J. ISMAEL

CLOSED CAUSAL LOOPS AND THE BILKING ARGUMENT

ABSTRACT. The most potentially powerful objection to the possibility of time travel stems from the fact that it can, under the right conditions, give rise to closed causal loops, and closed causal loops can be turned into self-defeating causal chains; folks killing their infant selves, setting out to destroy the world before they were born, and the like. It used to be thought that such chains present paradoxes; the received wisdom nowadays is that they give rise to physical anomalies in the form of inexplicably correlated events. I argue against the received wisdom. I can find nothing in them that argues against the possibility (even, the probability) of time travel.

The most popular examples of self-defeating causal loops are cases involving time travelers going back to kill their infant selves, or the grandfathers whose industrial success made possible the construction of their time machines, but simpler cases can be found that involve nobody *doing* anything, just inanimate systems behaving in familiar ways in accordance with known laws. Such cases are cleaner because they don't introduce intuitions about freedom and agency. The following set-up – due to John Earman, and which I'll refer to heretofore as an 'Earman rocket' – is an especially elegant example:

> Consider a rocket ship which at some space-time point x can fire a probe which will travel into the past lobe of the null cone at x. Suppose that the rocket is programmed to fire the probe unless a safety switch is on and that the safety switch is turned on if and only if the 'return' of the probe is detected by a sensing device with which the rocket is equipped. (Earman, p. 231)

If the rocket fires, then the probe travels into the past, is detected by the sensing device, the safety switch is activated, and it doesn't fire. If it doesn't fire, on the other hand, no probe is detected, the safety switch remains off and the rocket fires as programmed. So, the probe is fired, it appears, if and only if it is *not* fired.[1]

It used to be thought that we could stop there, i.e., that the mere supposition of the existence of Earman rockets was by itself a conceptual absurdity and we could rule out any physical hypothesis that allowed for their construction on purely logical grounds. Nowadays it is acknowledged that that isn't so; the argument contains a loophole. The resolution of the

Synthese **136:** 305–320, 2003.

apparent paradox given by Feynman and Wheeler, and defended especially eloquently by David Lewis" is widely accepted. But even among those who exploit it to defend the *possibility* of physical hypotheses that give rise to Earman-rocket-like systems, it is also widely held to be closeable at a relatively low cost in strength. For it is held that Earman rockets, even if they are not logical absurdities, present physical anomalies – a violation of the laws, an improbable conjunction of events, or some such thing – and we can assign the assumptions needed to escape contradiction, i.e., the *facts* about their behavior that follow from the supposition of their existence, arbitrarily low probability. Huw Price, for example, writes:

> let me record my view – similar to that of Horwich (1987, chap. 7) that the bilking argument . . . shows us that the hypothesis of time travel can be made to imply propositions of arbitrarily low probability. (Price, p. 278)

and Horwich states:

> time travel into the local past engenders closed causal chains that verge on being self-defeating, and this possibility can be precluded only by means of appropriate coincidences. (Horwich, p. 125)

From the point of view of someone interested in whether closed causal loops are a live physical possibility the weakened argument, if it is correct, is as almost good as the original; logical possibility is cold comfort, where probability is arbitrarily low.

The argument, in both versions, has the form of a *reductio*: we assume the conjunction of laws and spatiotemporal structures, and initial conditions that would permit the implementation of a self-defeating causal loop, derive a contradiction or anomaly, and then let the impossibility or improbability of the conclusion propagate back to impugn whatever part of the assumption seems most vulnerable.[2] The modalities involved can make the structure of the argument extremely confusing, but even once modal confusions are cleared up, the literature is extremely unsatisfying. For all the talk of strange coincidences and mysteriously coordinated behavior, I cannot find a discussion that plainly names any genuinely inexplicable correlations. I'll go over some preliminary attempts, say why they fail, modify Earman's example in a way designed to bring out troublesome correlations, but argue ultimately that they don't present genuine anomalies, nothing that can be used to argue against hypotheses that would allow for the construction of closed causal loops.

Lest it be thought that the bilking argument only has interest as an objection to a class of esoteric physical hypotheses, closed causal loops are less star-trekky than one would think. Godel's solution to the equations of General Relativity forced physicists take them seriously as nomological possibilities, tachyons have been explored as potential explanations for quantum non-locality, and it is not as clear as had once been thought that superluminal signaling or influence of certain kinds is impossible in Minkowski space-time. More importantly, however, the argument raises ancient questions about freedom, foreknowledge, and causal determination in a new and especially telling way, one that is free of some of the encumberances that complicate traditional presentations.[3] There is much to be learned about the relations between those concepts from reflection on these cases.

