philosophies

MDPI

Article

Back to the Present: How Not to Use Counterfactuals to Explain Causal Asymmetry

Alison Fernandes

Department of Philosophy, Trinity College Dublin, D02 PN40 Dublin, Ireland; asfernan@tcd.ie

Abstract: A plausible thought is that we should evaluate counterfactuals in the actual world by holding the present 'fixed'; the state of the counterfactual world at the time of the antecedent, outside the area of the antecedent, is required to match that of the actual world. When used to evaluate counterfactuals in the actual world, this requirement may produce reasonable results. However, the requirement is deeply problematic when used in the context of explaining causal asymmetry (why causes come before their effects). The requirement plays a crucial role in certain statistical mechanical explanations of the temporal asymmetry of causation. I will use a case of backwards time travel to show how the requirement enforces certain features of counterfactual structure *a priori*. For this reason, the requirement cannot be part of a completely general method of evaluating counterfactuals. More importantly, the way the requirement enforces features of counterfactual structure prevents counterfactual structure being derived from more fundamental physical structure—as explanations of causal asymmetry demand. Therefore, the requirement cannot be used when explaining causal asymmetry. To explain causal asymmetry, we need more temporally neutral methods for evaluating counterfactuals—those that produce the right results in cases involving backwards time travel, as well as in the actual world.

Keywords: causal asymmetry; time travel; counterfactuals; present; loewer

check for updates

Citation: Fernandes, A. Back to the Present: How Not to Use Counterfactuals to Explain Causal Asymmetry. *Philosophies* **2022**, *7*, 43. https://doi.org/10.3390/philosophies7020043

Academic Editor: Alasdair Richmond

Received: 18 February 2022
Accepted: 28 March 2022
Published: 9 April 2022

1. Introduction

A plausible thought is that we should evaluate counterfactuals in the actual world by holding the present 'fixed'. More precisely, some methods require the state of the counterfactual world at the time of the antecedent to match that of the actual world, outside the spatial area of the antecedent, when evaluating counterfactuals in the actual world—where this requirement is explicitly part of the recipe for evaluating counterfactuals [1–3]. There might also be some changes to the present required to satisfy the content of the antecedent, perhaps implied by context [3] (p. 26). Being in Uganda may, in ordinary contexts, imply that you are not in Spain. However, other than changes within the spatial area of the antecedent or directly required to satisfy the antecedent, the state of the universe at the time of the antecedent remains unchanged.

For example, say you wonder what would be the case were you to frolic on the university lawns at midnight (given that you do not in the actual world). According to 'altered states recipe' approaches [1–3], the relevant nearby counterfactual world that determines what counterfactuals are true is one in which either your bodily movements or location at midnight are *different* from what they are in the actual world. Your frolicking on the lawns at midnight may imply that you are not resting on the lawn or not in the lounge. However, the states, at midnight, of the lawns, the university buildings, the security guards, and the entire rest of the universe outside the spatial areas you occupy in the counterfactual world (and in the actual world) are *required* to be the same as what they are in the actual world. While future states outside the spatial area of the antecedent (and that do not directly concern the antecedent) may differ, depending on the rest of the machinery for evaluating counterfactuals, present states may not. Thus, any counterfactual of the form 'If you were to frolic on the university lawns at midnight, state *x* at midnight would obtain' (where *x* is different from what is the case in the actual world and does not occupy the spatial area of the antecedent or directly concern the content of the antecedent) is straightforwardly false.

Call an explicit requirement of this kind 'holding the distant present fixed'. This requirement looks reasonable. After all, evaluating counterfactuals seems to be about making minimal changes to actuality, and holding the distant present fixed seems to ensure minimal changes. However, while the requirement may produce reasonable results when evaluating counterfactuals in the actual world, it produces the wrong results in cases involving backwards time travel. The requirement implies strange counterfactual worlds, where there are changes to the past, but they are always 'put back' by the time of the present, by whatever means necessary. Assuming backwards time travel is possible, and such consequences unreasonable, holding the distant present fixed cannot be part of a completely general method of evaluating counterfactuals.

This result might seem trivial or uninteresting. After all, many approaches to evaluating counterfactuals were not designed for cases of backwards time travel. Indeed, defenders are sometimes explicit that their methods are not intended to cover cases involving backwards causation ([1], pp. 10–12; [2], p. 8) or only aim to recover counterfactuals true in the actual world [3] (pp. 21, 31). Thus, it is no wonder they produce odd results and no mark against them. Indeed, as far as I am aware, no one has even explicitly defended the idea that we should evaluate counterfactuals in backwards time travel scenarios by holding the distant present fixed. So, why does it matter if such approaches fail in these settings? Moreover, the project of determining the right way to evaluate counterfactuals in cases of backwards time travel might seem irrelevant to how we evaluate counterfactuals in the actual world. What could we possibly learn about evaluating counterfactuals in the actual world by considering time travel scenarios?

However, it turns out there is a project that relies on using a general method of evaluating counterfactuals—one that works in both backwards time travel scenarios and in the actual world. This is the project of using counterfactuals to explain temporal asymmetries in the actual world, particularly the temporal asymmetry of causation—why causes come before their effects at our world, hereafter 'causal asymmetry'. What motivates these accounts is less an *a priori* commitment to reducing causation, but the need to reconcile the temporal asymmetry of causation with temporal symmetry in the fundamental physical laws—see [4,5]. These accounts aim to trace causal asymmetry back to more fundamental physical asymmetries, using counterfactuals as a half-way step.

