The Asymmetry of Counterfactual Dependence

Christian Loew

ABSTRACT. A certain type of counterfactual is thought to be intimately related to causation, control, and explanation. The time asymmetry of these phenomena therefore plausibly arises from a time asymmetry of counterfactual dependence. But why is counterfactual dependence time asymmetric? The most influential account of the time asymmetry of counterfactual dependence is David Albert's account, which posits a new, time-asymmetric fundamental physical law, the so-called "past hypothesis." Albert argues that the time asymmetry of counterfactual dependence arises from holding fixed the past hypothesis when evaluating counterfactuals. In this paper, I argue that Albert's account misconstrues the time asymmetry of counterfactual dependence.

Contact Information

Christian Loew
Institute of Philosophy
University of Luxembourg
4366 Esch-sur-Alzette
Luxembourg

Email: christian.loew@uni.lu

Acknowledgments

Thanks to Merrie Bergmann, Siegfried Jaag, and three anonymous referees for this journal for very helpful comments and suggestions.

Abstract

A certain type of counterfactual is thought to be intimately related to causation, control, and explanation. The time asymmetry of these phenomena therefore plausibly arises from a time asymmetry of counterfactual dependence. But why is counterfactual dependence time asymmetric? The most influential account of the time asymmetry of counterfactual dependence is David Albert's account, which posits a new, time-asymmetric fundamental physical law, the so-called "past hypothesis." Albert argues that the time asymmetry of counterfactual dependence arises from holding fixed the past hypothesis when evaluating counterfactuals. In this paper, I argue that Albert's account misconstrues the time asymmetry of counterfactual dependence.

1. Introduction

A certain type of counterfactual that is typically expressed by subjunctive conditionals plays an important role in many areas of philosophy. Examples include "If the rock had not hit the window, then the window would not have shattered" and "If the match had been struck, then it would have lighted." These counterfactuals are thought to be intimately related to causation (Hitchcock 2001; Lewis 1986b), control (Albert 2015; Kutach 2011; Loewer 2007, 2012), and explanation (Lewis 1986a; Woodward 2003). Due to this connection, it is widely assumed that the time asymmetry of these phenomena arises from a time asymmetry of the relevant counterfactuals (cf. Lewis 1986c, 35–38; Loewer 2007, 324). But what accounts for the time asymmetry of counterfactual dependence?

The most influential account of the time asymmetry of counterfactual dependence is David Albert's (2000, 2015) entropic account, which adds a new, time-asymmetric fundamental physical law, the so-called "past hypothesis," to our existing physical theories. Albert argues that holding the past hypothesis fixed in counterfactual reasoning gives rise to a time asymmetry that, among other things, explains why we cannot control the past. We might then hope to similarly explain other asymmetries that are associated with counterfactuals, including causation and explanation (see Loewer 2012).

In this paper, I will argue that Albert's account misconstrues the time asymmetry of counterfactual dependence. I will show that holding the past hypothesis fixed in counterfactual reasoning is explanatorily redundant in Albert's own account of the time asymmetry of control. If his account explains why we cannot control the past conditional on the past hypothesis, then it would equally explain why we cannot control the past if the past hypothesis is not held fixed. Moreover, I will argue that not holding the past hypothesis fixed may even put us in a better position to account for the time asymmetries of control, causation, and explanation.[1]

2. Counterfactuals and the Past Hypothesis

To introduce the puzzle about the time asymmetry of counterfactual dependence, it will be helpful to have a framework for evaluating counterfactuals. A plausible method for evaluating counterfactuals is the *altered states recipe* (see Maudlin 2007, 21–34 and Paul and Hall 2013, 47–53). Paul and Hall illustrate the recipe for counterfactuals of the form "If C had not occurred, E would not have occurred:"

> [C]onstruct a counterfactual state of the world at time t as much like the actual state at time t as possible, save for the fact that C does not occur. Think of taking the actual time-t state of the world, and ringing carefully localized changes on it just sufficient to make it the case that C does not occur…. We then evolve the resulting state forward

1 Fernandes (forthcoming), Frisch (2007, 2010, 2014) and Price and Weslake (2009) also criticize Albert's account. My criticism is different and aimed at Albert's most recent statement of his theory (2015), which takes into account many of these earlier criticisms.

> in time, in accordance with the actual laws of nature. If the resulting history yields E, the conditional is false; otherwise it is true. (Paul and Hall 2013, 47–48)

Suppose Suzy throws a stone and the window shatters shortly thereafter. Now consider the counterfactual "If Suzy's throw had not occurred, then the window shattering would not have occurred." According to the altered states recipe, this counterfactual is evaluated as follows: We take a counterfactual state of the world that is as much like the actual state at the time of Suzy's throw as possible, save for the fact that Suzy does not throw the stone. We then evolve this counterfactual state forward in accordance with the fundamental laws of nature to see whether the window shattering occurs in this counterfactual history. If the window shattering does not occur, then the counterfactual is true; otherwise, it is false.[2]

The altered states recipe is only a blueprint for evaluating counterfactuals. A full account would need to flesh out what exactly makes a state be the most similar state to the actual one where the antecedent of a counterfactual is true (Paul and Hall 2013, 50). For example, there are numerous counterfactual states in which Suzy does not throw a stone. Which of these states is most similar to the actual state? Is it a state where everything in the region where Suzy's throw actually occurs is replaced by empty space? Or a state where Suzy is present but does something else? And if the latter, then what does Suzy do instead?

The altered states recipe, nonetheless, nicely illustrates the puzzle about the time asymmetry of counterfactual dependence. Most of our best candidates for the fundamental dynamical laws of physics are deterministic in both temporal directions.[3] According to these laws, the complete state of the world at any one time fixes a unique lawful future and a unique lawful past (Elga 2001, S318). Hence, we can take the state of the world at any one time and use the laws to evolve it both forward and backward to yield a complete history of the world. Moreover, deterministic laws entail that any counterfactual state that differs from the actual state yields not only a different future but also a different past. So, according to the altered states recipe, past events counterfactually depend on the present. How, then, can we account for the asymmetries that are associated with counterfactuals, in particular, the asymmetries of causation, control, and explanation?

