7. Conclusion

Did the April rain cause the forest fire? Does the physician who saves your life cause your death? Was the publication of *The Origin of Species* caused by the illness that retarded Darwin's writing? My explanation of the asymmetry between delaying and hastening is neutral about the answer to these questions. My claim that, in general, to delay an X is primarily to prevent an X, and only secondarily to bring about an X, is consistent with the thesis that such anaemic bringing about is not causing. But it is also consistent with the thesis, which I find attractive, that although such bringing about really is a kind of causing, the primacy of preventing in such causing distracts us from this fact, with the result that delayers tend not to be, for us, among the salient causes of what they delay.[28]

New College
Oxford OX1 3BN
UK

PENELOPE MACKIE

REFERENCES

Bennett, J. 1987: "Event Causation: the Counterfactual Analysis", in Tomberlin 1987, pp. 367-86.

——1988: *Events and their Names*. Indianapolis, IN: Hackett Publishing Company.

Lewis, D. 1986: "Events", in his *Philosophical Papers II*. Oxford: Oxford University Press, pp. 241-69.

Lombard, L. B. 1990: "Causes, Enablers, and the Counterfactual Analysis". *Philosophical Studies*, 59, pp. 195-211.

Mackie, P. 1991: "Causing, Enabling, and Counterfactual Dependence". *Philosophical Studies*, 62, pp. 325-30.

Tomberlin, J. E., ed. 1987: *Philosophical Perspectives, 1: Metaphysics, 1987*. Atascadero, CA: Ridgeview Publishing Company.

[28] Of the people who delayed the completion of this paper by providing comments on earlier versions I am especially grateful to Stephen Blamey, Robert Frazier, Alfred Mele, Hugh Rice, and Ralph Walker.

Agency and Causal Asymmetry [1]

HUW PRICE

Here are two features of causation that an adequate theory of causation might be expected to explain: (*i*) the causal relation is asymmetric—if A is a cause of B then B is not a cause of A; (*ii*) effects never (or almost never) occur *before* their causes. One of the tasks of a theory of causation is thus to explain the difference between causes and effects, to reveal the true point of the "arrow" of causation. Another is to explain why the arrow of causation is so well aligned with the arrow of time. The latter task obviously presupposes the former: if the causal relation were symmetric, it could have no preferred temporal orientation.

In this paper I want to criticise the leading contemporary attempt to address these issues, an approach advocated in different forms by David Lewis, David Papineau and Dan Hausman, among others. My main objection rests on the claim that this approach is unable satisfactorily to account for causal asymmetry in microphysics, given the time-symmetric nature of physical theory. Rather more briefly, the paper also urges the attractions of an alternative approach. This takes causal asymmetry to be perspectival in character, a manifestation of the fact that causal concepts originate in our experience as agents.

Both approaches would be superfluous if we were content to account for causal asymmetry as Hume did, namely by saying that the causal and temporal arrows are related by definition. On this view it is held that the fundamental causal relation is symmetric in time, and that we simply use the different terms *cause* and *effect* to distinguish the earlier and later members of a pair of events so related (so that the phrase *is a cause of* is simply shorthand for something like *is earlier than and causally related to*). It has often been noted that there is a heavy price to be paid for this convenience, however. As Hausman (1986, p. 143) puts it, this approach "rules out not only the existence of simultaneous or backward causation, but their logical possibility as well". It also precludes the project, attractive to many, of explicating temporal order in terms of causal order.

Recognizing the inadequacy of the Humean solution, a number of writers have recently been attracted to a more subtle approach. Their suggestion is that the asymmetry of the causal relation is not merely a conventionally bestowed image of temporal ordering, but rests on some intermediate asymmetry—on a temporally asymmetric feature of the world which is itself typically (though perhaps not

[1] I am grateful to David Braddon-Mitchell, Phil Dowe, John Forge, Dan Hausman, David Konstan, David Lewis, Robert McLaughlin, Peter Menzies, Michaelis Michael, David Owens and Michael Tooley for comments and discussions on these topics, and also to participants in seminars in Canberra, Sydney and Uppsala.

invariably) aligned with the temporal ordering. One such intermediary asymmetry has seemed especially promising. It is what has become known as the *fork asymmetry*: the fact that regular correlations between spatially separated events commonly turn out to be associated with joint correlations with some *earlier* event, but rarely, if ever, with a *later* event. More formally, when a probabilistic dependence between separated events A and B is screened off by the occurrence of a third event C, this screening event is found to lie in the common past of A and B, not in their common future: we find "future-directed open forks" in the world, but not past-directed open forks. [2]

In outline, the suggestion is thus that the asymmetry of cause and effect is to be explained in terms of the predominance of future-directed forks in the world. The fork asymmetry, or some closely related feature of the world, is *constitutive* of the distinction between cause and effect. The difference between causes and effects actually consists in some fact about the structure of the world, of the kind involved in the fork asymmetry. One way to develop this suggestion is that chosen by Hausman (1984, 1986). Hausman begins with a symmetric notion of causal connection. He then uses this symmetric notion to formulate a criterion which, given the fork asymmetry, seems to be satisfied by (what we intuitively think of as) causes but not by (what we intuitively think of as) effects. Arguably then, this property provides the crucial distinguishing characteristic of causes, as against effects.

However, I want to show that the fork asymmetry is not a sufficiently basic and widespread feature of the structure of the world to constitute the difference between cause and effect. In particular, this approach cannot make sense of our causal intuitions with respect to microphysics, or more generally what we might call "the physics of the very few". Here the fork asymmetry is absent because it depends on the statistical behaviour of large numbers of physical entities. The attempt to ground causal asymmetry on the fork asymmetry thus turns out to conflict with the reductionist intuition that macrocausation consists of a lot of little bits of microcausation, as well as with the fact that physicists are inclined to speak of asymmetric causation even in microphysics.

The main task of the paper is thus to argue that the leading contemporary approach to the problem of causal asymmetry is seriously flawed. As noted, however, I also want briefly to suggest an alternative approach, which originates in the proposal that causation should be explicated in terms of the notion of agency. Peter Menzies and I have urged the attractions of this general account of causation elsewhere.[3] Here I want to draw attention to a further advantage of the approach, namely that it seems to provide the only satisfactory account of causal asymmetry. However, as the main concern of the paper is with the deficiencies of the more popular account (based on the fork asymmetry), the bulk of the argument will be of interest even to those who find little attraction in the agency view.

[2] There is a useful discussion of the fork asymmetry in Horwich (1987).

[3] See Price (1991b), Menzies and Price (1992).

