Lewis, D. (1979). 'Counterfactual Dependence and Time's Arrow', *Noûs*, 13: 455–76. Reprinted in *Philosophical Papers, Volume 2* (1986). Oxford: Oxford University Press.

Russell, B. (1913). 'On The Notion of Cause', *Proceedings of the Aristotelian Society*, 13: 1–26.

Schaffer, J. (2001). 'Causation, Influence, and Effluence', *Analysis*, 61:11.

_____ (2003). 'The Metaphysics of Causation', *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.), http://plato.stanford.edu/archives/spr2003/entries/causation-metaphysics/.

Schaffer, J. (2004). 'Counterfactuals, Causal Independence and Conceptual Circularity', *Analysis*, 64: 284.

13

Causation, Counterfactuals, and Entropy

MATHIAS FRISCH

13.1 Introduction

Bertrand Russell famously argued that the notion of cause has no place in modern fundamental physics, where it has been replaced by the concept of functional dependency (Russell 1918). Fundamental dynamical laws specify how one thing follows after another, but since these laws are time-symmetric, they do not support an asymmetric distinction between cause and effect of the kind that appears to be part of our commonsense notion of cause. In a recent paper (Field 2004), Hartry Field has endorsed Russell's conclusion but has pointed to a problem resulting from Russell's thesis. Even if we were convinced by Russell's thesis, we cannot simply excise any 'weighty' asymmetric notion of cause from our conception of the world, since, as Nancy Cartwright has argued (Cartwright 1979), just such a notion seems to be essential for the distinction between effective and ineffective strategies: In deliberating which actions will further our goals we need to appeal to a robust distinction between causes and their effects, for, intuitively, we can influence the occurrence of an event by affecting the occurrence of its causes but not by influencing its effects. In fact, Field maintains that trying to reconcile the apparent need for causation in a theory of effective strategies with Russell's thesis is 'the central problem

I would like to thank audiences in Sydney, at Reed College, the University of Munich, and the University of Konstanz for their comments and criticisms on parts of various earlier incarnations of this paper. I want to thank especially Adam Elga, Gerhard Ernst, Doug Kutach, and Huw Price, for helpful discussions and Barry Loewer for extremely detailed comments and criticisms on an earlier draft of this paper.

in the metaphysics of causation'. What, then, is the source of the causal asymmetry? How do we locate causation within a fundamental universal physics with time-symmetric laws? And if invoking the concept of cause in fundamental physics did indeed prove to be a 'relic of a bygone age', how is it that we have come to understand the world in asymmetric causal terms?

In this chapter, I want to examine one recent attempt at providing answers to these questions: David Albert's entropy-account of causal influence (Albert 2000, ch. 6).[1] Albert appears to agree with Russell's conclusion that the fundamental dynamical laws of physics neither support nor require a 'weighty' concept of cause, yet he argues that the fact that the universe had an extremely low-entropy past can account for our possession of such a concept. That is, according to Albert, the causal asymmetry can be explained in terms of an asymmetry of physical initial or boundary conditions. Yet Albert denies that there is a completely general asymmetry of causation or causal influence. Rather, he claims that the causal asymmetry is grounded in a counterfactual asymmetry that exists for small, yet macroscopic hypothetical interventions of the kind for which we humans could in principle be responsible. Thus, if Albert were right, there is no tension between fundamental physics and the demands of a theory of effective strategies despite the fact that the micro-dynamical laws *alone* do not support an asymmetric notion of cause. For the causal asymmetry would turn out to be due to an asymmetry of initial conditions and not due to an asymmetry of the dynamical laws; and the asymmetry would arise precisely for those kinds of possible macro-interventions that play a role in a theory of effective strategies. Field's problem would be solved.

Similar to David Lewis (Lewis 1986a, 1986b), Albert advocates a counterfactual analysis of causation. Thus, at the heart of his account is an argument for an asymmetry of counterfactuals appealing to entropy considerations, which in turn is meant to explain the asymmetry of causation. Yet Albert's account has the advantage over Lewis's that it does not invoke the dubious notion of miracles that can be ranked with respect to their sizes. Moreover, while Albert adopts Lewis's core idea that facts about the present allow us to

[1] Interestingly, Field himself also suggests that the solution to the problem may lie in recognizing the importance of statistical regularities to the concept of causation, similar to those that arguably can account for the concepts of entropy and temperature. As Field points out, such a solution to the metaphysical problem has the consequence that the causal asymmetry is absent on the micro-physical level.

make inferences about the past in a way different from inferences about the future—Lewis's notion of *postdeterminants* is echoed in Albert's notion of *records*—his account does not rely on the problematic (and in fact provably false!) thesis of an asymmetry of overdetermination between the past and the future.[2] Thus, Albert's account might be understood as offering a defense of Lewis's overall project that avoids some of the latter's deep problems.

In this chapter I want to argue, however, that Albert's thermodynamic account of the causal and counterfactual asymmetries is problematic as well. Section 13.2 will be devoted to introductory remarks in which I will draw some distinctions that will be important in the subsequent discussion. In Section 13.3, I will summarize what I take Albert's argument for the time-asymmetry of counterfactuals to be. I will criticize the account in Section 13.4, where I will argue that, as it stands, Albert does not offer compelling reasons to accept his account of the causal and counterfactual asymmetries and that, in fact, there are good reasons to reject at least part of the account. In Section 13.5, I will briefly discuss a version of an entropy account, inspired by Albert's account, that Barry Loewer defends in this volume. I will end with a brief conclusion.

13.2 Counterfactuals and Causes

Albert, like Lewis before him, argues for a time-asymmetry of counterfactuals: In certain standard contexts the future counterfactually depends on the past, but the past does not counterfactually depend on the future. Yet one might appeal to considerations similar to those advanced by Russell to argue that all scientifically respectable counterfactuals are time-symmetric. Take a physical theory with time-symmetric dynamical laws that pose a well-defined initial value-problem. Then we can both predict and retrodict the evolution of a system governed by that theory, given the state of the system at a certain time. Moreover, the laws do not only allow us to derive the evolution of an actual system, but also allow us to determine how the evolution would have been different had the system's 'initial' state been different, where it makes no difference whether the 'initial' state occurs before or after the state in which we are interested. Thus, the laws seem

[2] For criticisms of Lewis's account see (Elga 2001) and (Frisch 2005).

:o support both forward-looking and backtracking counterfactuals equally: If the state of the system at the initial time were different, both its past and its future would have to be different. Just as the future counterfactually depends on the present initial state, so apparently does the past.

Now, both Lewis and Albert believe that at least in worlds as complex as ours we do not evaluate the truth of a counterfactual by simply letting the relevant counterfactual present state of the world evolve in accord with the dynamical laws. Lewis appeals to a complicated similarity metric between worlds and maintains that the closest counterfactual worlds to ours are 'miracle worlds' diverging from the actual world, in which the laws of the actual world are not exceptionless truths; while Albert argues, as we shall see in more detail shortly, that our inferences about the past are also constrained by the hypothesis of a low-entropy past. Yet there clearly are standard scientific contexts in which we draw inferences about states of a system at different times but in which our reasoning does not presuppose a rich and complex world and not one with thermodynamic features. In those contexts we draw inferences based on special, highly idealized circumstances in which the system in question can be represented as a relatively simple, perhaps purely macroscopic system—for example, as a mechanical or electromagnetic system. And, *pace* Lewis and Albert, the a appropriate procedure for drawing counterfactual inferences in such cases can simply be to investigate the evolution of possible states of the system, given certain initial or final values and the relevant dynamical laws. For example, in the context of examining possible trajectories of balls on a billiard table, it might be correct to assert the backtracking counterfactual that if a certain ball had gone into the corner pocket, then it would have to have been struck differently from the way it actually has been struck; just as it might be the correct thing to say that if the ball were struck differently, then it would roll into the corner pocket.

Thus, there certainly appear to be contexts for evaluating counterfactuals in which forward-looking counterfactuals are not privileged. Nevertheless, I think that Lewis and Albert are correct in claiming that there *also* is a sense in which we think that the future but not the past depends on the present and, hence, that there also are contexts in which counterfactuals are time-asymmetric.

We appear to take the past to be counterfactually independent of the present in contexts which we intuitively think of as causal. On this point,

I think, there is relatively widespread agreement, even though accounts of the precise relation between causal and counterfactual claims differ widely. One way to spell out the connection between causation and asymmetric counterfactuals is in terms of the notion of hypothetical interventions: Interventions into a cause influence the occurrence of its effects, but intervening into a putative effect cannot influence the occurrence of its causes (see Pearl 2000, Woodward 2003). One might then try to distinguish between asymmetric, intuitively causal contexts, on the one hand, and symmetric contexts, on the other, by invoking a distinction between *closed* and *partially open* systems. Counterfactuals associated with closed systems appear to be symmetric: Each set of possible initial conditions at a time defines a different closed system whose past and future evolution is given by the relevant dynamical laws. Systems with different initial states will in general have both different futures and different pasts. By contrast, counterfactuals associated with interventions from the outside into an otherwise closed system might be thought to be asymmetric, since interventions may be taken to affect only the future evolution of the system but not its past.[3]

Two aspects of this scheme are worth being made explicit. First, the scheme is most naturally spelled out as not advancing an account of the causal asymmetry. *Intervention* arguably is itself a causal notion and the account simply stipulates that interventions influence the causal 'future' of a system and not its causal 'past'. Second, it appears to be crucial to this way of thinking about causation that causal systems have an environment or 'outside' from which interventions can occur. This may suggest the Russellian claim that there is indeed no room for the notion of cause in a *universal* physics that aims to have models of the universe as a whole among its class of models.[4]

In sharp contrast with this scheme, Albert proposes an account of 'intervention' that applies to closed systems as well, thereby promising to provide a place for causal notions even within a universal physics. Hypothetical interventions, on Albert's account, are treated by postulating counterfactual initial states, where both future and past evolutions of the

[3] The notion of hypothetical interventions functions in some ways similar to Lewisian miracles, with the advantage that an interaction of the system with its environment takes the place of counterlegal time-evolutions.

[4] See, for example (Hausman 1998).

system in question are then determined by what Albert calls our *normal procedures of inference* (NPI), which include use of the fundamental dynamics, but do not invoke Lewis-style miraculous violations of the laws. Instead of appealing to a difference between open and closed systems, Albert's account suggests that the difference between symmetric and asymmetric counterfactuals is due to a difference between thermodynamic systems and non-thermodynamic systems. In the former case, Albert argues, a counterfactual present state that differs only locally from the actual present will be overwhelmingly likely to have evolved from a past identical to the actual past.

A central *explanandum* for Albert is, as he puts it, our 'fundamental conviction [...] that the future *depends on what happens now*—that the future depends on what we *do* now—in a way that the past does not' (Albert 2000, p. 125).[5] Thus, what he wants to 'get to the bottom of' (*ibid.*) is how it is that we have come to have a certain conception of the world. He takes for granted that our conception includes the notion of an asymmetric causal dependence of the future on the present and his account is meant to offer an explanation of this aspect of our conception.

