Causes, Randomness, and the Past Hypothesis

Mathias Frisch

1. Introduction

Why do we think of events in the present to be causally responsible for future events but not for events in the past? Why are there records of the past but not of the future? David Albert (2000) and Barry Loewer (2007) have argued that these temporal asymmetries ultimately have the same origin as the second law of thermodynamics and that it is possible to derive these and other temporal asymmetries from the core assumptions of a neo-Boltzmannian account of the thermodynamic asymmetry. These assumptions are the posit of a low entropy initial state of the universe, a probability postulate and a time-symmetric deterministic micro-dynamics. In this paper I want to investigate the prospects for such an 'entropy account' of the causal asymmetry.

I will begin in section 2 by examining one important role causal reasoning plays in physics: causal inferences allow us to gain information about the state of a system in the past, even when we do not have access to the system's full present state. As we will see, an implicit assumption in such inferences is an assumption of initial randomness. A similar assumption is also part of Albert and Loewer's account, in the form of the initial equiprobability postulate. The question, thus, is whether the randomness assumption should itself be thought of as a causal assumption or as part of a purely reductive account. On the one hand, the assumption appears to be a natural consequence of representing the world causally. On the other hand, the assumption is also central to Albert and Loewer's neo-Boltzmannian account of the thermodynamic asymmetry. Thus, in section 3 I will examine what role, if any, the other core assumption of the entropy account—the assumption of an extremely low-entropy initial state of the universe, which Albert calls the 'past hypothesis'—can play in underwriting the asymmetries of causation and of records. I will argue that extant arguments by Albert and Loewer for the role of the past hypothesis in a reductive account are not successful. I conclude that a convincing case for the claim that the causal asymmetry is reducible to thermodynamic considerations remains to be made.

2. Causal representations and initial randomness

Imagine that you are looking up into the night sky and, for some extended period, are observing the light emitted by a particular star. What licenses your inference that the light you are observing was indeed emitted by a star—that is, by a radiating source—rather than it being source-free radiation coming in from the infinite past?

David Lewis (1986) has argued, as part of his attempt to ground the causal asymmetry in a putative asymmetry of counterfactuals, that the present radically overdetermines the past in the sense that many past events have multiple localized traces in the present, each of which is *individually nomologically sufficient* for the occurrence of the past event of which it is a record or trace. Thus, according to Lewis's thesis, each individual packet of radiation observed at an instance is nomologically sufficient for the earlier existence of the star emitting the radiation. If Lewis were right, the question as to how we can justify our belief that the light was emitted would have a simple answer: The radiation observed by us together with the laws of electrodynamics imply the earlier existence of a star emitting the radiation. But Lewis is mistaken and there is no overdetermination of the kind envisaged by him (see Elga 2000 and Frisch 2005, ch. 7 for detailed criticisms of Lewis's claim). The existence of a localized packet of coherent radiation does not nomologically imply the presence of a source and is also compatible with the assumption that the radiation is source-free.

Contrary to Lewis's suggestion, physical dynamical laws such as the wave equation, which governs electromagnetic radiation phenomena, only allow us to make inferences from one time to another if we 'feed into' them a fully specified initial or final state. The wave equation determines a Cauchy problem, which as input requires data on an entire cross-section of the forward lightcone of the putative source to the determine the state of the world at the source's location. Given this feature of the laws, one might perhaps think that the only way we could infer the existence of the star would be from the complete state of the world on a full initial value surface.[1] But then we could never come to know whether the light was emitted by a star, since we never in fact are in possession of the data required. All we have available are the highly localized observations of the radiation field here on

---

[1] I am using the term 'initial value state to refer to states of a system that can be used both for prediction as well as for retrodiction, thereby avoiding the more cumbersome formulation 'initial and final value state'.

Earth (plus, perhaps isolated observations from space-based telescopes). Consequently, if our knowledge of physical systems was exhausted by what we can deduce from the dynamical laws together with knowledge of a system's initial state, then our knowledge of the physical world would be extremely impoverished.