In criticizing the proposal to explicate causal asymmetry in terms of the fork asymmetry I shall take as my main target what I think may justly be regarded as the best known version of this proposal, that of David Lewis.[4] It is true that Lewis does not base his account of causal asymmetry on the fork asymmetry as such. For one thing his approach is indirect, in that he explicates of causation in terms of counterfactuals. As he recognises, such an account may trace both the asymmetry and the predominant temporal orientation of causation to an asymmetry concerning counterfactuals:

> The way the future is depends counterfactually on the way the present is. If the present were different, the future would be different. ... In general the way things are later depends on the way things are earlier. Not so in reverse. Seldom, if ever, can we find a clearly true counterfactual about how the past would be different if the present were somehow different. (1986, p. 32)

The task of explaining causal asymmetry now becomes that of explaining the counterfactual asymmetry. Lewis attempts to explain the asymmetry of counterfactual dependence in terms of what he refers to as the asymmetry of overdetermination. As we shall see, this is closely related to the fork asymmetry. In effect, Lewis thus interposes a fourth arrow, an asymmetry of counterfactual dependence, between the arrow of causation and the arrow constituted by the fork asymmetry.[5]

Let us say, following Lewis, that A is a determinant of B if A is minimally sufficient for B, given the laws of nature. Then the claimed asymmetry of overdetermination consists in the fact that events typically have very few earlier determinants, but very many later determinants. As Lewis puts it:

> Whatever goes on leaves widespread and varied traces at future times. Most of these traces are so minute or so dispersed or so complicated that no human detective could ever read them; but no matter, so long as they exist. It is plausible that very many simultaneous disjoint combinations of traces of any present fact are determinants thereof: there is no lawful way for the combination to have come about in the absence of the fact. (Even if a trace could somehow have been faked, traces of the absence of the requisite means of fakery may be included with the trace itself to form a set jointly sufficient for the fact in question.) If so, the abundance of future traces makes for a like abundance of future determinants. We may reasonably expect overdetermination toward the past on an alto-

[4] See particularly Lewis (1986).

[5] Lewis himself is not altogether happy with this characterisation of his view as a version of the fork asymmetry approach to causal asymmetry. In correspondence he has pointed out to me that for one thing, the fork asymmetry is normally characterised in terms of probabilities, whereas his asymmetry of overdetermination is supposed to hold in a deterministic world in which all physical probabilities are zero or one. However, while I acknowledge that the issue deserves further clarification, I think that for present purposes what matters (and what will become clear) is that the two approaches are sufficiently close to be vulnerable to the same objection. A mark of their similarity is the fact that both take wave radiation to provide a paradigm example of their target asymmetry. (The similarity has previously been noted by David Papineau; see his 1985, p. 281.)

> gether different scale from the occasional case of mild overdetermination towards the future. (1986, p. 50)

What sort of fact about the world is the asymmetry of overdetermination? Lewis himself says that it

> is a contingent, *de facto* matter. Moreover, it may be a local matter, holding near here but not in remote parts of time and space. If so, then all that rests on it—the asymmetries of ... counterfactual dependence, of causation and openness—may likewise be local and subject to exceptions. (1986, pp. 50-1)

I want to agree with Lewis about the contingent character of the asymmetry of overdetermination; but to argue that even in these parts of time and space, it does not obtain to a sufficient extent to account for the asymmetry of causation.

The first task is to clarify the physical origins of this asymmetry. Why does it obtain around these parts to any extent? I shall try to throw some light on this issue by concentrating on a phenomenon that Lewis refers to as a special case of the asymmetry of overdetermination, namely the apparent asymmetry of radiation.

> There are processes in which a spherical wave expands outward from a point source to infinity. The opposite processes, in which a spherical wave contracts inward from infinity and is absorbed, would obey the laws of nature equally well. But they never occur. A process of either sort exhibits extreme overdetermination in one direction. Countless tiny samples of the wave each determine what happens at the space-time point where the wave is emitted or absorbed. The processes that occur are the ones in which this extreme overdetermination goes towards the past, not those in which it goes towards the future. I suggest that the same is true more generally. (Lewis 1986, p. 50)

This feature of radiative phenomena has been the subject of lengthy discussion in the physical and philosophical literature on time asymmetry. A particular concern has been the issue as to whether it comprises a non-thermodynamic temporal asymmetry in the world. One of those who have argued that it does—i.e., that there is a temporal asymmetry here which is not explained by thermodynamics—is Popper. Popper (1956) uses the example of a stone thrown into a still pond, which produces outgoing concentric waves on the water's surface. He points out that although this sort of occurrence is commonplace, we never observe the reverse process, in which incoming concentric waves arrive at the centre of a pond at just the right moment to expel a stone. The latter case is compatible with the laws of mechanics, which are time symmetric; but it never happens. Popper concludes that radiation thus exhibits a de facto temporal asymmetry, and argues that this asymmetry is not reducible to that of thermodynamics.

Like many other contributors to this debate, however, Popper fails to recognize the importance of the boundary conditions in this argument, and *their* connection with thermodynamics. The reason that outgoing waves are common is that the initial conditions that give rise to them are common. Solid objects are often so placed in the universe that they fall into ponds. And this can only be the case because our region of the universe is not in a state of thermodynamic equilibrium. If we consider a closed system which has reached thermodynamic equilibrium, then incoming flying stones will be just as unlikely as outgoing flying stones. Each could only occur as result of an incredibly unlikely quantum fluctuation, and pond surfaces will be undisturbed in either temporal sense for billions of years at a time.

In other words the apparent temporal asymmetry of radiation does depend on that of thermodynamics. We need thermodynamic disequilibrium in order to generate the conditions that make radiation appear to be asymmetric in time. The asymmetry depends on the fact that we have big disturbances (such as flying stones) in the initial conditions but not in the final conditions. When there are no big disturbances at either end, the situation is entirely symmetric. Note that this applies to other sorts of radiation, as much as to those on water surfaces. For example, the reason that electromagnetic radiation appears temporally asymmetric is that we have concentrated transmitters or sources of radiation—such things as stars and radio transmitters—but no corresponding receivers or sinks. Again, we only have such transmitters because the universe is very far from thermodynamic equilibrium; because entropy is much lower than its theoretical maximum. In a closed system in equilibrium, there would not be any such asymmetry. Concentrated transmitters would be just as rare as concentrated receivers.[6]

Thus in the case of radiation—the case that Lewis offers us as the paradigm example of overdetermination of the past by the future—the asymmetry depends on there being big disturbances as the source of the radiation. This means that it is an essentially macroscopic asymmetry. It depends on large-scale structural imbalances. Down at the microlevel, it will no longer be visible. Just as the pictorial characteristics of a printed picture disappear if we focus on the individual dots of ink that make it up, so the asymmetry of radiation disappears if we concentrate on the microstructure of the processes in which it is manifest.

I have argued that this is true of radiation. I want to suggest that it is equally true of other cases of overdetermination of the past by the future. Consider for example a universe consisting of a small number of Newtonian particles, moving in "random" fashion in some confined space—discs like ice hockey pucks moving without friction on the surface of a frozen sphere, perhaps. Suppose a particular event occurs at a particular space-time point: two particular discs collide at point X at time T. Is this event overdetermined in the future, any more than in the past? No: in either direction it is typically determined by nothing less than a complete description of the positions and momenta of all the discs at a particular time.