What would it be for an account to be successful in getting to the bottom of why we conceive of the world in time-asymmetric causal terms? One possible explanation of our belief in an asymmetry of causal dependence would be to give a genetic account of our conviction. The aim of such an account could *either* be to show how humans in general have actually developed a certain conception of the world *or* to offer a developmental story of how individual human beings come to acquire such a conception during childhood. Another kind of explanation of our conviction would be to propose a philosophical reconstruction that does not trace the actual mechanisms that have produced our conception but instead simply argues that being able to draw certain distinctions is advantageous to beings like us. If one could show that conceiving of the world in certain ways was to our advantage, that might go some way towards getting to the bottom of our having a certain conviction, since this might be part of an evolutionary account of why it is that we might have developed a certain conception. Finally, our concern might be epistemological and we might be interested in showing that, *contra* Russell, at least part of our causal conception of the

[5] The italics in this and in all subsequent quotes are in the original, unless otherwise noted.

world is justified or legitimate in that it can be grounded in considerations arising from fundamental physics.

Which of these different projects is Albert engaged in? Unfortunately Albert is not entirely clear on this issue. Partly his concern seems to be epistemological in that he appears to be interested in delineating to what extent our conviction in an asymmetry of dependence is justified. Understood in a certain sense, our conviction is clearly false, Albert would say: If we assume a deterministic physics with time-symmetric laws, then the past is determined by the present just as the future is. But Albert then argues that there also is a sense in which our conviction is supported by a kind of counterfactual reasoning we engage in, which in turn is grounded in the entropy-asymmetry. Yet there are also central aspects of Albert's account that are quite explicitly and unambiguously aimed at explaining how we actually go about making certain inferences—inferences involving records of the past—that are at the heart of our belief in an asymmetry of dependence. I will address the question to what extent Albert might succeed with these different projects in Section 13.4.

What exactly *is* our fundamental conviction? Despite the ease in which Albert moves from the main sentence in the above quotation to the sub-clause set off by dashes, they in fact express two quite distinct convictions. On the one hand, we believe in a general causal dependence: Events have causes and these causes are in an event's past, or at least not in its future—quite independently of whether or not we take the events in question to be under our control. On the other hand, there is our conviction that our own actions can influence the future but not the past. Different accounts of causation will differ on the proper relation between these two convictions. On many views, agency is itself a causal notion and the asymmetry characterizing our actions is merely a sub-class of a more general causal asymmetry. That is, on such a view an account that got to the bottom of our convictions would have to explain our belief in a general asymmetry of dependence and the belief in an asymmetry of agency would simply follow as a special case. By contrast, so-called *agency accounts* of causation, such as the ones defended by Menzies and Price (1993) and Ahmed (ch. 6 in this volume), argue that the notion of agency is primary and that our general causal convictions are the result of our tendency to project the asymmetry of agency onto the objects, in a manner similar to what is often held to occur in the case of secondary qualities.

Even though Albert himself does not raise this issue explicitly, it is obvious that his account squarely falls into the first category. For he argues that in worlds like ours 'more or less *any present feature of the world you can think of*' can amount to a '*causal handle*' on 'more or less any future one' (Albert 2000, p. 128), simply because the relevant counterfactuals that ground our concept of causal dependence come out true. That is, on Albert's account there is a broad class of counterfactuals—those with antecedents that postulate small macroscopic changes to the present—that are time-asymmetric because what he calls 'our normal procedures of inference' for evaluating such counterfactuals yield time-asymmetric results. Our procedures of inference presumably involve calculating probabilities of past or future macro states conditional on present macro states. This asymmetry accounts for the time-asymmetry of causal dependence and, in particular, the asymmetry of influence.

By contrast, the entropy account defended by Loewer (ch. 11 in this volume) privileges counterfactuals involving possible human interventions, similarly to agency accounts of causation. Loewer argues that statistical mechanical considerations entail in the first instance that so-called 'decision counterfactuals' are time-asymmetric. This asymmetry of decision counterfactuals is then 'spread over the objects': Other counterfactuals, not postulating human decisions, are time-asymmetric, on this account, in virtue of the fact that we take their antecedents to be the result of small micro events that are analogous in size to human decisions.

On Albert's account, the counterfactual 'If the billiard ball were 20 cm to the right of its actual location, then the past evolution of the balls on the table would have to have been different from the actual evolution' comes out false, because of the result of applying our normal statistical procedures of inference to the counterfactual antecedent state. By contrast, on Loewer's account, the counterfactual comes out false, since we treat the change in the ball's present location as if it were the result of a human decision (and intervention). In this chapter I will focus mainly on Albert's entropy account of causation. I discuss Loewer's account briefly in Section 13.5 and more fully in Frisch (forthcoming).

13.3 Albert's Argument

Instead of using the more familiar locution of 'the cause of an event', Albert introduces the notion of one event being a 'causal handle' for another. In light of the remarks in the preceding section it may come as somewhat of a surprise, however, that he suggests that one can introduce this notion even in the absence of any thermodynamic considerations. Albert points out that if we constrain the *remote* past of any physical system, then only very special alterations of the present can lead to a different *recent* past, while many alterations of the present may lead to a different future. To illustrate this point he asks us to consider a collection of billiard balls on a frictionless plane such that ball 5 is currently stationary with the additional constraint that ball 5 was moving 10 seconds ago. Given the additional constraint, that ball 5 has been involved in a collision in the past 10 seconds is determined by facts about the present state of ball 5 *alone*. That is, alterations to the present state of the balls *not* involving changes in the state of ball 5 cannot change the fact that ball 5 was involved in a collision during the last 10 seconds. (In fact, there will be many alterations to the present not involving ball 5 that will result in a present state inconsistent with the additional constraint on the past.) Yet there are many changes to the state of the balls not involving ball 5 that will result in a different future evolution of ball 5. From this Albert concludes that there are a far wider variety of 'what we might call *causal handles* on the future of the ball in question here, under these circumstances, than there are on its past' (Albert 2000, p. 128).

An obvious objection at this point is that evaluating counterfactual situations compatible with an apparently *ad hoc* time-asymmetric constraint on the past, but not the future, tells us nothing about the causal structure of the case and, in particular, cannot license any conclusions about an asymmetry of causal influence. Why are 'these circumstances' the right circumstances for assessing the causal structure? The asymmetry obviously is the result of imposing an asymmetric constraint on possible alterations. If instead we were only interested in possible changes to the present state of the system of billiard balls that are compatible with an additional constraint on the *future* evolution of the system without in any way constraining the past, then many more backtracking than forward directed counterfactuals would presumably come out true. Why then should we impose a constraint

only on the past? One answer might be that the past is fixed while the future is open. Alternatively, we could maintain that we ought to keep the causal history of the event fixed, but not its future effects. But obviously these answers would beg the question, if our aim is to give a counterfactual analysis of the notion of causal influence.

Ultimately, however, Albert is not interested in non-thermodynamic systems such as the idealized billiard balls. Rather, his aim is to argue that the counterfactual asymmetry arises in systems that are complex enough to exhibit thermodynamic features. Thus, Albert himself would probably agree that introducing the notion of a causal handle simply by reflecting on the consequences of imposing a time-asymmetric constraint on the motion of the billiard balls is somewhat misleading. The new ingredient in the case of thermodynamic systems is that there is a special asymmetric constraint on the past—the hypothesis of an extremely low-entropy initial state. This condition, which Albert calls 'the *past-hypothesis*', is a central assumption in standard accounts of the thermodynamic asymmetry that the entropy of a closed system never decreases. There it is needed to avoid the *reversibility objection* against the most straightforward attempt of accounting for the increase in entropy. While one can argue that the entropy of a given low-entropy macro-state increases, assuming an intuitively plausible probability distribution over micro-states and a Newtonian time-symmetric micro-dynamics, the same type of argument also allows us to conclude wrongly that the present macro-state is at a local entropy minimum and evolved from a higher entropy past. The undesirable retrodiction that entropy decreased in the past can be blocked, if the distribution of micro-states is conditionalized not only on the present macro-state but also on a low-entropy past and, ultimately, a low-entropy initial state of the universe.

Now, why should we keep the past-hypothesis fixed in assessing counterfactual changes to the present? How is this hypothesis different from the apparently question-begging assumption to hold the past state of billiard ball 5 fixed? We may, as Albert claims, have good inductive reasons to believe that the past-hypothesis is satisfied in the actual world. Yet there is much that we know inductively about the future in the actual world that we do not keep fixed in assessing counterfactual changes to the present. Albert's answer is that in assessing the truth of counterfactuals we need to consider other worlds that are in important ways like ours. In particular, the counterfactual worlds to which we appeal in assessing the results of

counterfactual changes to the present have to license the same normal procedures of inference as the actual world does. And Albert argues that these procedures rely crucially on assuming the truth of the past-hypothesis. If all counterfactual reasoning must rely on our normal procedures of inference from the state of the world at one time to the state at other times and these procedures presuppose the past-hypothesis, then all counterfactual reasoning must presuppose the past-hypothesis.

In somewhat more detail, Albert argues the following. In order to assess the truth of a counterfactual, where the antecedent involves some small yet macroscopic hypothetical alteration to the actual world, we need to look at counterfactual worlds that are like the actual world except for the small change and then use our normal procedures of inference to determine the past and future evolutions of such worlds. In the case of forward-looking counterfactuals, these procedures amount to taking the present macro-state of the world, assuming an equi-probability distribution over micro-states compatible with that macro-state and evolving the state forward in accord with the dynamical laws. Thus, counterfactuals such as 'If the light switch were flipped, then the light would go on' come out true (assuming that in the actual world the light is and remains off and I don't flip the switch), if most micro-states compatible with the initial macro-state evolve into micro-states corresponding to a macro-state such that the light is on.[6]

However, backtracking counterfactuals such as 'If the light switch were flipped, then the light would have to have been on prior to that' are not supported by our normal procedures of inference.[7] If we simply evolved the counterfactual present state backward in accord with the macroscopic

[6] Note that given his appeal to procedures of inference, one might think that, unlike Lewis, Albert is offering *assertibility* conditions of counterfactuals and not *truth* conditions. Yet Albert concludes his discussion of counterfactuals by saying 'And it follows—if all this is right—that the future does indeed counterfactually depend on what we do now, and the past [...] does not' (*TC*: 130). This suggests that Albert is committed to the view that whatever follows from our normal procedures of inference is true.

