# The Temporal Asymmetry of Counterfactuals

Terrance Tomkow & Kadri Vihvelin

In our last several chapters we have defended Jonathan Bennett's Simple Theory of Counterfactuals.[1] One consequence of Bennett's theory is that counterfactual *backtracking* – supposing that the past would be different if the present were different—is legitimate. We closed our last chapter endorsing backtracking writ large by arguing that nomological determinism entails *counterfactual determinism*.

> **Counterfactual Determinism**
>
> However the world happens to be: If the world had been different in any respect at any time then it would have been different in some respect at every moment in the future and the past.

In this chapter, we will consider how to reconcile this position with the undeniable fact that there *are* differences between forward and backtracking counterfactuals.

David Albert is surely right to say.

> What we think … is that the future counterfactually *depends* on the present— and (moreover) we think that the future depends on the present in a way that the past emphatically does not.[2]

Philosophers have labeled this commonsense contrast "The Asymmetry of Counterfactual Dependence." Adam Elga describes it thus:

> The asymmetry of counterfactual dependence is (roughly) that later affairs depend counterfactually on earlier ones, and not the other way around. This asymmetry seems to be a feature of our counterfactual talk (at least in many contexts), which is reason enough to seek an analysis of counterfactuals that reproduces it.[3]

To explain how the Simple Theory accounts for it we will first have to say —less roughly— what the asymmetry is. This is harder to do than it ought to be. It is not just that decades of

---

[1].Tomkow 2013, Tomkow & Vihvelin 2016.
[2] Albert 2000, p.125.
[3] Elga 2000, p. 313.

philosophical writing have muddied these waters—one expects that— but also that matters were murky at the source.

The source is, of course, David Lewis in "Counterfactual Dependence and Time's Arrow" (CDTA) where he summarizes the Temporal Asymmetry of Counterfactual Dependence thus:

> Consider those counterfactuals of the form "If it were that *A*, then it would be that C" in which the supposition *A* is indeed false, and in which *A* and *C* are entirely about the states of affairs at two times $T_A$ and $T_C$ respectively. Many such counterfactuals are true in which *C* also is false, and in which $T_C$ is later than $T_A$. These are counterfactuals that say how the way things are later depends on the way things were earlier. But if $T_C$ is earlier than $T_A$, then such counterfactuals are true if and only if *C* is true. These are the counterfactuals that tell us how the way things are earlier does not depend on the way things will be later." [4]

He offers this with some qualifications. He thinks the asymmetry is a contingent matter; he thinks time travel (which would make some backtrackers true) is logically possible. And he allows that people sometimes talk *as if* backtrackers were true, but he thinks these involve "non-standard" ways of assessing the similarity of worlds.

Even so, matters are far from clear. There are two ways in which any counterfactual may be false. Thus, where *A* and *C* are false:

(1) If *A* were the case, then it would be that *C*

might be false because:

(2) If *A* were the case, then it would (still) be the case that not-*C*.

or (1) might be false because:

(3) If *A* were the case, then it might be the case that *C* but it also might be the case that not-*C*.

In the case of (3) we are affirming that things would be different if *A* were true, but denying that they would have to be different in one specific way rather than another. Both (2) and (3) entail

[4] Lewis 1979, p.458.

that (1) is false, but they are inconsistent with one another. If (2) is true, (3) is false and vice versa.

Lewis says that any instance of (1) will be false if $T_A$ is later than $T_C$. But will these backtrackers be false because (2) or because (3)? It makes a difference between two very different asymmetry claims.

> **STRONG ASYMMETRY**
>
> Where *A* and *C* are false, "If *A* were the case, then it would be that *C* " is sometimes true if $T_A$ is earlier than $T_C$; however, if $T_A$ is later than $T_C$ then it will always be true that "if *A* were the case, then *C* would (still) not be the case."

> **WEAK ASYMMETRY**
>
> Where *A* and *C* are false, "If *A* were the case, then it would be that *C*" is sometimes true when $T_A$ is earlier than $T_C$; however, if $T_A$ is later than $T_C$ then it will sometimes be true that "If *A* were the case, then *C* might be true" but whenever this is so it will also be true that "If *A* were the case, then *C* might (still) be false.

Either would make backtrackers false, but they are different claims. The first tells us that

> **FIXED PAST**
>
> If the present were different in any way, the future would be different in quite specific ways, but the past would be exactly as it actually was.

Whereas Weak Asymmetry gives us:

> **INDETERMINATE PAST**
>
> If the present were different in any way, then the future would be different in quite specific ways; the past would be different too, but not in any very specific way.

Is Lewis arguing that the past is Fixed or that it is Indeterminate? The complicated answer is "both" and "neither."

The complication arises from the fact that on Lewis's account when we evaluate a counterfactual of the form of (1) we typically envisage worlds at which some miracle—some event whose occurrence would contravene the laws of our world— has occurred in the past prior to the antecedent. The miracle must be just "big" enough to bring about *A* without "gratuitous" departures from the way things actually are at $T_A$. Since the effects of past changes tend to

magnify and multiply over time, we can minimize gratuitous differences by thinking of miracles in the fairly recent past. However, if we make the miracle *too* recent we will have worlds with miraculous changes that are too "abrupt." That is why we don't just imagine worlds where the past is entirely fixed, and *A* miraculously becomes true at $T_A$. What we need, Lewis says, is to change the past a *bit* to provide an "orderly transition." Speaking of "if the match were struck it would have lit" he says:

> Right up to ***t***, the match was stationary and a foot away from the striking surface. If it had been struck at ***t***, would it have traveled a foot in no time at all? No; we should sacrifice the independence of the immediate past to provide an orderly transition from actual past to counterfactual present and future. That is not to say, however, that the immediate past depends on the present in any very definite way. There may be a variety of ways the transition might go, hence there may be no true counterfactuals that say in any detail how the immediate past would be if the present were different. I hope not, since if there were a definite and detailed dependence, it would be hard for me to say why some of this dependence should not be interpreted—wrongly, of course—as backward causation over short intervals of time in cases that are not at all extraordinary. [5]

So, Lewis's theory of counterfactuals requires *Weak Asymmetry* for the transition period leading up to *A*.

For how long will Weak Asymmetry reign? How long is the "transition period"? Lewis suggests that it affects only the "immediate past", but this is a misdirection. It will vary depending on the counterfactual, depending on how far back we must go to provide an "orderly past" to a counterfactual present. Thus,

> If JFK had run for a second term, he would have won.
>
> If dinosaurs ruled the earth, there would be no internet.
>
> If the cosmic background radiation were isotropic, there would be no stars.

would require revisions of history in ascending durations stretching back to the Big Bang if that's when the relevant miracle would have to be to ensure a "smooth" transition.

---

[5] Lewis 1979, p.463.

But Lewis also wants to say that we must suppose that *prior* to the miracle that initiates the transition period the past of the closest world is *exactly* like that of the actual world. He wants to say this because he holds that what makes for similarity of worlds is overall similarity of matters of particular fact at all times. If we could not just *assume* that the closest worlds were exactly like ours prior to the miracle—if we had to figure out how *A* might have come about with no miracle—then evaluating any counterfactual would require us to know too much about the distant past. Accordingly, Lewis insists that the past is Fixed for all times *prior* to the miracle.

Lewis' account thus requires us to think of the past as Fixed to a certain point in time and Indeterminate thereafter with the relevant point of time depending upon which counterfactual we are entertaining.

All this will appear as doubly complicated when we remember that Fixity and Indeterminacy are contraries. If (1) is false, then it will be because either (2) or (3) is true. But (2) and (3) cannot both be true, so any argument that any particular counterfactual is false on account of Fixity is also an argument that it is *not* false on account of Indeterminacy; and *vice versa*. Lewis must show us that whatever it is that makes the past of the antecedent indeterminate back to the time of the miracle does not operate prior to the time of the miracle. Conversely, he must explain why the reasons for holding the past fixed prior to the miracle don't require its being fixed thereafter.

Alas, Lewis does none of this. He doesn't distinguish the two asymmetry theses. He conflates them and muddles the arguments for both.

## II

The confusion sets in on the first page of CDTA.

> If the present were different, the future would be different… Likewise the present depends counterfactually on the past, and in general, the way things are later depends on the way things were 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. Such a counterfactual, unless clearly false, normally is not clear one way or the other. It is at best doubtful whether the way things are earlier depends on the way things will be later.

> Often, indeed, we seem to reason in a way that takes it for granted that the past is counterfactually independent of the present: that is, that even if the present were different, the past would be just as it actually is.[6]

Notice that what Lewis here announces as the thesis that "the past is counterfactually independent of the present" is the Strong Asymmetry Thesis, but the evidence he has offered for it supports Weak Asymmetry and thus is evidence *against* the Strong thesis. If the past really were counterfactually independent of the present in the Strong sense, then it would never be "unclear" how the past would be different if the present were different. It would always be clearly true that it would *not* have been different than it actually was.