As I will argue (Section 3), the project of explaining causal asymmetry requires a temporally neutral method of evaluating counterfactuals—one that does not, *a priori*, enforce temporal features of counterfactual structure, but instead allows that structure to be derived from contingent physical structure in the universe. However, in attempting to explain causal asymmetry, certain statistical mechanical accounts use methods of evaluating counterfactuals that hold the distant present fixed and rely crucially on this requirement [6–8]. However, because the requirement to hold the distant present fixed enforces features of the counterfactual structure *a priori*, it prevents these features of counterfactual structure being derived from physical structure. More particularly, the requirement to hold the distant present fixed presumes there is no counterfactual dependence of the distant present. This presumption not only produces unreasonable results in cases where we expect such dependence, including backwards time travel cases, but it prevents what counterfactual dependencies there are from being derived from physical structure, in the way explanations of causal asymmetry demand. Therefore, the purported explanations do not adequately trace causal asymmetry back to more fundamental physical asymmetries, and so do not explain causal asymmetry.

Altogether, my target here is not 'altered states recipe' approaches that hold the distant present fixed when evaluating counterfactuals in the actual world. My target is those that have adopted methods that hold the distant present fixed when explaining causal asymmetry in the actual world.

The paper proceeds as follows. In Section 2, I argue that the requirement to hold the distant present fixed implies unreasonable results in a scenario involving backwards time travel. This outcome may seem obvious, given that the requirement was not designed to be used in cases of backwards time travel. However, understanding why the requirement produces unreasonable results in time travel scenarios allows us to see why the requirement is illicit when used to explain causal asymmetry. In Section 3, I argue that the failure of the requirement in time travel scenarios not only shows that the requirement cannot be part of a completely general method of evaluating counterfactuals, but more importantly, the failure of the requirement compromises certain statistical mechanical explanations of causal asymmetry.

One may have other reasons for rejecting the requirement to hold the present fixed, such as wanting to allow for simultaneous direct causation ([1], p. 9; [9], p. 426), backtracking counterfactuals [10] (p. 340), or other contextual requirements concerning 'how the change is to be effected' [3] (p. 26). The time travel counterexample I give does not rely on these concerns. I adopt a broadly Lewisian approach to time travel [11]: I assume that backwards time travel is possible and requires backwards causation. While the counterexample could be presented using merely backwards causation (without backwards time travel), time travel allows the point to be put particularly vividly—particularly when agents and decision-making are introduced (Section 3). For defences of the possibility of backwards time travel, see [11–13]. For other attempts to use backwards time travel scenarios to argue against certain methods of evaluating counterfactuals and chances, see ([14], chapter 7; [15–18]. I adopt a broadly B-theoretic view of time, such that talk of the past, present, and future is to be treated indexically—these are times before, simultaneous with, or after a given reference frame (usually the time of the counterfactual antecedent).

2. Time-Travelling Tina

Here is a case involving backwards time travel. I will use the case to demonstrate how evaluating counterfactuals by holding the present fixed enforces features of counterfactual structure, leading to unreasonable results in cases of backwards time travel. In the next section, I use this examination to argue that the requirement cannot be part of a suitably general method of evaluating counterfactuals that could be used to explain causal asymmetry.

Tina at the Beach: Tina is working from home. It is the end of a hot day, and, while she is sorry to have stayed in, she comforts herself with the following thought: in just a moment she will jump into her time machine, travel back in time to the start of the day, and spend the same day at the beach. "No doubt I'm there right now," she thinks, "feeling the warm sun and hearing the waves lapping". Tina is deliberating, however, about whether to shave her beard before going. [1] Assume Tina has no knowledge of her state at the beach and that her state at the beach is not a cause or causal condition of the decision she makes now—Tina's decision and state at the beach are not parts of a causal loop.[2]

Stipulate that Tina's time machine works, so that she will travel in time (and space) from home (at time t_2) to the start of the day at the beach (time t_1) by reliable nomic and causal means, as represented in Figure 1. In Figure 1, 'Tina-at-home', 'Tina-arrived-at-beach', and 'Tina-tanned' refer to Tina at different spatiotemporal locations (assuming no particular metaphysics of persistence). Tina-at-home and Tina-tanned occupy different spatial locations, but the same temporal location. Tina's time travel may occur via curves in spacetime or other physical means. For simplicity, assume that Tina does not age significantly when she travels.

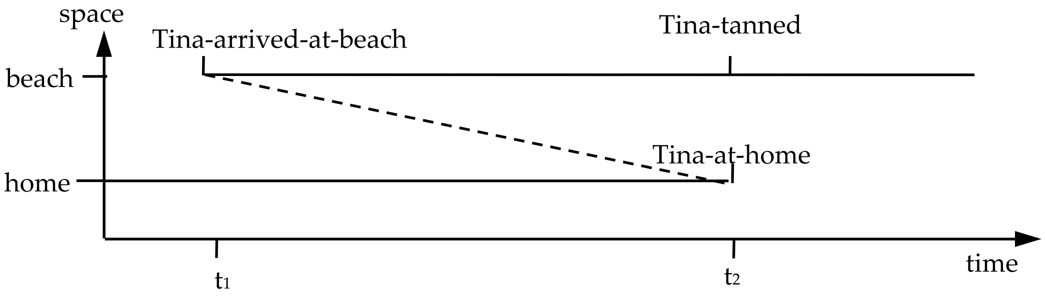

Figure 1. Tina's spatial location as a function of time. The dashed line indicates a form of time travel (whether continuous or discontinuous).