How, then, do we come to know that the light we observe has been emitted by a star? What in addition to the dynamical laws and our localized observations can ground this belief? The solution to the puzzle is that the inference to the existence of a star as source of the radiation is a paradigmatic example of a causal inference. The focused packets of radiation observed by us at different times are highly correlated with one another and we explain these correlations, which otherwise would seem extremely puzzling, by positing the existence of a single localized object, the star, as their common cause.

This kind of inference is one we employ rather frequently, both within and outside of science. Here is an example from outside physics, but with the same inferential structure: in a string of bank robberies the police discover similar kinds of little plastic toys left behind at each crime scene—a fact that is not made public by the investigators. The police infer from these discoveries that the robberies were committed by one and the same gang rather than by different groups that operate in complete independence from one another. This inference can be underwritten by an appeal to a principle of the common cause: onditionalizing on the common cause, the single gang posited to have committed the crimes, renders the different discoveries probabilistically independent of one another and raises the probability of discovering a toy at a crime scene.

Similarly, in the case of the star the values of the electromagnetic field variables and of the variables characterizing the state of the star satisfy a principle of the common cause that can license our inference to the existence of the star. Let us take as causal relata values of the variable $F(t,x)$ characterizing the field at a spacetime point and of the variable $S(t, x)$ characterizing the state of the source. The fields we observe at two different moments t_1 and t_2 are highly correlated and satsify

$$\Pr(F(t_1,x_1) \,\&\, F(t_2,x_2)) >> \Pr(F(t_1,x_1)) \times \Pr(F(t_2,x_2)).$$

Conditionalizing on the state of a star $S(t_{ret1}, x_{ret1})$ that has intersected the backward lightcone of our observation point at t_{ret1} and x_{ret1} increases the probability of observing the field at t_1:

$$\Pr(F(t_1,x_1)/\ S(t_{ret1}, x_{ret1})) > \Pr(F(t_1,x_1))$$

Since the state of the star at one time is highly correlated with its state at later times and $\Pr(S(t_{ret2}, x_{re2})/S(t_{ret1}, x_{ret1})$ is approximately equal to one, conditionalizing on $S(t_{ret1}, x_{ret1})$ also raises the probability $\Pr(F(t_2,x_2))$:

$$\Pr(F(t_2,x_2)/\ S(t_{ret1}, x_{ret1})) = \Pr(F(t_2,x_2)/\ S(t_{ret2}, x_{re2})) \times \Pr(S(t_{ret2}, x_{re2})/S(t_{ret1}, x_{ret1}) > \Pr(F(t_2,x_2)).$$

Finally, conditionalizing on $S(t_{ret1}, x_{ret1})$ renders the observed fields independent of each other:

$$\Pr(F(t_1,x_1)\ \&\ F(t_2,x_2)/\ S(t_{ret1}, x_{ret1})) = \Pr(F(t_1,x_1)/\ S(t_{ret1}, x_{ret1})) \times \Pr(F(t_2,x_2)/\ S(t_{ret1}, x_{ret1}))$$

The overall process is, of course, deterministic: initial fields in the remote past together with the field associated with the star nomologically determine what the observed fields will be. Yet, since we have observational access neither to the free fields prior to the putative emission events nor to the present fields on a complete initial value surface, we cannot set up a full-fledged initial- or final-value problem. Rather we somehow have to infer, based on our localized observations, what *both* initial fields *and* the fields associated with any sources might have been. Without any additional assumptions this would be impossible: our localized observations of the total field do not provide us with enough information to infer both the state of the source and the initial field and are compatible with multiple different combinations of initial fields and radiation fields. What we observe are relatively focused packets of radiation (that we interpret as light emitted by stars) within a background field that is approximately equal to zero (or at least very weak). It is consistent with this evidence not only that the radiation coming from a single direction is due to a star as its common cause, but also that there existed strong correlations among source-free initial fields in spatially distant regions at some remote time in the past that resulted in macroscopic fields converging onto the putative trajectory of the star, passing over it, and then rediverging, mimicking (for later observers) the presence of a star. The further back in time we followed such source free fields, the weaker these fields would be as the become more and more dispersed toward the past, originating in what ultimately might have been extremely delicately coordinated microscopic correlations among very distant field regions.