In general, the overdetermination of a past event by future events seems to provide a case of a future-directed fork, of the kind I mentioned at the beginning. The various future determinants are correlated with one another, having in common the fact that they imply the occurrence of the common past event for which they are determinants. Moreover, this sort of correlation seems to be connected with thermodynamics, in the sense that a state of extended correlation is a state of relative order, or low entropy. In practice we find that such states are preceded by states of lower entropy. So Lewis's asymmetry of overdetermination, like the fork

[6] These points are discussed in more detail in my (1991a) and (1991c).

asymmetry in general, would appear to be a product of thermodynamic disequilibrium. As before, this means that it is statistical and in that sense macroscopic in origin. It disappears when we focus too closely on the structure of the world.

This has important consequences for the asymmetry of counterfactual dependence, cashed in Lewis's terms. Perhaps the most striking way to illustrate these consequences is to consider a variant of the famous case of Nixon and the Bomb. An early objection to Lewis's account of counterfactuals was that (counterintuitively) it ought to make this sentence *false*:

(1) If Nixon had pressed the button, there would have been a nuclear war.

The argument was that a world in which Nixon presses the button, but in which some minute violation of the laws of nature then prevents a nuclear war, is much more like the actual world than one in which a nuclear war does take place. Lewis replies that the "miracle" required to bring history back into line, once the button has been pressed, will not be a small, local miracle: because the pressing is overdetermined by its effects, we need to assume a miracle which involves a corresponding range of widespread and diverse component miracles.

> To converge to [the actual world], a world where Nixon presses the button must break the links whereby a varied multitude of future conditions vastly overdetermine that he did not press it. The more overdetermination, the more links need breaking and the more widespread and diverse must a miracle be if it is to break them all. (1986, p. 50)

However, I have argued that the asymmetry of overdetermination is a product of macroscopic statistical considerations. If so, then the above argument ought to be unavailable to Lewis in a case in which we replace the macroscopic event of Nixon's pressing the button with some appropriate microscopic event. To vary the example as little as possible, let us concentrate on what goes on in Nixon's head. To make the point as vividly as possible, let us suppose that Nixon's brain is a thoroughly Newtonian mechanical system: his neurological state is completely determined by the positions and momenta of a huge number of Newtonian particles. (In the neurological context the appropriate Newtonian particle is of course the marble.)

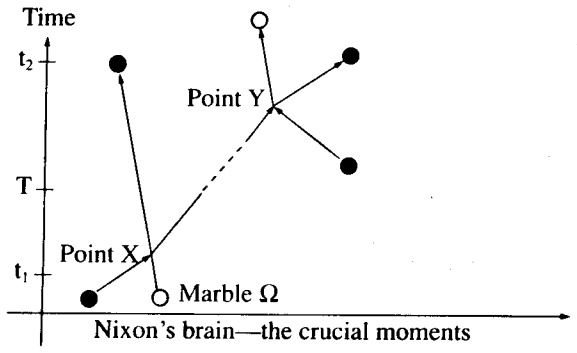


Diagram 1

We know that in the actual world Nixon's brain never occupied the particular state whose effect would have been to cause his finger to depress the final button. Perhaps it came close, however. Perhaps there was some particular marble, call it marble Ω, such that if Nixon had lost it at some time T in those dark final days before his resignation, he would have pressed the button. (If this seems implausible, think of the nail for want of which a kingdom was lost.) Now consider the counterfactual consequences of the supposition that Nixon did lose marble Ω at time T. It takes a very small miracle to get rid of a Newtonian marble; we simply need to assume that it ceases to exist at some point on its lonely trajectory between collisions. At time $t_1<T$, Ω sets off from Point X, where it has collided with some other marble (see Diagram 1); it sets off towards Point Y, where it is destined to make its next rendezvous at time $t_2>T$. A miracle intervenes, however, and it never arrives. As a result, Nixon presses the button and civilization as we know it ceases to exist.

I take it that against this background we would normally be inclined to assent to the following counterfactual proposition:

(2) If marble Ω had ceased to exist at T, there would have been a nuclear war.

The difficulty for Lewis is that in virtue of the microscopic nature of the counterfactual event in question, it is not overdetermined by its effects, in the manner required to generate an asymmetry for miracles. The miracle required to restore the actual course of history, in the event that Ω does cease to exist at T, is not diverse and widespread. It is simply that Ω should come into existence again between T and t_2, with whatever position and momentum it would have had if it had not ceased to exist in the first place. There are no other records or traces to adjust in the world. The fact that the action takes place between interactions means that Ω's absence goes unnoticed by the rest of the universe until t_2; and hence that God has at least this long to put a single local stitch in this tear in the fabric of the world, before the effects of Nixon's neurological aberration become dispersed.

Thus it seems that on Lewis's account, the counterfactual (2) should be regarded as false. Once we get below the statistical level, in other words, Lewis's account seems to imply that it ceases to be true that if the past were different, the future would be different. On the contrary, if the past were different the future would be the same; it is just that there would be some other difference in the past to counteract the difference initially supposed. This result immediately infects the theory of causality, as well. Even if according to Newtonian mechanics, marble arrangement *A* gives rise to marble arrangement *B*, it is not true on Lewis's theory that *A* causes *B*; for it is not true that if *A* had not happened then *B* would not have happened. *B* happens anyway, in the most similar worlds in which *A* is miraculously absent.

Note that I am not suggesting that it is a problem for Lewis simply that there may be worlds in which the asymmetry of counterfactual dependence and there-

fore the asymmetry of causation break down. As I noted earlier, Lewis himself readily acknowledges the possibility of such worlds. For example, he asks us to

> consider a simple world inhabited by just one atom.... You will doubtless conclude that convergence to this world takes no more of a varied and widespread miracle than divergence from it. That means, if I am right, that no asymmetry of counterfactual dependence prevails at this world. ... The asymmetry of miracles, and hence of counterfactual dependence, rests on a feature of worlds like [ours] which very simple worlds cannot share. (1986, p. 49)

The point of the present argument is not simply to confirm what Lewis here asserts, namely that such symmetric worlds are possible. It intended is to show that our world is very probably of this kind, once we get down to the microlevel. Hence an obvious objection is to point to its apparent dependence on a Newtonian model of the world. Given that our world is not Newtonian, doesn't this mean that the argument establishes no more than Lewis's own one-atom world? Isn't a Newtonian world just a more complicated example of a possible but non-actual world in which the relevant asymmetries break down?[7]

One way to reply to this objection would be to point out that if, contrary to present expectations, we discovered that we do live in a Newtonian world after all—or rather a world which is discrete and mechanistic in the Newtonian sense all the way down, so that even such phenomena as gravitation and electromagnetism came to be explained in these terms—this would not shake our conviction that if Nixon had lost the Ω marble then there would have been a nuclear war (given the assumed neurophysiological story, of course). It is important to keep in mind that the example does not depend on the details of Newtonian mechanics. The Newtonian model is simply the most familiar case of a physical theory whose basic structure involves discrete interactions. And quantum mechanics not withstanding, we do not *know* that our own world is not like that.