Consider the following counterfactuals:

1. If Tina-at-home were to shave, Tina-arrived-at-beach would be beardless.
2. If Tina-at-home were not to shave, Tina-arrived-at-beach would be bearded.

I claim these counterfactuals are true. First, they capture ordinary lawful and causal behaviour, where shaving at one time leads to be being beardless at later times, in a temporal order that aligns with causal order in the vicinity of the agent. There is no reason to expect violations of this ordinary behaviour, such as to avoid self-defeating causal loops—Tina's state at the beach is not a cause or causal condition of the decision she makes now. Second, 1 and 2 are the right counterfactuals for decision-making. Tina should decide whether to shave partly in light of whether she'd prefer to be bearded or beardless at the beach. Third, according to counterfactual accounts of causation, there must be counterfactual dependence of some of Tina's temporally earlier states on some of Tina's temporally later states for this to be a case of backwards time travel [11]. Thus, some counterfactuals of the form of 1 and 2 must be true.

Consider next the following counterfactuals:

3. If Tina-at-home were to shave, Tina-tanned would be beardless.
4. If Tina-at-home were not to shave, Tina-tanned would be bearded.

I claim these counterfactuals are also true. First, they capture ordinary lawful and causal behaviour, where shaving at one time leads to be being beardless at later times, in a temporal order that aligns with causal order in the vicinity of the agent. For the same reasons as above, there is no reason to expect violations to avoid self-defeating causal loops. Second, they are the right counterfactuals for decision-making: Tina should decide whether to shave partly in light of whether she would prefer to be bearded or not at the beach. Third, while there need not strictly be counterfactual dependence of Tina-tanned on Tina-at-home for this to be a case of time travel, there does need to be at least a chain of counterfactual dependencies linking Tina-at-home to Tina-tanned, under standard counterfactual accounts of causation [24]. While transitivity of counterfactual dependence is not guaranteed, there is no reason to expect violations due to features such as pre-emption.

However, counterfactuals 3 and 4 cannot both be true if counterfactuals are evaluated by holding the distant present fixed. Here is why. Say in the actual world Tina shaves and is therefore beardless at the beach. Under standard semantics [25], counterfactual 3 is then straightforwardly true. However, given the requirement to hold the distant present fixed, counterfactual 4 is false. Tina-tanned must be beardless, since Tina-tanned's state is in the distant present of Tina-at home. So, if Tina-at-home were not to shave, Tina-tanned would still be beardless. Mutatis mutandi if, in the actual world, Tina does not shave: counterfactual 4 is true, but counterfactual 3 is false. Because counterfactuals 3 and 4 cannot both be true, whether Tina-tanned is bearded cannot counterfactually depend on whether Tina-at-home shaves. More generally, if the distant present is held fixed when evaluating counterfactuals, there can be no counterfactual dependence of a time traveller's distant present states on her states now.

Any counterfactual changes to the past that do occur, moreover, must always be 'put back' by the time of the present, by whatever means necessary. For example, say Tina does not shave in the actual world. If so, the following counterfactual is true:

5. If Tina-at-home were to shave, Tina-arrived-at-beach would be beardless, but her beard would grow back, reattach, or otherwise return by later that day.

While this behaviour may not be inconsistent with fundamental physical laws, it is nevertheless inconsistent with ordinary macroscopic behaviour, and it cannot be explained by the need to avoid self-defeating causal loops, nor can the implications of counterfactual 5 be contained by Lewis' [26] distinction between what an agent causes, and what would merely be true. Because of the requirements of causal continuity in time travel, Tina can cause the seemingly miraculous behaviour of her beard.

One might argue that at least one of counterfactuals 3 or 4 must be false precisely because the area outside the antecedent should be held fixed when evaluating counterfactuals.[3] If so, take Tina's case to be a way of drawing out the surprising consequences of this view—it implies violations of ordinary lawful and causal behaviour, and failures of transitivity of counterfactual dependence. The view may also imply that Tina's deliberation is out of place, since her state at the beach does not depend counterfactually on her decision now. Tina's case therefore represents a choice point—either accept the strange consequences or reject the requirement when evaluating counterfactuals in cases involving backwards time travel.

While these results are not conclusive, I will assume, for the moment, that one should not accept these strange consequences. Therefore, the requirement to hold the present fixed cannot be part of a completely general method of evaluating counterfactuals, one that delivers appropriate results for all causal structures. As I noted in Section 1, this result might not seem too surprising, since many of those who defend the requirement to hold the distant present fixed only use the requirement when evaluating counterfactuals in the actual world [1–3]. The deeper point that Tina's case demonstrates, however, is that holding the distant present fixed enforces features of counterfactual structure. The requirement rules out simultaneous counterfactual dependence *a priori* and independently of what the rest of the counterfactual and causal structure of Tina's world is like. As I will argue in the next section, this deeper point implies that holding the distant present fixed cannot be used when evaluating counterfactuals on the way to explaining causal asymmetry—contrary to certain statistical mechanical accounts.