This schizophrenic slide from Weak to Strong pervades CDTA. In the opening paragraph, Lewis says:

> Today I am typing words on a page. Suppose today were different. Suppose I were typing different words. Then plainly tomorrow would be different also; for instance, different words would appear on the page. Would yesterday also be different? If so, how? Invited to answer, you will perhaps come up with something. But I do not think there is anything you can say about how yesterday would be that will seem clearly and uncontroversial true.[7]

Notice there are two questions here: "Would the past be different?" and "If so, how"? If the past were Fixed, the second question wouldn't require an answer. It would be clearly and uncontroversially true that *nothing* would be different. But it isn't. We think the past would have been different, but lack a clear answer to "How?" This is evidence for an Indeterminate, not a Fixed, past. And yet a few paragraphs later we find Lewis insisting that "we seem to reason" as if the past were Fixed.

> We evidently assume that even if our supposition about the present were true, the past would be no different. If I were acting otherwise just now… it is absurd even to raise the question about whether that past wrong would have taken place if I were acting otherwise now! More generally, in reasoning from a counterfactual supposition about any time, we ordinarily assume that facts about earlier times are counterfactually independent of the supposition…[8]

---

[6] Lewis 1979. p.456.
[7] Lewis 1979, p.456.
[8] Lewis 1979, p.457.

Notice that the question Lewis is here denouncing as “absurd” is precisely the question of his opening paragraph: "Suppose I were typing different words... Would yesterday also be different?" How did answers to that question go from being “unclear” and “controversial” to “absurd”?

In any case, it doesn’t seem that we can treat *either* the Strong or Weak Asymmetry thesis as true without exception. People engage in backtracking reasoning all the time. Thus:

What would Hilary have to have done to win the election?

is a perfectly sensible question which requires a backtracking answer:

If Hilary had won, she would have to have ...

And there are some backtrackers which are clearly and uncontroversially true:

(4) If JFK had run for a second term, he would have to have not been assassinated.

Lewis tries to dismiss such cases by arguing that they involve marginal or “non-standard” ways of resolving counterfactuals. He suggests that the “would have to have” signals the non-standard usage. But this is not plausible. Consider the unremarkable non-backtracker:

(5) If JFK had run for a second term, he would have won.

This is a claim about which reasonable people may disagree, but in ordinary contexts no one would propose that the closest worlds to the actual in which JFK runs again are ones in which JFK is assassinated but miraculously rises from the dead to run again. That would certainly have got him re-elected, but it is not a counterfactual situation relevant to the serious consideration of (5). We do not say this as a counterexample to Lewis’s theory. The problem with such a world, Lewis can say, is that it involves too *big* a miracle: better to imagine a world where some small, "inconspicuous" miracle—a gust of wind, a failure of nerve, a nod of the head—prevents the assassination, allowing an “orderly transition” to JFK’s re-election bid. But having said that, Lewis will have conceded that (4) is true and must be held true in any “standard” context in which (5) is held to be true or false.

(4) is an unequivocally true backtracker, but we might still think it a special case because not having been assassinated is *nomologically* necessary for running for president. It is surely that nomological connection that the “*would have to have*” signals. We might count such cases as

"exceptions that prove the rule," but we still need to know what the rule is. Is it Strong Asymmetry or Weak? Whatever our answer, we can add the burden of explaining why (4) is exceptional as a requirement on any explanation of asymmetry.

We might hope to get a clearer bead on Lewis's claim by looking at the explanation for asymmetry he offers us in CDTA. His argument there—that the future "overdetermines" the past—has been much discussed in the literature. Even so, it seems to us that its most fundamental problem has escaped notice. The problem is that the overdetermination argument—if it succeeds— it is a demonstration that the past is FIXED. That is, it purports to show why, when we consider a counterfactual $A > C$, the closest world must be *exactly* like the actual world prior to $T_A$. It leaves no room at all for an "orderly transition" period in which things would have been different (albeit in Indeterminate ways). Which is to say: Lewis's explanation of counterfactual asymmetry is inconsistent with his own theory of counterfactuals.

IV

To understand how things go wrong, we need to understand why Lewis's account of counterfactuals *needs* a transition period. This will take a few paragraphs because, for reasons we shall consider shortly, Lewis himself never really tells us.

First, consider what we may call the Naïve Theory of Counterfactuals. "Naïve" because it was the first proposal considered by Nelson Goodman, the first philosopher to think seriously about the topic[9]. The Naïve Theory says that in order to decide if a counterfactual $A > C$ is true all we have to do is to conjoin a description of the way the actual world is at $T_A$ with the supposition that $A$ is true and then see if $C$ follows from this new description of the world together with the laws of nature.

It's an appealing simple picture. So appealing that it has been rediscovered many times since Goodman first discussed it.[10] But it doesn't work.

To see why, notice first that when we prepare our new description of the world, we cannot just "add in" the supposition that $A$ is true. We are going to have to edit out the fact that $A$ is actually false. Otherwise we would have a contradiction. And while we are at it, we are also going to

[9] Goodman 1947.
[10] E.g. Maudlin 2007, Hall and Paul 2013.

have to– on pain of contradiction– edit out any facts which are actually true but are logically inconsistent with *A*.

But Goodman saw that even this would not be enough. The problem is that just removing not-*A* and its *logical* consequences from the description of the actual world would still leave in things that are *nomologically* incompatible with *A* and would still lead to contradiction given the laws of nature.

For example: in the actual world JFK did not run for a second term. What would have happened if he had? If we revise our description of the world *just* changing that fact that he did not run, we describe a world with a dead man campaigning for president. As we observed above, whatever might happen in such a world, it isn't relevant to answering questions about what would have happened if JFK had run a second time. To get things right, we will have to revise our description not just to exclude things that are *logically* inconsistent with JFK's second run but also to change anything else *that would have to be different* for the second run to happen. In Goodman's terms, we would have to eliminate any facts that are not "cotenable" with a second JFK run. Facts are not cotenable if one would be false if the other were true. [11]

But the problem of determining how things *would have been* different if JFK had run again is a problem of deciding what counterfactuals are true. We cannot, on pain of circularity, appeal to counterfactuals in our account of the truth conditions of counterfactuals. Goodman generalizes the worry:

> [the] really serious difficulty … confronts us. In order to determine the truth of a given counterfactual, it seems that we have to determine, among other things, whether there is a suitable S that is cotenable with *A* and meets certain further requirements. But in order to determine whether or not a given S is cotenable with *A*, we have to determine whether or not the counterfactual 'If *A* were true, then S would not be true' is itself true. But this means determining whether or not there is a suitable $S_1$, cotenable with *A*, that leads to — and so on. Thus we find ourselves involved in an infinite regress or a circle; for cotenability is defined in terms of counterfactuals, yet the meaning of counterfactuals is defined in terms of cotenability. In other words, to establish any counterfactual it seems that we first have to determine the truth of another. If so, we can never explain a

[11] Goodman 1947, p. 120.

> counterfactual except in terms of others, so that the problem of counterfactuals must remain unsolved.[12]

Goodman called this "the problem of cotenability" and despaired of solving it.

Lewis's transition period allows him to avoid Goodman's problem. His famous *Analysis 1* has it that:

> *Analysis 1*. Consider a counterfactual "If it were that *A*, then it would be that *C*" where *A* is entirely about affairs in a stretch of time $T_A$. Consider all those possible worlds ***w*** such that:
>
> (1) *A* is true at ***w***;
>
> (2) ***w*** is exactly like our actual world at all times before a transition period beginning shortly before $T_A$;
>
> (3) ***w*** conforms to the actual laws of nature at all times after $T_A$; and
>
> (4) During $T_A$ and the preceding transition period, ***w*** differs no more from our actual world than it must to permit *A* to hold.
>
> The counterfactual is true if and only if *C* holds at every such world ***w***.[13]

So while it begins with a miraculous divergence from the actual past, the closest world must conform to the actual laws of nature "at all times after $T_A$". And note that a world can't conform to the laws of nature *after* $T_A$ unless the way things are at $T_A$ is consistent with the laws of nature. If they are not, contradiction will result. The transition period is necessary because the laws of nature may not logically "permit" *A* to be true unless other adjustments are made at $T_A$ to remove nomologically inconsistent facts.

The Kennedy example illustrates why the transition will often require more than a short interval. To have him running for a second term in 1964 we must backtrack at least far enough to suppress all sorts of non-cotenable facts about that time — the grave in Arlington, the Zapruder film, all the records of an LBJ presidency… . And notice, too, that it is not enough to find *nomologically possible* worlds where Kennedy runs again. The closest worlds must be worlds that do not differ "gratuitously" from the actual world; that is, they must be worlds that differ from the actual world only as much as is needed to permit Kennedy to make a second run.