[7] This example is not Albert's own, who also follows Lewis in trying to stack the deck through his choice of examples. Albert considers the forward-looking counterfactual 'If the president pushed the button, there would be a nuclear explosion' and contrasts it with the backtracking counterfactual 'If the president pushed the button, then there would have been an explosion'. He says that there 'are (for example) no worlds *at all*, even *remotely* like our own, in which the [normal procedures of inference] translate small hypothetical present differences in the present position of anybody's finger into differences between a certain thermonuclear device's exploding or not exploding two minutes *ago*' (*TC*: 129–30). While it might well be that this particular backtracking counterfactuals comes out false, this of course does not show that backtracking counterfactuals are false in general. The passage continues as follows: 'And that (as I said before) is precisely because there is a past-hypothesis and not a future one. That (to put it another way) is because there are—vis-à-vis such things as the *past* explosion

regularities with which we are familiar, then the light's being off after the light switch was flipped presumably would have to have been preceded by the light's being on earlier. For generally, flipping the switch is accompanied by a change in the light's state from on to off or vice versa. And this fact would seem to allow us to draw inferences in a time-symmetric manner. We can, it seems, infer the *past* state of the light from its present state and the fact that the switch was flipped, just as we can infer its *future* state from its present state and the fact that the switch was flipped. But Albert argues that this inference would violate the presupposition that there are *records* of the past. In our example, records of the light's being off in the actual world might include the fact that there is no light that has recently escaped through the window and that the light bulb is relatively cold. Albert's notion of record plays a role analogous to that of Lewis's postdeterminants: A record is a relatively localized fact about the present from which we can infer the occurrence of some event in the past. If the present contains a record of the light's having been off, then we can infer that the light was in fact off. Like Lewis, Albert believes that we take many of the local facts we do hold fixed to be sufficient for the occurrence of certain (relatively localized) events in the past. Yet, unlike Lewis, Albert is well aware that there are no local facts about the present which *alone* are sufficient for the occurrence of some past event. Rather, inferences appealing to records, Albert holds, are always inferences from facts at two different times to a fact at a third time in between, since facts cannot function as records of the past unless we assume something about the more remote past that functions as a 'ready condition'.

Recall Albert's example of the collection of billiard balls. Albert points out that the fact that ball 5 is currently stationary functions as a record of the fact that the ball underwent a collision in the last ten seconds *given* the additional constraint that ball 5 was moving ten seconds ago. In other words, that the ball was moving 10 seconds ago functions as ready condition allowing us to record the ball's collisions in terms of its present state of motion. Without the ready condition we cannot infer whether ball 5 underwent a collision from the current state of ball 5 alone but would need to know the present state of the entire collection of balls.

---

of thermonuclear devices (or the lack of them)—such things as *records*, as *memories*' (*ibid.*) I will criticize equating the truth of the Past Hypothesis with the existence of records later.

What does all this have to do with the past-hypothesis? Ultimately, Albert claims, the single assumption that on its own can ensure that we can treat facts about the present as records of the past is the past-hypothesis. In treating a fact as record we need to presuppose that some ready condition in the more remote past was satisfied. But how do we know the latter fact? Again, we need some record of the ready condition's being satisfied, which in turn requires an even earlier ready condition. According to Albert, this regress ends with the past hypothesis, which functions as a first 'mother of all ready conditions'. That is, Albert's amended version of Lewis's thesis of overdetermination of the past by the present is that *given the past-hypothesis*, localized facts about the present are records of the occurrence of certain localized facts in the past. Thus, while the light bulb's being cold alone is not sufficient for the light's having been out, it is sufficient, or at least is overwhelmingly probable, on Albert's view, in conjunction with the assumption that our universe had an extremely low entropy past.

The structure of Albert's argument, then, is this. Albert argues for two claims:

(*i*) The past-hypothesis is true, if and only if there are records of the past.

(*ii*) If there are records of the past, then there is no counterfactual dependence of the past on the present.

From (*i*) and (*ii*) Albert's first conclusion follows:

(*iii*) If the past-hypothesis is true, then there is no counterfactual dependence of the past on the present.

If we add to this a counterfactual account of causation, we arrive at the final result that if the past hypothesis is true, then the future, but not the past, causally depends on the present—that is, in Albert's own terminology the present contains *causal handles* on the future but not on the past.

At the heart of the account is the idea that the past counterfactually depends on the present, if there is an actual present event c and actual past event e such that our normal procedures of inference license us to accept the counterfactual 'If c had not occurred e would not have occurred.' This raises the question as to how exactly according to our normal procedures of inference we evaluate the truth of counterfactuals. Yet given the question's

central importance to Albert's account, it is surprisingly difficult to find a precise statement of Albert's answer to this question. The most plausible proposal (which has been implicit in some of what I said above) appears to be that our normal procedures of inference involve calculating conditional probabilities. On this proposal the truth of the counterfactual 'If c had not occurred, e would not have occurred' is assessed, on Albert's account, by looking at the class of worlds that satisfy the past-hypothesis and whose present macro-states match the macro-state of the actual world as much as possible, given that c does not occur in those worlds. The past and future of these worlds is then determined by the micro-dynamical laws. If in most of these non-c worlds e does not occur, then the counterfactual is true, otherwise it is false.[8]

Thus, on what I take to be the most plausible reconstruction of Albert's account, the past does not counterfactually depend on the present, exactly if for all (suitably localized and small) actual present macro-events c and for all past macro-events e the following condition is satisfied: The conditional probability of es not occurring is extremely low, *given* the micro-dynamical laws, the past-hypothesis, and that the present is unchanged except for cs not occurring;[9] that is, if the past does not counterfactually depend on the present, then

$$(1)\ \Pr(\sim e/\sim c\&S_a\&PH) \approx 0$$

for all actual events c and e such that e is in the past of c, where PH is the past-hypothesis. Here I am taking events to be the goings-on in some region of space at a particular time. The event c is the complete actual present macro-state in some suitably small region of space. S_a is the remainder of the present macro-state of the world. Thus, $c\&S_a$ is the complete present macro-state of the world. Also I am assuming that $\sim c\&S_a$ is nomically possible—that is, that the occurrence of c is not implied by S_a together with synchronic constraints imposed by the laws.

[8] A similar scheme for assessing assertibility conditions for counterfactuals is advocated in (Kutach 2002).

[9] One might worry that since conditional probabilities come in degrees it is not clear how this can result in conditions for the truth or falsity of backtracking counterfactuals. But since the relevant thermodynamic probabilities are usually either absurdly small or very, very close to one, this might perhaps license Albert's conclusion that 'the future does indeed counterfactually depend on what we do now, and the past [...] does not'.

13.4 Criticism

13.4.1 Records and Ready Conditions: The Puzzling Regress

Let us assume for the moment that Albert can indeed establish claims (*i*) and (*ii*). Would that be enough to explain our fundamental conviction that the future depends on the present, but the past does not? Our conviction, on Albert's account, is due to how we reason counterfactually. According to (*ii*) there is a connection between counterfactual reasoning, on the one hand, and memories or records, on the other. Quite plausibly, such a connection can indeed explain why there are certain standard contexts in which we take the past to be counterfactually independent of the present. We have memories and records of the past but not the future; and this fact might very will be the reason for why we take the past to be counterfactually independent of the present. For arguably, we keep the past, but not the future, fixed in assessing the results of counterfactual changes to the present, precisely because we have memories of the past.

Now, even the connection between the existence of memories and the fixity of the past is not completely obvious and straightforward. For it does not strictly follow from the fact that we keep all those events in the past fixed of which we have memories or records that we should keep the *entire* past fixed when reasoning counterfactually. Hypothetical changes to the present do not affect what we *know* to have happened, we might think, but what about past events of which there are no records or traces? Or past events of which we do not know there to be traces? Or past events whose traces are unknown to us? It is conceivable that these cases could have been treated differently in our reasoning practices, and an account that promises truly to get to the bottom of our fundamental conviction ought to offer an explanation for why our conviction apparently does not distinguish between them. Why do we believe that the entire past is independent of what we do now and not just those events of which we have memories? But perhaps the fact that we take the entire past to be counterfactually independent of the present can simply be explained by a tendency of ours to generalize from those events of which we have records or memories—and which we therefore keep fixed—to the past as a whole.

More problematic than Albert's account of the role of records in counterfactual reasoning is his appeal to a low-entropy past to explain our

inferential practices involving records. According to Albert, 'the reason there can be records of the past and not of the future is nothing other than that it seems to us that our experience is confirmatory of a past-hypothesis and not a future one' (Albert 2000, p. 118). And indeed, that the past-hypothesis holds is, for Albert, strictly equivalent to the fact that there are records and memories of the past; the second claim is just the first when we 'put it another way' (Albert 2000, p. 130). Thus, Albert's account might be intended to offer at least a partial explanation of why it is that we engage in reasoning based on records and in causal reasoning. First, such reasoning would be impossible in a world without a past-hypothesis; and second, the fact that our reasoning based on records can be reconstructed as arguments that assume the past-hypothesis as premise shows why it is advantageous to possess certain conceptual distinctions or to have the ability to draw certain inferences. Albert appears to argue that our ordinary reasoning practices appealing to records happen to latch on to an important feature of the world—a feature that becomes evident once our inferential procedures are reconstructed as involving an appeal to the past-hypothesis. This purported insight might then be cited as part of an evolutionary explanation for why beings like us might have developed certain reasoning practices.[10]

But this does not yet fully capture what Albert says that the role of the past-hypothesis in our reasoning is. As we have seen, Albert argues that every inference based on records is, as a matter of fact, an inference from two times—from a record and an earlier ready condition—to a time in between: 'The sort of inference one makes from a recording is [...] from *two* times to a *third* which lies *in between them*' (Albert 2000, p. 117). Thus, anyone using a fact about the present as record of the past needs to possess some information about the more distant past as well. The puzzle for Albert then is 'how it is that we can ever manage to *come by* such information' (Albert 2000, p. 118). This additional information needed to draw inferences based on records cannot be obtained by means of retrodiction, Albert points out, because otherwise such inferences would ultimately be reducible to standard retrodictions from 'initial data' and the laws. Thus, he argues, the fact that some ready condition obtained itself has to be known through a further record—a record which in turn requires knowledge of a second

[10] Loewer (ch. 11, in this volume) is much more explicit about the fact that he wants to give an account of our concept of cause that shows its evolutionary advantages.

ready condition. This, according to him, leads to a regress which can be blocked *only if* there is 'something we can be in a position to *assume* about some other time', where this other time 'must be *prior in time* to everything of which we can potentially ever *have* a record, which is to say that it [that is, the 'something' we have to be in a position to assume] can be nothing other than the initial macrocondition of the universe as a whole' (Albert 2000, p. 118). Thus, according to Albert there is a puzzle concerning how we '*come by*' information regarding ready conditions, or how ready conditions are 'established' and this puzzle results in a regress that can only be blocked by positing the past-hypothesis as 'mother of all ready conditions'.