3. Upshots for Explaining Causal Asymmetry

The first, less interesting, upshot of Tina's case is that the requirement to hold the present fixed cannot be part of a completely general method of evaluating counterfactuals. The lack of such a method will not matter for some projects. For example, altered states recipe approaches standardly hold the distant present fixed but defenders of these approaches typically aim to merely recover counterfactuals that are true of the actual world [3] (pp. 21, 31), and are sometimes explicit that their methods are not intended to cover cases involving backwards causation ([1], p. 10; [2], p. 8). Defenders can respond by simply limiting the scope of their accounts. Defenders might also respond to Tina's case by altering the requirement, such that the present is only held fixed by default, and introducing additional causal stipulations. For example, the distant present might be held fixed except for states that are in the causal future of the antecedent. See [27,28] for related proposals.[4]

The second, by far more important, upshot of Tina's case concerns how we explain causal asymmetry in the actual world. Certain statistical mechanical accounts ([6], Ch. 6; [7]; [8], pp. 234–236; [29]) use counterfactuals evaluated by holding the present fixed to explain causal asymmetry. These accounts cannot make use of the first response above. They aim to use a method of evaluating counterfactuals that, combined with the physical structure of a given world, delivers that world's causal structure. Defenders cannot assume, without explanation, that there is no backwards causation in a given case prior to determining what counterfactual method to apply. It is the counterfactual method itself that is supposed to deliver the fact that there is no backwards causation. Because these accounts aim to explain causal structure using only non-causal features of reality, they also cannot take the second response above and include causal stipulations in the method. Moreover, as we will see, these accounts only succeed if the whole distant present is held fixed when evaluating counterfactuals. It is not enough if the distant present is held fixed by default.

The underlying concern for counterfactual-based explanations of causal asymmetry that use the requirement is that holding the distant present fixed cannot be part of a suitably temporally neutral method of evaluating counterfactuals—one that can explain causal asymmetry. To use counterfactuals to explain causal asymmetry, it is essential that the method not illicitly 'build in' features of counterfactual structure. It is a familiar point that one cannot evaluate counterfactuals by holding the past fixed when explaining causal asymmetry [24]—such a method is illicit because it rules out counterfactual dependence of the past on the present (ruling out backwards causation), independently of what the underlying physical structure of a world is like. It makes explaining causal asymmetry 'too easy'. While the requirement to hold the distant present fixed is not temporally asymmetric, it does similar harm—it enforces features of the counterfactual structure and prevents its being derived from the underlying physical structure.

Tina's case demonstrates how this occurs. Even though the causal structure of Tina's world suggests there should be interesting forms of simultaneous counterfactual dependence, relevant to her decision-making, the requirement to hold the distant present fixed rules out these dependencies *a priori* and enforces a different counterfactual structure that excludes simultaneous counterfactual dependencies—independently of the physical structure of her world. The problem raised by Tina's case is not merely that the results of the requirement are unintuitive, but that the requirement rules out simultaneous counterfactual dependence *a priori*, preventing features of counterfactual structure being explained in physical terms.

In the remainder of the paper, I will use Tina's case to argue more directly against the above statistical mechanical explanations of causal asymmetry. While, for simplicity, I will focus on Loewer's account [7], I aim to undermine what has seemed to be a promising general approach to explaining causal asymmetry.[5]

Loewer adopts a statistical mechanical method for evaluating counterfactuals. The method is used to evaluate counterfactuals where the antecedents are decisions of agents, rather than events or states more generally, and where the consequents are typically probabilities of macrostates, rather than events or states more generally. Macrostates are states of systems characterised using macroscopic language—'a mug of water at boiling point' specifies a macrostate, whereas a description of the location and velocity of all the water molecules specifies a microstate.

Here is Loewer's method for evaluating counterfactuals, put roughly. To evaluate what would be the case, were an agent to decide other than they do in the actual world, consider a partially specified counterfactual world that has the same macroscopic state as the actual world at the time of the decision, t, that started out in the same macroscopic state as the actual world, that has the same fundamental dynamical laws as the actual world, but where the microscopic state of the agent's brain at t is different—the microscopic state is such that the agent decides **D1**. Apply a statistical postulate that, roughly put, takes complete microscopically specified counterfactual worlds compatible with the above partial characterisation to be each equally probable. The probability distribution over microscopically specified counterfactual worlds implies probabilities of macrostates at any given time—which are the counterfactual consequents.

For example, to determine what would happen, were you to decide to frolic on the lawns (given that you do not in the actual world), consider a partially specified world that has you deciding, at t, to frolic on the lawns, but that has the same macrostate as the actual world at t, the same fundamental dynamical laws, and that started out in the same macrostate. If, in most of the microscopically specified counterfactual worlds consistent with this partial specification, you do frolic on the lawns at t1, then the counterfactual 'If at t you were to decide to frolic on the lawns, the probability of your frolicking on the lawns would be high' is true.