Yet such highly correlated initial free fields strike us as radically improbable and, thus, we invariably infer that the correlated packets of radiation we observe are emitted by a star. That is, an implicit assumption in our inference to the star as common cause of the observed field values is that the initial fields are effectively random—or rather as random as possible, given our observational evidence. Thus, instead of having to posit a precise value for the initial fields, which we cannot know, we can nevertheless infer something about the history of the radiation field, if we make the weaker assumption that the initial fields at some point in the remote past were effectively random. Without knowledge of full initial or final conditions our inferences cannot be based solely on the fully deterministic laws, but given the initial randomness assumption we can probabilistically infer the existence of the star as most likely explanation of our observations. What is more, since the randomness assumption allows us ignore the initial fields, we can even appeal to part of the dynamical laws, the 'retarded' or 'causal' Green function specifying the field associated with a source in its past, to infer more detailed information concerning the state of the star without having to plug that function into the full deterministic equation of motion.

How then should we think about the probabilities in our inference to a common cause? In the first instance the probabilities in question can be thought of as epistemic probabilities. On this understanding the randomness assumption partly reflects our ignorance of the precise initial conditions: we presuppose, whatever the precise values of the initial fields are, that it is extremely unlikely that they contain strongly correlated spatially distant disturbances that are coherently focused on spacetime points in their future. Note that this is a time-asymmetric assumption: we do not assume—and would be wrong to do so—that final fields contain no strong correlations between spatially distant regions. But one might also want to assign a more objective status to the randomness assumption or even treat it as lawlike constraint, equivalently to the way Albert and Loewer propose to treat the postulate of an initial equiprobability distribution of microstates in the foundations of statistical mechanics.

My aim is not here to resurrect a principle of the common cause as a metaphysical principle which states that whenever two simultaneous events are correlated and neither of the two events is a cause of the other, then the two events must be joint effects of a common cause. Such a principle is arguably false—a point that has been much discussed in the literature (for an overview of such arguments, see Arntzenius 2010). Rather all we need in order to account for our inference to the existence of a star is an epistemological version of the principle, according to which observed correlations often, but by no means always, allow us to infer a common cause in their past. While our expectation generally is that correlations can be explained by a localized common cause, there may be cases where no such explanation is possible (as appears to be the case for certain quantum mechanical phenomena) or cases where a separate cause explanation of the correlations is superior to any putative common cause explanation. The inference to a localized common cause, thus, ought to be thought of as involving a comparison between the common cause explanation and potential separate-cause explanations (see Sober 1984). Elliott Sober proposes that such a comparison between different hypotheses takes the form of a comparison of likelihoods. In our case two possible explanations of our observations are that the correlations among the fields observed at different times are due to the star as common cause or that the observed fields are the forward evolutions of completely independent earlier initial free fields. The first hypothesis renders the observed fields much more likely:

$$\Pr(F(t_1,x_1)\ \&\ F(t_2,x_2)/\ S(t_{ret1},\ x_{ret1})\ \&\ \textit{random weak initial fields}) \gg$$
$$\Pr(F(t_1,x_1)\ \&\ F(t_2,x_2)/\ \&\ \textit{random weak initial fields})$$

Thus, we accept the first hypothesis. Note that the probabilities in question are strictly between 0 and 1, since it is nomologically compatible with the randomness assumption that there are strongly correlated but very small disturbances in the free fields in the remote past that at later times become much larger and converge on the (putative) worldline of the star, either to mask the fields associated with a star or to mimic the field of a star where there is none. Yet such correlations are extremely improbable.