But what exactly is the puzzle? I think the only way to interpret what Albert actually says without doing too much violence to the text is to read him as raising a puzzle concerning how we *actually* reason. If inferences based on records are in fact inferences from two times to a third time in between, then nobody can actually draw inferences from putative records without having information concerning the relevant ready conditions and Albert's account is meant to explain how we arrive at this additional information. Albert introduces the role of ready conditions by discussing the billiard ball example that we encountered above: the fact that ball 5 is currently stationary is *by itself* not a record of anything and knowing this fact alone does not put us in a position to infer anything about the ball's past. The present state of ball 5 only becomes a record of past collisions when it is conjoined with the additional fact that the ball was moving ten seconds ago. Thus, we simply cannot draw any inferences about the past (beyond mere guessing) based solely on the 'record state' of ball 5 and without having some belief or making some assumption about the ready state of our 'recording device'. The example illustrates Albert's view that the record relation is a function from two variables—a record variable and a ready state variable—to a variable representing the recorded state. To compute the value of the variable representing the recorded state we need the values of *both* other variables as inputs. If we are concerned, as Albert is, with the question as to how we can make inferences about times other than the present based on what we know about the present, then it might indeed seem quite puzzling how we, as a matter of fact, come to possess information about the value of variables representing earlier ready conditions.

How else might we understand Albert's puzzle? Perhaps an alternative interpretation, of the puzzle is as a puzzle about justification. On this

interpretation, Albert's worry is not how we as a matter of fact establish or come by certain ready conditions but how we *justify* our assumptions about ready conditions and, therefore, our inferences based on records. Admittedly, this second interpretation does some violence to the text. The questions as to how we ever *manage to come by the information* that ball 5 was moving ten seconds ago and how a ready condition is *established* seem straightforwardly to be concerned with the question as to how we as a matter of fact can obtain certain information about the past rather than with how we justify our beliefs about past ready conditions. Nevertheless, I want to explore this second interpretive option in what follows as well.

Whatever the correct interpretation of Albert's puzzle is, it is clear that the role the past-hypothesis is meant to play in its solution is one *in addition* to being a necessary condition for the possibility of records and in addition to explaining why reasoning based on records can be valuable in virtue of its tracking certain features of the physical world. For whatever the role of the past-hypothesis as 'mother of all ready condition' is in solving his puzzle, it is a role that Albert worries we may instead, albeit mistakenly, be tempted to assign to the ready state of our brains during normal sense perception (see Albert 2000, p. 118, footnote 6). But the claim that it is a necessary condition for successful reasoning based on records that our brain be in some appropriate ready state does not conflict with the claim that the past-hypothesis is a necessary condition for the possibility of records. And if Albert's aim was (merely) to establish that, as Loewer puts it in Chapter 11 (p. 323) of this volume, 'the information expressed by SM-counterfactuals [and by records] is important for us because it tracks the statistical mechanical probability distribution' (Loewer this volume), then an appeal to the ready state of the brain as mother of all ready conditions, as alternative to the past-hypothesis would not only be mistaken, as Albert claims, but would simply be nonsense. By contrast, an assumption about the condition of our brains might at least with some initial plausibility be offered as an alternative answer both to the question as to how we as a matter of fact go about to establish ready conditions and to the question as to how we might justify our inferences based on records: we as a matter of fact come by information about the past by consulting our memories and we justify beliefs about the past by appealing to the veridicality of our memories. Albert's footnote appears to be directed against such a view.

Albert's paradigm example of a record, as we have seen, is the present state of motion of a billiard ball, which can function as a record of past collisions, given the ball's earlier state of motion as 'ready condition'. It is obvious in this example that any inference to the recorded state has to be an inference from both the record state and an earlier ready state. Yet this is less obvious for a different kind of record—our memories of past events. In the former case, what the content of the record is—what the record state is a record *of*—is not given by the record state alone but only in conjunction with the ready state. By contrast, memories are intentional states and it is part of the content of a memory what it is a memory of. Thus, there is an important difference between the state of the billiard ball as record and our memories: in the former case but not in the latter a ready state is necessary to fix the content of the record.

One might think that, nevertheless, even inferences based on memories are inferences from two times to a third in between. Even in the case of memories, one might hold, an appeal to a ready condition is implicit, since we only infer p from a memory that p when we believe the memory to be veridical; and we believe a memory to be veridical only if we believe that our brain was in some appropriate ready state when we perceived that p. Yet even if this line of reasoning were correct, the appeal to ready conditions would fulfill two quite distinct roles in the two cases. In the cases of memories its only job is to support the claim that our memories are veridical, while in the billiard case the ready condition is needed to fix the content of the record. I doubt, however, that we generally do reason from our memories along these lines. Generally, it seems to me, we simply assume that our memories are veridical and are, as a matter of fact, neither concerned with establishing that our brains were in an appropriate ready condition nor interested in coming by the relevant information about the appropriate ready conditions. Only in situations where *doubts* about the veridicality of our memories arise do we appeal to information that suggests that our brains were in appropriate ready states at the time when we perceived whatever it is we take ourselves to remember. That is, unless there are doubts about the veridicality of our memories, the inferences in which we actually engage make no appeal to ready conditions and simply have the form 'I have a memory that p; therefore, p.'

At the very least, then, how we think of the puzzle concerning ready conditions has to be qualified in important ways. Yet one might think

that my considerations leave Albert's main claim intact: there still appears to be a puzzle how we come by relevant information about earlier ready conditions. In the case of the billiard balls the information is needed even to fix the content of the record state and to be able to draw any inference at all from that state; and in the case of memories it is needed at least when we are asked to justify our belief in the veridicality of a particular memory.

What then is Albert's solution to the puzzle? First, Albert's negative claim is that there is simply *no* solution to the puzzle at the first level, as it were. Each ready condition itself needs to be established by means of a further record, which requires yet another ready condition. This leads to a regress, which, as Albert says, 'obviously' would go on ad infinitum. It is impossible to stop the regress for any particular record short of positing a 'mother of all ready conditions' at the time of the early universe: 'this mother', he says, '*must* be *prior in time* to everything of which we can potentially ever have a record' (Albert 2000, p. 118, first italics are mine). Thus, Albert's second, positive claim is that the puzzle can be solved by an appeal to the past hypothesis as mother of all ready conditions.

But as an account of how we actually reason Albert's account surely must be mistaken. Most people (and, I want to submit, all people most of the time) do not assume any initial macro condition of the universe when they engage in reasoning appealing to records of the past—let alone *the* actual initial macro condition. And most people are not at all in a position to assume what that initial condition might be. According to Albert, reasoning appealing to records is possible precisely since 'it seems to us that our experience is confirmatory of a past-hypothesis.' (Albert 2000, p. 118) But to most people it does not seem that this is so, for the simple reason that most people do not possess the relevant concepts. For something to seem to us *that* it is thus and so, we must be able to form a coherent thought concerning it being thus and so.[11] But quite plausibly most people do not grasp the content of the past-hypothesis or of the second law of thermodynamics, and in fact the content of their experience may seem to them to be

[11] What, on Albert's account is the relation between our being in a position to assume the past-hypothesis, our in fact assuming the past-hypothesis, and it seeming to us that our experience is confirmatory of the past-hypothesis? Albert does not tell.

confirmatory of a quite different past than that postulated by modern cosmology. Nevertheless, most people arguably engage in reasoning appealing to records.[12]

If Albert's account fails to provide an answer to the question as to how we obtain information about ready conditions in cases like that of the billiard balls fails, what then is the answer to the puzzle? It seems to me that if all records were like the position of the billiard balls there would be no solution to the puzzle and reasoning from records would simply be impossible. In the case of the billiard balls, Albert's puzzle seems inevitably to result in a regress. How else could we possibly come by the necessary specific information about the ready state, if not either by means of retrodiction or through yet another record? But there are records that differ from this case in crucial respects. Consider, first, record states with intentional content, such as memories. For such records we need to be in a position only to assume that the ready state *whatever it may have been* was a state of a kind that normally results in veridical records. It seems less plausible that such a very general assumption about ready states needs to be established by means of a further record. More plausible, I want to suggest, is that the general veracity of our memories is one of our most fundamental convictions that underwrites much of our reasoning about the world and that, when challenged, we support our belief that a certain memory is veridical by appealing to how this memory *coheres* both with other memories we have and our current perceptions. These other memories might be of earlier states of affairs, but often they will concern other events simultaneous with the event remembered.

Moreover, the regress can be blocked even for many record states without intentional content. Often, it seems to me, when we take certain facts about the present to function as records of the past, we assume that the recording system in question had earlier been in its 'normal' or 'typical' state, which

[12] Given how obvious it is that an appeal to the past-hypothesis cannot be part of our actual reasoning practices involving records, one may wonder if I have misinterpreted Albert. And, indeed, in personal discussions Albert has insisted that he is not proposing his theory as an account of our actual reasoning practices. Yet it is difficult for me to see how else one might interpret the passages cited above. The puzzle concerning ready conditions raised by the billiard ball example and expressed by the question as to how we 'come by' information about ready conditions appears unequivocally to be a puzzle about how we do as a matter of fact obtain information relevant to drawing inferences based on records.

functions as ready condition. And what we assume the normal state of the system to be is given by our past experiences with systems of the kind at issue. Thus, when we observe diverging ripples on a pond, we take this to be a record of the fact that some wave source, like a stone, had just broken the surface of the water. Here we implicitly assume that there were no coherent concentrically converging waves present on the pond prior to the source; and we make this assumption not necessarily because we have a record of the pond's having been still, but because our past experience with wave media suggests that there are no coherent waves in the absence of sources.

One worry about this sketch is that in appealing to past experiences of what constitutes a 'normal' ready state of a system, the account relies on our already having information about the past, yet how this is possible is precisely what is at issue. How can we, in the present, have information about the 'normal' states of systems in the past? Don't we already have to rely on records or memories of past experiences to assume certain 'normal' past states as ready conditions? But I want to suggest that (at least in principle) we can think of information about past normal states as being inferred from the totality of our present experience. Certain assumptions about the past are part of the best explanation of the totality of our present experience (including our memories) and of the success of our predictions about how future observations will turn out. It seems that at least sometimes we come by information about the past by positing past states that best account for our present beliefs, experiences and memories.

Moreover, we can use the same considerations to *justify* beliefs about the past. That is, our beliefs in normal ready states (such as the belief that there are no coherent waves on ponds prior to the presence of wave sources) are supported by their 'conspicuous success [...] in making *predictions* about how *future* particular observations are likely to *come out*, [...] and because [they manage] to render various of our *other* most fundamental convictions (about the veracity of memories [...]) compatible with one another'. Here I have simply adopted Albert's explanation of how our belief in the past-hypothesis is grounded in an inference to the best explanation (Albert 2000, p. 119).[13] Thus, I want to propose that we can use the very

[13] Just before the passage quoted is, by the way, the only place in the section on pp. 117–19 where Albert speaks of the question of the justification of our beliefs. Albert here asserts both that what *justifies* our belief in the past-hypothesis is an inference to the best explanation and that *we believe* in the hypothesis for these reasons.

same inferential procedures that might be used to support a belief in the past-hypothesis to support beliefs in much more recent ready states: In both cases our assumptions about the past are supported by the fact that they make the best sense of aspects of our present experience and are predictively successful.