More precisely, on Loewer's method, the conditional 'If at t I were to decide **D1** then the probability of **B** would be *x*' is true just in case Pr(**B**/**M**(t)&**D**(t)) = *x*, where **B** is a macroscopic event, **M**(t) is the macroscopic state of the world at t, **D** is the decision at t (compatible with the macrostate **M**(t)), and Pr is a chance function evaluated using statistical mechanical probabilities.[6] For example, were a particular decision **D1** to be made at t, an event **B** would occur with high probability, just in case **B**'s probability is high, given **D1**(t)**,** and given the macrostate of the actual world at the time of **D** remains as it is in the actual world (that is, given **M**(t)).

In both the rough and precise formulation, because the probabilities used to derive counterfactuals involve conditionalising over the entire macroscopic state of the world at the time of the antecedent, that is, reasoning to a partially specified world that has the same macrostate as the actual world, the macroscopic state of the distant present is 'held fixed'.[7]

Using this method, Loewer derives a temporal asymmetry in what decision-counterfactuals are true. Assume (a) that the present contains vastly more 'macro signatures' of the past than the future. Macro signatures are local states that render other local states highly probable, given statistical mechanical probabilities [7] (p. 318)—they are often referred to in the literature as 'records' [6] (chapter 6). Records include states such as memories, recordings, fossils, etc. If we evaluate counterfactuals by holding the present macrostate fixed, any macro signatures of the actual past contained in the present also remain the same in counterfactual worlds. Because counterfactuals and macro signatures are both evaluated using the same statistical mechanical probabilities, the events they are macro signatures of in the past will also remain unchanged in counterfactual worlds. Therefore, Loewer argues, there will be no counterfactual dependence of past events recorded in the present on present decisions. Any such counterfactual dependence is ruled out by the macro signatures contained in the present.

Assume also (b) that, given our biology, small changes in our brain state can in general be probabilistically correlated with large changes elsewhere. Taking (a) and (b) together, Loewer argues there can often be significant counterfactual dependence of the future on present decisions, but that past states will not depend counterfactually on present decisions. If so, there is a temporal asymmetry in what decision counterfactuals are true—one that might explain a more general causal asymmetry ([7] (p. 297); [31]).

There are various objections one might raise to Loewer's account.[8] My point for the moment is that its success requires holding the whole distant present fixed. Here is why. Assumption (b) allows that small changes in brain states can, in general, be probabilistically correlated with macroscopic changes in the world at the same time. Say we allow that some parts of the macroscopic present outside the antecedent may be different in counterfactual worlds. Then, by (b), small changes in brain states can be counterfactually correlated with changes in the present—including changes to the macro signatures contained in the present. Because the macro signatures contained in the present will sometimes be different in counterfactual worlds, the events they record in the past will be different as well. Thus, even when there are macro signatures of the past state contained in the present, there can still be significant dependence of past events on present decisions. Moreover, given correlations between decisions and past states are macroscopic, they are the kind that we could come to know about, therefore they cannot be ruled out as unknowable ([7], p. 318; [29] p. 127). Without the requirement to hold the whole distant present fixed, Loewer's account does not begin to rule out backwards causation. The requirement to hold the distant present fixed thus plays a crucial role in Loewer's explanation of causal asymmetry.

However, as we saw with Tina's case, holding the distant present fixed enforces features of counterfactual structure, independently of the physical structure of a world. Thus, it cannot be used when deriving what the counterfactual structure of a given world is like from its underlying physical structure. One might allow Loewer to use the requirement when deriving further features of counterfactual structure, if he provided some physical justification for the requirement—some explanation of why it accurately reflects the pre-existing counterfactual structure of a given world. At no point, however, does Loewer offer a physical justification for the requirement. There is no attempt to explain, using physical features, why there is no counterfactual dependence of the distant present at our world. Because Loewer assumes, *a priori* and without physical justification, that there is no simultaneous counterfactual dependence at our world, his purported explanation fails to adequately trace causal asymmetry back to physical features of our world. At a crucial point, Loewer's account assumes, rather than explains, what is required to derive causal asymmetry.

One might be tempted to respond that, insofar as holding the distant present fixed is part of a reasonable method of evaluating counterfactuals at our world, there can be no harm employing it when explaining causal asymmetry at our world. To be clear, I am not objecting to the method as a reasonable method of evaluating counterfactuals in our world. Indeed, the method may be reasonable in worlds in which all causal relations are aligned in the same temporal direction. However, its use in explaining causal asymmetry is a distinct issue. In an explanatory context where we are trying to derive causal structure from non-causal physical structure via counterfactuals, we cannot assume counterfactual structure that is in no way derived from physical structure.

One might instead attempt to justify the requirement to hold the present fixed and explain why it is part of a reasonable method of evaluating counterfactuals at our world. If this could be done in non-causal terms, then perhaps the requirement could still be used when giving a non-causal explanation of causal asymmetry, but the prospects do not look promising. The reason why the requirement seems to produce reasonable results in our world is because causal relations in our world are always temporally aligned. For this reason, there are no combinations of backwards and forwards causal influence that would produce simultaneous counterfactual dependence of the kind found in Tina's case. However, if we could explain why all causal relations were temporally aligned (in non-causal terms) we would have done most of the work in explaining causal asymmetry. Indeed, as we will see below, Loewer's account comes close to simply assuming there are at least some forwards causal relations. This assumption, combined with the claim that all causal relations are temporally aligned, would be enough to explain causal asymmetry without making use of the requirement to hold the present fixed or Loewer's macro signature-based explanation above.