Traditional formulations of the common cause principle are restricted to correlations among simultaneous events. But the principle can easily be broadened to include the kind of correlations we have been discussing. For correlations among timelike related events, such as our successive observations of starlight, there is, in principle, a third kind of hypothesis one ought to consider, namely that the first observation is a cause of the second observation. But because there is no plausible mechanism for such a direct causal link between our observations in our case, this hypothesis can be rejected as radically implausible.

Since the full classical theory governing radiation phenomena is deterministic, it predicts that the correlations between different values of the field will also be screened-off from one another by events in their future. Such future screening-off events, however, will in general not be localized events, but will correspond to widely spread out regions in phase space. Thus, it is crucial that the randomness assumption allows us temporally asymmetrically to infer the existence of a *localized* common cause in the past.

A similar kind of inference to a localized common cause also plays a role elsewhere in cosmology. I said above, that the initial randomness assumption ought to be expressed as stating that initial fields are as random as possible. Yet according to modern cosmology the initial fields cannot be completely random, since there exists a weak cosmic microwave background radiation, which appears to be almost isotropic. Cosmologists want to resist having to postulate that these background radiation is approximately the same throught regions that could have had no causal connection. Here inflationary cosmology is meant to help by offering an account that provides a common cause for the background radiation: according to the theory, the early universe underwent exponential growth within the first 10^{-30}s after the Big Bang, allowing the radiation to have originated in a small causally connected region.

I want to draw two conclusions from our discussion so far: first, causal reasoning is widespread in physics. I have only investigated a single example here, but the discussion readily generalizes. Only very rarely do we know the state of a system on an entire Cauchy surface and can rely on the dynamical laws to make inferences about the past. Instead we need to exploit correlations among different observations to infer past common causes. Second, and more important for our present purposes, a necessary condition for the reliability of causal inferences is the assumption of initial (micro-) randomness. That is, we need to assume that the values of all variables on which the value of the correlated variables depends in addition to a common cause C are randomly distributed. Without that assumption there could be correlations among these variables characterizing an earlier initial state of the system at issue that resulted in correlations among the effect variables even in the absence of a common cause or that render the effect variables probabilistically independent despite the presence of a common cause. (see Arntzenius 2010)

Once we have acknowledged the central role of an initial randomness assumption in causal reasoning, two strategies present themselves concerning the place of causal representations in physics. First, one might take the initial randomness assumption to be itself an intrinsically causal assumption. On this view the fact that causal representations play a successful role in reasoning in physics need not be reducible to any non-causal features of the world (or of us as human agents investigating the world). Second, one might take the assumption to be part of a set of non-causal assumptions to which causal representations can ultimately be reduced.

That the randomness assumption is itself a causal assumption is *prima facie* quite plausible. In a world that can be successfully represented by common cause structures we would expect the values of variables associated with spatially separated regions of space to be randomly distributed unless they are linked through a common cause in their past. Thus, the initial randomness assumption and the assumption that physical systems can be represented by time-asymmetric causal structures can be seen as mutually supportive: Only under the assumption of initial randomness will correlations be a successful guide to causal structures; and in the absence of common causes in the past we expect the values of certain variables to distributed randomly. Moreover, the pair of assumptions fits well with our explanatory practices: we generally search for explanations of correlations by earlier factors. That is, certain correlations strike us as utterly mysterious, unless they can be accounted for in terms of an earlier common cause, while pointing to any later (either localized or non-local) screening-off condition does nothing to remove our sense of puzzlement. By contrast, we do not similarly think that randomly distributed values of variables call for an explanation in terms of earlier factors.

The second strategy is the strategy that Albert and Loewer pursue in their attempts to reduce the causal asymmetry to the same principles that are at the hear of their neo-Boltzmannian account of the thermodynamic asymmetry. Unlike philosophers of physics, such as John Earman (2011), who maintain that causal notions are hopelessly vague expressions of a philosopher's metaphysics, Albert and Loewer believe that a kind of causal reasoning plays an important role in how we come to know about the world. Yet they do not take causal representations to be fundamental and argue that the asymmetry of the causal relation—the fact that we take the future but not the past to causally to depend on the present—is ultimately reducible to the asymmetry of thermodynamics.