1. EARMAN ROCKETS

The reason there is no contradiction in supposing there are Earman rockets is that the firing of such a rocket doesn't follow *logically* from its non-firing, nor its non-firing from its firing; they follow causally from one another and causal followings are weaker than logical ones. What is true is that, holding fixed the laws that govern the process, if *everything works as it should*, the rocket fires *iff* it doesn't fire. And we can deduce therefrom that – again, holding fixed the laws – everything never works as it should. We can deduce from the description of the rocket and the laws which are supposed to govern it that something goes wrong, somewhere along the way, every time. Some contingency arises to spoil things, some bug in the program keeps the rocket from firing or some malfunction interrupts the process after the firing, before it is inhibited by the safety switch. If I get a grant from the NSF, set up 70 Earman rockets, and hire lab technicians to sit and press the buttons that activate their programs once a minute, 16 hours a day, for 17 long years, the result will certainly be that each time any of the buttons is pressed, some kind of system failure or malfunction – a different one, perhaps, every time – occurs.

That is the only consistent way of filling in the story.[4] If there are closed causal loops, and it is possible to set up Earman rockets, attempted launches will always fail, and the explanation of particular failures will look the same as explanations of cases in which causal chains intrinsically like them laid along an open time-like curve (i.e., cases which are pointwise similar, but differ topologically, cases in which a rocket is programmed to fire a probe which is detected by a sensing device which activates a safety switch which is wired up to prohibit the launch of a rocket, but in which the rocket whose launch is prohibited by the activation of the switch is not *the very same one* whose launch kicked off the whole chain of events) fail to come out right: *viz.*, the program fails, the probe goes off in the wrong direction, an electrical malfunction keeps the safety switch from inhibiting the firing . . . or *some* such thing. Considered individually, there will be nothing anomalous in the explanations; they will have precisely the same form and status as explanation of failures in topologically ordinary cases.

It is almost irresistible to suppose, however, that there is something anomalous in the cases considered *collectively*, i.e., in our *unfailing* lack of success in Earman rocket launchings. Horwich writes, for example:

> we know, on the basis of purely semantic reasons, that attempts at [implementing self-defeating causal chains] will invariably fail. But we recognize that there is considerable strangeness in this – something *ad hoc* and unsatisfying about explaining the repeated failures in terms of [individual and unrelated malfunctions] . . . it is implausible that such mishaps would occur so faithfully over and over again. (Horwich, p. 121)

But that is not right either; there is nothing *ad hoc* or mysteriously coincidental in the scenario, as described, for we never succeed in ordinary cases (cases of non-self-defeating rocket launchings, cases consisting of intrinsically similar sequences of events laid along open time-like curves) in which we fail, either. And it is built into the description of the class of cases that we are considering, when we consider Earman rocket launchings (and self-defeating causal chains, quite generally), *that* they are failures, in the same way it is built into the description of the class of cases in which I don't manage to get ahold of my mom on the telephone *that* they are unsuccessful attempts at reaching her. For, when we describe a self-defeating causal chain, we sneak in, under the guise of the first event, a description of the *last* which is incompatible with the success of the operation as a whole. Earman rocket firings are successful just in case they result in non-firing, so by considering just those chains in which either the rocket is not fired (which is to say, those in which something went wrong with the program) or those which *begin* (which is to say, end) with firing (i.e., those in which something malfunctioned between the firing and the activation of the safety switch), we consider only unsuccessful cases. And just as in the case of my failed attempts at mom- calling, there are diverse and unrelated explanations of the individual failures, but nothing spooky or coincidental about the fact that *all* founder. So far, we have seen nothing paradoxical or even strange about the behavior of Earman rockets, nothing to suggest that their existence would create any sort of physical anomaly.