Can Loewer justify holding the present fixed in other terms? He might be thought to do so indirectly, via an assumption about decisions [7] (p. 317):

> given the macrostate **M** of the world (including the agent's brain) the various decisions that are available to her are all equally likely. Decisions are thus indeterministic relative to the macro state of the brain and environment prior to, and at the moment of, making the decision. This indeterminacy captures the idea that which decision one makes is 'open' prior to making the decision.

While Loewer acknowledges that the assumption that 'each possible decision is equally likely is certainly false' he '[doesn't] think this simplification affects the account' ([7], p. 317, n. 39). The crucial assumption, however, is not that decisions are equally likely, but that no single decision is highly probable, given local states in the present or past. Loewer assumes, in other words, that there cannot be macro signatures of available decisions in the distant present or past. He assumes there cannot be simultaneous or previous states that reflect what decision is made in the present. If so, there cannot be cases, like Tina's, where her decision (to shave or not) is reflected in her state at the beach (beardless or not) and where holding fixed the distant present (such as whether she is beardless) leads to strange results—where changes to her beard in the past must always be 'put back' by the present. Moreover, a lack of significant correlations between available decisions and macro states in the distant present might be thought to justify holding the distant present fixed.

However, Loewer's assumption about available decisions is deeply problematic. First, as others have noted [32], the assumption is temporally asymmetric. Loewer assumes that small changes in our brain states can (in general) be probabilistically correlated with large changes elsewhere (assumption b), above). However, he also assumes that our available decisions are not probabilistically correlated with past states. If so, assuming our available decisions are probabilistically correlated with any states at other times, they will be probabilistically correlated with future states—and we have assumed that there are at least some forwards causal relations.[9] However, since Loewer assumes available decisions are not probabilistically correlated with past states, and such correlations are a requirement for counterfactual dependence, Loewer rules out backwards causal relations that are as direct as the forwards causal relations. A further problem is that if we justify the requirement to hold the present fixed by assuming all causal relations are temporally aligned (as above), the assumption that there are some forwards directed causal relations allows one to derive that all causal relations are forwards directed—independently of the counterfactual and probabilistic structure of the world.

A second problem with Loewer's assumption is it is false of decisions we make in the actual world. Provided we sometimes reliably make decisions in response to particular macroscopic events, there can be macro signatures of our decisions in the past and distant present. For example, my decision to play certain piano keys may be a macro signature of what notes I have already played [30] (p. 31), and so of what notes a sound recording in the present contains. Loewer could retreat to the position that the assumption about available decisions is merely a 'fiction' [7] (p. 317) or 'myth' [29] (p. 127). While this response deals with the second problem, it does not address the first, or a third.

The third problem is that the assumption is unreasonable, even as a fiction. The assumption implies that agents like Tina do not even fictionally count as having available decisions, merely because they take there to be states in the distant present that are macro signatures of their decisions. However, given Tina herself has no records now of her state at the beach, there is nothing in her knowledge of the past to prevent her employing the fiction. Moreover, in Tina's case, there are macro signatures in the distant present of her decisions precisely because she can control the past. Having control of the past should not rule her out as having available decisions in the fiction, and believing she has control of the past should not prevent her employing the fiction.

I suspect Loewer's assumption might have looked reasonable because we confuse direct and indirect control. It may be true that our direct control of the present is limited to a small local area, such as our brain states, but this does not imply that our indirect control of the present is similarly limited. To assume we cannot indirectly control the present is to assume something about the causal and counterfactual structure of the entire rest of the world—namely that we cannot control the distant present using the past. However, as I have argued, we are not entitled to this as an assumption when explaining causal asymmetry. While it may be true that our indirect control in the actual world is limited to the future, this cannot be simply assumed when explaining why this is so.[10]

At this point, one might be tempted to give up on the project of explaining causal asymmetry. If one does not use counterfactuals to derive causal structure from non-causal structure, one can adopt a causal method of evaluating counterfactuals ([3,13,27,28,34]). Causal approaches could be used to derive counterfactuals in the actual world. They could also be used to deliver the intuitive results in time travel cases like Tina's, provided they didn't hold events in the causal future of the antecedent 'fixed'.[11] However, adopting only a causal approach to evaluating counterfactuals would mean giving up on the project of explaining causal asymmetry in counterfactual terms. For reasons explored [4–6] and elsewhere, this would be to abandon an otherwise promising approach to explaining causal asymmetry in scientific terms and unifying a range of temporally asymmetric phenomena.[12]

Moreover, there are alternatives. To explain an asymmetry of decision counterfactuals in the actual world, what we need to do is rule out counterfactual dependence in cases where an agent's decision is (or is taken to be) evidence of, or probabilistically correlated with, the states she is responding to—cases like the piano player above. We do not need to rule out counterfactual dependencies in other kinds of cases. Ruling out counterfactual dependence concerning an agent's responses is the approach taken by [9,35–37]. According to these broadly physicalist agent-based approaches, the evidence the agent has while deliberating prevents their decision being a means to raise the probability of states they are responding to. For this reason, the piano player's past playing does not depend counterfactually on their decision now—or at least not in a way that would amount to the agent's decision influencing, controlling, or causing their past playing.