---

[12] Goodman 1947, p. 121.
[13] Lewis 1979, p.462.

This *avoids* Goodman's problem. But does it *solve* it? If we try to take Lewis's account of counterfactuals beyond the metaphors of "smooth" and "orderly" what can we say in answer to the question "How long must the transition period be?" If the answer is, "long enough so that ***w*** at $T_A$ does not include any features of the actual world that *would have to have been different* if *A*", the theory turns circular.

If Lewis had a different answer to Goodman's worry, he never gave it.

V

Keeping all this in mind, let us return to Lewis's overdetermination thesis. Though he thought it gave the right results, Lewis was not satisfied with *Analysis 1*. The problem was that *Analysis 1* simply stipulates that the past, prior to the transition, must be Fixed and Lewis wanted to explain the asymmetry, not just assume it by definitional fiat. He needed a temporally neutral account of similarity which gave the same results as *Analysis 1*.

In his earlier work, Lewis sometimes talked as if similarity of worlds is simply a matter of match in matters of particular fact. If we just measured similarity by overall match of past, present and future facts we would have a temporally neutral account of similarity. But could such a similarity measure give us the same results as *Analysis 1*? An argument from Michael Slote, Kit Fine, and others[14] seemed to show that it could not. Fine's example was:

If Nixon had pushed the button, there would have been a nuclear holocaust.

This seems to be true. But, Fine argued, if all that matters to counterfactual similarity are matters of particular fact and if miracles are allowed, why isn't the closest world to ours one where some small miracle – say, a firing of a decisive neuron— causes Nixon to push the button but then another miracle prevents the missiles from firing. The result would be a world which would more resemble the actual world in the future than one destroyed in a nuclear war. If this is allowed, the upshot would be that nothing much would make a difference to anything since any miraculous divergence miracle could be offset by other "convergence miracles" that would bring things back to the way they actually were.

[14] Fine 1975, Slote1978, Bennett 1974, Bowie 1979, Jackson 1977, Schlossberger 1978, Tichy 1976, Creary & Hill 1975.

Lewis answer to the "Future Similarity objection" in CDTA was that similarity in matters of particular fact is *not* enough. In comparing worlds, it is also important to consider the number and size of the miracles involved. The fewer and smaller the miracles required to make the antecedent true, the better even if the worlds end up greatly different in matters of particular fact. He proposed a prioritized list. In comparing worlds for similarity:

> (1) It is of the first importance to avoid big, widespread, diverse violations of law.
> (2) It is of the second importance to maximize the spatio-temporal region throughout which perfect match of particular fact prevails.
> (3) It is of the third importance to avoid even small, localized, simple violations of law.
> (4) It is of little or no importance to secure approximate similarity of particular fact, even in matters that concern us greatly.[15]

At the actual world, $w_0$, Nixon does not push the button but at a nearby world $w_1$

> the laws of $w_0$ are violated in some simple, localized, inconspicuous way. A tiny miracle takes place. Perhaps a few extra neurons fire in some corner of Nixon's brain. As a result, Nixon pushes the button. With no extra miracles events take their lawful course and the worlds $w_0$ and $w_1$ go their separate ways.[16]

And even though $w_0$ and $w_1$ will go very different ways "in matters that concern us greatly" (one has a nuclear holocaust!) we won't find a world more similar than $w_1$ because holocaust is what happens when things take their "lawful course" after Nixon pushes the button. To suppress these later differences we would need many more miracles to eliminate all the traces of the button pushing.

> There are his fingerprints on the button. Nixon is still trembling, wondering what went wrong—or right. His gin bottle is depleted. The click of the button has been preserved on tape. Light waves that flew out the window, bearing the image of Nixon's finger on the button, are still on their way into outer space. The wire is ever so slightly warmed where the signal current passed through it. And so on, and on and on.[17]

---

[15] Lewis 1979, p. 469.
[16] Lewis 1979, p. 469.
[17] Lewis 1979, p. 469.

Wiping out all these traces would require a lot of small miracles or one very big one.

This, generalized, is Lewis's explanation of why convergence is always more miraculous—in the sense of requiring more and bigger miracles—than divergence. He puts it in terms of what he calls "determinants":

> Any particular fact about a deterministic world is predetermined throughout the past and postdetermined throughout the future. At any time, past or future, it has at least one *determinant*: a minimal set of conditions jointly sufficient, given the laws of nature, for the fact in question. (Members of such a set may be causes of the fact, or traces of it, or neither.) The fact may have only one determinant at a given time… Or it may have two or more...each sufficient by itself. If so, it is *overdetermined* at that time.[18]

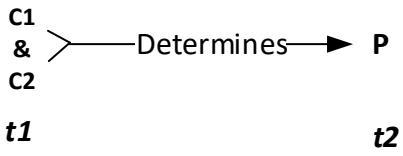
Our world, Lewis tells us, happens to be one at which the pre-determinants of every fact are greatly outnumbered by its post-determinants. At our world, the future *overdetermines* the past. This "asymmetry of overdetermination" means that worlds that are different from ours at any time can be more similar to our world in their pasts than they can be in their futures. To suppress those future differences in the way the Future Similarity objection proposes will always require more miracles.

> Extreme over-determination of earlier affairs by later ones …may well be more or less universal at a world like ours. 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 the determinants thereof; there is no lawful way for the combination to have come about in the absence of the fact. If so, the abundance of future traces makes for a like abundance of future determinants.
>
> That would explain the asymmetry of miracles. It takes a miracle to break the link between any determinant and that which it determines. … 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. [19]

[18] Lewis 1979, p. 473.
[19] Lewis 1979, p. 473.

As we noted above, Lewis's overdetermination argument has been much discussed in the philosophical literature but almost[20] without exception this literature has focused on the question of whether Lewis's CDTA weighting system really succeeds in capturing the asymmetry stipulated in *Analysis 1*. They are reworkings of the Future Similarity Objection. What has gone unnoticed is that Lewis's overdetermination argument is actually *inconsistent* with *Analysis1*.

To see this, let's assess some arbitrary counterfactual "If *P* were not the case then …." where *P* is a fact about some time $t_2$. Already, Lewis's claims about the way things are "at our world" tells us a good deal about *P*. He has told us that, at every time prior to $t_2$, there is some set of facts about that time which pre-determine *P*. Suppose, then, that at $t_1$ *P*'s pre-determinant is comprised of two facts which we'll label $C_1$ and $C_2$.

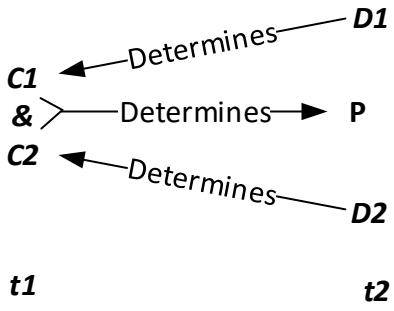



Lewis invites us to speak of $C_1$ and $C_2$ as "causes". To say that they together pre-determine P is to say:

$$\Box_N (C_1 \& C_2 . \supset P)$$

Lewis has also told us that every "particular fact" about every time will have many post-determinants. This will be true of *P*, but it will also be true of $C_1$ and $C_2$ since they too are "particular facts". Let us suppose, then, that at $t_2$ $C_1$ and $C_2$ each have at least one post-determinant; call them $D_1$ and $D_2$. (Lewis says that every particular fact will have *many* post-determinants at every future time, but, as we shall see, one apiece will be problematic enough.)

[20] The exception we have in mind is Huw Price 1997. Price accepts Lewis argument for macro-physical events but argues that there is no over-determination at the microphysical level.



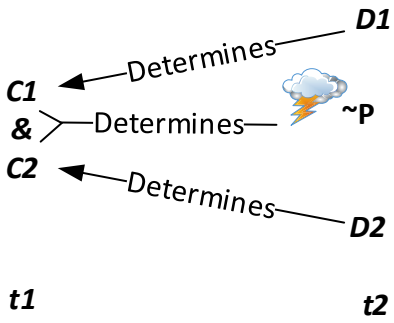

Once again, since determination is a relation of nomological sufficiency, the laws will tell us:

$$\Box_N (D_1 \supset C_1)$$

and

$$\Box_N (D_2 \supset C_2)$$

We were considering a counterfactual of the form "If *P* were not the case then….". To do so, Lewis directs us to imagine a small miracle that prevents $C_1$ and $C_2$ from causing *P* and to see what lawfully unfolds after $\boldsymbol{t_2}$.