2. MODIFIED EARMAN ROCKETS

There is, moreover, so far as I can see, no way to locate a difficulty without adding something to the description of the case. So let's do that; in particular, let's equip the Earman rockets with 'hazard' buttons that manually override their safety switches so that they can't function to stop the launch in the ordinary way, and let's set the lab technicians launching rockets, depressing hazard buttons or leaving them up, as they like. Sometimes (very often, we can suppose, if we are good engineers) when the hazard switch for a rocket is depressed, it will function as it should; the rocket will fire, the probe will be detected by the sensing device, and the safety switch will be activated before the hazard switch overrides it. As before, however, when the hazard button is *not* depressed, things will, without fail, go awry. We can run the rockets for a year with their hazard buttons depressed, with scarce a malfunction; let the button up, and there will without doubt, not be a *single* success no matter how long we run the thing. If there is an anomaly, it consists in the fact that a press of the hazard button – a button which should not, one would think, affect the working of the machine until the moment immediately before the signal reaches the launch – turns a smoothly functioning apparatus into one bedeviled by scattered and unrelated malfunctions that frustrate any attempt to make it work. No improvement in the quality of our materials, the ingenuity of our engineers, or the training of our technicians will make the tiniest difference to our success rate.

Nor does it matter *when* the hazard button is elevated; we can keep it down while the rockets are launched and have our technicians decide whether to leave it or release it sometime down the road. Things will often go as planned right up until the moment of release, but we can be quite sure, if the hazard button is working properly, no matter how close it is to launch time, once it is let up, some event will come 'out of nowhere', so to speak – an event with a causal history of its own, completely independent of the technician's decision – to prevent the launch. We can be sure of it. The apparent coincidences, which I said weren't to be found in the original scenario, are present here and easily identified: technological malfunction is evidently correlated with, but not caused by, the position of the hazard button.[5]

3. COINCIDENCES

Here, more explicitly, is why there were no coincidences in the original scenario, and hence why the modification is needed. Given a set of particulars, and a pair of properties, X and Y, pertaining to them,[6] Y is <u>correlated</u> with X in case there is a higher incidence of Y's among the X's than among the non-X's. A correlation between X and Y is <u>coincidental</u> if they are neither analytically tied to one another (e.g., *being-a-bachelor* and *being-male*) nor *causally* linked (e.g., *having-cancer* and *being-*

a-smoker).[7] In the scenario with the unmodified Earman rocket, there are no coincidental correllations, because the correlated properties (*viz., being-an-attempt-to-implement-a-self-defeating-causal-chain* and *being-thwarted-by-system-malfunction*), are not analytically independent. Being self-defeating is not analytically independent of being thwarted because saying that a launch is thwarted is just a way of saying that it is defeated, and defeating *oneself* is just a special way of getting defeated. If we call the chain of events that our Earman rockets were designed to implement, intrinsically characterized, an 'Earman chain',[8] what is needed for coincidental correlation is a condition that is *analytically independent* of success and that distinguishes the sometimes successful from the invariably *un*successful Earman chains. The job of the hazard switch is to divide attempts to implement Earman chains into classes (those that occur when it is depressed and those that occur when it is up), any difference in incidence of failure between which (so long as the hazard switch is causally isolated from the rest of the set-up, i.e., so long as the position of the hazard switch can be treated as a free variable) must be a matter of chance.[9] *Being-an-attempt-to-implement-an-Earman-chain-while-the-hazard-button-is-up*, unlike *being-an-attempt-to-implement-a-self-defeating-causal-chain*, is analytically independent of success, and any uncaused correllation between it and system failure will have to be, it would seem, chalked up to coincidence.

In Earman's original scenario, the safety switch was there only to provide the last link which turns an ordinary Earman chain into a self-defeating loop; it is activated *if and only* if the probe is detected by the sensing device, which is to say, *if and only if* the rocket launches and there are no malfunctions. In the modified set-up, the switch gets activated if, but not *only* if, the probe is detected by the sensing device, allowing it to be manually overidden and putting the button that effects the manual override in a separate room to emphasize its causal isolation from the rest of the set-up. It could just as well have been put in a separate country or on the Earman rocket itself; the important thing is that without a comparison between what happens to the rocket when the button is up and what happens when it is down (or – what amounts to the same thing, since an Earman rocket without a safety switch is functionally the same as an Earman rocket with a safety switch and a depressed hazard button – a comparison between what happens to an Earman rocket with a safety switch to one without),[10] there are no unexplained correlations. Without a comparison between what happens when some free variable takes two different values, the set-up presents no identifiable anomaly.