However, the more important point is that the example is only offered by way of an intuition pump to what is in any case a plausible hypothesis about the actual world, namely that the asymmetry of overdetermination is statistical and in that sense macroscopic in origin. If we accept this hypothesis then Lewis's analysis appears to entail that there is no causal asymmetry in microphysics.

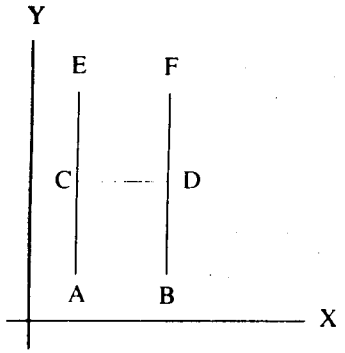
Moreover, although the above modification of the Nixon case provides a graphic and familiar framework within which to make this point, it does not put the difficulty for Lewis's account in its most pressing form. In the original Nixon case the claim was that miraculous reconvergence may yield a world which is more similar to the actual world than is a world in which the consequences of an initial divergence are allowed to develop—to which Lewis replied that it is the difficulty of the reconvergence that supplies the required dissimilarity. The above

[7] As Lewis puts it (in correspondence): "the world of [this] example differs from the world we take to be ours in more ways than meet the eye: it's not just a world of classical physics! Do marbles act as gravitational sources? Do they interact with the electromagnetic field? If so, then after a marble miraculously disappears, you won't get reconvergence just by putting the marble back (with appropriate position and velocity)".

argument has been that reconvergence need not be difficult in microphysical cases, and hence that the original objection may be resurrected. But to concentrate solely on the possibility of reconvergence would be to miss the more basic lesson, namely that Lewis's asymmetry of miracles fails for microphysical events: in general it is just as easy to *converge* to a world in virtue of a microphysical miracle as it is to *diverge* in virtue of such a miracle.

Consider for example the following gravitational case (depicted in Diagram 2). Let P be a freely falling uncharged elementary particle which is actually at point C at time T. Consider the miracle that puts P at D (a short distance in the direction X from C) at T. In the actual world P's spatial path is ACE. In one near world it is ACDF, with a discontinuity between C and D. In another near world it is BDCE, again with a discontinuity. Is there any difference in the "ease" of the miracles involved in these two worlds? No, for the gravitational field theory which governs the motion of P is time-symmetric.[8]

Diagram 2



The example does not depend on the fact that it involves a particle. It could be expressed in terms of the gravitational field alone, making the miracle a small discontinuous change in the field at T. Nor does it depend on the fact that we have not introduced electrodynamics. All it depends on is that the miracle concerned is too insignificant to need the help we actually get from thermodynamics. In our world we get big returns on small miracles: because he loses a neuron, Nixon presses the button and all hell breaks loose—again, for the want of a nerve a kingdom is lost. This would not happen in a world in thermodynamic equilibrium. More importantly, because even in the absence of equilibrium it is an asymmetry which essentially involves events big enough to display the relevant statistical characteristics, it does not show up if we just concentrate on little events.

To illustrate the difficulty the example leads to for Lewis's theory, suppose now that the particle P is a photon, arriving via point *A* from (say) a distant gal-

[8] Indeed, strictly we haven't yet been told which is the past and which is the future—i.e., in effect, in which direction the particle concerned is travelling—but there is nothing in physics that would enable us to use that information if it were provided.

axy. Its presence at E gives rise to some effect Φ. Hence if P had been at D at T, Φ would not have happened. For Lewis the truth of this counterfactual requires that the most similar worlds in which P is at D at T are worlds in which Φ does not occur. But a world in which P emerges on a slightly different course from the distant galaxy, traverses the path BD, but then shifts left at T so as to bring about Φ, seems no less similar to the actual world than the one in which P shifts right at T, thus preventing Φ. Convergence is as easy as divergence at this level, so that there is no asymmetry of miracles.

This objection is similar to one based on what Lewis (1986, p. 56) calls the *Bennett world*. The Bennett world is what we get by adding to a counterfactual future—itself the result of a small divergence from the actual world—the past derived by extrapolating backwards in accordance with the laws of the actual world. (In effect, in terms of our Diagram 2, it is the world in which the particle follows the path BDF.) The point of the Bennett world is that it is a world apparently very similar to the actual world to which, by definition, miraculous convergence is easy (at least once). However, whereas the Bennett world uses the original miracle to define a non-actual world to which convergence is easy, the present argument uses the reverse of the original miracle, so that convergence is to the actual world.

One consequence of this difference is that Lewis's reply to the Bennett world objection does not seem applicable to the present case. Lewis claims that the Bennett world is in fact quite unlike the actual world, on the grounds that after the time of the miracle in question, it contains many traces that seem to indicate its history is that of the actual world, whereas its true history is likely to be very different. Why is it likely to be different? Because (Lewis tells us)

> there is no reason to think that two lawful histories can, before diverging, remain very close throughout a long initial segment of time. To constrain a history to be lawful in its own right, and to constrain it also to stay very close to a given lawful history for a long time and then swerve off, is to impose two very strong constraints. There is not the slightest reason to think the two constraints are compatible. (1986, p. 57)

Whatever the merits of this last claim, the argument it serves to support seems unavailable to Lewis in the present case, for here convergence is to a world which is the actual world.

In defence of Lewis, however, it may seem that a variant of the reply will work in the present case. Could it not be argued that the world defined by the path BDCE is strange (and therefore more dissimilar than it seems to the actual world), in virtue of the fact that at time T all evidence that its past is different from that of the actual (ACE) world has vanished, apart from the aberrant position of the particle P itself? It shows misleading apparent traces of a past it does not have, and no traces of a past it does have—again, a miraculous cover up. Perhaps this provides the respect in which the world of ACDF is closer to the actual world than that of BDCE.