Albert (2000, 2015) provides the currently most influential account of the time asymmetry of counterfactual dependence.[4] The main innovation of his account is that it

---

2 I prefer the altered states recipe to the Lewis-Stalnaker framework, which evaluates counterfactuals in terms of similarity between possible worlds, for reasons given in Paul and Hall (2013, 42–49). I suspect that everything I say in this paper can be translated into a possible worlds framework.

3 Most discussions of counterfactuals make this assumption (see Albert 2000, chap. 6; 2015, chap. 2; Elga 2001, 315; Lewis 1986c, 37; Paul and Hall 2013, 27). The puzzle about the time asymmetry of counterfactual dependence would still arise for indeterministic laws, as long as these laws specify probabilities in both temporal directions (cf. Field 2003, 437). See Maudlin (2007, 28–31) on how to adapt the altered states recipe for indeterministic laws.

4 Lewis (1986c) is another influential theory, but see Elga (2001), Field (2003), and Frisch (2005, chap. 8) for decisive criticisms. Albert's account can be seen as a

assumes the so-called "past hypothesis" as a fundamental law of nature in addition to the fundamental dynamical laws. The past hypothesis is

> a new and explicitly *non*-time-reversal-symmetric fundamental law of nature … to the effect that the universe had some particular, simple, compact, symmetric, cosmologically sensible, very low-entropy initial macrocondition. (Albert 2015, 5; italics in the original)

Three features of the past hypothesis will be important for what follows: First, since the past hypothesis is posited as an additional fundamental physical law, it rules out histories of the world that do not conform with it as nomologically impossible. Second, the past hypothesis is time asymmetric because it applies to the initial state of the universe. Third, the past hypothesis specifies that the initial state of the world has certain macroscopic properties (such as low entropy) but leaves open its exact microscopic properties.

Albert argues that his account of the time asymmetry of counterfactual dependence is compatible with most methods for evaluating counterfactuals as long as they do "not introduce any asymmetry between the past and the future over and above the asymmetry which is introduced by the past hypothesis" (Albert 2015, 42). To have something concrete to work with, I will treat Albert's proposal as a way of supplementing the altered states recipe, which meets Albert's condition because it introduces no asymmetry into how counterfactuals are evaluated.

An implicit restriction in the altered states recipe is that when we look for the most similar state to the actual state that makes the antecedent of some counterfactual true, we only consider counterfactual states that are compatible with the fundamental dynamical laws. After all, the recipe tells us to evolve the modified state forward and backward in accordance with the fundamental laws, and this is only possible if the modified state is compatible with the fundamental dynamical laws. We can then add the past hypothesis to the altered states recipe by assuming that the most similar state that makes the antecedent true also must be compatible with the past hypothesis.[5]

Even with the past hypothesis added to the altered states recipe, the past still counterfactually depends on the present. If the fundamental dynamical laws are deterministic, as Albert assumes they are, then any state that differs from the actual present state in any way at all lawfully entails a different past (cf. Loewer 2012). So it still needs to be shown how evaluating counterfactuals in accordance with the past hypothesis gives rise to the time asymmetries of causation, control, and explanation.

---

development of Lewis's theory that avoids these problems (see Loewer 2007).

5 Albert also posits a "statistical postulate" that supplies a probability distribution over microstates compatible with the past hypothesis. It says (roughly) that every microstate is equally likely. He argues that eligible counterfactual microstates need to be such that the "associated macrohistories are assigned reasonable probability values by the statistical postulate" (Albert 2015, 41–42). This constraint will not matter for my criticism of Albert's account.

3. The Time Asymmetry of Control

Albert argues that understanding counterfactuals in terms of the past hypothesis accounts for the "time asymmetry of control," that is, the fact that agents like us can control the future but not the past. He says little about the other asymmetries, but it is implied that they can be explained along similar lines (see Loewer 2012, 132). In this and the next section, I will criticize Albert's account of the time asymmetry of control by arguing that his appeal to the past hypothesis is explanatorily redundant.

Albert's account of the time asymmetry of control has two parts. First, he argues that the past hypothesis limits influence over the past such that "any creature which we might imaginably be tempted to treat as an agent … seems likely to be in a position to influence much about the future and next to nothing about the past" (Albert 2015, 44). Second, he argues that this limited influence does not allow us to control the past because "whatever opportunities there may be to influence the past are either rare or impractical or invisible or in some other way beside the point" (2015, 52).

With regard to the first part, Albert adopts what he calls the "fiction of agency," whose upshot is that we have "*direct* and *unmediated* and freely exercised control" (italics in the original) only over a small part of the world's present state, such as regions of our brains or the locations of our hands and feet (2015, 42). For simplicity, I will assume that the small region we have direct control over concerns events in our brains that I will call "decisions." Nothing hinges on this assumption, as long as the region of the present that we have direct control over is small.

Albert further assumes that we can influence past and future events only to the extent that they counterfactually depend on the present region over which we have direct control. Suppose I face a decision between snapping my fingers and not snapping my fingers. We can determine the extent to which my decision influences the past as follows. Take two complete states of the world that are as much like the actual state of the world at the time of my decision as possible. The two states, however, differ from each other in that in one I decide to snap my fingers and in the other I decide not to snap my fingers. My decision can thus be understood as a choice about which of these two states to make actual. We then take each of the two complete states of the world and evolve it backward in accordance with the fundamental physical laws, yielding a complete past history of the world. My decision influences those past events that occur in one history but not in the other.