But do the correlations in the new set-up constitute a real anomaly? Certainly, they can be made to look strange; we can arrange a thousand Earman rockets side by side, each manned by a lab assistant who launches his rocket with the hazard button up or down at his whim, and do run after run of the experiment. What we observe – and it is important here that I am not reporting an *experimental* result; things *couldn't* go otherwise – is that, although things almost always proceed without a hitch when the button is down, in every case in which it is up, some foiling event – an event with a perfectly determinate and independent causal history, an event which was *bound* to happen and (if our rockets were well-designed) bound to occur for reasons that had as little to do with the position of the hazard button as the price of tea in China – occurs. When the hazard button is up, the loop is closed and there are malfunctions; when the hazard button is down, the loop is open, severed at the last instant by the diversion of the signal meant to inhibit the launch, and everything often goes as it should. It doesn't matter *when* the button is released (i.e., where the releasing event lies along the closed curve that includes the launch); all that matters its position at (the moment immediately preceding) take-off. If it is up and the rocket leaves the pad, there is, with certainty, an impending mishap.[11] Perhaps the correlation will be perfect, i.e., perhaps we will *always* succeed when the button is up, *never* when it is down. More likely, the correlation will be imperfect; sometimes hazard buttons malfunction, sometimes Earman chains on open curves fail, and we will have only a correllation. The difference isn't important; so long as Earman chains on open curves are *somewhat* reliably successful, and so long as our hazard button functions *reasonably* well, suspicion of either a miraculous coincidence or some hidden causal connection linking the position of the hazard button to the motley malfunctions will seem inescapable.

Notice that the recipe for generating the correlations is quite general; all we need suppose is that there are closed, time-like curves, that the laws are local and undiscriminating,[12] and that there exists pair of events, A and B, joined by a reliable process (i.e., an easy to implement, often successfully realized chain of causes, beginning with A and ending with B). For, if these hold, we can take the A-to-B process and lay it along a closed curve, tying its ends together in a self defeating way by passing the chain connecting B to A through one that defeats A,[13] and add something that can play the role of a hazard button to sever the B-to-A connection, making sure that its position is causally isolated from the rest of the set-up. The incidence of foiling events will in that case, with certainty, be correlated with the position of the hazard button. Again, the strategy is to find a way of identifying the class of closed chains by a condition C (the ‘position’ of a ‘hazard button’) that is analytically independent of their closure, giving us a correlation between C and failure. Closure isn’t itself analytically independent of failure, but so long as C is analytically independent of closure, the correlation between C and failure will appear coincidental; it will relate distinct events with independent causal histories.

C $\rightarrow$ closure, and
closure $\rightarrow$ failure, wherefrom
C $\rightarrow$ failure

So there you sit, (supposing you have taken over control of one of the rockets), finger on the hazard button, experiencing almost invariable success when you leave it down, foiling events coming apparently out of nowhere in each case that it is released. You know with certainty, before you even know *whether* you are going to release the button, that *if* you do, a quantum fluctuation, a whim of the gods, the product of diffuse conditions present at the Big Bang is bound to occur.[14] *Some* mishap is bound to occur, you know that it is, and the basis of your knowledge isn’t knowledge of conditions that necessitate any particular one causally.

4. ANALYSIS

Let’s figure out what’s going on here. There are two sorts of causal hypotheses that might be thought to explain the correlations: hypotheses that recognize hidden causal connections between the position of the hazard button and foiling events, or hypotheses that treat the global topology of the chains on which they occur as itself – in a way that no local, undiscriminating laws can take into account – *causally* responsible for the failures. Neither, unfortunately, is any help here for a reason that is simplicity itself: the correlations were deduced from a description of which such hypotheses were no part. If it is *contradictory*, rather than merely false, to suppose that Earman chains initiated by launchings that take place while the hazard button is elevated do not contain foiling events, the occurrence of a foiling event has to be contained in our description of the case *on the supposition only of such causal relations as are stipulated in that description*. There are no two ways about it. If we did *not* sneak information about their outcomes into our description of Earman launches that occur while the hazard button is elevated, we could not be, as we in fact are, certain that there would be anything but ordinary statistics of failure in cases in which the button is elevated; we could not know, without doing any experiments, just *what* those statistics would be.