However, it seems to me very doubtful that this reply will be sufficiently general in its impact to deal with the endless potential counterexamples to Lewis's theory to be found in a time symmetric microphysics. Lewis requires in effect that almost all microevents at any given time T have the property that if God alters them slightly, and runs history back accordingly, the resulting history is very different from the actual history of world up to T; though of course the traces of this different history will have all but vanished by T, given that at that stage the world in question so closely resembles the actual world. In other words, Lewis requires that such a world would *typically* be one in which significantly more historical evidence was misleading than is actually the case. But our ordinary use of historical evidence appears to depend on the assumption that this is not so. For suppose it were the case that most minor variations of this kind would result in worlds whose historical evidence was significantly more misleading than we take our actual evidence to be—i.e., that almost all worlds that "look very much like ours" at the present time are *deceptive* worlds, in this sense. Then what grounds would we have for resisting the conclusion that our world is likely to be one of the deceptive worlds? After all, what we know about the world we inhabit is essentially that it is one of the worlds that now looks very much like *this*. If in the grand scheme of things almost all such worlds are deceptive worlds, would it not be presumptuous of us to assume that ours is one of the exceptions?[9]

In summary then, I suggest that Lewis's asymmetries of overdetermination and of miracles do fail in microphysics. So too therefore do the asymmetries of counterfactual dependence and of causation, if these are analyzed in Lewis's terms.

What is wrong with this conclusion? Could we not treat it as an important discovery about the world—a discovery made possible by Lewis's analysis—that there is no microscopic causal asymmetry?

I think that there are two main objections to this suggestion. One stems from reductionist intuitions. Many of us think that there is something fundamental about microphysics. For one thing, we like to think that higher-level properties and relations obtain in virtue of fundamental physical properties and relations. For another, we like to think that big physical things and events are simply collections of little physical things and events. As a result, we are attracted to the idea that macroscopic causation is constituted by a lot of microscopic causation. One aspect of this intuition is that causal connections between temporally separated events decompose "horizontally" into chains of more immediate causal connections; another is that these immediate connections, if not already primitive, decompose "vertically" into a complex of microphysical causal relations. So if there is no temporal asymmetry in these microphysical relations, it is hard to see how putting a lot of them together could make any difference. Some part of

[9] This point may also tell against Lewis's answer to the Bennett world objection. Note that it would be no use appealing in reply to an anthropic principle, as cosmologists are inclined to do to try to explain features of the universe that seem both improbable and fortuitous. In the present case the relevant near worlds are all worlds in which we exist and prosper to much the same extent as in the actual world; and hence if most of these worlds are deceptive worlds, our prosperity cannot be held to depend on our world's being non-deceptive.

the reductionist story will have to go, if we are to save asymmetric macrocausation.

It might be suggested that we could do justice to these reductionist intuitions without microcausation. Perhaps macrocausality depends on microstructure, but not specifically on micro*causal* structure.[10] But it seems to me that this fails to do justice to our intuitions. Compare the case of mass. According to Mach's Hypothesis the mass of a body depends on its relation to the universe as a whole. Mass thus becomes very strongly extrinsic. All the same, it remains essentially true that the mass of a larger body depends on the masses of the smaller bodies that comprise it. It seems to me that we expect causation to behave in a similar way.

Suppose however that we are prepared to give up causal reductionism. I think that there is another reason to regard the implications of Lewis's account for microphysical causation as unsatisfactory. This is that physicists themselves take for granted causal and counterfactual asymmetry, in the microworld as much as in the macroworld. No physicist doubts that the excitation of an electron may be said to be caused by an incident photon, for example; or that the later emission of a photon (temporally the mirror image of the former event) is an effect of the excitation that precedes it. Or at any rate, if any physicist does doubt such claims, it is because he or she doubts the applicability of the notion of causation in quantum mechanics, and not because we are dealing with microscopic, sub-statistical events.

One response to this observation about how physicists talk about causation would of course be to dismiss it, to regard it as no more than a product of a widespread misconception about the nature of causality. However, I think that this would be a particularly inappropriate defence of Lewis's account of causal asymmetry. For Lewis relies on such ordinary intuitions about causality at various crucial points. He insists for example that our search for the correct account of similarity between possible worlds be guided by what we already take to be the truth about what is to be analyzed in terms of this relation: "We must use what we know about counterfactuals to find out about the appropriate similarity relation—not the other way round" (1986, p. 43). So Lewis is poorly placed to reject expert opinion on these matters.[11]

Thus it seems to me that the apparent asymmetry of overdetermination is insufficiently general to account for the asymmetry of causation. To cope with reductionist intuitions, and to make sense of the use physicists make of causation in micro-domains, we need some other account of causal asymmetry. The

[10] Suggestions of this kind were put to me in correspondence by David Lewis and by David Owens, and in person by Michaelis Michael.

[11] Philosophers less committed to respecting commonsense intuitions have more freedom at this point, of course. However, one of the virtues of Lewis's treatment of causation is that it draws our attention to the fact that to dispense with (asymmetric) causal talk in physics would almost certainly be to dispense with (asymmetric) counterfactual talk as well. It is doubtful whether causal eliminativists have fully appreciated the extent of this loss.

argument appears to count not simply against Lewis's proposal, but also against other recent attempts to explain causal asymmetry in terms of the fork asymmetry, such as those of Hausman (1984, 1986) and Papineau (1985).[12] In effect, these writers cover the same ground as Lewis but by a more direct route. They attempt to explicate the asymmetry of causation in terms of an asymmetry in the world that parallels Lewis's asymmetry of overdetermination; but they do so directly, without Lewis's detour through counterfactuals. The detour may have advantages to compensate for its greater length, of course. That issue aside, Hausman's and Papineau's accounts may be expected to inherit the defects of Lewis's starting point, namely that the non-causal asymmetry they rely on is not sufficiently general to account for microcausation.

What are the alternatives? One might return to the Humean idea that causes are distinguished from effects merely in virtue of their order in time, so that it is simply analytic that causes precede their effects. But this has the disadvantages we noted at the beginning. Temporal precedence does not seem to provide the right sort of asymmetry to be what constitutes causal asymmetry. It fails on several counts, one of them being that like the asymmetry of overdetermination, it is not sufficiently widespread: in cases of simultaneous causation we appear to have causal asymmetry without temporal precedence. But if causal asymmetry is not simply temporal precedence, and if the point of the microphysical case is that at that level there is no other sort of temporal asymmetry in the physical structure of the world, then what in the world could possibly constitute causal asymmetry?

One possible answer is that what constitutes causal asymmetry is something over and above the aspects of the world with which physics is immediately concerned. This will be the view of a strongly realist non-Humean account of causation, of the sort advocated by Michael Tooley (1987). According to a view of this kind it is not surprising that we do not find causal asymmetry within microphysics. Strictly speaking, we don't find it within any sort of physics. However, the trouble is that the direction of causation then becomes epistemologically inaccessible: we simply have no way of knowing whether our ordinary ascription of the terms cause and effect is correct or back to front. The present argument shows that we cannot settle the matter by looking at microphysics. The only temporal asymmetry that could possibly constitute evidence one way or the other is the macroscopic thermodynamic asymmetry, and here the best explanation seems to lie not in causal connections between particular events (a path that would in any case lead us back to microphysics) but in cosmological constraints on the boundary conditions of the universe.

