C/T) \geq \text{Total}(E/T)$:

Theorem 4: If TCE is a conjunctive fork with vertex C , then $\text{Total}(C/T) \geq \text{Total}(E/T)$ if $\text{Prior}(C) \geq \text{Prior}(E)$.

In other words, one can expect to know more about the past event C than about the future event E , after observing the state of the trace T , if entropy increases in the main process from C to E .

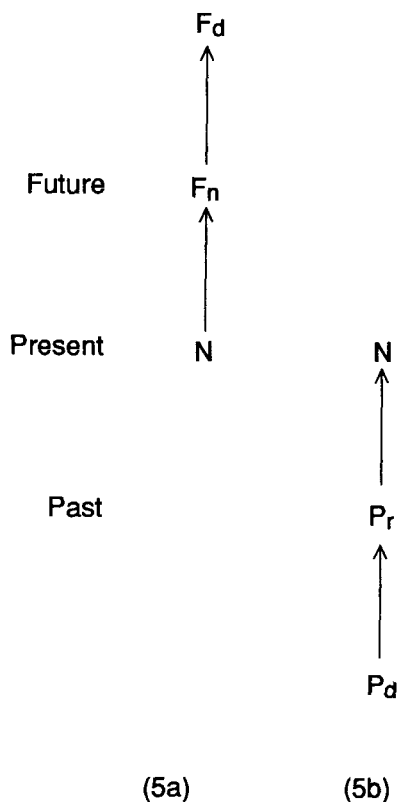
The approximate Second Law of Thermodynamics says that entropy never decreases in closed systems. However, we do not think this shows that Theorem 4 is the key to understanding why we know more about the past than about the future. For one thing, many processes are not thermodynamically closed. Furthermore, when entropy is given a nonthermodynamic interpretation, there is no guarantee that entropy will increase.¹⁵ Still, Theorem 4 is not without its interest. We have here a sufficient condition that connects the idea of a conjunctive fork with the problem of why we know more about the past than about the future. Cases not covered by this sufficient condition will be discussed in the next section.

We wish to emphasize that Theorems 3 and 4 apply to conjunctive forks open to the past just as much as they apply to conjunctive forks open to the future. No temporal assumption figures in their statement or proof. For this reason, these theorems do not, by themselves, give rise to a *temporal* asymmetry. Matters change, however, when they are brought into contact with Theorems 1 and 2. For if a common effect rarely forms a conjunctive fork with its joint causes, while a common cause often forms a conjunctive fork with its joint effects, then the asymmetries established in Theorems 3 and 4 do give rise to *temporal* asymmetries.

Theorems 3 and 4 apply to any triplet of events such that one of them renders the other two conditionally independent. We have thought of these results as applying to the case in which it is a common cause that screens off. But other structures can exemplify this pattern as well. Consider the causal process depicted in Figure 5a from present (N) to near future (F_n) to distant future (F_d). If the process is Markovian, then F_n screens off N from F_d . In this case, learning the state of N can be expected to change one's opinion about F_n more than it will change one's opinion about F_d (Theorem 3). And when entropy increases with time, observing the present allows one to know more about the near future than about the distant future (Theorem 4).

¹⁵ In Barrett and Sober [1], we discuss entropy's relevance to the issue of temporally asymmetric inference. Some examples of how entropy behaves in nonthermodynamic situations (e.g., in evolving Mendelian populations) are discussed there.

Figure 5



Symmetrical remarks apply to the case (depicted in Figure 5b) in which one observes the present (N) and makes inferences concerning both the recent past (P_r) and the distant past (P_d). If the intermediate node screens off the two end points from each other, then the present is more informative about the recent past than about the distant past (Theorem 3). And when entropy decreases with time, observing the present allows one to know more about the recent past than about the distant past (Theorem 4).

Change in entropy provides a sufficient condition for various epistemological asymmetries. In the case of conjunctive forks open to the future, entropy increase in the main process suffices for C to be better known than E . But sufficiency is not necessity. We turn now to cases of declining entropy. What epistemological asymmetries do they exhibit?