What we have here is a sort of puzzle; on the one hand, it is clear that, since the correlations are deduced from our description of the situation, no details – no matters of fact, or relations between matters of fact – that weren't included, implicitly or explicitly, in that description can play any role in the explanation. On the other hand, all the usual suspects – the sorts of connections that are ordinarily the sources of correlation – seem to have been explicitly *ex*cluded.[15]

The nature of the puzzle points us in the direction of the solution. We have to see how we built into our description of Earman launches that happen while the hazard button is down, that they are, if successful, self-defeating (hence unsuccessful, and hence foiled by impending events). It is this first step (if successful, self-defeating) that is substantive, the others (hence unsuccessful, hence foiled ...) just draw out the implications of self-defeatingness. So how *do* we know that Earman rocket launches that take place while the hazard button is depressed are, if successful, self-defeating? We made only two stipulations about causal relations when we described the situation; we stipulated that the causal histories of foiling events are (typically) independent both of the position of the hazard button and of one another, and we stipulated that depressing the hazard button breaks the causal connection between the 'first' and 'last' events in the Earman chain. The reason that the position of the hazard button is correlated with failure is that the existence of a foiling event is a causally necessary condition of launch, given closure, and an elevated hazard button *indicates* (because it causes) closure. The position of the hazard button, that is to say, indicates *that* the launch signals the impending occurrence of a foiling event; it signals that the causal relations in virtue of which the launch couldn't have occurred but for the occurrence of a foiling event, actually obtain.

The situation is causally structured like ordinary cases of *retro*diction, cases in which we infer from the occurrence of an event, that the causally necessary conditions for its occurrence obtain. The only real novelty is in the temporal relations between cause and effect. The same is true of other cases of self-defeating causal chains; what characterizes them as a class, and what lends the appearance of anomaly, is the inevitability of failure even while there is the *possibility* of success.[16] In my bid to kill my infant self, for instance, there is the same certainty of failure, though failure be neither logically nor causally necessitated by the facts that obtain at the moment that I stand over my crib, gun in hand, finger on the trigger. The inevitability is a purely epistemic one, perfectly compatible with the logical and nomological possibility of success. And it arises for the simple reason that in the context of closed curves, some of the causes of an event lie in its future (its *absolute* future, its forward light cone).

5. CONCLUSION

Suppose all of this is correct; does anything remain of the bilking argument? If there are closed causal loops, it would be possible to set up systems like Earman rockets designed to implement self-defeating causal chains, and then . . . what? Contradictions would be true? No. The physical laws would be violated? No. There would be mysterious coincidences? No. There would be correlations that were not explicable in a perfectly ordinary way, i.e., explicable in a way in which we are accustomed to explaining correlations even outside the context of closed loops? Not so far as I can see. The *only* physical peculiarity lies in the temporal relations between correlated events that are related as cause to effect, relations that make for a novel combination of epistemic and metaphysical modalities, i.e., that provide, as I have said, for the *certainty* of failure while there is still the *possibility* of success. In the context of closed causal loops, we are able to predict the future in the absence of the types of necessitation that are ordinarily conditions of its possibility, and that is interesting, but not surprising, and not something that argues against the existence of closed loops. Such loops introduce no paradoxes or anomalies, just abnormalities in the way information is spread around.

It may be that we cannot tolerate *too* much novelty on a global scale in how information is spread around without losing our grip on our causal concepts; probably, that is to say, the conditions of applicability of our causal concepts include non-trivial constraints on the availability of information (e.g., if foreknowledge of the right sort were ubiquitous enough, it may be that there would be nothing left of the distinction between cause and effect). If there are such constraints, however, revealing them is going to be harder work than simply pointing out that closed causal loops would let us construct self-defeating systems; it is going to have to take the form of a careful examination (of just the sort that Horwich, Lewis, and more recently, Price have begun) of the temporal asymmetries inherent in our causal and epistemic concepts. They do not appear to be violated by the local abnormalities involved in the operation of simple self-defeating scenarios like Earman rockets, for we can give apparently consistent descriptions of those scenarios that, as strange as they seem, involve no evident violations of law, no inexplicable coincidences or improbable events.

6. ADDENDA

Two quick addenda; one to address what seems to me the real intuition behind the bilking argument, and one to look a little more closely at Horwich's discussion.