But now there is a problem. It is no part of this story that the miracle that prevents *P* will also somehow prevent $D_1$ or $D_2$. Indeed, the crux of Lewis's argument is that the post-determinants of any fact are so "many and diverse" that they *cannot* all be erased by a small localized miracle. So even if a miracle prevents *P* from being true at $\mathbf{t_2}$ it will not prevent the post-determinants of its causes from being true. But if they are true, the world at $\mathbf{t_2}$—the time of the antecedent of our counterfactual—is not nomologically possible. After the miracle, it will be the case that:

$$D_1 \,\&\, D_2 \,\&\, \sim P$$

But if that's so, applying the laws yields a contradiction:

$\Box_N (D_1 \supset C_1)$

$\Box_N (D_2 \supset C_2)$

$\Box_N (C_1 \& C_2 . \supset P)$

$\therefore \Box_N (D_1 \& D_2 . \supset P)$

$\therefore P$

$\therefore P \& \sim P$

The post-determinants of the pre-determinant of *P* are not nomologically consistent with ~P. In other words, ~P is not *co-tenable* with $D_1$ & $D_2$. If Lewis's overdetermination thesis is correct, this will be true for every *P*.

Thus, for instance, take the actual world, $\boldsymbol{w}_0$, where Nixon did not push the button and no missiles were launched. Lewis said that the closest world, $\boldsymbol{w}_1$, is one where "a tiny miracle takes place. Perhaps a few extra neurons fire in some corner of Nixon's brain" getting Nixon to push the button. But if the pre-miracle past of $\boldsymbol{w}_1$ is like the past of $\boldsymbol{w}_0$, it must include the pre-determinant causes that *prevented* those extra neurons from firing at $\boldsymbol{w}_0$. Those causes will have left traces in future times when and after the button is pushed. Traces so many and diverse, we have been assured, that no one tiny miracle could have suppressed them all. Thus, even as the button is being pushed there are facts that make it nomologically impossible that any button pushing should have come to pass. They will be as nomologically inconsistent with the buttons being pushed as Kennedy's lying in Arlington is with his campaigning for a second term.

To make changes in the world sufficient to allow the antecedent of our counterfactual, *P*, to be false we could lengthen our transition period: moving the miracle backward in time before $t_1$ to prevent $C_1$ or $C_2$. But now our problem reasserts itself. $C_1$ and $C_2$ must themselves have pre-determinants and those pre-determinants must be post-determined by other facts at $t_1$; facts, again, so many and diverse that no single miracle could remove them all. Backtracking further— eliminating the pre-determinants of those post-determinants — will not help. Overdetermination means that the causes of every pre-determinant we miraculously prevent from

having its determined effect will have other post-determinants going forward that will nomologically require that the effect occurs. And so on, and on, and on.

One of the reasons the overdetermination argument has beguiled so many readers is Lewis' artful use of the word "traces". In a postscript to CDTA he claims.

> "My argument for an asymmetry of miracles … relied on an empirical premise: at a world like ours, everything that happens leaves many and varied traces, so that it would take many .. miracles working together to eradicate those traces and achieve reconvergence."[21]

And this, he says, is something everybody knows.

> … everyone believes in the ubiquity of traces… Consider detective stories. The background against which they are to be read is common knowledge. And part of that background knowledge is the assumption that events leave many and varied traces. Else the plots would not make sense. We are supposed to marvel at the skill of the detective in spotting and reading the traces. We are not supposed to marvel that the traces are there at all.[22]

It is indeed common knowledge that everything that happens leaves many traces, and as we shall soon see, this fact *is* relevant to asymmetry. But these common sense "traces" are just causal upshots, not Lewis's *post-determinants*. Remember, post-determinants are sets of facts which are not just *evidence* of some prior fact; they are *nomologically sufficient conditions* for it. Lewis needs them to be so for it to require miracles to eliminate them. Not even Sherlock Holmes discovered *post-determinants*. If he had, he would never have had to discover more than one. A criminal might hope to erase the ordinary traces of his crime, but he could not hope to eliminate a Lewis "trace".

> 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.)[23]

Traces like these are more than clues, they are *nomological guarantees* of past facts. In most detective fiction, the clues initially point to many suspects some of whom later prove innocent.

---

[21] Lewis 1986, p.66.
[22] Lewis 1986, p.66.
[23] Lewis 1979, p.474.

There is no intelligible fiction in which suspects are innocent despite *post-determinants* of their guilt.

Lewis closes CDTA with the claim that his asymmetry of overdetermination explains Popper's famous radiation asymmetry.

> "An asymmetry noted by Popper is a special case of the asymmetry of overdetermination. 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 over determination 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 process that occur are the ones in which this extreme overdetermination goes toward the past, not those in which it goes toward the future. I suggest the same is true more generally." [24]

The picture Lewis seems to have in mind looks like this:

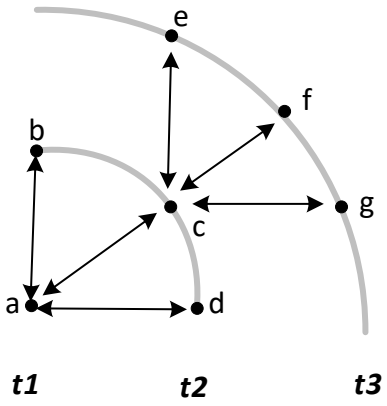



*Figure 1*

Where each dot represents a sample of a wave at a time, the letters stand for local facts about those samples and the arrows represent relations of determination amongst those facts. So, in the picture, *a* pre-determines each of *b*, *c* and *d* while *b*, *c*, *d* each determines and together overdetermine *a*. The overdetermination of *a* by its post-determinants is supposed to mean that even though forward- trackers like:

$$\sim a > \sim c$$

[24] Lewis 1979, p. 475.

are true, no backtrackers like

$$\sim c > \sim a$$

will be true.

But this doesn't work. The problem has once again to do with co-tenability. What is the closest world to Figure 1 at which ~*c*? We can certainly imagine a small miracle just before $t_2$ erasing *c*, and we could agree that it would take a lot of miracles to restore *e*, *f* and *g* absent *c*. But *c* itself is a "trace" of *a* and our miracle will not make the other $t_2$ traces of *a* go away. The laws represented by all those arrows have the consequence that:

$$\Box_N (a \equiv b \equiv c \equiv d \equiv e \equiv f \equiv g)$$

At any world with laws like this it will be nomologically *impossible* that *b* or *d* be true if *c* is not. To minimize miracles and get a nomologically possible state of affairs where ~c "smoothly" evolves at $t_2$ we will need a backtracking transition period. We will need to go to a world where ~a. But, of course, at a world where ~a there are no waves at all.

The same regress operates for any wave sample and any time. Given the laws of the Figure 1 world, all the waves at every time must stand or fall together. No asymmetry has been explained because the picture does not allow asymmetry.

## VI

Lewis's explanation for Strong Asymmetry falls apart. Which is just as well because, apart from the requirements of his own theory of counterfactuals, there is no reason to believe the past is counterfactually Fixed.

It is often said that Lewis needed a Fixed Past to sustain his counterfactual analysis of causation. The thought is that without the Strong Asymmetry his counterfactual account of causation could not explain the temporal asymmetry of causation. Lewis himself may once have thought this was so, but as his metaphysics evolved, he found a way to jettison the requirement for a Fixed past. Already, in CDTA, we see him denying that the changes in the transition period count as caused because they are not "definite and detailed." He would later build this into his metaphysics by way of his theory of *events*. His account of causation, he said, was only a theory

of *event* causation and he stipulated that events cannot be too "disjunctive." [25] This meant he could concede that if the window weren't broken *something* would have been different previously but still insist that whatever this might be —"Suzy would have changed her mind, or missed, or not thrown hard enough…" was too disjunctive to count as an event. Weak Asymmetry is consistent with the idea that backtracking always results in this sort of disjunction—"it might have been this way or might have been that way …". So, for the purposes of analyzing causal direction, Weak Asymmetry, together with his doctrine of events, are all Lewis's account of event causation required.

In the end, the *only* reason Lewis has for insisting on a Fixed Past stems from his theory of counterfactuals itself: his assumption that counterfactuals must be evaluated in terms of the overall similarity of worlds throughout their histories. Recall that this is why he held that the closest worlds must contain miracles. The argument went that if the closest worlds must obey the laws of nature—if they have no miracles— then determinism would require that they have different pasts back to the beginning of time. How then can we evaluate counterfactuals if to do so we have to reconstruct and compare entire world histories? A "recent" miracle allows us to *assume* that the closest worlds have pasts exactly like ours at least up to the time of the miracle.

This is a good argument against a particular theory of counterfactuals — one which says that the closest world must have the same laws as ours *and* which measures similarity by closeness of fact across all times. It's a good argument because no credible theory of counterfactuals can have the result that counterfactuals are unknowable because they would require knowing too much. But this epistemological point counts not at all against Bennett's Simple Theory.