These response-focused accounts rule out backwards counterfactual dependence in the piano player case and provide an alternative route to explaining causal asymmetry in the actual world, but they do not vindicate a general requirement to hold the distant present fixed when evaluating counterfactuals. They allow for simultaneous counterfactual dependence in Tina's case. Precisely because they do not hold the distant present fixed, they do better when applied to worlds with complex causal structures. These methods work because they are sensitive to local probabilistic and counterfactual dependencies, and do not employ global requirements such as holding the present fixed. While these approaches still face difficult choices concerning precisely what to hold fixed, particularly in cases involving causal loops (see [18] for discussion), the local nature of these approaches makes them better candidates for explaining, in physical terms, why our world has the causal structure that it does.

There are no doubt other alternatives to explore. Regardless, the lesson of Tina's case is that if we are to explain even a global causal asymmetry in our world in physical terms, the method used cannot employ the global requirements to 'hold the distant present fixed'. Such a requirement presumes features of counterfactual structure and prevents their being explained in physical terms. A promising alternative is to use methods that are more sensitive to local structure, including those that only rule out dependencies in case where an agent's decisions are responses to events.

Funding: This research received no external funding.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The author declares no conflict of interest.

Notes

[1] The case will not rely on gender-swapping or anything like that.

[2] Plausibly, her having no knowledge of her state at the beach is a requirement on her reasonably deliberating about her decision [19]. One might argue that causal loops involving the agent's decisions are unavoidable in cases of backwards time travel, even if Tina travels into the far distant present. If Tina's case involves a causal loop, it is perhaps less surprising that methods of evaluating counterfactuals fail. See [11,13,20–23] for discussion for some of the difficulties evaluating counterfactuals in cases involving causal loops. My concerns with holding the distant present fixed are unrelated to causal loops.

[3] Ideally, one would also want some independent motivation for the requirement. The standard Lewisian motivation [24] is to recover our intuitive judgements, but that is no help when the intuitions are in question or favour an alternative.

[4] While [3] (pp. 30–31) might be thought to attempt a non-causal solution, his solution is temporally asymmetric and has a causal flavour, particularly regarding his talk of 'infection'.

[5] Loewer [29] uses probabilities in place of counterfactuals, but uses the same requirements for how each are evaluated and takes counterfactual structure to derive from probabilistic structure [29] (p. 132). Albert's original account [6] (chapter 6) is ambiguous but is often interpreted as holding the distant present fixed [30] (p. 27). Albert confirms (private communication) that this is what he had in mind. Kutach [8] (pp. 234–236) uses the requirement to explain why we cannot influence the past by means of our forwards influence. While Kutach accepts that this asymmetry of influence may not hold in time travel scenarios [8] (p. 229), he does not take the requirement itself to be problematic when explaining temporal asymmetries.

[6] Statistical mechanical probabilities are derived from taking the Lebesgue probability measure over microstates compatible with the low-entropy macrostate of the early universe—the 'Past Hypothesis'—and conditionalising over later macrostates [7] (p. 317).

[7] Whether the macrostate [7] or the microstate [29] outside the antecedent is held fixed will not matter to my arguments. If holding the macrostate fixed is problematic, as Tina's case suggest, then holding the microstate fixed is also problematic.

[8] What if there are no macro signatures of a past (or future) event contained in the present? Loewer [7] (p. 318) responds that, in that case, the past (or future) event will not probabilistically depend on the present decision. Loewer's explanation of the asymmetry fails, however, if the decision is the only record of the past event in the present [30]. I discuss this kind of case below.

[9] At least on a standard Lewisian counterfactual account of causation [24]. Again, Loewer is not explicit about the precise relation between counterfactuals and causal relations.

[10] Could Tina's case be ruled out because it implies violations of thermodynamic asymmetries? It is controversial whether time travel (along time-like curves) implies such violations [33] (p. 137), but, to make this response sufficiently general, one would need to argue that backwards causation implies violations of thermodynamic asymmetries, and it is precisely to be shown, not assumed, that the direction of causation is to be explained in statistical mechanical or thermodynamic terms when giving statistical mechanical explanations of causal asymmetry.

[11] Causal methods of evaluating counterfactuals face difficulties dealing with causal structures such as causal loops—see [11,13,20–23] for discussion. Note that the standard ways of dealing with counterfactuals in cases of causal loops will not help Loewer either—standard accounts presume either temporal asymmetry [11,21–23] or are causal [13,20]. See [18] for discussion.

[12] A similar point holds for accounts that use probabilities rather than counterfactuals to explain causal asymmetry [29]. Causal methods of evaluating probabilities cannot be used if the project is to explain causal asymmetry using probabilities.