Now, some of what Albert says suggests that we can use this inference procedure *only* to justify our belief in the past-hypothesis and that it is strictly impossible to infer anything about more recent ready states in this manner. For, as we have seen, he holds that in order to establish *any* ready condition there '*must* be something we can' assume, which '*must* be *prior in time* to everything' of which we can have a record (Albert 2000, p. 118, the first two emphases are mine). But it is difficult to see why this inference technique supposedly works 'globally' and for the totality of our experience, but not more 'locally' for certain aspects of our present experience. Why should we be able to infer a low-entropy state of the early universe as best explanation of the totality of our experience, yet be unable to infer that ponds are 'normally' still as best explanation of our present experiences (including putative memories and records) as far as they relate to small bodies of water?

On Albert's behalf, one might try to point to the fact that the past hypothesis is needed to guard against Loschmidt's reversibility objection: Unless we assume a low-entropy past, the most likely past evolution is a fluctuation out of equilibrium. But again, it is unclear why we cannot use more 'local' inferences to the best explanation to establish merely that the more recent past ought to have been one in which entropy behaved appropriately such that ripples on a pond could indeed function as records of past wave sources.[14] In addition, Albert might argue that the past-hypothesis is 'more fundamental' in that belief in it can provide the best, most unified explanation of the totality of our experience. For, he maintains, everything we know of the world can be deduced from what we know about the world's present macrocondition; 'the standard microstatistical rule; the dynamical equations of motion; the past-hypothesis' (Albert 2000, p. 119). But even if Albert could show that the past-hypothesis *can* function as the 'mother of all ready conditions', this does not imply that, as he claims, the

[14] To what extent the asymmetry of wave phenomena is related to that of thermodynamics is a difficult question. For a discussion of this issue see (Frisch 2005).

past-hypothesis *must* be assumed in any inference involving records. Thus, inference procedures similar to the one to which he appeals ought to be able to provide us with direct information (not relying on knowledge of the 'mother') about the 'mother's offspring' as well and Albert's regress argument can be blocked at its very first steps.

Thus, it appears that we do not need to appeal to the past-hypothesis in justifying our beliefs at least about the recent past. Finally, there is also a worry about appealing to the past-hypothesis as ultimate justification of our beliefs about the past. Albert presents the following skeptical problem. According to Loschmidt's reversibility objection, the most likely evolution of the present state of the world is as a fluctuation out of equilibrium. How is it then, that we could ever come to have any reason to believe in the past-hypothesis? One might think that we can know of a low-entropy past simply through records or memories of the past. Yet any putative record is itself already part of the present state and, thus, just like anything else in the present appears to be most likely the result of anti-thermodynamic fluctuations. Albert's solution to the problem is two-fold. On the one hand, he argues that we can have inductive grounds for accepting the past-hypothesis. For we can make predictions from the assumption that the universe had a low-entropy past which we can confirm. On the other hand, and as Albert says, 'more profoundly' (Albert 2000, p. 119), the past-hypothesis is supported by the fact that it renders our fundamental convictions, such as the veracity of memories, correct.

Yet it is difficult to see how the past-hypothesis could be confirmed by successful predictions if it cannot be confirmed (and in fact is apparently disconfirmed) by the sum-total of our experiences at any one moment in time. For once the predicted outcome is realized, our reasons for taking that outcome to be the result of a prior prediction are again based on records or memories. If, as Albert suggests, our experience *now* provides us with no reasons to accept the past-hypothesis, then neither does our experience a few minutes *later*, after we have conducted an experiment to confirm our predictions.[15]

Thus, we are left with Albert's second solution: our reason for believing the past-hypothesis is that it is needed to account for some of our fundamental convictions, such as the veracity of memories. This spells

[15] I owe this point to discussions with Dan Parker.

trouble for Albert's theory both as an account of our actual reasoning practices and as an epistemological justification of these practices. If we assume, with Albert, that our memories are veridical, then, it seems, at the very least we assume what the contents of our memories are. Thus, we take ourselves to be in a position to draw inferences about the past without actually having to first postulate the past-hypothesis as ready-condition. Moreover, the account cannot provide a justification for our inference practices either. For if our justification for positing the past-hypothesis is that it renders our memories veridical, then we cannot in turn justify the veracity of our memories by appealing to the past-hypothesis, on pain of circularity.[16]

In this section I have argued that Albert's account fails as an explanation of our fundamental conviction that the future, but not the past, depends on what happens now. *Assuming* the past-hypothesis is a necessary condition neither of our drawing inferences based on records nor of our being in a position to justify these inferences. Yet despite the fact that Albert's argument for the necessity of assuming the past-hypothesis in inferences about the past plays a prominent role in his overall discussion, his core thesis that our counterfactual and causal reasoning can in principle be recovered from positing the past-hypothesis relies rather on the claim that the past-hypothesis is sufficient for the reliability of records and, ultimately, for the asymmetry of causal dependence. It is to the sufficiency claims expressed in (*i*) and (*ii*) above to which I want to turn next.

13.4.2 Records and the Past-Hypothesis

As we have seen, Albert maintains that local facts about the present can function as records of the past, if we can assume certain facts about the remote past as 'ready condition', and that 'the initial macro-condition of the universe as a whole' can function as the 'mother of all ready

[16] Albert seems to be of two minds about how much weight to attach to the fact that we believe in the veracity of our memories. On the one hand he suggests that a belief in the veracity of our memories is among our 'most fundamental convictions' that even plays a role in justifying our belief in the past-hypothesis. On the other hand, he denies, as we have seen, that a belief in the veracity of memories can function as a 'mother of all ready conditions'. As one reason for this he offers the observation that the evidence of our senses can be overridden (see Albert 2000, p. 118, footnote 6). But the fact that it is possible for a *particular* memory to be overridden by other evidence is compatible with the claim that we generally assume our memories to be veridical and that this assumption ultimately 'grounds' all our reasoning based on records.

conditions'. (Albert 2000, p. 118). From this he immediately and without further argument concludes the following:

> And so it turns out that *precisely* the thing that makes it the case that the second law of thermodynamics is (statistically) true throughout the entire history of the world is *also* the thing that makes it the case that we can have epistemic access to the past which is not of a predictive/retrodictive sort; the reason there can be records of the past and not the future is nothing other than that it seems to us that our experience is confirmatory of a past-hypothesis but not of a future one. (Albert 2000, p. 118)

And a little further on he says:

> [E]verything we can know of the past and present and future history of the world can be deduced, in its entirety [...] from the following four elements: what we know of the world's present macrocondition [...]; the standard microstatistical rule; the dynamical equations of motion; the past-hypothesis. (p. 119)

Hence, Albert takes the claim that 'the initial macro-condition of the universe as a whole' functions as ready condition to imply premise (*i*)—the claim that the past-hypothesis can play the role of 'mother of all ready conditions'.

But to conclude from the claim that the initial macro-condition of the universe can function as ready condition that the past-hypothesis alone is such a ready condition is a *non-sequitur*. For the past-hypothesis provides us with significantly less than a full specification of the initial macro-state of the universe. All the past-hypothesis asserts is 'that the world first came into being in whatever particular low-entropy highly condensed big-bang sort of macro-condition it is that the normal inferential procedures of cosmology will eventually present to us' (96). And clearly whatever it is that cosmology will eventually present us with, this will fall far short of a complete account of the initial macro-state of the universe. Thus, Albert owes us an argument for why the broad constraints on the early universe posited by the past-hypothesis (as opposed to a full specification of the macro-state of the early universe) are sufficient to ensure that local facts about the present can function as records of the past.

To illustrate this point, we might imagine a slightly amended version of Albert's billiard balls. Let us assume that ball 5 is currently *moving* and was *stationary* five seconds ago. Further, let us imagine that the balls are moving on a table with very weak frictional forces. Given the ready condition that

ball 5 was stationary, the fact that the ball is currently moving functions as a record of a collision in the last five seconds. The ready condition in this case exactly specifies the value of one of the system's state-space variables. But obviously it does not follow from the fact that the ball's having been *stationary* can function as ready condition that also the claim that the system of balls was in a *low-entropy* initial state can function as ready condition. From the fact that the system of balls was in a low-entropy state we can conclude that the most likely evolution of the system of balls was one that is thermodynamically normal and, hence, that the ball's currently moving is not due to random 'anti-frictional' forces exerted by the table. But without the further assumption that the ball was stationary five seconds ago, we cannot exclude the possibility that the ball has been moving without collisions for more than five seconds.

Loewer (ch. 11, in this volume) stresses that in addition to a low-entropy constraint we need to posit that the state of the early universe also satisfied certain symmetry constraints (without specifying what these constraints are). But again the billiard ball example suggests that such an additional constraint still falls short of what is needed. Even if we were told that the initial low-entropy state of the system of balls (more than five seconds ago) was the highly symmetric special macro-state when the balls are racked, we could not infer from the fact that ball 5 is currently moving that it underwent a collision in the last five seconds.

As far as I can tell, Albert does not offer any argument that in the case of the universe as a whole the assumption of a low-entropy past alone can function as ready condition. Moreover, Albert's discussion suffers from the fact that he does not distinguish clearly between the claim that the past-hypothesis is *sufficient* for the reliability of records and the claim that it is *necessary* in various senses. In the previous section I criticized Albert's claim that assuming the past-hypothesis is a necessary condition of drawing inferences based on records. In the text Albert moves from this claim without any discussion to the claim I quoted above that, 'anyway, [...] everything we can know of the past' can be deduced from the past-hypothesis in the standard way—that is, to the sufficiency of the past-hypothesis for the reliability of records. There is one additional set of considerations Albert advances, namely that there would be no reliable records in worlds that do not satisfy the past-hypothesis. In any such world, Albert says, the most probable way in which putative records originate

would be as results of random fluctuations from a maximal-entropy state and, hence, they would not be correlated with the relatively low-entropy states of which they are supposed to be records. But this argument can at most show that the past hypothesis is necessary for the existence of records.

Thus, there are three distinct theses we ought to distinguish: First, that *assuming* the past-hypothesis is a necessary condition of our actually drawing inferences based on records. Second, that the fact that the past-hypothesis *holds* is a *necessary* condition for the reliability of records. And third, that the past-hypothesis is a *sufficient* condition for the reliability of records. The second thesis is arguably true; the first, I have argued, clearly false; and Albert provides no argument for the third.

Yet we can try to imagine what kind of considerations one might advance in support of the sufficiency thesis. How, we might ask, could it be that local facts about the present are associated with a past different from that of the actual world? One way to construct such a situation is to postulate some small yet macroscopic change to the actual present and then evolve the resulting state backward in time. The effect of such local changes will in general be that *other* local facts are no longer associated with the same past events with which they were associated in the actual world: such present facts constitute fake records, as it were. In the amended billiard ball example ball 5 is presently moving in the actual world, and this is associated with the ball having undergone a collision in the past five seconds. But there can be changes to the state of balls *other* than ball 5 that, if we evolve the state of the balls backward in accord with the laws, will result in a history where ball 5 did not undergo a collision in the last 5 seconds. We might say that in the corresponding counterfactual world the fact that ball 5 is presently moving constitutes a 'fake' record of its past evolution. Of course in such counterfactual worlds the ready condition that ball 5 was at rest five seconds ago is not satisfied. The crucial question is whether the past-hypothesis would likewise not be true in such a world.