Bennett's theory does not measure similarity of worlds by matches of fact across all times but only at $T_A$, the time of the antecedent. On his account, we need pay no more attention to the past than we need to decide questions about present similarity. The closest worlds are assumed to have achieved their state at $T_A$ lawfully; they contain no miracles. Given determinism, that means that their pasts will differ from the actual past at every moment. They may differ greatly, but we do not have to know what those pasts would have to have been like to compare them with the actual world at $T_A$. And given that they must have achieved their $T_A$ state lawfully, there is

[25] "Events", in Lewis 1986, 241–269.

no danger of those worlds containing facts nomologically incompatible—that is, which are not cotenable — with the truth of the antecedent.

It is true that Bennett's Simple theory does not give us a psychological explanation of how we arrive at our counterfactual opinions. That is, it does not tell us how we imaginatively construct our picture of what the world would be like if *A* were true when we know that *A* is false. But Lewis' theory doesn't do that either. In deciding if $A > C$ is true, both Lewis and Bennett would have us think of what the world would be like at some time $T_{A-}$ prior to $T_A$: a world which, at $T_{A-}$ and after, obeys the laws of nature but a world just different enough from the actual world at $T_{A-}$ so that *A* will lawfully come about with the minimum nomologically possible difference between this world and the actual world at least from $T_A$ forward. Neither Lewis or Bennett tell us how to think up such a world. But once we have, Lewis invites us to assume that it came about as a miraculous departure from a past exactly like the actual past prior to $T_{A-}$, while Bennett invites us to assume that it has some un-miraculous history or other— as it must have given that the world at $T_{A-}$ is required, on both theories, to be nomologically possible. Epistemologically and psychologically the theories are on a par.

What is most important for present purposes is that because Bennett's theory ties counterfactuals to laws it allows us to give an explanation of the *real* temporal asymmetry of counterfactuals.

V

The real asymmetry is *Weak.*

If Lewis were typing different words on that day, who can say what would have been different the day before? Lewis is right; there is no clear answer. But if we assume determinism is true – and, remember, Lewis assumes determinism throughout CDTA[26]— then we will want to insist, in the spirit of Leibniz, that if Lewis were doing otherwise, *something* in the past would have to have been different. Indeed, if we heard someone claiming that something could have happened otherwise even if the past were exactly the same, we would take them to be *denying* determinism.

Lewis might have decided to write about something else, or decided to go for a walk instead or he might have been interrupted by a phone call... We cannot say how the past would be different

[26] Lewis 1979, 460-461.

not because we cannot backtrack, but because there are too many equally good backward tracks to take.

Backtracking counterfactuals are almost always false not because the past wouldn't have been different but because the past might have been different in so many ways. Suzy throws a rock and breaks a window. It is clearly true that if she hadn't thrown the rock the window would not be broken. But if the window weren't broken now, would that be because Suzy didn't throw the rock? Maybe, but maybe not. She might have thrown the rock but missed, or not thrown it hard enough, or it might have been deflected by a passing bird.... There are lots of ways things might have gone.

The explanation of why this is typically so has two parts:

The first, and most obvious part, is to note that, as a matter of contingent fact, almost everything that happens has multiple causes. When we consider how things would have been if things had happened differently we suppose not —as Lewis claims— that all those causes would remain in place but have miraculously failed to produce the effect but rather— as the laws and Bennett require—that at least one of those causes would have failed to occur. That there are multiple causes means there are many ways things might have gone differently.

In the other direction, this explains why forward looking counterfactuals are so often true and specific. Every present fact is the upshot of combinations of past facts *every* one of which is, in the circumstances, necessary for that upshot. That upshot in turn is, in its circumstances, necessary for any number of future upshots, *none* of which would occur in its absence. That is, there is typically a *disjunction* of nomologically possible ways that something might have failed to come about but a *conjunction* of effects that would not have come to pass if it didn't. Which is why forward trackers succeed where backtrackers fail. Many different things *might* have prevented Suzy from throwing that brick but if she hadn't thrown it, some things that are now true would clearly be different: there would be no glass on the floor, the dog would not be howling, the police would not have been called...

That first part of our answer suggests why, given some counterfactual $A > C$, we should expect to find more than one lawful history which might have brought it about that $A$. But that isn't quite enough. It does not explain why many of these different histories can be expected to be more or less equally similar to the actual world where $\sim A$. On Bennett's theory we measure

similarity by comparing worlds at $T_A$, the time of the antecedent, so however many different ways it might have come about that *A*, if one of these is much more like the actual world at $T_A$ than any of the alternatives then that is the way things *would* have been if *A*; no "might" about it. The second part of our explanation is needed to explain why this will not typically be so.

We get this part of the explanation by noticing that Bennett's similarity criteria can be formulated in the following way: The closest *A* world to the actual world is one whose past is different enough from the actual world to bring it about that *A* but otherwise leaves the fewest *traces* that distinguish it from the actual world at $T_A$. "Traces" not in Lewis' technical sense of "determinants" but in the ordinary sense of "causal upshot". Having framed things in this way we can put Lewis's core insight to work: At our world, *everything* that happens has multiple causes in the past and each of those causes in turn has multiple effects on the present. Whatever changes in the past would lawfully (that is to say, non-miraculously) bring it about that *A* will make for *other* differences at $T_A$. That is why we are typically left with ties and tradeoffs in which *A* worlds would be most like the actual world at $T_A$.

Suzy threw a brick and broke the window. What would the world be like now if the window was not broken? It would surely depend upon why it was not broken. Was it because Suzy decided not to throw the brick? Did her throw fell short? Or did the brick hit a passing bird? Each of these different pasts would have consequences in the present. In worlds where she didn't throw she may now be striding ruefully away or frozen in indecision. In worlds where she missed she may be cursing her aim and looking for something else to throw. In worlds where her brick hit a passing bird, she may be wracked with grief. And so on. Which of these is most like the actual world here and now where Suzy did throw the brick, and the window is broken?

Generalized, this explanation answers the problem which began this essay: How to reconcile Nomological and Counterfactual Determinism with the temporal asymmetry of counterfactuals.

*Nomological* Determinism requires that every nomologically possible *total* state of the world have only one nomologically possible history and future. *Counterfactual* Determinism says that if the world were different in any way at any time, it would be different at every time. The problem, it seems, is that if we assume, as Bennett requires, that any of the alternative ways the world *might* be must have the same deterministic laws as the actual world, then every alternative way the world might be must have a specific and unique past and future. But now if every

nomologically possible alternative world state has a unique nomologically guaranteed history how can there be any uncertainty about how the past would be different if the present were different?

The answer is that while it is true that if the world were different now in any respect, it would have to be different in *some way or other* at every time, there is typically no one specific way in which things *would have to be* different. That is so because there is no one way the world would have to be *now* if any fact about it now were otherwise.

This explains why counterfactual backtracking so often fails and we can now also understand why it sometimes succeeds.

Bennett's theory requires that counterfactual situations obey the laws of nature. Suppose then that:

(6) $\Box_N (A \supset C)$

(6) tells us there is no nomologically possible world at which *A* is true, and *C* is false. If so, it follows that at any nomologically possible world, the nomologically possible worlds most similar to it at which *A* is true are also worlds where *C* because there are no nomologically possible worlds where *A* and ~*C*. And if the closest *A* world to *any* nomologically possible world is a world where *C*, it must be true that at every nomologically possible world, $A > C$.

So, for Bennett:

(7) $\Box_N (A \supset C) \equiv \Box_N (A > C)$

And this will be so whatever times the antecedent and consequent are about.

This explains why the backtracking:

Kennedy is re-elected in 1964 > Kennedy is not assassinated in 1963

is true. It is true because its antecedent is *nomologically* sufficient for its consequent and hence the counterfactual itself is nomologically true. *That* is what we are signaling when we say that if Kennedy were elected he *would have* to have avoided assassination.

There are more examples where that one came from. It is worth reflecting on just how *many*.