7. FREEDOM AND THE BLOCK UNIVERSE

What the bilking argument really brings out is the old familiar clash between our ideas of personal freedom and a conception of our willful actions as natural events. The dissatisfaction one feels (and I feel it too) with the reconstruction given above of "I can, but won't, bilk", is really dissatisfaction with the reconstruction of "X is possible" applied to willful actions treated as natural events in an atemporal framework. The real force of the bilking argument derives from the intuition that, whatever can be said about the rest of the world, I am free; *I can bilk*. It is the intuition that there is more to saying that I can *do* something than simply that its occurrence is causally compatible with water under the bridge.[17] It is the intuition that there is more to saying that, as I stand there face to face with my younger self, it is *in my power* to give myself a kiss on the forehead or a punch in the nose, to tell myself the winning lottery numbers or the secret that will win me the Nobel Prize.[18]

If there is a difficulty with reconstructing our notions of personal freedom in an atemporally conceived universe, i.e., an incompatibility between 'I *can* X', as it is understood within that framework, and I *won't* X', it is masked by our typical ignorance of the latter. It is masked by the fact that, because of contingent facts about the way information is spread, and because of the way we keep the modalities evolving in concert with knowledge,[19] we typically don't find ourselves in a position to simultaneously *assert* both.[20] Any such incompatibility will come to the fore in situations involving foreknowledge, and that is why we confront it when we consider closed causal loops; such loops provide channels for the propagation of information backwards in time.

Aristotle certainly thought that there was an incompatibility; his Sea-Battle argument presents the dilemma: accept fatalism or reject the atemporal conception of time. But – and this is the important point for our purposes – if there is one, it has nothing specifically to do with the existence of closed loops. The difficulty of doing justice to the notion that we can bilk is just a special case of the difficulty of doing justice to the notion that we can *do* anything in an atemporal universe; it is the difficulty of making out a sense of possibility, in that context, that is adequate to our notions of personal freedom. Mere causal compatibility with water under the bridge, which is all that is needed for physical purposes, doesn't seem to capture it.

8. HORWICH

Horwich's discussion is especially confusing; he asserts again and again, in passages like the following, that the repeated failure of bilking attempts involves strange coincidences, unexplained corellations, unlikely conjunctions of events.

> misgivings arise, I think, when we focus our attention on cases of attempted autofanticide – cases where someone *tries* to realize a self-defeating causal chain. We know of course that success is impossible. Yet *repeated failure is nevertheless surprising and disturbing, for it constitutes a violation of well-established regularities*. (Horwich, p. 119–120)

then later,

> Now we know, on the basis of purely semantic reasons, that attempts at [killing one's infant self or doing something one remembers was not done] will invariably fail. But we recognize that *there is considerable strangeness in this – something ad hoc and unsatisfying about explaining the repeated failures in terms of changes of mind, guns, misfiring, and so forth.* (Horwich, p. 120–121)

and again,

> . . . bilking attempts are usually attempts to perform actions that are ordinarily quite easy to perform. Failure is indeed possible for a variety of reasons; but *an indefinitely long string of failures, correlated with circumstances leading up to time travel, constitutes a very striking coincidence*. (Horwich, p. 122)

And the alleged improbability of repeated failure is used as the central premise in the argument that dominates the rest of the discussion.

You won't find, however, many explicit characterizations of the events in whose correlation coincidence is supposed to consist. I could find just one, and it occurs only after the main discussion when Horwich is considering some anticipated objections. "If there were a regular practice of travel into the past", he writes,

> then there would have to be a correlation between, for example, *the time traveler having an intention to bump off the child who lives at his old address* and the existence of *circumstances that will frustrate this intention*. And we know enough about human motivation (specifically, about the factors that might produce this bilking inclination) and the kinds of phenomena that could cause this plan to fail (amnesia, gun-jamming, brilliant surgeons, etc.) to claim that any such correlation would be an improbable coincidence. (Horwich, p. 125)

But this is wrong; the characterization of the relevant events is ambiguous, and neither way of resolving the ambiguity yields a coincidence. Notice, first, that since there can only be one successful case of killing per world, we have to talk about, and compare the incidence of success in, attempted killings across possible worlds (no such comparison is necessary in the case of the Earman rocket, since there can be both multiple successful and failed launches of Earman rockets per world, though the successful ones, of course, have to be laid along open time-like curves). If we suppose that Horwich's time traveler is actual and grew up at, say, 65 Brown Rd., we can interpret him as *either* asserting that there is a correlation between a time traveler in a given world having an intention to bump off a child living at 65 Brown Rd. and the existence of circumstances that are bound to frustrate that intention *or* asserting that there is a correlation between a time traveler in a given world having an intention to bump off the child who lives at the place *he* grew up and the existence of circumstances that frustration *that* intention.