Adam Elga, in a somewhat different context, has presented an argument that suggests that the past-hypothesis would indeed not be satisfied in most counterfactual worlds resulting from localized macroscopic changes to the actual world (Elga 2001).[17] Elga points out that the time-evolution of the actual world *toward the past* is thermodynamically extremely unlikely.

[17] Loewer (ch. 11, in this volume) stresses that the argument presented by Elga is originally due to Albert.

(This is most easily seen if we imagine that the direction of time were flipped.) Moreover, the evolution toward the past is extremely sensitive to small changes in the micro-state of the world: most small changes will result in worlds that evolve in thermodynamically normal ways toward the past—that is, worlds that violate the past-hypothesis and behave anti-thermodynamically in the normal time-sense. For example, in the case of the billiard balls it is probable that changing the position of any of the balls will, through small changes in the gravitational force, disturb the normal thermodynamic behavior of thermodynamic sub-systems in the vicinity. Further and further into the past, more and more sub-regions of such a world will be 'infected' by the anti-thermodynamic behavior with the result that the remote past of the world will have high entropy.

The upshot is that localized macroscopic changes to the present result both in an anti-thermodynamic past *and* in 'fake' records of the past. But this again is not enough to establish (*i*). One might think that the argument is simply this: If there are fake records in a world, then the past-hypothesis is false in that world. Thus, taking the contrapositive, if the past-hypothesis is true, then there are no fake records. Now, one may doubt that the Elga-Albert considerations do in fact show that in worlds with fake records the past-hypothesis does not hold (since there might be ways to construct worlds with fake records without violating the past-hypothesis). But even if we grant this step in the argument, the conclusion does not follow, for the argument is not valid in a probabilistic context.

Let us grant that the probability of the past-hypothesis given that there are fake records is extremely small. That is, formally:

$$(2)\ \Pr(PH/R\&{\sim}e) = \varepsilon,$$

where R is a record of some event e in its past in the actual world. (2) says that the probability of the past-hypothesis is extremely small given that the record R is present without the event e of which it is a putative record having occurred. From this we would like to conclude (*i*)—that is, that the past-hypothesis ensures the reliability of records; or, in other words, that the probability of an event e not occurring, given the past-hypothesis and record R of its occurrence, is extremely small:

$$(3)\ \Pr({\sim}e/R\&PH) = \delta.$$

Yet, (3) does not follow from (2). In fact, as a simple application of Bayes' theorem can show, in order to get from (2) to (3) we need to assume as additional premise that

(4) $\Pr(PH/R) \geq \Pr(\sim e/R)$.

But (4) is false. The left-hand side of (4) is the probability that a world had an extremely low-entropy past, given certain information about its current macro-state. This probability is absurdly small, as Loschmidt's reversibility objection has taught us. The right-hand side is the probability that a past event e did not occur given the presence of a putative record of e (and nothing more!). This probability will in general not be all that small. In fact, Albert's own argument relies crucially on the assumption that this probability is in general not small, for it is precisely this assumption that makes an appeal to ready conditions necessary in the first place and that ultimately is supposed to underwrite the connection between records and the past-hypothesis. If $\Pr(e/R)$ were close to one, then an inference from R to e would not need to involve an appeal to a ready condition prior to e.

For (3) to be true, it would have to be the case that the probability of PH is much, much lower, given fake records, than the probability of PH given the record state. But the Elga-Albert argument cannot establish this claim. Of course it is extremely improbable, given the present state of a world that differs from the actual world by a small macro-change, that this world satisfies the past-hypothesis. Yet that the actual world evolved from an extremely low-entropy state, given its present macro-state, is similarly improbable. This, after all, is just the reversibility objection that is circumvented by simply postulating a low-entropy past. That a small counterfactual macro-change to the present will again result in a micro-state which evolved from a low entropy macro-state is no less probable than the low-entropy past of the actual world.

12.4.3 Records in Counterfactual Worlds

The second premise of Albert's argument is the claim that if there are records of the past, then there is no counterfactual dependence of the past on the present. In discussing an example of a putative case of backward counterfactual dependence, Albert supports this claim by saying that the past could not have been different since a different past would have to

have left traces or records that ought to be part of the present. Since by assumption the counterfactual present is identical to the actual present except for a small, local change, the counterfactual present contains traces of the actual past but not of any counterfactual past events. Hence, the past does not counterfactually depend on the present. This argument relies on (3)—that is, that the probability of any past event e not occurring, given its present records R and the past-hypothesis PH is very small. We have just seen that the claim that the past-hypothesis alone can ensure the reliability of records is problematic, but for present purposes I want to grant that claim and see what follows from it.

According to premise (*ii*), (3) implies (1), or equivalently

(5) $\Pr(e/\sim c\&S_a\&PH) \approx 1$.[18]

As in the case of premise (*i*), however, Albert provides no argument for (*ii*). If in (3) δ were strictly equal to zero, then (5) would indeed follow and (*ii*) would not need to be introduced as an independent premise but would simply be a consequence of the probability calculus. Since, however, macro-states are only probabilistically given in terms of the underlying micro-states and their dynamics, we need a justification for the move from (3) to (5). It may be plausible that conditionalizing on the entire state S_a of which the records R are a part does not change the probability of e, that is to say, that (3) implies

(6) $\Pr(e/S_a\&PH) = 1-\delta$.

But what is less clear is why conditionalizing on $\sim c$ as well should not significantly affect the probability of e. Again, Albert owes us an argument.

There is one particular class of events for which it is perhaps most obvious that there is indeed the need for an argument here—those events e that are complete macro-states of cross sections of the backward light cone of c in the relatively recent past of c and where there are macro-laws governing the system that are near-deterministic. In such cases, e determines the occurrence of c with 'thermodynamic certainty,' as it were:

(7) $\Pr(c/e\&PH) \approx 1-\varepsilon$, with $\varepsilon \approx 0$.

[18] Intuitively, the difference between (3) and (5) is this. According to (3), in the vast majority of worlds in which R and PH occur, R is a reliable record of e. According to (5), R is a reliable record of e even in the majority of those counterfactual worlds in which c does not occur.

Cleary there are systems like this—systems that appear to behave deterministically and non-chaotically on the macro-level. There are, of course, also macro-systems which behave chaotically or in which the macro-dynamics is probabilistic (such as coin tosses); and there arguably are systems for which we cannot write down any macro-dynamics. Here I want to focus on systems which we model in terms of a deterministic macro-dynamics.[19] As a concrete example, we can once more think of a version of Albert's billiard ball example: As event c we can pick the current velocity and position of ball 5 and as e we can choose the macro state of the world five seconds ago in a sphere centered on the current location of ball 5 with a diameter of five lightseconds.[20] In particular, e includes the state of the entire collection of billiard balls five seconds ago. Records of aspects of e might be, among other things, the positions and velocities of balls other than ball 5 and light waves that were reflected by the collection of balls five seconds ago. Our macro-dynamics gives us *ceteris paribus* laws, according to which in modeling the billiard balls we can ignore everything in the past light-cone aside from what happens on the billiard table itself.

Since the probability of c is completely determined by events in its backward light cone, e will screen off c from any event outside of the light cone of c. In particular,

(8) $\Pr(c/e\&PH) = \Pr(c/e\&S_a\&PH)$.

Moreover, for many systems for which (7) holds, $\Pr(\sim e/\sim c\&PH) \approx 0$ ought to hold as well. For example, if counterfactually ball 5 were at the other end of the table now, then it seems dynamically unlikely that it would have been at its actual past location a short while ago. However, it follows from (5) and the definition of conditional probability that

(9) $\Pr(e\&\sim c/S_a\&PH) \gg \Pr(\sim e\&\sim c/S_a\&PH)$.

That is, it is much more probable that the actual past event e occurs without the actual present than that neither the past nor the present are those of

[19] My aim here is not to show that Albert's normal procedures of inference fail in *all* cases, only that there are standard cases where it looks as if the procedures do not yield the result Albert needs.

[20] Is it okay to consider events of that size, given that Albert's focus is on small macro-changes for which we might take ourselves to be responsible? The answer is: Yes, since the hypothetical change is small—a change to the state of ball 5. And we are simply asking whether changes to the occurrence c could have as a consequence changes to the occurrence of e, just as Nixon's pushing the bottom would have large consequences in the future.

the actual world. And this is so, even though the occurrence of e makes it dynamically extremely improbable that c does not occur and there are no purely dynamic constraints that make $\sim e\&\sim c$ improbable.

More intuitively, we can put this result this way. Premise (*ii*) assumes that we treat inferences based on records as much more reliable than predictions and retrodictions based on the dynamics. The dynamics alone would predict that a locally different present macro-state would in general have resulted from a different past macro-state. According to (*ii*), however, any such retrodiction is overridden, as it were, by the assumption that records are reliable. Yet one might worry that this gets things backwards: No matter how reliable our records are, they never can be more reliable—and will in general be far less reliable—than any inferences we can draw based on a complete macro-state, the past-hypothesis, and the dynamics. Of course the conjunction of (5), (7) and (8) does not contain a contradiction. Yet it is far from obvious (and requires an argument) why the record condition (3) ought to commit us to (5), and hence (9).[21]

Think about the collection of billiard balls. Intuitively, it seems that if the position or velocity of ball 5 were different, then the overall state of the system of billiard balls would have to have been different five seconds ago. Contrary to this, Albert's account would claim that since there are records of the state of the balls five seconds ago, including the present state of the remaining billiard balls, it is overwhelmingly likely that the macro state of the system of balls would have been exactly the same as the actual state. This claim strikes me as highly counterintuitive. Once again, we might test our intuitions by taking the racked up billiard balls as stand-in for the past-hypothesis. Let us assume that after the break the billiard balls move for ten seconds. And let us then ask what our dynamics predicts for the state of the balls just five seconds ago, given that we assume that ball 5 came to rest somewhere else from where it actually did, the positions of the other balls remain unaltered, and the 'past-hypothesis' that the balls ten seconds ago were racked up. I take it that Albert's intuition is that something thermodynamically 'odd' must have happened to the balls in order for ball 5 to end up at a macroscopically distinct present location and quite plausibly this intuition is correct. But what is difficult to see is why the most plausible

[21] Of course Albert maintains that the micro-dynamics and the statistical postulate *alone* make all the wrong retrodictions. But what is at issue here is whether the dynamics and statistics further constrained by the past-hypothesis still are less reliable than our records of the past.

past evolution of the counterfactual system of billiard balls is supposed to be one that *exactly* matches the actual macro-evolution until immediately before the present and only then diverges in some thermodynamically unexpected way. It might seem more plausible that our dynamics will tell us that when there is an apparent 'tension' between our records, as given by the location of all the billiard balls except for ball 5, and the present state of ball 5 itself, then there will be some kind of 'trade-off' where the relatively not-too-distant past of the counterfactual system will be different from that of the actual system in ways not entirely compatible with the reliability of records.