At a deterministic world, there must be,, at every time past and future at least one nomologically sufficient condition— a determinant in Lewis' sense— for every present fact. So if at the actual world *A* is false then it must be that every time in the actual past and future there is a fact *D* such that:

$$\Box_N (D \supset \sim A)$$

Contraposition gives us:

$$\Box_N (A \supset \sim D)$$

Which, given the equivalence (7), gives us.

$$\Box_N (A > \sim D)$$

And it doesn't matter whether *D* is in the past or future of $T_A$. Thus we arrive at Counterfactual Determinism in its strongest form: If the present were different then not only *would* the past and the future be different, they would *have to be.*

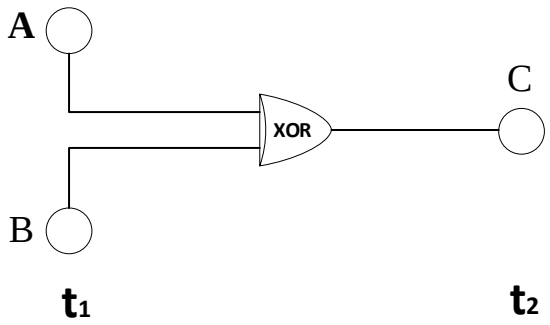
This backtracking does not subvert our explanation of asymmetries. As we observed, the determinants of any particular fact— e.g. Suzy throws a rock— will typically be a large conjunction of independent facts, and that means there will be a large disjunction of ways the world might be if that determinant were false. Still, a more or less infinite number of backtrackers from every fact about every time is a *lot* of backtracking. How then to account for the consensus view that backtracking is rare, that "the future depends on the past in a way the past does not depend on the future"? How did Lewis get away with marginalizing backtracking as a mug's game?

The explanation is, in part, that we rarely find ourselves backtracking or forward tracking to determinants because we don't know what the determinants are. Remember "determinants" in this sense are conditions which are nomologically *sufficient* for an outcome. In real life, we simply cannot tell the *whole* nomological story about why *anything* happens. We cite some salient facts but then invoke *ceteris paribus* clauses to gloss over the myriad other necessary conditions. As Nancy Cartwright points out, this is as true of scientists in the lab as the man on the street: Newton's Laws guarantee exactly where the billiard ball will go but only assuming the

table is level, the balls perfectly round, there is no wind, no earthquake shakes the table and who knows what else.[27]

So we don't backtrack or forward track in this way because we can't. We never know enough. But there is more to it than that. We don't miss this knowledge because we are not *interested* in this *kind* of counterfactual dependence.

Some counterfactuals are nomologically necessary, but not all. To see this think about one of our simple circuit-diagram worlds[28]:



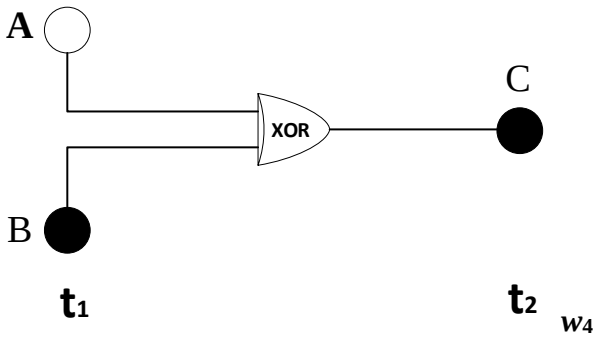

***w3***

This world has one law

$$L_{XOR}: \quad \Box_N (C \equiv (A \oplus B)$$

At ***w3*** it is true that $A > C$, but its truth is nomologically contingent because $A$ is not nomologically sufficient for $C$. Here is a world with the same laws but different facts on the ground, ***w4***:

[27] Cartwright 1980.
[28] We introduced these little intuition pumps in Tomkow & Vihvelin 2015




At $w_4$, $A > C$ is false. Indeed, at $w_4$, $A > \sim C$.

The small set of worlds governed by $L_{XOR}$ are interesting in another respect having to do with backtracking. Notice that at $w_3$ the backtracking:

$C > A$

is not true. If $C$ were true at $w_3$, it *might* be because $A$ is true or it might be because $B$ is true. Likewise, where $C$ is true— as it is at $w_4$— it is not the case that:

$\sim C > \sim B$

Since a world where $B$ & $A$ is every bit as much as like $w_4$ at $t_2$ as one in which $\sim B$ & $\sim A$.

It's tempting to say that there is no backtracking in these worlds, but that cannot be right: Wherever there is forward determinism there will be true backtrackers. At $w_3$ it is true that:

$C >. A \oplus B$

And at $w_4$ it is true that:

$\sim C > \sim(A \oplus B)$

Both these backtrackers are *nomologically* true in these worlds. At all the possible worlds that obey $L_{XOR}$, the nomologically *contingent* counterfactuals always track temporally forward.

Now think about paradigm cases of temporal counterfactual asymmetry, the kind that make us want to say that the past does not depend on the present the way the future does. You will find they are *always* cases of nomologically *contingent* counterfactuals. Significantly, wherever the language of counterfactuals intersects the language of causation—wherever we argue that this event caused that event because that event would not have occurred without this one— we are always invoking a nomologically contingent counterfactual relation.

Corresponding to the distinction between nomologically and non-nomologically contingent counterfactuals we have a distinction between nomological and non-nomological counterfactual dependence:

$$\text{Q non-nomologically depends on P.} =_{df}. \; P \,\&\, Q \,\&\, (\sim P > \sim Q) \,\&\, \sim\Box_N (\sim P > \sim Q)$$

versus

$$\text{Q nomologically depends on P.} =_{df}. \; P \,\&\, Q \,\&\, (\sim P > \sim Q) \,\&\, \Box_N (\sim P > \sim Q)$$

*Nomological* counterfactual dependence is no respecter of time's direction. In a deterministic world, it tracks backwards and forwards indiscriminately. At our world, it is nomologically *contingent* counterfactuals that follow Time's Arrow. Because Bennett's theory rests the truth conditions of counterfactuals on laws, not miracles, we are now in a position to say what the laws of our world must be like for this to be so.

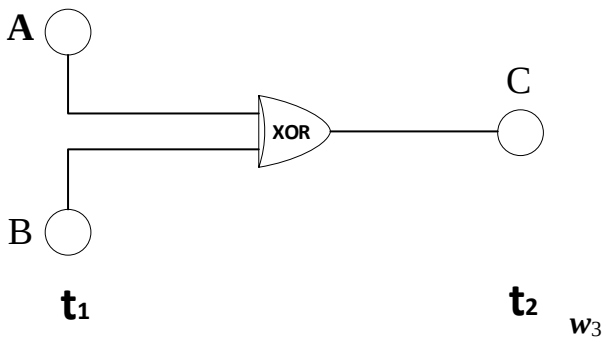
## VI

So let us ask again: why does the past not depend on the present in the way the present depends on the past, remembering that it is nomologically *contingent* counterfactual dependence that we have in mind.

Our answer above had two parts. The first part was to observe that, at our world, every effect has many causes. The second part was that, at our world, every cause has many effects. The first part explains why, when we consider worlds where some present fact fails to obtain, there are always a variety of nomologically possible ways that the failure might have come about. The second part means that every past difference will make for many present differences so the alternative pasts will make for alternative presents, no one of which is most similar to the actual present.

Framed in this way, our answer used the language of causation. But the asymmetry here does not arise from any assumed direction of causation but from the structure of the laws themselves.

We observed above that at worlds like $\boldsymbol{w_3}$, the nomologically contingent counterfactuals track in only one direction no matter what the initial conditions.




At first, this might seem puzzling. After all, its governing law:

$$L_{XOR}: \quad \Box_N (C. \equiv. A \oplus B)$$

expresses a material equivalence between conditions at $t_1$ and $t_2$; how does this equivalence give rise to an asymmetry? The answer is that at these worlds the future is a *logically irreversible* function of the past.

Laws express functions and functions can have directions. Some functions are "reversible"[29] in the sense that, given their values, one can always infer their arguments. Negation is the paradigm example. Some functions are only partially reversible: one can infer their arguments for some values but not others; conjunction and disjunction are examples. Exclusive disjunction, as exemplified in $L_{XOR}$, is irreversible for all values. Among mathematical functions some, like mathematical negation, are reversible; others, like multiplication and addition, are not.

The laws of nature describe some states of the world as functions of other states of the world; they tell us that if the world is thus then it will be so. But where the law is logically irreversible they cannot tell us how the world must have been in order for it to be so. It is because $L_{XOR}$ embodies an irreversible function that we cannot backtrack from the state of *C* to the initial conditions of *A* and *B*.

At the actual world, the laws that describe how physical states change over time— the dynamical laws— are logically irreversible. For example, Newton tells us that:

[29] sometimes termed "invertible"

$$\boldsymbol{v_f = v_i + at}$$

Allowing us to infer that an object with a velocity of $\boldsymbol{v_i}$ undergoing a constant acceleration $\boldsymbol{a}$ will have a velocity of $\boldsymbol{v_f}$ after time $\boldsymbol{t}$ has passed. But we cannot perform the inference in reverse. Knowing only the velocity of some object at some time $\boldsymbol{t_f}$ we cannot infer its velocity or acceleration at any earlier time even if we assume that acceleration to have been constant. *Ceteris paribus*, any object might have had a different initial velocity but been subject to the same acceleration for the same time, and worlds where $\boldsymbol{a}$ and $\boldsymbol{t}$ have the same values as they actually do are more similar to worlds where they are different. At those worlds $\mathbf{v_f}$ must be different. Accordingly, we know that if $\boldsymbol{v_i}$ were different then $\boldsymbol{v_f}$ would be different. But we cannot say that if $\boldsymbol{v_f}$ were different $\boldsymbol{v_i}$ would have to be different. For any $\boldsymbol{v_f}$ there are countless combinations of intial velocity, mass, and force which are solutions to the equation. At bottom, the source of the asymmetry is the irreversabilty of addition.