VI. Declining Entropy in the Main Process

In this section we examine cases in which entropy declines in the main process from C to E ; the goal is to determine when this constraint permits the past to be better known than the future. Subtracting equation (7) from equation (6), we obtain:

$$(8) \quad [\text{Total}(C/T) - \text{Total}(E/T)] = [\text{Prior}(C) - \text{Prior}(E)] + [\text{Mutual}(C,T) - \text{Mutual}(E,T)].$$

The rightmost bracketed term is known to be positive by Theorem 3. The first bracketed term on the right side is by assumption negative; that's what decreasing entropy means. It follows, then, that the sign of the left side will depend on which of the two terms in brackets on the right side is larger. This appears to be a complex matter.

The probability distribution of a conjunctive fork TCE has five degrees of freedom. If we choose values for $P(C_1)$, $P(T_1/C_1)$, $P(T_1/C_0)$, $P(E_1/C_1)$, and $P(E_1/C_0)$, the entire probability distribution is determined. That is, all eight quantities of the form $P(T_i \& C_j \& E_k)$ can be computed from the five values just given.¹⁶ Moreover, the five values can be chosen independently. This creates, in effect, a five-dimensional problem.

We do not have an analytical solution to this problem. However, we investigated the behaviour of (8) via computer simulations. For notational convenience, we label the five values as follows:

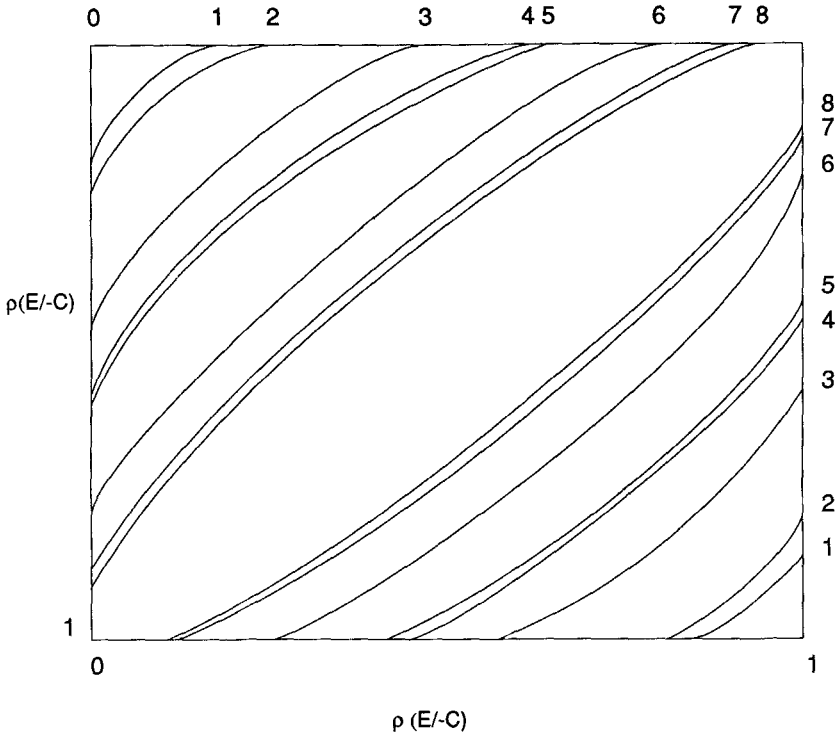
$$\begin{aligned} c &= P(C_1) \\ t &= P(T_1/C_1) & x &= P(E_1/C_1) \\ u &= P(T_1/C_0) & y &= P(E_1/C_0) \end{aligned}$$

The variables c , x , and y govern the main process (from C to E) in the fork; c , t , and u govern the branch process that leads from C to the trace T . We began by fixing c ; initially, we set $c = 0.5$. Since at this value the prior information about C — $\text{Prior}(C)$ — attains its minimum value (i.e. the entropy at C is maximal), $c = 0.5$ insures entropy non-increase from C to E . We then treated t and u as parameters to be varied; for each assignment of values to t and u , we computed the left side of (8) throughout the unit square in the x - y plane (that is, for all values of x and y in a discrete array).