The ambiguity is a familiar one; the expression "the child that lives at his old address" can be read either as a rigidified or as an unrigidified description of the object of the time traveler's homicidal intentions. If the former, it refers to the child living at the counterpart of the address of the place that our actual time traveler grew up, and there is no correlation between the having of the relevant homicidal intentions and the existence of conditions bound to frustrate them, for the child at 65 Brown Rd. often perishes at the time traveler's hands. If the latter, it refers to the child living at the place that the homicidal counterpart of our time traveler at that world grew up – which is to say, to the child that *he* once was – and in that case there is correlation, but no coincidence. The reason Horwich doesn't put his finger on an anomaly, if this is correct, is that he fails to make the crucial distinction between unremarkable constant failure of attempts to implement self-defeating causal chains and remarkable positive correlation between failure and *independent* conditions that distinguish the sometimes successful from the invariably unsuccessful cases.

ACKNOWLEDGEMENTS

I am gratefully indebted to Richard Healey, to an audience at Arizona State University, and also to an anonymous reviewer for very helpful comments.

NOTES

[1] So, the argument is that:

(i) if [ ], there would be closed causal loops, and
(ii) if there were closed causal loops, (granted clever technicians, availability of materials, and so on), there would be Earman rockets.

The modalities can be confusing and are seldom explicit. The argument will rule out only the *conjunction* of laws, spatiotemporal structures, and contingent conditions that permit the construction of Earman rockets, and it will attach to the conjunction whatever brand of unaccceptability attaches to the supposition of the existence of Earman rockets. If Earman rockets are logically impossible, the conjunction will be logically impossible; if they are nomologically impossible, it will involve a violation of the laws, and if they are merely improbable, it will have only a low probability.

One famous example of a modal confusion is a view that has been attributed (fairly or unfairly) to Godel, that, on the one hand, the existence of Earman rockets is a conceptual absurdity, and, on the other, it is only the failure of contingent conditions (e.g., availability of materials) that keeps us from constructing them. The error is evident; if the conditions in question are genuinely contingent, there are both logically and nomologically possible worlds in which they obtain, and in some of those there are Earman rockets. But if Earman rockets are a conceptual absurdity, there are no such worlds. Either the conditions in question are not, after all, contingent, or Earman rockets are not impossible devices.

[2] The idea here is that if we derive a physical impossibility, we show that physical laws are violated in the assumption that there exist closed causal loops; if we derive a physical implausibility, we show that the supposition that there are closed causal loops involves the assumption of a physical fact of low probability.

[3] There are no gods or persons, or other types of potentially *unnatural* objects, just physical systems, behaving in familiar ways, in accord with known laws.

[4] Without, that is, resorting to hypotheses involving branching universes and the like.

[5] It might seem that all of this misses the point because the difference between the two cases is modal: the rocket set-up placed on an open time-like curve *can*, and on a closed time-like curve *can't*, succeed. The reply (Lewis', and also Horwich's, and again, I think it is exactly right) is to deny that there is a modal difference between the two cases. To say that a launch *can* succeed is to say that there is nothing that has already occurred, to keep it from doing so; it is to say that it is compatible with the water that has already passed, so to speak, under the bridge, when the possibility is assessed. This is true of both scenarios, provided we have a way of characterizing 'water under the bridge' that excludes the problematic events. This will not always be possible; in Godel universes, for instance, in which there is a closed time-like curve through every point, it clearly will not, and where it is not, the right reaction may be to simply deny that our causal concepts have any applicability; see Section 5.

The response is not *ad hoc*; it is a straightforward application of the view of modality that goes with the conception of time in the background: the 'atemporal', or 'block universe', conception embodied in our classical physical theories and assumed by Lewis, Horwich, and Earman. The only thing it can mean, in that context, to say that some future event which does not occur nevertheless *can*, is that nothing *yet* rules it out. Uneasiness about the response, about the claim that an Earman rocket launching *can* succeed (or, for instance, that a time traveler *can* kill his infant self) – skepticism, that is, about whether we can have got the meaning of 'can' right, if either of these things come out true – is really uneasiness about the conception of possibility that goes together with the atemporal conception of time, and should be directed at it. See Addenda.

[6] I use 'properties' and 'types of event' interchangeably.

[7] The causal links needn't, of course, be direct; they may involve common causes, or a chain – as attenuated as you please – of intervening causes. I appeal freely to causal concepts, remaining relatively neutral about the *nature* of causation; any analysis that combines nomological determination with additional conditions (e.g., counterfactual dependence, temporal precedence), i.e., anything that fits the schema 'A caused B just in case A nomologically determines B and c', will do.