Now, ultimately the question whether or not it is a consequence of the dynamics, the statistical postulate, and the past-hypothesis that worlds that differ from the actual world locally and macroscopically are overwhelmingly likely to have had exactly the same macroscopic past as the actual world should not be a question that is settled through a battle of intuitions. Whether Albert's thesis is right is a question for the relevant physics. My aim in arguing that Albert's thesis is counterintuitive is only to urge that there is indeed still the need for a physical argument. That is, my aim is to combat the impression that once we realize the importance of the past-hypothesis as additional constraint on the dynamics, then the counterfactual independence of the past from the present follows almost immediately and that there really is no need for a detailed physical argument for why records of the past, in a sense, 'trump' what might be suggested by a macro-dynamics.

13.4.4 Possible Replies: Transition Periods and Degrees of Counterfactual Dependence

In the last section I have suggested that a more careful examination of the relevant physics might show, at the very least, that local changes to the present would have to have been preceded by a transition period during which the counterfactual past differed macroscopically from that of the actual world. Could Albert not simply concede this point without abandoning his account? Perhaps Albert need not show that, as far as small macroscopic changes to the present are concerned, there is absolutely no counterfactual dependence of the past on the present. A weaker claim might be sufficient, namely that whatever counterfactual dependence of the past on the present there is, it is much less and dies off much faster

than any dependence of the future on the present. That is, one might try to argue that neither the record condition (5) nor the counterfactual independence claims (1) hold for *all* past events and that events in the very recent past of c will exhibit some counterfactual dependence on c, but that nevertheless there is a significant asymmetry in the degrees of counterfactual dependence that is sufficient to account for our asymmetric notion of causal dependence.[22]

However, this defense faces serious problems of its own. First, nothing in my criticism relies on the fact that e is an event in the very recent past. As long as e is recent enough for there to be a macro-dynamics that is near-deterministic linking e and c, the arguments go through.

Second, the defense proposes to replace what appears to be a sharp and precise distinction with a qualitative and gradual difference. According to our common sense notion of causation we think that our actions can influence the future but have *no influence at all* on the past. What a defender of an entropy-account would have to explain is how we have come to believe in this sharp distinction, despite the fact that, according to the account, there is some counterfactual dependence of the past on small interventions into the present. In particular, it is not sufficient simply to argue that there might be some 'transition' period that is necessary for a world with a past that perfectly agrees with the macroevolution of the actual world to evolve in a nomically possible—even if thermodynamically abnormal—way into a counterfactual present state. For it is of crucial importance *how long* such a transition period needs to be. If the transition periods were on the order of fractions of a second, then perhaps it would be somewhat plausible to suggest that such a limited counterfactual dependence of the past on the present could somehow give rise to our fundamental conviction of a strict independence of the past from the present.

Yet the suggestion seems to be that during a transition period an extremely rare anti-thermodynamic fluctuation takes place that carries a past state that is macroscopically indistinguishable from the actual past into a counterfactual present state and it is not clear that such fluctuations could occur fast enough to render the transition period negligible. How long, for example, would it take for anti-frictional forces on a billiard table to move a ball that came to rest at one end of the table to the other end? In order

[22] This reply was suggested to me by Adam Elga and Doug Kutach.

for a ball that was at rest at one end of the table to end up at rest at the other end, the ball has to start rolling across the table and eventually slow down again to come to rest. The first part of the ball's trip has to be the result of anti-frictional behavior, while the second part will be in accord with normal, frictional behavior. Since anti-frictional behavior is simply the time-reverse of normal frictional behavior, the ball will take just as long for the first part of its trip as for the second part. Hence the total time it will take for a ball to travel across the table due to a combination of anti-frictional and frictional forces will be roughly twice the time it takes for a ball to roll for half the length of the table and then come to a stop. On a realistic billiard table this time is probably on the order of a few seconds—too long to be negligible. For if the transition period were that long, there ought to be, according to Albert's account, 'causal handles' on the relatively recent past of the kind that we should take ourselves to be able to exploit: just as we think that our actions now have effects even a few seconds into the future, we ought to think that our actions have effects at least a few seconds into the past. Moreover, most likely a few seconds will not turn out to be the upper limit on the time of such anti-thermodynamic transitions.

If this is right, then an appeal to anti-thermodynamic transition periods can be of no help for Albert's account. Even if a careful examination of the underlying physics revealed that the most probable history of a counterfactual world was one in which the hypothetical change to the present was due to a thermodynamically abnormal fluctuation immediately preceding the present state, the periods of mismatch in the past would probably be too long to support our fundamental conviction that the entire past is counterfactually independent of any of our decisions in the present.

13.5 Counterfactuals and Decisions

In this volume, Barry Loewer also defends an entropy account of counterfactuals. While Loewer's account is closely related to Albert's account, there are also important differences between the two accounts. One such difference concerns the question how we evaluate the truth of counterfactuals. According to Albert's account there is a single procedure for evaluating counterfactuals given by our 'normal procedures of inference'. Loewer, by contrast, proposes two distinct kinds of truth conditions—one kind

for what he calls 'decision conditionals' and another for counterfactuals positing small macro changes to the world. The truth conditions for the latter are different from—and in fact incompatible with—the truth conditions suggested by Albert's 'normal procedures of inference'. I want to end this paper with a few remarks concerning Loewer's account, which I discuss more fully in Frisch (forthcoming).

Loewer apparently endorses Albert's account of the role(s) the past-hypothesis plays in the production of records. Similarly to Albert, Loewer holds that the past-hypothesis is *necessary* for the production of records—the *PH* 'allows for the production of localized macro records' in that it removes an obstacle to there being such records (this volume, p. 303). In addition, the past-hypothesis 'plays the role of a ready state for our universe'—that is, it is what Albert calls the 'mother of all ready condition'. Since Loewer further agrees with Albert that without positing the relevant ready states 'we are not justified in making' inferences based on records and that a ready state together with the dynamical laws and the record state entails the recorded state, he presumably also agrees with Albert that the past-hypothesis as first ready-state is necessary for *justifying* inferences based on records and that the past-hypothesis is a *sufficient* condition for the reliability of records.[23]

Loewer's account makes explicit an intermediate step in the argument from statistical mechanics to a counterfactual asymmetry—a step that Albert does not discuss but may well be implicit in Albert's account. Loewer argues that a statistical mechanical account implies a tree-structure for possible macro histories, according to which there are many small differences in the micro conditions at a time t not involving changes in the macro state at t that result in different macro futures, but there are only very few such differences that result in different macro pasts. Here is how Loewer characterizes this tree-structure:

> The SM probability distribution embodies a way in which 'the future' (i.e. the temporal direction away from the time at which PH obtains) is 'open' at least insofar as *macro states* are being considered. Since all histories must satisfy the PH they are very constrained at one boundary condition but there is no similar

[23] Loewer stresses that the past-hypothesis and the dynamical laws are not 'sufficient to account for the existence of recording systems let alone the particular records that have been formed in our world' (this volume, p.304). But this is true of any ready condition: no ready condition is sufficient for the existence of a particular record. Rather, what Albert's account of records claims is that a ready condition and the laws in conjunction with a record are sufficient for the recorded state.

constraint at other times. It is true that (almost) all histories eventually end up in an equilibrium state (there is a time at which almost all histories are in an equilibrium state) but this is not a constraint it is a consequence of the dynamics and the PH and it is not very constraining (almost all states are equilibrium states). Another feature of the SM distribution when applied to the macro state of the kind of world we find ourselves in is that the macro state of the world at any time is compatible with micro states that lead to rather different macro futures. For example, conditional on the present macro state of the world the SM probability distribution may assign substantial chances both to its raining and not raining tomorrow. On the other hand, there is typically much less branching towards the past. The reason is that the macro states that arise in our world typically contain many macroscopic signatures (i.e. macro states/events that record other macro states/events) of past events but fewer macroscopic signatures of future states/events. (Loewer this volume, pp.302–3)

This passage raises several questions. First, the asymmetry embodied by the tree structure is not strict; Loewer says only that there is 'typically *much less* branching towards the past'. Yet the asymmetry he ultimately wants to derive is strict: according to our folk notion of cause, there is absolutely no causal dependence of the past on the present and there are no true back-tracking counterfactuals in causal contexts. One question, then, is whether Loewer can recover these strict asymmetries from the merely quantitative differences between the amount of future and past branching. One way in which this might be achieved is if all (or at least the overwhelming majority of) branches toward the past have negligible probabilities.

A second question is how exactly we are supposed to understand the claim that possible macro histories exhibit a tree structure. The passage above suggests that the branching structure is a *consequence* of the existence of records: 'the reason', Loewer says, for the relative absence of past branches is the existence of macro records. Yet immediately after this passage Loewer continues by saying that 'Albert shows how the assumption of the PH (and the consequent branching structure) *allows* for the production of localized macro records of past events' [my italics] (this volume, p.303), which suggests that the tree structure *entails* the possibility of records.[24]

No matter how precisely the tree structure and its relation to the possibility of records are ultimately to be understood, it is not immediately

[24] I discuss prospects and problems for both these readings in detail in Frisch (forthcoming).

obvious how any such asymmetry follows from the three assumptions that Loewer, following Albert, identifies as constituting the micro statistical account—that is, the dynamical laws, the past-hypothesis, and the assumption of an equi-probability distribution over all micro states compatible with the Big Bang initial state of the universe. Indeed one might worry that if anything the Second Law implies an upside-down tree structure. Since states closer to equilibrium occupy vastly larger regions of phase space than macro states very far from equilibrium, it follows from Liouville's theorem that there will be many different non-equilibrium states very far from equilibrium that evolve into the same state closer to equilibrium in the future. Thus, there appear to be many more changes to the micro state of a system close to equilibrium that are associated with different pasts further away from equilibrium than there are changes to the micro state of a system far from equilibrium that are associated with different futures closer to equilibrium. Loewer points out that the *PH* constrains possible past evolutions, while there is no similar constraint on future evolutions. But constraining the past to low-entropy states entails that there will be a vastly *larger* number of possible past states than of possible future states, which have higher entropies and occupy vastly larger regions of phase space.