Albert argues that due to the past hypothesis, our influence over the past is extremely limited. He argues that the past hypothesis gives rise to the time asymmetry of records, that is, the fact that we have records of the past but not of the future. Records are localized events, such as photographs, memories, or footprints, that allow us to infer with high probability that certain events at other times (viz., the events remembered, photographed, etc.) have occurred.[6] It is not obvious why events like photographs or memories are reliable records of the past. The fundamental dynamical laws allow many ways for these events to come into existence such that they would not be correlated with the past events they appear to represent.

---

6 See, e.g., Albert (2000, chap. 6), Frisch (2014, 224–28), and Horwich (1987, chap. 5) for more discussion of records.

For example, they allow that my current self and all photographs of my alleged childhood are the result of a fluctuation from a state of higher entropy. In this case, all of these photographs would be radically misleading. Albert argues that the past hypothesis (together with the dynamical laws and plausible statistical assumptions) makes it very likely that most records of the past are reliable. This point is controversial, but I will grant it for what follows.[7]

Most past events, according to Albert, do not depend on our present decisions because they have many current records that exist outside of our decisions. For example, my past trip to Paris is currently recorded in my memories, photographs, and credit card receipts. A counterfactual difference in my decision, which according to the myth of agency only corresponds to a small difference at the present time, leaves these records intact. So if we take a counterfactual state compatible with the past hypothesis that differs from the actual state only with respect to my decision and evolve it backward in accordance with the fundamental physical laws, then these records assure that it is very probable that my past visit to Paris would still have occurred. Thus, my past visit to Paris does not counterfactually depend on my current decision. The same holds for all other past events of which there are sufficiently many present records that exist outside of my decisions (cf. Loewer 2012, 128). At the same time, the future depends on our present decisions more extensively because there is no 'future hypothesis' and so most future events do not have present records.

Holding fixed the past hypothesis (that is, only considering counterfactual states that are compatible with it) is crucial for this account. It is not enough that there are actual present records of me having been in Paris. These records also need to remain reliable in the counterfactual scenario where I make a different decision. Elga (2001, S323–24) shows that if we evaluate counterfactuals only in accordance with the fundamental dynamical laws—and without the past hypothesis—then any small counterfactual change to the present likely will lawfully entail a counter-entropic past. Given such a counter-entropic past, alleged present records of the past would be radically misleading (more on this later). Holding the past hypothesis fixed makes such a counter-entropic past, in which current records are unreliable, highly unlikely (Loewer 2007, 315–16). So evaluating counterfactuals in accordance with the past hypothesis limits our influence over the past because it ensures that records are counterfactually stable.

4. The Redundancy of the Past Hypothesis

The next step for Albert is to show that we cannot use the limited influence over the past that the past hypothesis does allow to control the past. In this section, I will argue that Albert's reasons for why we cannot use this influence to control the past make his appeal to the past hypothesis explanatorily redundant. If these reasons explain why we cannot use the influence over the past that we would have despite holding the past hypothesis fixed to control the

7 Albert is not always clear about whether the past hypothesis (together with the statistical postulate) is sufficient for the reliability of (many of) our current records, but, as far as I can see, his account of the time asymmetry of control requires this assumption. See Frisch (2007, 2014) for discussion of Albert's account of records.

past, then they equally explain why we could not use the more extensive influence that we would have without holding the past hypothesis fixed to control the past.

Albert's account allows that we can influence past outcomes that have no current records or where our decision itself is a record of the outcome. I will follow Albert in calling the former cases "Elga cases" and the latter cases "Frisch cases," after the philosophers who discovered them. Albert needs additional reasons for why we cannot control the past in Elga and Frisch cases, where our decisions do influence the past.

Elga cases concern circumstances where some past outcome has no present record. Suppose all macroscopic records of the existence of Atlantis have since vanished into microscopic noise. It is then compatible with the past hypothesis that my present decision to, say, snap my fingers influences whether Atlantis did or did not exist (Albert 2015, 47). Suppose that if I had decided to snap my fingers, then Atlantis would not have existed, but if I had decided to not snap my fingers, then Atlantis would have existed.[8]

Albert's reason for why we cannot control the past in Elga cases generalizes as follows:

Ignorance. I cannot control an outcome that, in certain circumstances, depends on my decision if "I can have no way whatsoever of knowing, and I can have no grounds whatsoever for suspecting, when it is that [these circumstances] actually obtain!" (Albert 2015, 48)

Ignorance explains why I cannot control whether Atlantis existed. Even if Atlantis's existence does in certain circumstances counterfactually depend on my decision, I still have no way of knowing when I am in these circumstances and in which way it depends on my decision. In particular, I have no way of knowing whether Atlantis would have existed if I had decided to snap my fingers but not otherwise or whether Atlantis would not have existed if I had snapped my fingers but would have existed otherwise. So I cannot use my influence over Atlantis's existence to further a particular desired end. Ignorance, however, does not apply to all cases of influence over the past.

Frisch cases concern circumstances where an agent's present decision itself is a record of some past outcome. Frisch (2007, 2014) gives an example of such circumstances that I will quote in full:

> While playing a piano piece that I know well, I am unsure whether I am currently playing a part of the piece that is repeated in the score for the first or the second time. I decide to play the second ending rather than repeating the part. Many of the notes I play, of course, I play without choosing or deciding to play them. But in the case I am imagining the question what notes to play next has arisen, and I consciously choose to play the second ending. Since I have learned from experience that when I play a piece I know well my decisions to play certain notes are good evidence for where I am in the piece, my present decision not to repeat the part constitutes good evidence for a certain past event—my having already played the part in question once. We can even

8 Albert (2015) and Loewer (2012) attribute this case to Adam Elga.

> imagine that I have a vague and unreliable memory of having already played the part. My decision to play the second ending, then, can constitute additional evidence for the reliability of my memory. (Frisch 2014, 222)

My decision to play the second ending in this case is a record of me having already played the first ending. Moreover, Frisch stipulates that there are not sufficiently many other present records of which notes I have played in the past to screen of its dependence on my current decision. My decision then influences which notes I have played in the past.[9]

Ignorance does not apply to Frisch cases. Since my decision is a record of whether I have already played the first ending, I can have evidence that whether I have already played it depends on my decision. For example, I might have learned from past instances where there were independent records that my decision to play the second ending is reliably correlated with whether I have already played the first ending (Frisch 2014, 222). So I can know that which notes I have already played depends on my decision in these circumstances.