The logical irreversibility of physical laws should not be confused with the temporal reversibility of some physical processes. A process which involves a physical system changing between states $S_1$ and $S_2$ is reversible in this sense if the laws allow state $S_1$ to proceed or follow $S_2$.[30] Thus, a movie of billiard balls colliding shows the balls obeying the laws of physics whether we run the film forwards or backwards. This *physical* reversibility is consistent with the *logical* reversibility of the laws. Inverting the value of $\Delta\boldsymbol{v}$ to $\Delta\text{-}\boldsymbol{v}$ is not the same thing as inferring $\boldsymbol{v_i}$ and $\boldsymbol{a}$ from $\boldsymbol{v_f}$.

The dynamical laws of quantum mechanics—the complicated laws that describe how quantum states evolve over time—are likewise logically irreversible for the same uncomplicated reason: they involve irreversible mathematical functions.

The logical irreversibility of functions means we cannot backtrack from their values to their arguments. Where the functions are dynamical laws, mapping initial conditions onto later states, irreversibility means we cannot backtrack from present to past.

A more complex question is how to reconcile the logical irreversibility of laws like this with Determinism. Determinism, as we have been understanding it, is a temporal two-way street. It

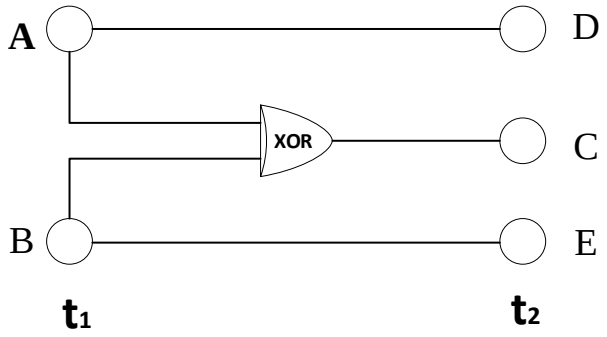
[30] It is sometimes said that, according to physics, all processes are "reversible" in this sense. This is a confusion as David Albert and Sean Carroll point out. Albert 2000, Chap.1. Carroll 2016, Chap.7.

requires that the total state of the world at all times must be deducible from the laws and a complete description of the world at any one time past or future. Worlds like $w_3$ and $w_4$ are deterministic in only one temporal direction. We can infer their futures from their pasts but not *vice versa*. This is so, of course, precisely because their simple governing law, $L_{XOR}$, is logically irreversible. By definition, any logically irreversible dynamic law will be indeterministic in one temporal direction. $w_3$ is an indeterministic world because its total states are connected by one irreversible law. But $w_3$ is a very simple world. In the real world, dynamic laws like Newton's describe relations between *partial* states of the world from moment to moment.

Is it possible for worlds with some logically irreversible laws connecting their partial states to nevertheless be overall deterministic in both temporal directions? The short answer is that it must be possible since ours is one of those worlds. A slightly longer answer goes like this.

Even where a function is irreversible in the sense that we cannot infer its inputs given its outputs, we may be able to infer *something* about its inputs from those outputs. Thus, at a world like $\mathbf{w_4}$ we can know that if *C* is true at $t_2$, then *A* and *B* must have different truth values at $\mathbf{t_1}$ even if we cannot say what those values are. So long as past facts are related to present facts by more than one one-way function—that is, so long as the past leaves many independent *traces* in the present— it is possible for the *sum* of present facts to tell us the whole story about the past. A conjunction of irreversible functions can add up to a reversible one.

The other thing to note is that not all the laws of the actual world are logically irreversible. Conservation laws are so called because they tell us what must be the same at every time. Given the right mix of laws of both sorts, it is possible for there to be worlds whose laws are *globally* reversible – that is fully deterministic – but *locally* irreversible/indeterministic. $w_5$ may be the simplest possible exemplar:




$\Box_N (A \equiv D)$

$\Box_N (C .\equiv. A \oplus B)$

$\Box_N (B \equiv E)$

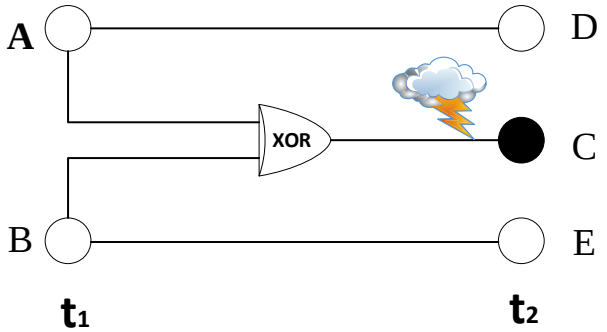
$w_5$

$w_5$ is deterministic in both temporal directions. Given a description of the total state of the world at one time we can, from its laws, deduce its condition at the other.

As we would expect, it is logically reversible laws –think of them as conservation laws — that make for nomologically true forward and backtracking counterfactuals. For example, it is true that $A > D$ and $D > A$ because it is true that $\Box_N (A \supset D)$ and $\Box_N (D \supset A)$.

On the other hand, at $w_5$ $A > C$ is true but only contingently so. $A > C$ wouldn't be true if $B$ were true. $A > C$ is a forward tracking counterfactual. What about the backtracking $C > A$? What is the world most like $w_5$ at $t_2$ where $C$ is true?

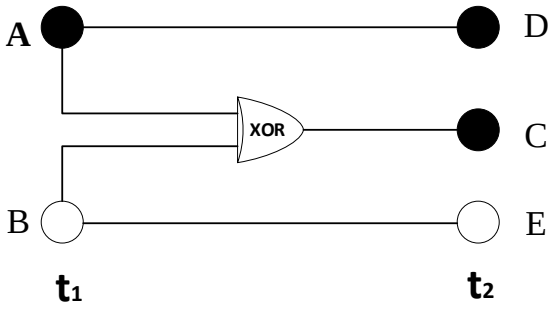
Lewis would have us imagine a miracle just before $t_2$.




$w_5$*

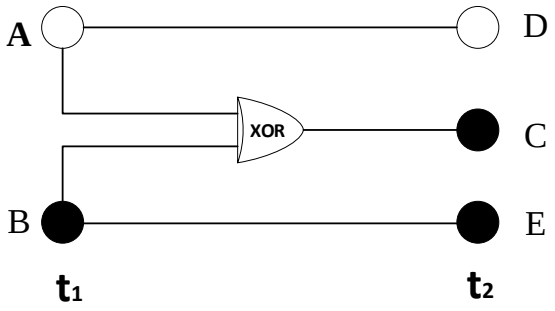
But this doesn't work because the state of the world at $t_2$ at $w_{5*}$ is not nomologically possible according to the laws that govern $w_5$. $C$'s being true is not *cotenable* with both $D$ and $E$ being false.

Bennett's theory requires worlds without miracles and says moreover that when we are comparing worlds for similarity what is relevant is similarity at the time of the antecedent. When evaluating the backtracking $C > A$ the time of the antecedent is $t_2$. There are two legal worlds where $C$ is true.




***w6***

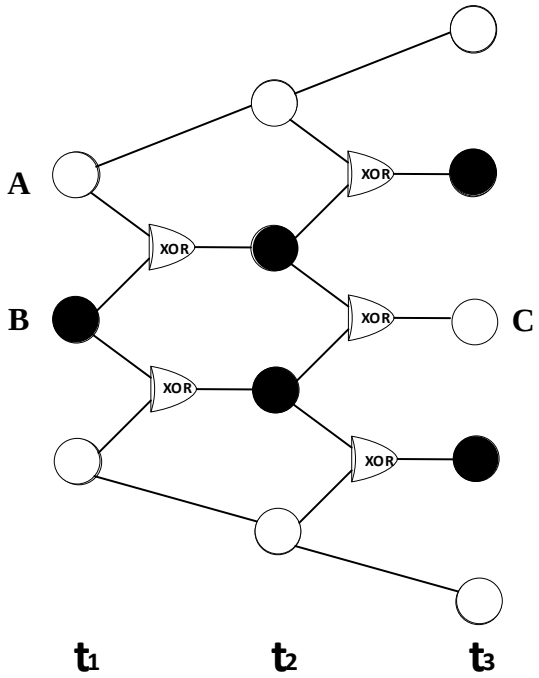
and




***w7***

But, at time $t_2$, each of $w_6$ and $w_7$ is as much like $w_5$ at $t_2$ as the other, so we have no reason to choose between them. If *C* were true then *A* might have been true as in $w_6$, but it equally well it might not, as in $w_7$. This is so both because the truth or falsehood of *C* is dependent upon multiple facts about the past and also because each of those facts has more than one upshot in the present. Notwithstanding the fact that these worlds are deterministic in both temporal directions, their futures depend upon their pasts in a way their pasts do not depend upon their future.