Figure 6 depicts the results of eight such simulations. In each case, $[\text{Total}(C/T) - \text{Total}(E/T)]$ is positive within elliptical region running from lower-left to upper-right; within this region the *past* is better known than the *future*. A negative value for this quantity represents cases in which the *future* is better known than the *past*; these correspond to areas outside the ellipse.

¹⁶ Five values suffice rather than seven because of the screening-off condition. For example, $P(T_1 \& C_1 \& E_1) = P(T_1 \& E_1 / C_1)P(C_1) = P(T_1 / C_1)P(E_1 / C_1)P(C_1)$, where the last equality follows because C screens off T from E .

Figure 6



The eight cases depicted in Figure 6 correspond to the following assignments. In each, recall that we set $c = 0.5$:

t	u
1. $P(T_1/C_1) = 1.0$	$P(T_1/C_0) = 0.2$
2. $P(T_1/C_1) = 0.9$	$P(T_1/C_0) = 0.1$
3. $P(T_1/C_1) = 0.99$	$P(T_1/C_0) = 0.49$
4. $P(T_1/C_1) = 0.75$	$P(T_1/C_0) = 0.25$
5. $P(T_1/C_1) = 0.87$	$P(T_1/C_0) = 0.37$
6. $P(T_1/C_1) = 0.98$	$P(T_1/C_0) = 0.78$
7. $P(T_1/C_1) = 0.6$	$P(T_1/C_0) = 0.4$
8. $P(T_1/C_1) = 0.79$	$P(T_1/C_0) = 0.59$

Moving down the list increases the region within which the future is better known than the past.

The major axis for all these ellipses is the same.¹⁷ The minor axis varies depending on t and u . The ellipse gets fatter as $|t-u|$ gets larger — that is, as the strength of the correlation between C and T increases. When C and T are highly correlated, $\text{Total}(C/T)$ exceeds $\text{Total}(E/T)$ for ‘most’ values of x and y .

¹⁷ This major axis may be shifted by selecting a value for $P(C_1)$ other than 0.5.

We hasten to add that ‘most’ does not mean ‘all’. Even when T and C are strongly correlated, there still are regions in which the future is better known than the past. This will be true when $P(E_1/C_1)$ and $P(E_1/C_0)$ are either both close to 1 or both close to 0. In other words, when $P(E_1)$ is close to 1 or close to 0, the future will be better known than the past, if $P(C_1) = 0.5$. This is entirely intuitive; if one is *a priori* quite certain about the future and *a priori* quite uncertain about the past, observing a highly reliable trace may not reverse this asymmetry.¹⁸

The computer evidence strongly suggests the following conclusions. The minor axis of the elliptical region increases monotonically with $|t-u|$, and also with $|(t+u)/2 - 1/2|$.¹⁹ Whenever C and T are well-correlated ($|t-u|$ is near 1) or either $u = P(T_1/C_0)$ is near 0 or $t = P(T_1/C_1)$ is near 1, $\text{Total}(C/T)$ exceeds $\text{Total}(E/T)$ for ‘most’ values of x and y .

Traces may vary in their reliability. A ‘good’ trace, we might say, is highly correlated with the event of which it is a trace. Ordinary usage also assigns a causal meaning to the word ‘trace’. For T to be a trace of C , C must cause T . Even if we had excellent indicators of the future, these would not be *traces* of the future, properly speaking.

This causal requirement (coupled with the assumption that cause must precede effect) does *not* entail that a trace allows us to know the past better than we know the future. Although T is a trace of C but not of E , it does not follow just from this that $\text{Total}(C/T) > \text{Total}(E/T)$.

Nevertheless, a *good* trace in a conjunctive fork will enshrine this epistemological asymmetry, except when the past is highly uncertain *a priori* and the future is highly certain *a priori*. The goodness of a trace comes in degrees; so does the *a priori* degree of certainty attaching to past and future. The results represented in Figure 6 show how these quantities may be traded off against each other while preserving the temporal asymmetry.

We do not claim that ‘most’ conjunctive forks open to the future are such that $\text{Total}(C/T) > \text{Total}(E/T)$. We do not know how to quantify the idea of ‘most’ in this setting; in addition, it is unclear how often the trace in a fork is a good one. However, we are more confident about the pattern that obtains for those conjunctive forks with good traces that human beings bother to discuss.