In response to Frisch cases, Albert spells out two further circumstances in which we cannot use our influence for control.[10] He defends two conditions such that if at least one of these conditions is met in a particular case, it is not a case of control. He argues that each condition alone suffices to explain why we cannot control the past in Frisch cases. The first condition is:

Correlation. I cannot control an outcome that depends, in certain circumstances, on my decision if there are lawlike correlations between my decisions and the decisions of another (possible) agent.[11]

The main idea seems to be this: If my decision is a record of some past outcome (just as the decision of Frisch's piano player is a record of what notes she has played in the past), then there could be another agent whose decision also is a record of the same outcome. In this case, since each agent's decision is correlated with the same past outcome, the agents' decisions would also be correlated with each other (Albert 2015, 45). Albert argues that an agent could not then consistently treat her decision as being under her control since it could be correlated with another agent's independent decision. So the pianist in Frisch's example could not simultaneously treat herself as having control over her present decision and treat

---

9 One might question the inference from my decision being one of very few records of what notes I have played in the past to my decision thereby *influencing* what notes I have played. Albert (2015, 48), however, grants this inference, and so I will do the same at least for the sake of the argument. Fernandes (forthcoming) and Kutach (2011, 252) also accept the inference.

10 Albert credits Fernandes (forthcoming) with talking him out of a different response to Frisch cases.

11 See Albert (2015: 49). It is not fully clear whether Albert thinks that the other agent needs to be actual or merely conceivable. However, if the other agent needs to be actual, then the reason is unlikely to apply to all Frisch cases since there likely will not always be such an agent.

that decision as a record of what notes she has played in the past. Else, she would have to acknowledge that her decision could be correlated with another agent's independent decision, which, according to Albert, is incompatible with "any fiction (that is) about those two decisions, as the free and spontaneous acts of two separate and autonomous agents" (Albert 2015, 49). I will say more about this argument in Section 5, but for now I want to move on to Albert's other reason.

Albert's other reason for why we have no control in Frisch cases is:

Fragility. I cannot control an outcome that depends, in certain circumstances, on my decision if the dependence "turns out to have none of the robustness, and none of the flexibility, and none of the utility, that come along with familiar idea[s] of what it might amount [to] to exercise 'intentional control.'" (Albert 2015, 51)

Albert points out that in ordinary cases of control, for example, when my decision influences whether a door is open, the dependence holds across a wide range of deliberational contexts. For example, if I had decided to open the door because I wanted to go to the bank instead of the office, the opening of the door would still have depended on my decision. Albert argues that if there is an influence that only obtains in very specific deliberational contexts, then it does not count as control.

Fragility is supposed to explain why Frisch's pianist example is not an instance of control over the past. Albert points out that having already played the first ending depends on my decision to play the second ending only in very specific deliberational contexts, and so the dependence cannot be utilized for practical purposes:

> [I]f (for example) I make any attempt at exploiting that counterfactual dependence in the service of making a *profit*; if (say) I play the second ending because somebody offers me a million dollars to bring it about that I have already played the first one, then *it simply isn't going to work*—because in *that* case the evidential connection between my decision to play the second ending and my already having played the first is going to be *broken*, is going to be *screened off*, by my having made that decision for the money. (Albert 2015, 50; italics in the original)

I will eventually question whether Albert's reliance on Fragility is compatible with his own provision against introducing unexplained temporal asymmetries, but I want to postpone this discussion until Section 5. For now, it is only important that Albert needs Ignorance and at least one of Correlation and Fragility to account for the time asymmetry of control.

Holding fixed the past hypothesis then is not needed for Albert's account to explain why we cannot control the past. I will argue that if Ignorance, Correlation, and Fragility explain why we cannot control the past if we hold the past hypothesis fixed in our account of counterfactuals, then they would equally explain why we cannot control the past if we do not hold the past hypothesis fixed. The past hypothesis restricts counterfactual histories to ones that start with the same initial macrostate as the actual universe. If counterfactuals are evaluated without the past hypothesis, then the past can counterfactually depend on our decisions a lot more than otherwise because the universe then could have started in a

different macrostate. Moreover, if Albert is right about the role of the past hypothesis in assuring the reliability of records (as I assume he is), then without the past hypothesis even events that do have many present records outside of our decisions can counterfactually depend on our decisions.

But Ignorance, Correlation, and Fragility would still explain why we cannot use the more extensive influence over the past that we would have if the past hypothesis is not held fixed to control the past. We can still divide cases where our decisions influence the past into two classes: cases where your decision is a record of the outcome in question (Frisch-style cases) and cases where your decision is not a record of the outcome in question (Elga-style cases). Let us start with the latter cases. Without the past hypothesis, past outcomes can depend on your decision even if these outcomes have multiple macroscopic records that exist outside of your decision. For example, there are a lot of current records of Rome's existence. Yet without the past hypothesis, the past existence of Rome may still counterfactually depend on your current decision, say, to snap or not snap your fingers. Would this give you control over Rome's existence?

If Albert is right that Ignorance explains why you lack control over the existence of Atlantis in the original Elga case, then it equally explains why you would lack control over the existence of Rome in this case. If we do not hold the past hypothesis fixed, then the existence of Rome may depend on your decision to snap your fingers despite having multiple present records. But you still could not use this dependence to further any desired ends because when you make the decision, you do not know in which way Rome's existence depends on it. It might be that if you were to decide to snap your fingers, Rome would not have existed. But it equally might be that Rome would not have existed if you were to decide to not snap your fingers. Of course, it seems absurd that Rome's existence would depend on your decision at all. But Albert already entertains the equally unintuitive possibility that Atlantis's existence and other past events may depend on your decisions.