[8] All we are given is an intrinsic description are the types of, and local temporal relations between, events in the chain (i.e., with details filled in: 'rocket of such and such a type sends probe on such and such trajectory towards sensing device of such and such construction . . . '); the global topology of the chain - in particular, whether it is closed or open – isn't specified.

[9] To be perfectly explicit: let **X** be *being-an-attempt-to-implement-an-Earman-chain-while-the-hazard-button-is-up*, let **Y** be *being-thwarted-from-successful-completion-by-a-system-malfunction*, and if we have any luck in keeping malfunctions from occurring when the link between the return of the probe and the launch is broken by the hazard button, there will be a positive correllation between **X** and **Y**. How strong the correlation is will depend on how much luck we have preventing malfunction in the non-self-defeating cases, and the stronger it is, the more improbable it will be.

[10] The advantage of having the hazard switch is that instead of comparing what happens to a pair of rockets which are supposed to be similar save that one has and one lacks a safety switch, we can compare what happens to one and the same rocket when its switch is in two different positions.

[11] To be more explicit about how terms are being used: taking for granted the notion of the local direction of a time-like curve, we say A is in B's past just in case $B \neq A$ and there is a forward looking time-like path from A to B, and B is in A's future just in case A is in B's past. It follows that every event on a closed time-like curve is in the future *and* the past of every other.

[12] Which is to say, they don't discriminate between duplicate events.

[13] Parity, to the extent that it holds, allows us derive the existence of a possible B-to-A process from that of an A-to-B one, and it needn't be added as a separate premise.

[14] So long, that is, as we count malfunctions of the hazard button as foiling events.

[15] I am idealizing in supposing each of the foiling events to have a fully determinate causal history, independent of the position of the hazard button and independent of the causal histories of the others. In real situations, the causal histories of events that occur so closely in space and time are bound to be tied up with one another, and to that extent, these independences are unrealistic. That we can derive the correlations without assuming them, however, means that we cannot cite such dependence as there may *actually* be in explanation of the derived correlations. Again, no details not included in the description from which the correlations were derived can play a role in their explanation.

[16] Both of these notions, possibility and certainty, are time-relative. Possibilities, at least physical possibilities, change with time in the sense that an event that was possible at one time (or, if you prefer, possible relative to the facts that comprise the world's up until that time, or the propositions that describe those facts; these are different ways of dividing the work, on any of them, the possibilities change with something that changes with time) may not be possible at another. It may no longer be, though it once was, possible to save my

marriage, for instance, or win the Arizona lottery, or ace the test. Certainties change not only with time, but from one person to another. We abstract from the latter by focusing on the case of a single person.

Ordinarily, we can be certain of the occurrence of past events that are contingent, or even chancy, in the sense that their non-occurrence was possible relative to all of the facts that obtained before they occurred, but, obviously, relative to the facts that have occurred up till the moment of certainty, non-occurrence is no longer possible. That is water under the bridge. And we can be (more or less) certain of future events that are causally necessitated by water under the bridge, if we know what has passed and the causal laws. What is special about the case of self-defeating causal chains is the certainty of failure while there is still the possibility of success in the absence of the sorts of causal relations that are its ordinary source.

[17] In *external* time.

[18] And in my power in a sense in which it is *not* in my power to make a radioactive atom decay at a particular time.

[19] It is no accident that we think of the modalities evolving in the way that we do; it is no accident, that is, that we typically have knowledge of the past and not the future, on the one hand, and that we think of freedom as applying only to future actions, on the other.

[20] If $\sim$X obtains, it obtains atemporally, and so if possibly X ever obtains, them $\sim$X and possibly X sometimes simultaneously obtain, whether or not we know it. Evolving possibilities are reconstructed in an atemporal framework as relations to times, so the right way to say this is that if $\sim$X obtains, it obtains atemporally, and so if there is some t such that possibly X at t, then there is some t such that $\sim$X at t and possibly X at t.

REFERENCES

Earman, J.: 'Implications of Causal Propagation Outside the Null-Cone', *Australasian Journal of Philosophy* **50**.

Horwich, P.: 1987, *Asymmetries in Time*, MIT Press, Cambridge.

Price, H.: 1996, *Time's Arrow and Archimedes' Point*, Oxford University Press, Oxford.

University of Arizona
Department of Philosophy
Tucson, AZ 85719
U.S.A.
E-mail: jtismael@u.arizona.edu