We usually apply the word ‘trace’ to events that we think are good traces. What is more, we usually bother to gather the evidence provided by traces only when we are not already certain about the states of C and E . So in practice, we usually make inferences within conjunctive forks only when we believe $P(C_1)$ and $P(E_1)$ are not extreme and when we believe that C and T are reasonably highly correlated. For cases of this sort, the trace allows us to know more about the past than about the future.

¹⁸ An example: let C be the diet you ate at age 10, T your present medical records that describe that diet, and E whether you will be dead by age 90. We are rather uncertain about C but highly certain about E before we look at the medical records. This difference may remain in force even after we look at the records, depending on how reliable the records are as traces of your earlier diet.

¹⁹ This latter quantity measures the shift of the $[u,t]$ interval away from ‘dead centre’ of $[0,1]$.

VII. Conclusion

The question 'why do we know more about the past than about the future?' needs to be regimented. Otherwise, it will not be clear what the appropriate paired comparisons ought to be. The two of us know more about what we will eat for dinner tomorrow than we know about what Boswell and Johnson ate on any given night. But surely this is to compare apples and oranges.

In this paper we have regimented the problem of comparison by envisaging a main process that extends from past to present to future. A present trace, which branched off from that main process some time in the past, is the object whose state we are able to observe. On the basis of that observation, we ask whether the *past state of the main process* is better known than the *future state of the main process*. This is not the only way to make the comparative problem precise, but it is an important one to consider.²⁰

The resulting structure is a fork open to the future. The earliest event we called the 'cause' (C) of the main branch effect (E) and the trace (T). The causal language here is incidental to some of the results we obtained. When TCE is a conjunctive fork open to the future, certain informational asymmetries exist. The fact that C is the common cause of T and E may be important, but it is not a condition on which Theorems 3 and 4 depend.

It is also an incidental fact that E occurs in the future. E could be any event on the main process that happens after C . Because of our interest in temporal asymmetry, we assumed that E occurs after both C and T . What this shows is that the informational asymmetries do not depend on the futurity of E , but on E 's topological relationship to C and T .

We have explored some of the epistemological consequences of conjunctive forks open to the future. But for this to reveal some real epistemological asymmetry, we had to consider whether there are conjunctive forks open to the past and whether such forks, were they to exist, would enshrine asymmetries that are mirror images of the ones just mentioned.

Theorems 3 and 4 apply to any conjunctive fork. If common effects usually form conjunctive forks with their joint causes, full symmetry would be restored. But such forks are rare, or so we argued in connection with Theorem 1. In contrast, we took Theorem 2 to show that common causes often form conjunctive forks with their joint effects. Altogether, these four theorems point to a genuine epistemological asymmetry. Conjunctive forks of one kind are common, whereas conjunctive forks of the other kind are rare. And the common kind does show why we often know more about the past than about the future.

Still, there are reasons to be circumspect about thinking that this analysis is the key to the general epistemological problem. Indeed, we are not confident that there is a single key; perhaps the temporal asymmetry in our knowledge is a mixture of several different phenomena (Sober [16]).

²⁰ Another case to consider involves a process that extends from past to present to future. We observe the present state of the process and, on that basis, retrodict the process's past and predict the process's future. This 'topology' of the inference problem is explored in Barrett and Sober [1].

One limitation is that this analysis focuses exclusively on conjunctive forks. This has more generality than might first appear, in that it also encompasses causal chains in which the intermediate node screens off the end points from each other.²¹ However, there also are *nonconjunctive* forks as well as other topologies whose inferential properties we have not considered.

A second limitation is that the framework used here requires a somewhat definite probability model of the various processes involved. We formulate an inference about the past and an inference about the future. We then ask under what circumstances the past is better known than the future. The answer depends on relationships that obtain among the relevant probabilities. If entropy in the main process increases with time, then a simple asymmetry results (Theorem 4); if entropy declines, a more complex analysis is required. For examples in which it is possible to say what happens to the entropy, our results apply. But when this is something we cannot judge, our results do not show why the world is epistemologically asymmetrical.