In Frisch-style cases your decision itself is a reliable record of the past outcome. If we do not hold fixed the past hypothesis, there can be Frisch-style cases even in circumstances where the past outcome has many records that exist outside of your decision. But if Correlation and Fragility explain why we cannot control the past in the original Frisch case, then they also would explain why we cannot control the past in these additional cases. Correlation applies to the original Frisch case because my decision to play the second ending is a record of me having already played the first ending. It then would also apply to these other Frisch-style cases where my decision also is a record of the relevant outcome. The same is true for Fragility. Albert argues that Fragility applies to Frisch's pianist case because in a deliberational context where I decide to play the second ending for a different reason, the different reason would screen off the original correlation; so my decision would no longer be a record of me having already played the first ending. Albert does not defend why this screening off happens, but we have reason to believe that if it happens in the original Frisch case, then it still would happen in the additional Frisch-style cases that we would get without holding the past hypothesis fixed. (In fact, I will argue shortly that Fragility is more plausible if our account of counterfactuals does not hold the past hypothesis fixed.)

So holding fixed the past hypothesis is doing no work in Albert's account of the time asymmetry of control. The account rests on the plausibility of Ignorance, Correlation, and

Fragility. If these conditions explain why we cannot control the past conditional on the past hypothesis, then they also explain why we cannot control the past if the past hypothesis is not held fixed. So Albert misconstrues the time asymmetry of control. It does not arise from holding fixed the past hypothesis but from the further conditions he imposes on control.

5. A Problem for Albert's Account

In the rest of the paper, I will argue that we may even get a better account of the asymmetry of control and related asymmetries if we do not hold the past hypothesis fixed when evaluating counterfactuals. I will point out a problem that exists for Albert's account and show that we can get around this problem by evaluating counterfactuals without the past hypothesis. My argument has two parts: First, I will argue that Albert can deal with Frisch cases only if Fragility applies to these cases. Second, I will argue that Albert has no explanation of why Fragility applies to Frisch cases, but there is such an explanation if we do not hold the past hypothesis fixed in counterfactual reasoning.

Albert needs Fragility in order to have a plausible reply to Frisch cases. Ignorance does not apply to Frisch cases because in these cases we can know that the past outcome depends on our decisions. So Albert can only rely on Correlation or Fragility to deal with Frisch cases. Correlation says that we do not have control over a decision if the decision is (or might be) lawfully correlated with another agent's decision. But Correlation is not plausible as a constraint on control for two reasons. First, it would prove too much. Consider two experienced firefighters. When they enter a burning building, they pick up on many subtle clues and act accordingly. It is then plausible that their respective decisions are correlated. For example, if one firefighter decides to leave the burning building because she senses that it is not safe, it is very likely that the other firefighter will decide the same. In such cases, where two experts make decisions based on the same evidence, there plausibly are lawlike correlations between their decisions.[12] If Correlation were to rule out control, the firefighters would lack control over their decisions in these circumstances. Yet it seems clear that they could have control.

Second, it is not clear what the rationale for Correlation is in the first place. Albert's "fiction of agency" grants that an agent having control over her decisions is compatible with determinism. So control is assumed to be compatible with the fact that our decisions are a logical consequence of events in the distant past and the laws of nature. But if we can maintain control in light of determinism, it is hard to see why we cannot also maintain control in light of lawful correlations between our decisions and those of other agents.[13] So

12 See Holton (2006) who argues that experts' decisions can function like measurement devices. Frisch (2007, 2014) cites Holton as the inspiration for his pianist case.

13 Albert's "fiction of agency" is ambiguous. Does it require that we can rationally maintain the *belief* that we have control over our decisions (even though the belief might be false)? Or does it require that we actually have such control? My criticism of Correlation applies either way. If the fiction is compatible with determinism, then it should also be compatible with lawlike correlations between our decisions and other agents' decisions.

Correlation cannot explain why we cannot control the past in Frisch cases or any other cases. The burden of explaining why we cannot control the past in Frisch cases thus falls to Fragility.

Fragility is a plausible requirement for control. It says that we cannot control outcomes that, in particular circumstances, counterfactually depend on our decisions if this dependence would not continue to hold in deliberational contexts where we make the same decision for a different reason. As Albert points out, in Frisch's pianist case, having already played the first ending counterfactually depends on the pianist's decision to play the second ending only if that decision is made for a specific reason. The dependence, therefore, is extremely fragile and so the pianist cannot use it to promote desired ends.

Albert's account of why Fragility applies to Frisch cases, however, presupposes an unexplained time asymmetry. Albert argues that Fragility applies to Frisch cases because the relevant evidential correlations between our decisions and past outcomes are counterfactually fragile. He points out that in the pianist case, if I were to make the same decision for a different reason, my decision would no longer be a record of having already played the first ending. At the same time, Albert emphasizes that there is no such fragility with regard to evidential correlations between our decisions and future outcomes. For example, if I were to base my decision to open some door on a different reason, my decision typically would still be correlated with the door being open in the future (Albert 2015, 50). This asymmetry is meant to account for why Fragility rules out control over the past in Frisch cases but does not rule out control over the future. I think Albert is right about the existence of this asymmetry of fragility. But what explains this asymmetry?