References

1. Collins, J.; Hall, N.; Paul, L.A. Counterfactuals and Causation: History, Problems, and Prospects. In *Causation and Counterfactuals*; MIT Press: Cambridge, MA, USA, 2004; pp. 1–58.
2. Paul, L.A.; Hall, N. *Causation: A User's Guide*; Oxford University Press: Oxford, UK, 2013.
3. Maudlin, T. *The Metaphysics within Physics*; Oxford University Press: Oxford, UK, 2007.
4. Field, H. Causation in a Physical World. In *The Oxford Handbook of Metaphysics*; Loux, M.J., Zimmerman, D.W., Eds.; Oxford University Press: Oxford, UK, 2003; pp. 435–460.
5. Fernandes, A. The Temporal Asymmetry of Causation. *MS.* **2022**.
6. Albert, D. *Time and Chance*; Harvard University Press: Cambridge, MA, USA, 2000.
7. Loewer, B. Counterfactuals and the Second Law. In *Causation, Physics, and the Constitution of Reality*; Price, H., Corry, R., Eds.; Oxford University Press: Oxford, UK, 2007; pp. 293–326.
8. Kutach, D. *Causation and Its Basis in Fundamental Physics*; Oxford University Press: Oxford, UK, 2013.
9. Price, H.; Weslake, B. The Time-Asymmetry of Causation. In *The Oxford Handbook of Causation*; Beebee, H., Menzies, P., Hitchcock, C., Eds.; Oxford University Press: Oxford, UK, 2009.
10. Kutach, D. The Physical Foundations of Causation. In *Causation, Physics, and the Constitution of Reality*; Price, H., Corry, R., Eds.; Oxford University Press: Oxford, UK, 2007; pp. 327–350.
11. Lewis, D. The Paradoxes of Time Travel. *Am. Philos. Q.* **1976**, *13*, 145–152.
12. Arntzenius, F.; Maudlin, T. Time Travel and Modern Physics. In *The Stanford Encyclopedia of Philosophy (Winter 2013 Edition)*; Zalta, E.N., Ed.; Metaphysics Research Lab., Stanford University: Stanford, CA, USA, 2013; Available online: https://plato.stanford.edu/archives/win2013/entries/time-travel-phys/ (accessed on 18 February 2022).
13. Wasserman, R. *Paradoxes of Time Travel*; Oxford University Press: Oxford, UK, 2018.
14. Price, H. *Time's Arrow & Archimedes' Point*; Oxford University Press: New York, NY, USA, 1996.
15. Tooley, M. Backward Causation and the Stalnaker-Lewis Approach to Counterfactuals. *Analysis* **2002**, *62*, 191–197. [CrossRef]
16. Wasserman, R. Lewis on Backwards Causation. *Thought* **2015**, *4*, 141–150. [CrossRef]
17. Cusbert, J. Backwards Causation and the Chancy Past. *Mind* **2018**, *127*, 1–33. [CrossRef]
18. Fernandes, A. Time Travel and Counterfactual Asymmetry. *Synthese* **2021**, *198*, 1983–2001. [CrossRef]
19. Fernandes, A. Freedom, Self-Prediction, and the Possibility of Time Travel. *Philos. Stud.* **2020**, *177*, 89–108. [CrossRef]
20. Vihvelin, K. What time travelers cannot do. *Philos. Stud.* **1996**, *81*, 315–330. [CrossRef]
21. Smith, N.J.J. Bananas enough for time travel? *Br. J. Philos. Sci.* **1997**, *48*, 363–389. [CrossRef]
22. Sider, T. Time travel, coincidences and counterfactuals. *Philos. Stud.* **2002**, *110*, 115–138. [CrossRef]
23. Ismael, J. Closed Causal Loops and the Bilking Argument. *Synthese* **2003**, *136*, 305–320. [CrossRef]
24. Lewis, D. Counterfactual Dependence and Time's Arrow. *Noûs* **1979**, *13*, 455–476. [CrossRef]
25. Lewis, D. Causation. *J. Philos.* **1973**, *70*, 556–567. [CrossRef]
26. Lewis, D. Are We Free to Break the Laws? *Theoria* **1981**, *47*, 113–121. [CrossRef]
27. Edgington, D. Counterfactuals and the Benefit of Hindsight. In *Cause and Chance: Causation in an Indeterministic World*; Dowe, P., Noordhof, P., Eds.; Routledge: New York, NY, USA, 2004; pp. 12–27.
28. Schaffer, J. Counterfactuals, Causal Independence, and Conceptual Circularity. *Analysis* **2004**, *64*, 299–309. [CrossRef]
29. Loewer, B. Two accounts of laws and time. *Philos. Stud.* **2012**, *160*, 115–137. [CrossRef]
30. Frisch, M. Does a Low-Entropy Constraint Prevent Us from Influencing the Past? In *Time, Chance and Reduction: Philosophical Aspects of Statistical Mechanics*; Ernst, G., Hüttemann, A., Eds.; Cambridge University Press: Cambridge, UK, 2010; pp. 13–33.
31. Loewer, B. The Mentaculus Vision. In *Statistical Mechanics and Scientific Explanation: Determinism, Indeterminism and Laws of Nature*; Allori, V., Ed.; World Scientific: Singapore, 2020; pp. 3–29.

32. Beebee, H. Causation, projection, inference and agency. In *Passions and Projections: Themes from the Philosophy of Simon Blackburn*; Johnson, R.N., Smith, M., Eds.; Oxford University Press: New York, NY, USA, 2015; pp. 25–48.
33. Callender, C. *What Makes Time Special*; Oxford University Press: Oxford, UK, 2017.
34. Bennett, J. Counterfactuals and Temporal Direction. *Philos. Rev.* **1984**, *93*, 57–91. [CrossRef]
35. Blanchard, T. Causation in a Physical World. Ph.D. Thesis, Rutgers University, New Brunswick, NJ, USA, 2014.
36. Albert, D. *After Physics*; Harvard University Press: Cambridge, MA, USA, 2015.
37. Fernandes, A. A Deliberative Approach to Causation. *Philos. Phenomenol. Res.* **2017**, *95*, 686–708. [CrossRef]