Assuming that possible macro histories did indeed exhibit the kind of tree structure postulated by Loewer, can that underwrite a counterfactual asymmetry? It seems to follow immediately from the tree structure that Loewer's decision counterfactuals (this volume, Section 11.3) are time-asymmetric. Yet it is less clear that the asymmetry can be extended to non-decision counterfactuals. Loewer proposes the following truth conditions for counterfactuals positing small macro changes:

> 'If $A(t)$ had been true then the chance of B would have been x' is true iff t' is the latest time at which a divergence from the actual macro history similar in probability to a decision event can occur and $\Pr(B/A(t)\&M(t')) = x$. (Loewer this volume, p. 320)

Here $M(t')$ is the complete macro state at t'. The difference between these truth conditions and those suggested by Albert's account is that the probability of the consequent of the counterfactual is conditionalized on the macro state at some time *earlier* than the antecedent $A(t)$ rather than on the macro state at t. Generally the conditional probability of an event will change, if we conditionalize it on the complete macro state at different

times and, hence, it will not in general be the case that $\Pr(B/A(t)\& M(t')) = Pr(B/A(t)\& S_a(t))$, for $t' < t$. According to Loewer's proposal the truth conditions for counterfactuals postulating small macro changes to the world, thus, are not given by Albert's 'normal procedures of inference'.

One problem for Loewer's proposal is that for macro systems that are *not* extremely sensitive to changes in the system's micro state—and there are many such systems that appear to evolve deterministically on the macro level—there will be backtracking counterfactuals and counterfactuals relating what intuitively are joint effects of a common cause which come out true according to the truth conditions. Consider the following example. In the actual world I am standing alone with an axe in a forest, with no other axe-wielding person in the vicinity. I do not swing my axe and no tree falls. Let the counterfactual event B be that I swing the axe at a tree and the event A be that the tree falls. Does my swinging of the axe counterfactually depend on the falling of the tree? Arguably there is no relatively probable decision-sized event D compatible with the actual macro history later than my possible decision to swing the axe that would result in the tree falling.[25] Hence the time t' of the relevant macro state on which we ought to conditionalize is the time of my contemplating whether to swing the axe. And the right probability to consider in evaluating the truth of the counterfactual 'If the tree were to fall, then the event that I would have swung the axe has probability p' is $\Pr(B/A$ & *the macro state of the world at the time when I contemplate*). But this probability arguably is quite large! For given that the tree falls and that at the time I was contemplating whether to swing the axe no other putative causes of the tree's falling were nearby, it is highly probable that I did indeed swing the axe. But that my swinging of the axe caused the tree to fall and not the other way around is the kind of judgment that is paradigmatically causal and any account of causation would have to get examples like this one right. How, then, do we arrive at the causal asymmetry, given that the relevant backtracking counterfactual is true?

The worry can perhaps be brought into sharper focus by contrasting Albert and Loewer's account with Lewis's theory. On Lewis's account, unlike on Loewer's, the counterfactual 'If the tree fell, I would have to

[25] I want to emphasize that this example depends in no way on the fact that it involves a human action that ultimately is due to a decision. One can readily construct structurally identical examples not involving human actions.

have swung the axe' comes out false. It comes out false, since according to Lewis's prescription the possible world in which the tree falls that is closest to the actual world is one that diverges from the actual world *immediately before* the tree falls and *after* my not swinging the axe. That is, on Lewis's account, when we evaluate backtracking counterfactuals, the miracle needs to be introduced temporally *between* the occurrence of the antecedent event and its putative earlier effect. Yet if a system is macroscopically relatively stable, then, on Loewer's proposal, we may have to go back in time quite far until we reach a time at which a relatively probable micro 'miracle' is sufficient to alter the system's macro evolution. In all such cases—that is, in any of the vast number of causal claims concerning relatively stable macro systems—there will be backtracking counterfactuals with high probabilities on Loewer's account. To be sure, Loewer does not offer a worked out account of how precisely counterfactual claims are related to causal claims—even though he does maintain, echoing Albert's notion of a causal handle, that in cases where the probabilities in the consequent vary significantly with changes in the antecedent, the events in the antecedent provide 'a kind of "handle" ' on the consequent (this volume, p. 320). Yet to the extent that Loewer's counterfactuals are meant to match Lewis's the account appears to fail.

There is a closely related worry concerning situations that we intuitively take to involve two events that have a common cause. Consider the counterfactual 'If the sound associated with an axe striking a tree occurred, then the event that the tree will fall has probability p.' Arguably, the latest decision-sized micro event that could result in the sound might again be my decision to strike the tree and the probability p might be quite large. Thus, in the case of counterfactuals linking two effects of a common cause, Loewer's counterfactuals also do not in general match Lewis's, since on Lewis's account the counterfactual 'If there were an axe-striking sound, then the tree would fall' is false, because the account involves introducing a miracle *between* the occurrence of the sound and the striking of the axe. Loewer suggests that the problem of backtracking counterfactuals might be solved by characterizing 'causal priority in terms of the temporal direction of control by decisions' (Loewer this volume, p.321). But this proposal would be of no help with the problem of counterfactual dependence between effects of a common cause.

13.6 Conclusion

I want to sum up what I have argued. It is one of our fundamental convictions that the future depends on the present in ways in which the past does not. Albert's account is meant to account for that conviction by delineating an inference procedure according to which the future counterfactually depends on small local changes to the present macro state but the past does not. This inference procedure involves the past-hypothesis, the micro-dynamical laws, and a statistical postulate. One of my aims in this paper was simply to get clearer on the question as to what exactly it could be to get to the bottom of this conviction of ours. There are at least three distinct projects in which Albert might be engaged. The first project is that of trying to account for how we actually proceed in making certain inferences. As we have seen, Albert appears to argue that in making inferences from records, we make inferences from two times to a time in between; and in order to come by the information about the past involved in making such inferences we must assume the initial state of the universe and it must seem to us that our experience is confirmatory of the past-hypothesis. I have argued that as an account of our actual inferential practices Albert's account is inadequate. If it indeed were the case that inferences based on records only were possible if we assumed the past-hypothesis, then most of us are not in a position to reason from records. The second project is epistemological and is an attempt to provide a justification for our fundamental conviction. However, to the extent that Albert's account is meant to justify the veracity of our memories—an assumption that on his account underlies all our counterfactual inferences—it is circular, since it is our belief in the veracity of our memories itself that, according to Albert, justifies our belief in the past-hypothesis.

The third project Albert might be engaged in is one of offering a philosophical reconstruction of our counterfactual reasoning practices. Albert, on this reading might only be claiming that postulating the past-hypothesis can ensure that records are reliable and that the past is counterfactually independent of the present, even though we need not assume the past-hypothesis when we actually reason about the past. To be sure, on this third reading Albert's worry how we actually come by the information about the past needed in reasoning from records remains unanswered. Nevertheless,

the fact that our reasoning practices could be reconstructed in the way suggested by Albert would obviously be extremely interesting and might play a role in an evolutionary explanation of our concept of cause. Yet I have argued that on this third reading the project is problematic as well.

First, it is highly doubtful that, as Albert claims, the past-hypothesis alone can in principle ensure that records are reliable. Even if standard arguments from statistical physics were able to show that if records were unreliable, then the past-hypothesis would most likely be violated, it does not follow that if the past-hypothesis is true, then records are probably reliable. At the very least, there is a *lacuna* in the argument here. And I have suggested that when we try to gauge the prospects of Albert's claim by thinking about relatively small, isolated thermodynamic systems as toy-examples, then it seems highly unlikely that this *lacuna* will ever be filled.

Second, I have criticized Albert's claim that, given the past-hypothesis, the existence of records ensures that the past is counterfactually independent of small macro changes to the present. As I have shown, Albert's account presupposes that, in the case of systems that are governed by a nearly-deterministic macro dynamics, the reliability of records trumps, as it were, what should be expected in light of the dynamics plus the past-hypothesis. Again, Albert owes us an argument for why we should accept his claim that it is a consequence of our physics that records strictly take precedence here. More plausible, I have suggested, might be that there is some trade-off between the reliability of records and our macro dynamics, which will have the effect that at least the relatively recent past of counterfactual systems will macroscopically differ from that of the actual system. In order to settle this question a more detailed argument appealing to the relevant physics is desperately needed.

I have argued, moreover, that Albert could not simply reply to my worries by acknowledging the existence of an anti-thermodynamic transition period during which a past identical to the actual past evolves into a counterfactual present, without thinking carefully about the length of such transition periods. For the existence of such a transition period constitutes a counterargument to entropy accounts unless it can be shown that the transition period is very, very short. One sometimes gets the impression from reading defenders of entropy accounts of causation that they assume that thermodynamic fluctuations can very quickly and within time periods that are of negligible duration ensure that a past identical to the actual past

evolves into a present that differs from the actual present locally yet macroscopically. Yet if thermodynamic fluctuations are simply the time-reverse of normal thermodynamic behavior, then this assumption is unwarranted. The relaxation times of thermodynamic systems can be very long. And just as it will take a very long time for Napoleon's boots to fully decompose (to cite one of Albert's own examples) so it would take a very long time for the boots to form anti-thermodynamically out of primordial 'goo'.

Finally, I have made some brief remarks suggesting that Loewer's recent entropy account of counterfactuals does not, as it stands, fill in the argumentative gaps in Albert's entropy account of causal influence. Unfortunately, then, Field's puzzle concerning the role of causation in fundamental physics still awaits a solution.

References

Ahmed, A. (this volume). 'Agency and Causation'.

Albert, D. (2000). *Time and Chance*. Cambridge, Mass.: Harvard University Press.

Cartwright, N. (1979). 'Causal Laws and Effective Strategies', *Noûs* 13: 419–38.

Elga, A. (2001). 'Statistical Mechanics and the Asymmetry of Counterfactual Dependence', *Philosophy of Science* 68 (Proceedings): S313–S324.

Field, H. (2004). 'Causation in a Physical World', in M. Loux and D. Zimmerman (eds), *Oxford Handbook of Metaphysics*. Oxford: Oxford University Press.

Frisch, M. (2005). *Inconsistency, Asymmetry and Non-Locality: Philosophical Issues in Classical Electrodynamics*. New York: Oxford University Press.

——(forthcoming). 'Causal Asymmetry, Counterfactual Decisions and Entropy', *Philosophy of Science*.

Hausman, D. (1998). *Causal Asymmetries*. Cambridge, UK; New York: Cambridge University Press.

Kutach, D. (2002). 'The Entropy Theory of Counterfactuals', *Philosophy of Science* 69: 82–104.

Lewis, D. (1986a). 'Causation', in *Philosophical Papers*. Oxford: Oxford University Press.

——(1986b). 'Counterfactual Dependence and Time's Arrow', in *Philosophical Papers*. Oxford: Oxford University Press. Original edition, Noûs, 13 (1979).

Loewer, B. (this volume). 'Counterfactuals and the Second Law'.

Menzies, P and Huw Price (1993). 'Causation as a Secondary Quality', *British Journal for the Philosophy of Science* 44 (2): 187–203.

Pearl, J. (2000). *Causality*. Cambridge: Cambridge University Press.

Russell, B. (1918). 'On the Notion of Cause', in *Mysticism and Logic and other Essays*. New York: Longmans, Green and Co.

Woodward, J. (2003). *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.