At some point, Albert seems to suggest that the dependence of past events on our decisions is fragile because our reasons for a decision "screen off" evidential correlations between the decision and past outcomes (Albert 2015, 50). But this assumption would already presuppose a time asymmetry in our agency. An agent's reasons for a decision plausibly screen off correlations between the decision and past outcomes only if these reasons occur in the immediate past of the decision: as (causal or nomological) intermediaries between the decision and the past outcome. And while reasons for action may always precede decisions in agents like us, Price (2007) points out that we can conceive of agents whose reasons occur in the immediate future of their decisions. Such agents' reasons, hence, would not screen off correlations between their decisions and past outcomes. So if fragility is merely a consequence of the fact that reasons for action precede our decisions, then Albert's account would at most explain why agent whose deliberation is time asymmetric cannot control the past. Yet Albert's goal is to explain for "any creature which we might imaginably be tempted to treat as an agent" (Albert 2015, 44) why this creature could not control the past.[14]

---

14 There is a legitimate project that uses asymmetries in our local environment to explain our asymmetric perspective as agents and then uses this asymmetric perspective to explain why *we* cannot control the past (see, e.g., Fernandes forthcoming and Price and Weslake 2009). But this is not Albert's project. Albert wants to explain why *any* agent would be unable to control the past due to asymmetries in our world's objective counterfactual structure (cf. Blanchard 2016, 261).

Albert is explicit that a full account of the time asymmetry of control cannot presuppose any unexplained time asymmetry. He points out that our algorithm for evaluating counterfactuals should "not introduce any asymmetry between the past and the future over and above the asymmetry which is introduced by the past hypothesis" (Albert 2015, 42). Moreover, when he introduces his "fiction of agency," he emphasizes that we "want to make certain—as with the algorithm for evaluating counterfactuals—that whatever particular fiction we choose does not introduce any new asymmetries of its own between past and future" (Albert 2015, 43). So if Fragility is supposed to explain why we cannot control the past in Frisch cases, then it has to follow from evaluating counterfactuals in accordance with the past hypothesis that any dependence of past outcomes on agents' decisions in these cases is fragile. In the next section, I will argue that Albert's account does not have the resources to establish this fragility. However, we do get such an account if we do not hold the past hypothesis fixed when evaluating counterfactuals.

6. Counterfactuals without the Past Hypothesis

How could it follow, without any time-asymmetric assumptions about agents, that the dependence of past events on agents' decisions in Frisch cases is always fragile? Here is a natural proposal: A counterfactual difference in an agent's reason for a decision would plausibly be reflected in at least some small difference in present circumstances. This assumption is not time asymmetric since it is plausible regardless of whether the reason is in the past or in the future of the decision. Now suppose we can show that any small difference in present circumstances would make it that the dependence of past outcomes on the agent's decision would no longer hold, where the same is not true for the future. It would then follow that in the special case where a present difference is due to a difference in the agent's reason for the decision, the dependence also would no longer hold. Hence, we would have a story of how the required fragility in Frisch cases follows from a general asymmetry of counterfactual dependence.[15]

There is no straightforward way of getting this fragility in all Frisch cases if we hold the past hypothesis fixed. The past hypothesis, remember, was introduced to assure that records are counterfactually stable. In particular, Albert's account of why our influence over the past is extremely limited presupposes that most current records of past events are *not* counterfactually fragile against small differences in the state of the world outside of the records. For example, it presupposes that the memories and credit card receipts of my past trip to Paris still would have been records if I had made a different decision right now.

Why then should counterfactual dependence be fragile in Frisch cases? Albert grants that in Frisch cases a past outcome counterfactually depends on the agent's decision because the decision is a record of the outcome. As we have just seen, holding fixed the past hypothesis is supposed to ensure that records are counterfactually stable against small differences in the state of the world outside of the records. Moreover, a counterfactual difference in an agent's reason for a decision will (at least typically) only amount to a small

[15] There might be other time-symmetric ways of showing that the dependence of past outcomes on our decisions in Frisch cases is always fragile, but this way is very natural.

difference in present circumstances. So if we hold fixed the past hypothesis, we expect that if the agent had made the same decision for a different reason, her decision would have remained a record of the past outcome. Hence, the outcome still would have depended on the decision.

There might be some difference between ordinary circumstances and Frisch cases such that records are counterfactually stable in ordinary circumstances but not in Frisch cases. In particular, in ordinary circumstances past outcomes tend to have many records, whereas in Frisch cases they only have few other records in addition to the agent's decision. However, a main purpose of holding the past hypothesis fixed is to ensure the counterfactual stability of records. So some substantial argument would be required to show that it succeeds to do so in ordinary circumstances but not in Frisch cases. At the very least, no story has been provided of how we get fragility in Frisch cases if we hold the past hypothesis fixed. By contrast, I will argue that the required fragility follows naturally if we evaluate counterfactuals without holding the past hypothesis fixed.

Following the altered states recipe, we evaluate counterfactuals by taking the complete state of the world in which the antecedent of the counterfactual is true that is most similar to the actual state and evolving it backward and forward in accordance with the fundamental dynamical laws. Holding the past hypothesis fixed would require us to ignore any states that are incompatible with the actual initial macrostate of the universe as candidates for the most similar state. I propose that we give up this restriction such that the most similar state to the actual state in which the antecedent of a counterfactual is true may be one that lawfully entails a different initial macrostate. In the following, I will argue that allowing these additional states as candidates for the most similar state provides additional resources for explaining why we cannot control the past.

Counterfactual dependence still is time asymmetric even if we do not hold the past hypothesis fixed. The actual present state of the universe is time asymmetric in that it lawfully entails a very special, low-entropy macroscopic past but it does not lawfully entail a low-entropy macroscopic future. We can make this time asymmetry more precise by bringing in considerations from Statistical Mechanics (see, e.g., Albert 2000; Elga 2001; Horwich 1987). The entropy of our universe (and of relatively isolated subsystems) tends to increase toward the future but it is extremely unlikely to decrease. For example, heat spontaneously flows from hotter bodies to colder bodies and gases expand throughout their available volumes, but we never observe the time-reverse behavior.