The difficulty here is not that entropy is a concept that applies only to chambers of gas. Entropy is not *that* limited a concept; it is well-defined whenever probability is.²² Rather, the problem arises when we are unable to say anything about the probability distributions attaching to events *C* and *E*. Bayesians will not feel the bite of this point, since they think that such probabilities are always well-defined. But for those who reject this Bayesian idea, the fact that our treatment requires a probability model may represent a real limitation.

However, we do not believe that this limitation is *egregious*. How much it matters will vary from one application to another. The probabilities we have talked about are often empirically determinable. Once a scientist subsumes a token event under some suitable type-description, it often will be possible to infer probabilities from observed frequencies.²³

We have approached the problem of why we know more about the past than the future within a probabilistic framework. This is at variance with many philosophical explications of the concept of knowledge, but is more in accord with how philosophers sometimes approach the concept of justified belief. It may be doubted whether a piece of papyrus gives us knowledge of the Pharaoh's court or whether a fossil gives us knowledge of an extinct animal's morphology.²⁴ Such doubts presuppose a strong reading of the concept of knowledge that is quite

²¹ The results of this paper apply to such chains when we imagine ourselves to observe the state of one of its endpoints. The case in which we observe the state of the midpoint is treated in Barrett and Sober [1].

²² This point is developed in Barrett and Sober [1] in connection with Earman's [4] and Horwich's [8] discussions of entropy.

²³ We believe that our analysis is consistent with one standard criticism of Bayesianism. Critics often insist that a hypothesis has a probability only if it describes the possible outcome of a chance process. They therefore deny that scientific laws have probabilities. Note that the statements that receive probabilities in our treatment describe *events*; they do not express universal generalizations.

²⁴ For example, Dretske's [3] theory of knowledge says that the state (*f*) of a fossil provides knowledge of the state (*a*) of an ancient ancestor only if the fossil would not have been in state *f* if the ancestor had not been in state *a*. We think that there is a strong conventional (or

alien to the kind of evidence that the sciences are able to provide. We do not take this to show that such strong readings of the knowledge concept are *mistaken*, but that they are *irrelevant*. Those who think that knowledge imposes such strong requirements should view this paper as discussing, not knowledge, but justified belief. The question is not why we *know* more about the past than about the future, but why we have *better justification* for our beliefs about the past than for those that are about the future.

It may be objected that the approach we have adopted cannot hope to explain the temporal asymmetry in our knowledge, since that knowledge was temporally asymmetric long before anyone thought to apply a probability model. We reject this criticism. We believe that we have identified a family of probability asymmetries in the world.²⁵ Conjunctive forks *in nature* have this property: when we learn about them, we will find ourselves better able to know the past than the future. Our results provide some justification for the commonsense idea that we know more about the past than about the future, even though the framework we have used is not part of common sense. Just as Reichenbach suspected, it is asymmetries in the causal structure of the world that engender temporal asymmetries in what we know about the world.²⁶

University of Wisconsin, Madison

Received September 1990

Revised March 1991

Appendix: Proof of Theorem 3

We show that $M(T,C) > M(T,E)$ if TCE is a conjunctive fork with vertex C . We follow Van Rijsbergen's [19] proof. Recall that

$$M(T,C) = \sum_{i,j} P(T_i \& C_j) \log [P(T_i \& C_j) / P(T_i)P(C_j)]$$

$$M(T,E) = \sum_{i,k} P(T_i \& E_k) \log [P(T_i \& E_k) / P(T_i)P(E_k)].$$

Where H is the entropy function, let

$$A = H(T) - M(T,C)$$

$$B = H(T) - M(T,E).$$

²⁴ continued . . .

contextual) element in how this counterfactual should be interpreted. However, on one of its natural readings, the condition is *not* satisfied. Relative to that reading of the counterfactual, Dretske's theory would have us conclude that the fossil does *not* provide knowledge of the ancestor's morphology. The sciences typically confront situations in which a given body of data has more than one possible explanation. If this renders the sciences incapable of providing *knowledge* (properly so called), so much the worse for the concept of knowledge.