In my view a far more appealing alternative is to accept that in a certain sense causal asymmetry is not in the world, but is rather a product of our own asymmetric perspective on the world. We ourselves are strikingly asymmetric in time. We remember the past and act for the future, to mention but two of the more obvious aspects of this asymmetry. It does not seem unreasonable to expect that the

[12] David Owens (1992) also bases an account of causal asymmetry on the fork asymmetry.

effects of this asymmetry will have come to be deeply entrenched in the ways in which we describe our surroundings; nor that this entrenchment should not wear its origins on its sleeve, so that it would be easy to disentangle its contribution from that part of our descriptions we might then think of as perspective-independent. After all, there is a familiar model for such entrenchment in the notion of the secondary qualities. The ordinary categories of colour, smell, taste and so on are now recognized to be in part a product of the peculiarities of our own sensory apparatus. However, not only is the appreciation of this point far from a trivial intellectual achievement; its implications and precise formulation are very much a matter for current philosophical debate.[13]

The analogy with the familiar secondary qualities thus serves to meet two sorts of objection to the proposal that causal asymmetry might be perspective-dependent. The first objection is that if causal asymmetry were perspectival, this would already be obvious to us, and accordingly we would be less inclined to regard the cause–effect distinction as an objective matter. The second objection follows on, claiming that this view would therefore conflict with the practice of physicists, who do treat this distinction as having objective significance. (This point would of course be particularly damaging, if sustained, given that I have claimed that it is an objection to Lewis's view that it fails to validate physical practice.) In reply, I take it that the lesson of the secondary qualities is firstly that it is far from easy to discern the distortions that originate in the peculiarities of our own perspective on the world; and secondly that even when they are discerned, it is far from clear what we should do about them—far from clear, in particular, that the right thing would be to abandon the language of the secondary qualities, at least in scientific discourse. It is not clear that this is a serious option.[14]

But what feature of our perspective could it be that manifests itself in the cause–effect distinction? The answer may lie in the agency or manipulability theory of causation in general. The core of this theory is the view that we acquire the notion of causation in virtue of our experience as agents. Roughly, to think of *A* as a cause of *B* is to think of *A* as a potential *means* for achieving (or making more likely) *B* as an *end*. There are a number of stock objections to this approach to causation, which no doubt account for the fact that it has never been particularly popular. In my view these objections are greatly overrated. As Peter Menzies and I have argued (Menzies and Price, 1992), it turns out that they are closely analogous to, and no more forceful than, a range of objections that might be made against standard theories of colour. Once it is appreciated that causation is analogous to a secondary quality, in other words, the usual objections to the agency theory lose their force.[15]

[13] See for example Wright (1992), Johnston (1989, 1992), Pettit (1991) and Price (1991d). One focus of these discussions has been the question how best to represent the subjectivity apparently characteristic of the traditional secondary qualities; another, the issue how much of our discourse actually exhibits this or related forms of subjectivity.

[14] Even if it is an option with respect to the traditional secondary qualities, it may not be with respect to other concepts, which—although displaying an analogous kind of subjectivity—are such as to play a more significant role in science. In my view causation is one such concept, and probability another. For a brief discussion the subjectivity of the latter concept (and of the notion of entropy), see the first part of Price (1991a).

However, the present point is not that the agency account of causation lacks the disadvantages with which it has usually been saddled. It is rather that such an account has a significant and largely unrecognized advantage: it is particularly well placed to explain the nature of causal asymmetry, and its prevailing orientation in time. For it is able to say that the asymmetry of causation simply reflects (or better, perhaps, *projects*) that of the means–end relation. Causes are potential means, on this view, and effects their potential ends. The origins of causal asymmetry thus lie in our experience of doing one thing to achieve another; in the fact that in the circumstances in which this is possible, we cannot reverse the order of things, bringing about the second state of affairs in order to achieve the first. This gives us the causal arrow, the characteristic alignment of which with the temporal arrow then follows from the fact that it is normally impossible to achieve an *earlier* end by bringing about a *later* means.[16]

These remarks are merely introductory, of course. It needs to be explained how the means–end relation comes to have these characteristics—in particular, how it comes to have such a striking temporal orientation. A natural suspicion is that such an explanation would itself need to appeal to the asymmetry of cause and effect, thus invalidating the proposed account of causal asymmetry. However, it seems to me that this suspicion is bound to be unfounded, if the alternative is that causal asymmetry is to be explicated in terms of some objective asymmetry in the world, as it is by Lewis, Hausman and Papineau. For in this case the latter asymmetry will itself be available to explain the asymmetry of the means–end relation. Reference to causation in such an explanation will simply be construed as indirect reference to the underlying objective asymmetry; in which case the explanation does not make any essential or ineliminable reference to causation at all.

In other words, here is a plausible strategy for explaining the asymmetry and temporal orientation of agency: first explain it making full use of the ordinary notions of cause and effect; then rewrite the explanation, so that at those points at which it appeals to the asymmetry of cause and effect, it refers instead to the kind of objective physical asymmetry that Lewis, Hausman and Papineau take to constitute causal asymmetry. The result will be an explanation of the relevant features of agency which does not itself appeal to causal asymmetry.

[15] As Menzies and I note, the agency theory of causation admits a considerable number of variants, ranging from the view that our capacities as agents give us access to an objective causal relation, via projectivist and dispositional accounts, to an outright error theory. The argument of the present paper leaves most of these options open. At the end of the paper I briefly indicate the reasons for my own preference, which is for a form of projectivism.

[16] The status of backward causation then turns on the question as to whether there are or could be exceptions: cases in which we could coherently act for past ends. Dummett (1964) has defended the conceptual possibility. A number of writers (including myself, in my 1984) have suggested that quantum mechanics might provide some actual examples.

Let us explore this idea a little further. It is a familiar point that we know more about the past than we do about the future, and that we normally deliberate and act for the sake of the future but not the past. Moreover, it seems natural to explain these asymmetries in terms of that of causation: essentially, in terms of the observation that knowledge is an effect and actions are causes. Developing these ideas may be expected to yield a formal model of what is essential to our status as knowers, deliberators, and agents in the world. According to the agency theory, it will be this model that a creature must instantiate if it is to develop and possess the concept of causation—for causal concepts depend on these features in much the same way that colour concepts depend on our colour vision.