The Boltzmannian tradition in Statistical Mechanics tries to explain the unlikeliness of entropy decrease toward the future in terms of the scarcity of the relevant microstates. We can represent all microstates that are compatible with the current macrostate of our universe in an abstract space called "phase space." Each point in this space represents a fully specified microstate of the world. It turns out that the overwhelming majority of microstates compatible with the current macrostate, including the overwhelming majority of microstates in the immediate vicinity in phase space to the actual microstate, lawfully entail that entropy increases toward the future. Microstates that lawfully entail that entropy decreases toward the future, by contrast, are extremely rare. If we then add probabilistic assumptions that say (roughly) that each possible microstate is equally likely, it follows that it is extremely unlikely that our universe is in a microstate such that entropy will decrease toward the future.

It does not matter for my argument whether this explanation succeeds. It only matters that microstates that lawfully entail that entropy decreases toward the future are in fact extremely rare. Moreover, the same is true for the past. Since the fundamental dynamical laws are (small possible exceptions aside) time-reversal invariant, microstates compatible with the current macrostate that lawfully entail that entropy was lower in the past are also extremely rare. But there is an important difference between past and future: The entropy of our universe will be higher in the future. So the actual macroscopic future is typical in the sense that most microstates that are compatible with the current macrostate would also lawfully entail a macroscopic future of higher entropy that, at least over a short time-period, is macroscopically similar to the actual future. But entropy was lower in the past. So the actual macroscopic past is *atypical* since the overwhelming majority of possible microstates, including the overwhelming majority of microstates in the immediate vicinity in phase space to the actual microstate, would lawfully entail a macroscopic past of higher entropy that is radically different from the actual, low-entropy past.[16]

This atypicality of the past history of our universe entails that the macroscopic past of systems is extremely sensitive to small changes in their end state. Most possible microstates, including the ones in the immediate vicinity of the actual microstate in phase space, would lawfully entail a very different macroscopic past. So almost any small difference in the actual microstate will yield a microstate that lawfully entails a macroscopic past that is extremely different from the actual past. Elga (2001) illustrates this sensitivity with the example of an egg that is cracked and cooked in a frying pan. Most alternative microstates that differ from the actual microstate of the world at the time when the egg is cooked would lawfully entail a macroscopic past where the egg never got cracked but was cooked all along. (Such a past strikes us as strange because all cooked eggs that we actually observe have been cracked and cooked. But the actual past nonetheless is atypical in that most possible microstates would lawfully entail a different macroscopic past.)

The sensitivity of systems' past histories to small changes in their end state entails that any dependence of past events on our current decisions is extremely fragile. Most microstates that are similar to the actual state but in which some decision of mine is different lawfully entail a different macroscopic past in which many actual past events would not have occurred. So if we do not hold the past hypothesis fixed, many past macroscopic events plausibly counterfactually depend on my current decisions. This dependence, however, is extremely fragile: while these events actually counterfactually depend on my decision, if circumstances were even slightly different, then (in the resulting counterfactual

16 See Elga (2001). Albert (2000, 151) uses the same fact for a different purpose. A foundational question in Statistical Mechanics concerns why entropy was not also higher in the past. This is the so-called "reversibility objection." Albert's past hypothesis, which is time asymmetric, is intended to solve this problem by motivating a "loop-sided" probability measure according to which entropy is overwhelmingly like to increase toward the future but not toward the past. This possible role of the past hypothesis in explaining the entropic time asymmetry, however, leaves open whether we should hold it fixed when evaluating counterfactuals (see Section 7).

circumstances) the events in question would no longer counterfactually depend on my decision.

Take any past macroscopic event D that counterfactually depends on a given agent's decision. Because of the atypicality of the actual macroscopic past, it is extremely likely that D also depends on numerous other present events outside of the agent's decision. More specifically, any counterfactual difference in the present microstate almost certainly would have resulted in a radically different macroscopic past that does not include D. But in such counterfactual situations where D does not occur due to some difference other than the agent's decision, D also no longer counterfactually depends on the agent's decision. This is for the following reason: Only a tiny proportion of possible microstates lawfully entail a macroscopic past that includes D. So if we take the counterfactual situation (where present circumstances outside of the decision are different and D does not occur) and change it such that the agent's decision does not occur, the resulting microstate almost certainly also would lawfully entail a past history that does not include D. Hence, if circumstances outside of the agent's decision had been even slightly different, then in the resulting counterfactual situation D would not have occurred regardless of the agent's decision. So, macroscopic past events that actually counterfactually depend on an agent's decision would no longer counterfactually depend on the decision if present circumstances were even slightly different.

This fragility explains why we cannot control the past in Frisch cases. Let me remind you of Albert's Fragility condition. Suppose whether your front door opens counterfactually depends on your decision to open the door. But now also suppose (unrealistically) that if your reason for making the decision had been in any way different, the door would not have opened regardless of whether you had decided to open it or not. Albert's Fragility condition entails that we would lack control in such cases because the resulting dependence is not stable across deliberational contexts (Albert 2015, 51).

The above considerations show that if we evaluate counterfactuals without holding the past hypothesis fixed, then any dependence of the past on our decisions is fragile in this sense. Many macroscopic past events counterfactually depend on agents' decisions. But if circumstances outside of the agent's decision were even slightly different, this dependence would no longer hold. Moreover, differences in an agent's reason for a decision amount to at least a small difference in the present circumstances outside of the agent's decision. Hence, any dependence of the past on an agent's decision is extremely fragile against changes in the agent's reason. So if we do not hold the past hypothesis fixed, then Fragility explains why we cannot control the past in Frisch cases.[17]

Moreover, this fragility in how the past counterfactually depends on the present promises deeper insight into the time asymmetries of causation and causal explanation. Woodward (2006, 2) points out that "we tend to regard causal claims that are highly sensitive as defective, nonstandard, or at least importantly different from less sensitive causal claims," where a causal claim "is sensitive if it holds in the actual circumstances but would not

17 Eckhardt (2006) and Kutach (2011) also use fragility considerations to explain why we cannot control the past. Horwich (1987) and Frisch (2014) discuss similar physical time asymmetries.

continue to hold in circumstances that depart in various ways from the actual circumstances." The fragility of counterfactual dependence in the backward direction entails that if there were causal relations in the backward direction, the corresponding causal claims would be highly sensitive in Woodward's sense. Almost any small change to the present would make it be that a purported past effect of some current event would not have occurred, and so the corresponding causal claim would not have been true. Causal claims in the backward direction would thus be sensitive in Woodward's sense. This fact then is the beginning of an explanation regarding why, despite counterfactual dependence of earlier events on later events, we do not regard later events as causes of earlier events and why we do not cite later events to causally explain earlier events. More work would be needed to show that the proposed account of the time asymmetry of counterfactual dependence is ultimately viable, but it has some very attractive features.