²⁵ In arguing for the various asymmetries discussed in this paper, we have taken the concept of cause and our everyday and scientific descriptions of causes at face value. Our arguments therefore do not address various conventionalist theses about causality or about the appropriate description of causes.

²⁶ We thank Ellery Eells, Malcolm Forster, and the anonymous referees of this Journal for useful suggestions.

Then

$$A = \sum_{-i,j,k} P[T_i \& C_j \& E_k] \log P(T_i / C_j)$$

$$B = \sum_{-i,j,k} P[T_i \& C_j \& E_k] \log P(T_i / E_k).$$

That C screens off E from T means that

$$P(T_i / C_j \& E_k) = P(T_i / C_j), \text{ for each } i, j, k.$$

This entails that

$$A = \sum_{-j,k} P(C_j \& E_k) [\sum_i P(T_i / C_j \& E_k) \log P(T_i / C_j \& E_k)].$$

A and B may be compared using Kraft's inequality, which is the following result from information theory:

$$\sum_{i,j,k} P(T_i / C_j \& E_k) \log P(T_i / C_j \& E_k) \geq \sum_{i,j,k} P(T_i / C_j \& E_k) \log P(T_i / E_k).$$

This entails that $A \leq B$, which means that

$$H(T) - M(T, C) \leq H(T) - M(T, E).$$

So $M(T, C) \geq M(T, E)$.²⁷

REFERENCES

1. M. Barrett and E. Sober, 'Is Entropy Relevant to the Asymmetry between Retrodiction and Prediction?', *British Journal for the Philosophy of Science*, forthcoming.
2. N. Cartwright, 'Causal Laws and Effective Strategies', *Nous* 13 (1979) pp. 419-437.
3. F. Dretske, *Knowledge and the Flow of Information* (Cambridge, MA: MIT Press, 1982).
4. J. Earman, 'An Attempt to Add a Little Direction to "The Problem of the Direction of Time"', *Philosophy of Science* 41 (1974) pp. 15-47.
5. E. Eells, *Probabilistic Causality* (Cambridge: Cambridge University Press, 1991).
6. E. Eells and E. Sober, 'Probabilistic Causality and the Question of Transitivity', *Philosophy of Science* 50 (1983) pp. 35-57.
7. M. Forster, 'In Defence of a Causal Theory of Mental Representation', unpublished manuscript.
8. P. Horwich, *Asymmetries in Time* (Cambridge, MA: MIT Press, 1987).
9. A. Khinchin, *Mathematical Foundations of Information Theory* (New York: Dover, 1957).
10. H. Reichenbach, *The Direction of Time* (Berkeley: University of California Press, 1956).
11. B. Skyrms, *Causal Necessity* (New Haven: Yale University Press, 1980).
12. E. Sober, *The Nature of Selection: Evolutionary Theory in Philosophical Focus* (Cambridge, MA: MIT Press, 1984).
13. E. Sober, 'The Principle of the Common Cause' in J. Fetzer (ed.), *Explanation and Causation: Essays in Honor of Wesley Salmon* (Dordrecht: Reidel, 1988), pp. 211-228.
14. E. Sober, *Reconstructing the Past: Parsimony, Evolution, and Inference* (Cambridge, MA: MIT Press, 1988).
15. E. Sober, 'Independent Evidence About a Common Cause', *Philosophy of Science* 56 (1989) pp. 275-287.
16. E. Sober, 'Temporally Asymmetric Inference in a Markov Process', *Philosophy of Science*, forthcoming.
17. P. Suppes, *A Probabilistic Theory of Causality* (Amsterdam: North Holland Publishing Co, 1970).
18. B. Van Fraassen, 'The Charybdis of Realism: Epistemological Implications of Bell's Inequality', *Synthese* 52 (1982) pp. 25-38.
19. K. Van Rijsbergen, 'A Discrimination Gain Hypothesis', *Proceedings of the 6th International ACM Conference on Information Retrieval* (London: Association for Computing Machinery, 1983) pp. 101-104.

²⁷ Forster [7] proves a stronger result. He shows, not just that T can be expected to provide at least as much information about C as it does about E, but that each possible state of T in fact provides at least as much information about C as it does about E.