As it stands, the model will make free use of the notions of cause and effect. However, it seems that we will be free to regard references to these notions as place-holders for references to the physical asymmetries in the world on which the existence of such asymmetric entities actually depends. Indeed, a good indication of how this stage of the account might go is to be found in the approach to various temporal asymmetries recommended by Horwich (1987). Horwich suggests that the fork asymmetry underpins what he calls the *knowledge asymmetry*: the fact that "we know more about the past than we do about the future" (Horwich 1987, p. 201). The connection turns on the fact that "the processes that give us knowledge about the past are typically [future-directed open forks]" (Horwich 1987, pp. 201-2). Notice that there is no talk of causation here; what we use instead are the notions of process and correlation. Hence Horwich may legitimately go on to argue that the knowledge asymmetry underlies asymmetries of explanation and hence of causation itself. At this stage the account he offers differs from the agency theory of causation in several respects, but these differences needn't concern us here. The important point is simply that Horwich's project illustrates the way in which the existence of knowledge, deliberation and action may plausibly be held to rest on physical asymmetries in the world, and hence not to depend in any essential way on concepts such as causation (which it is suggested they be invoked to explain).

Finally, why will such an approach not be vulnerable to a version of the objection we have raised to Lewis's account of causal asymmetry, namely that it cannot account for the asymmetry of microphysical causation? Because agents are essentially macroscopic, and depend on the very thermodynamic asymmetry which is the source of the various physical asymmetries to which writers such as Lewis, Hausman and Papineau appeal. In other words, this route to an explanation of the asymmetry and temporal orientation of agency would invoke only the "correct"—i.e., extensionally correct—part of its opponents' theory of causal asymmetry.

I have suggested that with such an account in hand, the agency theory of causation has the basis for a superior account of causal asymmetry: an account that succeeds not only where Lewis's also appears to succeed, in the macroscopic case; but also where Lewis's fails, in the microscopic case. Indeed, the agency theory seems to offer the only viable account of causal asymmetry. Its great disadvantage may seem to be that it makes causal asymmetry an anthropocentric matter. I suggest that we acknowledge this consequence, but deny that it is a disadvantage. Its effect is merely to put causation in its proper metaphysical perspective, as something like a secondary quality.[17] As in the case of the more familiar secondary qualities, the shift in perspective may make us feel metaphysically impoverished, in losing what we took to be an objective feature of the world. The feeling should be short-lived, however. After all, if what we appear to have lost was illusory anyway then our true ontological circumstances are unchanged—and yet we will have made a direct gain on the side of epistemology, as we came to understand the source of the illusion.

The issue as to whether and in what sense an agency view of causation actually does entail that causation is illusory will depend on precisely how the account is developed. As I noted above, there are number of possibilities. These range from the view that our capacities as agents simply give us privileged access to a fully objective causal relation in the world, via projectivist accounts and theories modelled on dispositional treatments of colour, to an outright error theory. The argument of the present paper leaves most of these options open. The one option that seems to me ruled out by the present argument is that agency merely gives us privileged access to a fully objective *and asymmetric* causal relation—though it remains possible that it gives us access to a symmetric quasi-causal relation, so that only the asymmetry is perspectival.[18]

My own preference is for a form of projectivism. However, the case for this preference is lengthy, and necessarily engages with contemporary debates which have little specifically to do with the case of causation.[19] Hence it would be inappropriate as well as quite impractical for me to try to press the case within the confines of the present paper. All the same, both friends and foes of perspectivalism may be interested in some pointers to the territory ahead, as I see it. (Friends may wonder whether they and I are going to be able to agree on the best route into the promised land—while foes may be curious as to which particular part of the subjectivist swamp will eventually claim me.)

Briefly then, I reject the error theory in part because I doubt the availability of the kind of notion of truth that seems required to make it coherent. The error theorist wants to say for example that "Smoking causes cancer" is false. In saying this she is presumably not supporting the tobacco lobby—i.e., not *denying* that smoking causes cancer, in the way that the cigarette manufacturers did for so long. She agrees that the statement that smoking causes cancer is correct, and its denial incorrect, by the ordinary internal standards of causal discourse. But she lays claim to a higher standpoint, and with it a more substantial notion of truth,

[17] Perhaps the surprising thing is that this conclusion does not already seem intuitively plausible. I think it is an indication of the extent to which twentieth century physics has muddied the waters concerning both causation and temporal asymmetry that contemporary physicalists don't find it natural to deny that such an asymmetric relation as causation is an intrinsic feature of the physical world.

[18] David Braddon-Mitchell develops a view of this kind; see his (1992).

[19] Particularly the debate referred to in footnote 13.

in terms of which "Smoking causes cancer" is indeed false (or so she asserts). I doubt that there is any such external standpoint to be had, and certainly my own preferred account of truth seems to me to leave little room for such a view.[20]

Turning to the (qualified) physicalist view that agency simply gives us access to a symmetric physical relation in the world—a relation that may be thought of as a posteriori identical to the symmetric core of the causal relation—I reject this view largely because the proposed "identification" seems to me to change the subject. Admittedly I feel related qualms about much less contentious a posteriori identity claims, such as that of water and H_2O. I also concede that in some cases—certainly water, possibly causation—changing the subject may well be a sensible move; existing usage is not sacrosanct. But on the other hand I would argue that the philosophical attractiveness of such a move is often overestimated: elsewhere (Price, 1992) I defend a form of metaphysical pluralism, from the standpoint of which the appeal of such identity claims is greatly diminished. Indeed, to the pluralist it appears that both reductionism and eliminativism typically rest on something akin to a category mistake. Once we appreciate the distinctive functions of causal discourse, in other words, we come to appreciate that its legitimacy does not depend on its ability to be recast in physicalist terms.

What are these distinctive functions? It is here that the agency theory makes its entrance, claiming that causal discourse reflects the agent's peculiarly deliberative perspective on the world. Projectivist accounts of other philosophical topics provide a familiar model at this point. For one thing this is an *explanatory* account of causal discourse, in the way that emotivism is an explanatory theory of evaluative discourse. Just as the emotivist explicates moral judgement not by analyzing its distinctive content but by telling us what it does (viz, that it expresses evaluative attitudes), so this projectivist version of the agency theory explicates causal judgement in terms of the deliberative psychology on which it rests.[21] In other words, this is not an account of what causation *is*, but an account of how we come to speak in causal terms.

Projectivism is normally a variety of non-factualism, however, and of course there is a good deal of intuitive resistance to the suggestion that causal claims are not genuinely factual claims. The version of projectivism I favour meets this challenge by rejecting the traditional assumption that there is a well-defined factual–non-factual boundary in language, arguing instead that to the extent that the distinction can be drawn at all, it is a matter of degree—and no part of language is purely *factual*, in the relevant sense.[22] This suggestion seems to me to respect

[20] The account concerned is developed in my (1988); see also my (1990). The error theorist might of course try to couch her theory as a denial of the *existence* of causation, rather than of the truth of causal claims. I think a Quinean account of ontological commitment counts against this variant; see my (1992).

[21] Perhaps it is Ramsey who really deserves the credit for this view. As he puts it: "From the situation when we are deliberating seems to ... arise the general difference of cause and effect" (1978, p. 146).