7. Conclusion

Albert's account in terms of the past hypothesis is the most influential current account of the time asymmetry of counterfactual dependence. In this paper, I have shown that Albert's appeal to the past hypothesis is problematic. I have argued that holding fixed the past hypothesis when evaluating counterfactuals is explanatorily redundant and possibly even counter-productive when trying to account for control and related asymmetries.

My argument undermines one central reason for Albert's claim that the past hypothesis is a fundamental physical law. A plausible constraint on fundamental physical laws is that they determine the nomological possibilities and thereby, together with facts about what actually happens, fix our world's counterfactual structure (Hall 2012). So a good reason for treating the past hypothesis as a fundamental physical law would be if holding it fixed were necessary to explain the asymmetries associated with counterfactuals. But I have argued that we are in at least as good a position to explain these asymmetries if we do not hold the past hypothesis fixed.

Albert provides other reasons for treating the past hypothesis as a fundamental physical law, including its role in grounding the time asymmetry of Thermodynamics and the reliability of records. Assessing these reasons and weighing them against the above considerations is beyond the scope of this paper. But I hope to have shown that the role of the past hypothesis in accounting for the time asymmetries in our universe needs to be reconsidered.

References

Albert, David. 2000. *Time and Chance*. Cambridge, MA: Harvard University Press.
---. 2015. *After Physics*. Cambridge, MA: Harvard University Press.
Blanchard, Thomas. 2016. "Physics and Causation." *Philosophy Compass* 11:256–66.
Eckhardt, William. 2006. "Causal Time Asymmetry." *Studies in History and Philosophy of Modern Physics* 37:439–66.
Elga, Adam. 2001. "Statistical Mechanics and the Asymmetry of Counterfactual Dependence." *Philosophy of Science* 68:S313–24.
Fernandes, Alison. Forthcoming. "Time, Flies, and Why We Can't Control the Past." In *Time's Arrows and the Probability Structure of the World,* ed. Barry Loewer, Brad Weslake, and Eric Winsberg. Cambridge, MA: Harvard University Press.
Field, Hartry. 2003. "Causation in a Physical World." In *Oxford Handbook of Metaphysics*, ed. Michael J. Loux and Dean W. Zimmerman, 435–60. Oxford: Oxford University Press.
Frisch, Mathias. 2005. *Inconsistency, Asymmetry, and Non-Locality*. Oxford: Oxford University Press.
---. 2007. "Causation, Counterfactuals, and Entropy." In Price and Corry 2007, 351–96.
---. 2010. "Does a Low-Entropy Constraint Prevent Us from Influencing the Past?" In *Time, Chance, and Reduction*, edited by A. Hüttemann and G. Ernst, 13–33. Cambridge: Cambridge University Press.
---. 2014. *Causal Reasoning in Physics*. Cambridge: Cambridge University Press.
Hall, Ned. 2012. "Humean Reductionism about Laws of Nature: Director's Cut." Unpublished manuscript.
Hitchcock, Christopher. 2001. "The Intransitivity of Causation Revealed in Equations and Graphs." *Journal of Philosophy* 98:273–99.
Holton, Richard. 2006. "The Act of Choice." *Philosophers' Imprint* 6:1–15.
Horwich, Paul. 1987. *Asymmetries in Time*. Cambridge, MA: MIT Press.
Kutach, Douglas. 2011. "The Asymmetry of Influence." In *The Oxford Handbook of Philosophy of Time*, ed. Craig Callender, 247–75. Oxford: Oxford University Press.
Lewis, David. 1986a. "Causal Explanation." In *Philosophical Papers*, vol. 2, 214–40. Oxford: Oxford University Press.
---. 1986b. "Causation." In *Philosophical Papers*, vol. 2, 159–213. Oxford: Oxford University Press.
---. 1986c. "Counterfactual Dependence and Time's Arrow." In *Philosophical Papers*, vol. 2, 32–52. Oxford: Oxford University Press.
Loewer, Barry. 2007. "Counterfactuals and the Second Law." In Price and Corry 2007, 293–326.
---. 2012. "Two Accounts of Laws and Time." *Philosophical Studies* 160:115–37.
Maudlin, Tim. 2007. *The Metaphysics within Physics*. Oxford: Oxford University Press.
Paul, L.A., and Ned Hall. 2013. *Causation: A User's Guide*. Oxford: Oxford University Press.
Price, Huw. 2007. "Causal Perspectivalism." In Price and Corry 2007, 250–92.

---, and Richard Corry. 2007. *Causation, Physics and the Constitution of Reality*. Oxford: Oxford University Press.
---, and Brad Weslake. 2009. "The Time-Asymmetry of Causation." In *The Oxford Handbook of Causation*, ed. Helen Beebee, Christopher Hitchcock, and Peter Menzies, 414–46. Oxford: Oxford University Press.
Woodward, James. 2003. *Making Things Happen: A Theory of Causal Explanation.* New York: Oxford University Press.
---. 2006. "Sensitive and Insensitive Causation." *Philosophical Review* 115:1–50.