The same kinds of dependencies can operate over extended times. As here:



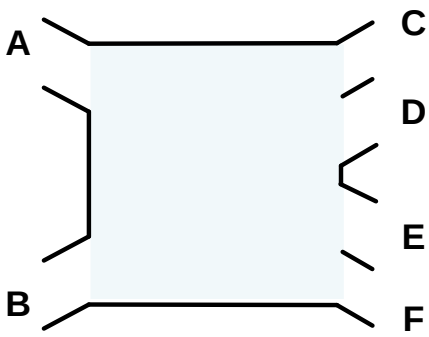

***w8***

This world too is deterministic in both temporal directions but all of its contingent counterfactual dependencies track from left to right. At ***w8*** it is true that *A* > *C* and the counterfactual is nomologically contingent. But it is not true at ***w8*** that *C* > *A* because there is no single way *C* might be true which is closer to the way things actually stand at ***w8*** at ***t3*** than other ways *C* might be true.

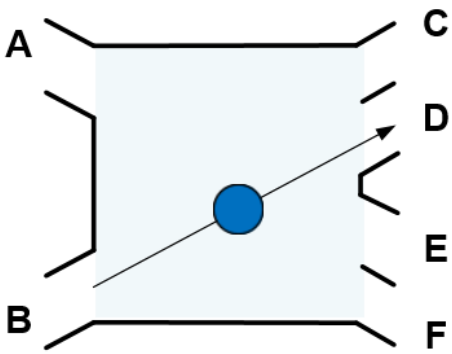
Notice something else here. At **w8** the consequences of *B*'s being true at ***t1*** extend and spread through time. At each successive time the number of features that contingently counterfactually depend on *B* increases. It is in structures like this, we suggest, that we see the explanation of Popper's radiation asymmetry.

If you are uncomfortable with circuit diagrams, here is another example of such a system with worlds built from a simple Billiard Ball computer[31].

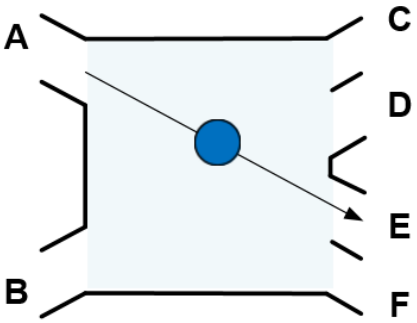
[31] https://en.wikipedia.org/wiki/Billiard-ball_computer



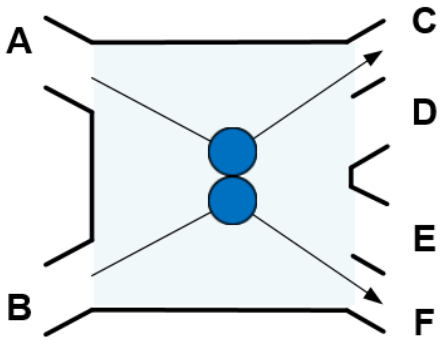

$w_{B1}$




$w_{B2}$




$w_{B3}$




$w_{B4}$

These worlds are governed by these laws:

$$\Box_N (E \equiv (A \,\&\, \sim B)$$

$$\Box_N (D \equiv (B \,\&\, \sim A)$$

$$\Box_N (C \,\&\, F \equiv (A \,\&\, B)$$

These simple worlds illustrate four points:

First: Note that all these worlds are deterministic in both temporal directions. If you know what balls went in, you know what balls came out and *vice versa*.

Second: There are contingent counterfactuals true at each of these worlds. For example, at $w_{B1}$ it is true that if $A > E$ but not at $w_{B2}$. At $w_{B3}$ it is true that if $B > C$ but this is not true at $w_{B1}$. Nevertheless, at all of these worlds the *contingent* counterfactual dependencies always track from left to right. If we ask, say, what would happen at $w_{B4}$ if $C$ were false the answer is that it might equally be that not-$A$ or that not-$B$. There are true backtrackers going the other direction at each of these worlds, but they are in every case *nomologically* true.

Third: the processes depicted are all physically reversible in the sense that we could run a film of this system running forward or backward without violating any of its laws. But..

Fourth: If we strip the arrows out of the picture and imagine the balls as traveling from right to left as readily as from left to right there is still a significant *modal* difference between the left and right-hand sides. For example: at $w_{B2}$ the forward tracking $A > C$ is true. So, reverse the film and run $w_{B2}$ in reverse. Now $A > C$ is a backtracker, but it remains true and contingent. Likewise, the forward tracking $C > A$ remains true, but not nomologically contingent. Reversing the order of these states does not reverse the direction of contingent counterfactual dependence.

In these salient respects, we offer this simple Billiard Ball world as a model of the actual universe. At the very least, it illustrates how there can be counterfactual asymmetry in a two-way deterministic world. But we think it shows much more. In future chapters, we will try to convince you that the relation we have been awkwardly calling "contingent counterfactual dependence" is, in fact, causation. The temporal direction of contingent counterfactual dependence is the direction of causation and, we will argue, of time itself.

# REFERENCES


Albert, D. Z. 2000. *Time and Chance*. Cambridge: Harvard University Press.

Bennett, J. 1974. "Counterfactuals and Possible Worlds." *The Canadian Journal of Philosophy* 4*:* 381-402.

Bowie, G. L. 1979. "The Similarity Approach to Counterfactuals: Some Problems." *Nous* 12: 477-498.

Cartwright, N. 1980. "Do the Laws of Physics state the Facts?" *Pacific Philosophical Quarterly* 61: 75–84.

Carroll, S. 2016. *From Eternity to Here*. New York: Dutton.

Creary, L. & Hill, C. 1975. "Review of Lewis." *Philosophy of Science* 42: 341-344.

Elga, A., 2000. "Statistical Mechanics and the Asymmetry of Counterfactual Dependence." *Philosophy of Science* 68 (Supplement): 313–324.

Fine, K. 1975. "Review of Lewis." *Mind* 84: 451-458.

Goodman, N., 1947. "The Problem of Counterfactual Conditionals," *Journal of Philosophy* 44: 113–128.

Jackson, F. 1977. "A Causal Theory of Counterfactuals." *Australasian Journal of Philosophy* 55: 3-21.

Lewis, D. 1979. "Counterfactual Dependence and Time's Arrow." *Noûs* 13: 455–76.

——, 1986. *Philosophical Papers: Volume II*, Oxford: Oxford University Press.

Maudlin,T. 2007. "A Modest Proposal Concerning Laws, Counterfactuals, and Explanations." In *The Metaphysics within Physics*. New York: Oxford University Press.

Noordhof, P. 2004. "Prospects for a counterfactual theory of causation." In *Cause and Chance: Causation in an Indeterministic World*, ed. P. Dowe and P. Noordhof, 188–201. Routledge: London.

Paul, L. A., & Hall, E. J. 2013. *Causation: A Users guide*. Oxford: Oxford University Press.

Price, H. 1997. *Time's Arrow & Archimedes Point*. New York: Oxford University Press.

Richards, T. 1975. "The Worlds of David Lewis." *Australian Journal of Philosophy* 53:105-118.

Schlossberger, E. 1978. "Similarity and Counterfactuals." *Analysis* 38 1978: 3-27

Slote, M. 1978. "Time in Counterfactuals." *Philosophical Review* 87: 3–27.

Tichy, P. 1976. "A Counterexample to the Stalnaker-Lewis Analysis of Counterfactuals." *Philosophical Studies* 29: 271-27.

Tomkow,T. 2013. *The Simple Theory of Counterfactuals*. http://tomkow.typepad.com/tomkowcom/2013/07/the-simple-theory-of-counterfactuals.html

Tomkow,T., & Vihvelin K. 2016. *Determinism*. http://vihvelin.typepad.com/vihvelincom/2016/07/determinism.html.

Tomkow,T., & Vihvelin K. 2015. *Counterfactuals: The Short Course*. http://vihvelin.typepad.com/vihvelincom/2015/08/counterfactuals-the-short-course.html

Tooley, M. 2003. "The Stalnaker-Lewis approach to counterfactuals." *The Journal of Philosophy* 100: 371–77.