[22] See my (1988) and (1992). As I note there, the rejection of the view that there is any genuinely factual part of language is perhaps the major respect in which my approach differs from Blackburn's quasi-realist projectivism (for which see particularly his 1984).

the intuition that causal talk is not significantly less factual than other parts of scientific discourse, while retaining the projectivist's commitment to the principle that the best way to understand causation is to ask not what causation *is*, but what talk of causation *does*—what function it serves in our lives.

The ability to take on board the insights of perspectivalism without endorsing non-factualism or an error theory has been one of the advantages claimed for recent generalisations of the Lockean dispositional treatment of secondary qualities.[23] While I welcome the perspectivalism of these accounts—particularly that of Pettit (1991), who takes it to have global application—I feel that they develop the perspectivalist's pragmatism in the wrong direction. Like the physicalist, they concentrate too much on *analyzing* causation (or whatever) and too little on *explaining* our *talk* of causation.[24] Again, it is the projectivist who puts the emphasis where it should be.

These brief remarks will win no immediate converts for global projectivism, of course. Scepticism about the enterprise would be entirely healthy at this point. All the same, I hope that they serve to indicate that to accept the main thesis of the paper—that is, that only a perspectival theory can adequately account for the asymmetry of causation—is not to cast oneself into an uncharted philosophical wilderness. There are several paths on offer (one of them in my view a good deal more promising than the others). Moreover, the recognition that causal asymmetry may well be perspectival seems to provide even the sceptic with a motive to explore the territory further.

Traditional and Modern Philosophy
University of Sydney
NSW 2006
Australia

HUW PRICE

References

Blackburn, S. 1984: *Spreading the Word*. Oxford: Oxford University Press.
Braddon-Mitchell, D. 1992: "Could Causation be Symmetrical?". Forthcoming.
Dummett, M. 1964: "Bringing About the Past". *Philosophical Review*, 73, pp. 338-59.
Haldane, J. and Wright, C. eds. 1992: *Reality, Representation and Projection*. Oxford: Oxford University Press.
Hausman, D. 1984: "Causal Priority". *Nous*, 18, pp. 261-79.
——1986: "Causation and Experimentation". *American Philosophical Quarterly*, 23, pp. 143-54.
Horwich, Paul 1987: *Asymmetries in Time*. Cambridge, Mass.: MIT Press.
Johnston, Mark 1989: "Dispositional Theories of Value". *Proceedings of the Aristotelian Society*, Supplementary Volume 63, pp. 139-74.
——1992: "Objectivity Refigured", in Haldane and Wright (1992).

[23] Particularly by Pettit (1991). See also Wright (1992) and Johnston (1989, 1992).

[24] I develop this and related objections to the dispositional strategy in my (1991d).

Lewis, David 1986: "Counterfactual Dependence and Time's Arrow", in his *Philosophical Papers*, Vol. II, Oxford: Oxford University Press, pp. 32-66. Originally published in 1979 in *Nous*, 13, pp. 455-76.

Menzies, P. and Price, H. 1992: "Causation as a Secondary Quality". *British Journal for the Philosophy of Science*, forthcoming.

Owens, David 1992: *Causes and Coincidences*. Cambridge: Cambridge University Press.

Papineau, David 1985: "Causal Asymmetry". *British Journal for the Philosophy of Science*, 36, pp. 273-89.

Pettit, Philip 1991: "Realism and Response-Dependence". *Mind*, 100, 3, pp. 587-626.

Popper, K. 1956: "The Arrow of Time". *Nature*, 177, p. 538.

Price, Huw 1984: "The Philosophy and Physics of Affecting the Past". *Synthese*. 16, pp. 299-323.

——1988: *Facts and the Function of Truth*. Oxford: Basil Blackwell.

——1990: "Why 'Not'?" *Mind*, 99, 2, pp. 221-38.

——1991a: Article review of K. G. & J. S. Denbigh, *Entropy in Relation to Incomplete Knowledge* (Cambridge University Press, 1985) and Zeh, H. D., *The Physical Basis of the Direction of Time* (Springer-Verlag, Berlin, 1989). *British Journal for the Philosophy of Science*, 42, pp. 111-144.

——1991b: "Agency and Probabilistic Causality". *British Journal for the Philosophy of Science*, 42, pp. 157-76.

——1991c: "The Asymmetry of Radiation: Reinterpreting the Wheeler-Feynman Argument". *Foundations of Physics*, 21, 8, pp. 959-75.

——1991d: "Two Paths to Pragmatism", in Menzies, P., ed., *Response-Dependent Concepts* (Working Papers in Philosophy, No. 1), Canberra: Philosophy Program, Research School of Social Sciences, ANU, pp. 46-82.

——1992: "Metaphysical Pluralism". *Journal of Philosophy*, forthcoming.

Ramsey, F.P. 1978: "General Propositions and Causality", in Mellor, D. H. ed., *Foundations: Essays in Philosophy, Logic, Mathematics and Economics*. London: Routledge & Kegan Paul, pp. 133-151.

Tooley, Michael 1987: *Causation: A Realist Approach*. Oxford: Oxford University Press.

Wright, Crispin 1992: "Realism—the Contemporary Debate: Whither Now?", in Haldane and Wright (1992).

Time and the Anthropic Principle

JOHN LESLIE

1. Brandon Carter's Doomsday Argument

Prima facie, we should prefer theories which make our observations fairly much to be expected, rather than highly extraordinary. Waking up in the night, you form two theories. Each has a half chance of being right, in your estimation. One, that you left the back door open, gives the chances as one in ten that the neighbour's cat is in your bedroom. The other, that you shut it, puts those chances at one in ten thousand. You switch on the light and see the cat. You should now much prefer the first theory.

Consider next your observed position in time. If the human race is going to last for at least a few thousand more centuries at its current size, let alone at the much larger size which it would attain if it spread through its galaxy, then you are a very exceptionally early human, perhaps among the earliest 0.01%. But if the race is due to end shortly—which, when one thinks of the ozone layer, H-bombs, etc., can seem not particularly unlikely—then you are fairly unexceptional: you live in a period which a human observer could quite have expected to experience. Thanks to recent population growth, roughly 10% of all humans who have so far been born are still alive today. Now, shouldn't this influence you? May not the rather unexceptional position which you would occupy in human population history, if that history were soon to end, give you some grounds, grounds reinforcing those got through considering the ozone layer and H-bombs, for thinking that it will indeed end fairly shortly?

This paper will suggest that it does give you such grounds. If the world is deterministic then the grounds can be very disturbingly strong. If it is radically indeterministic, and if the indeterminism is of a sort likely to affect how long the human race will last, then they may be considerably weaker, but still worrying.

Many find it paradoxical that one could learn anything in this way from one's observed temporal placement. However, the reasoning I have just sketched can seem natural to people familiar with "anthropic" reasoning in cosmology. In fact, it was first sketched by Brandon Carter, the applied mathematician who invented the words "anthropic principle".

Mind, Vol. 101 . 403 . July 1992