One of the core principles of the thermodynamic account is a probability postulate—the assumption that all microstates compatible with the initial macro-state of the universe were equiprobable. This assumption, which ensures that the temporal evolution of the universe from its initial macro-state is overwhelming probability 'typical' and exhibits thermodynamically normal behavior plays a role analogous to the assumption of initial micro-randomness in my discussion above. Thus Albert, Loewer and I are in agreement on the importance of such an assumption to the success of causal representations. In the next section I want to what role in addition the assumption that the universe began its life in an extremely low-entropy state, what Albert dubbed 'the past hypothesis', can play in a reductive account of the causal relation.

3. Albert and Loewer's entropy account

The central assumptions of Albert and Loewer's reductive account are (i) the 'past hypothesis', according to which the universe began its life in whatever extremely low-entropy macro-state with which cosmology eventually presents us with, (ii) a probability postulate, which posits an equiprobability distribution over all micro-states with this initial macro-state, and (iii) a time-reversal invariant deterministic micro-dynamics. These assumptions, they argue, entail several asymmetries, which they take to be intimately connected to the causal asymmetry:

1. A temporal asymmetry of records or traces, which consists in the fact that there are localized records of the past but not of the future;
2. a closely related temporal asymmetry of knowledge, which is meant to capture the fact that we can in some sense know more about the past than about the future;
3. a temporal asymmetry of influence or control, which consists in the fact that we take ourselves to have some control over the future but to have absolutely no control over the past;
4. and, finally, a temporal asymmetry of counterfactual dependence and the asymmetry of causal influence.

In the previous section I have argued that we can acquire knowledge of the past state of a physical system, given merely localized data, through causal inferences. By contrast, Albert and Loewer focus in the first instance on the notion of records or traces of the past. Yet from my perspective this is merely a shift in emphasis, since inference appealing to records or traces can be thought of as a special case of causal inferences. Traces exhibit the same kind of fork asymmetry that is at the core of inferences to a common cause. Multiple records of the state of a system (such as the light emitted by a star recorded at multiple locations simultaneously) are all separate effects of a common cause and even the interaction of a single recording device with a system is an example of a causal fork.

In a recent talk, Albert described what he takes to be a crucial difference between our inferences toward the past and toward the future as follows: present traces of the past allow us to infer facts about the state of a system at some past time t_p without any knowledge of the history of the system between t_p and the present. Since traces can be relatively isolated from the system at issue after the recording interaction took place, the reliability of a trace may be completely independent of the evolution of the system after the recording interaction. By contrast, any prediction concerning the future state of a system at a time t_f crucially depends on the evolution of the system between the present and t_f.

While Albert did not make this connection, the asymmetry he described is exactly the asymmetry associated with causal forks that are open toward the future. Any correlations between a present effect E and the occurrence of an earlier cause C are independent of what occurs causally after C along other causal routes distinct from that leading from C to E. By contrast, whether some later effect E' occurs, does not depend merely on the occurrence of a single present cause C' and the route from C' to E', but depends on the occurrence of other causes of E' as well. We can express this asymmetry in interventionist terms: Interventions in the causal future of C on causal routes other than that linking C and E cannot affect whether or not the occurrence of E is a reliable record of the occurrence of C. By contrast, interventions in the causal past of E' on causal routes distinct from the one leading from E' to C' can interfere with any correlations that might otherwise exist between E' and C'. Thus, in order to predict E' on the basis of the occurrence of C' we need to know what occurs on other causal routes leading into C', while other causal routes leading away from C are irrelevant to the reliability of the correlation between E and C.

Under what conditions can a localized event E be a reliable record of an earlier event C that caused it? Albert has convincingly argued that records are always, at least implicitly, inferences from two times to a time in between. In addition to the recording event E we also need to make an assumption about the 'ready state' of the recording system prior to its interaction with the recorded system. In the previous section I have arrived at a similar conclusion. Inferences from correlations among different effects (playing the role of records) to a common cause as the recorded state depend on an assumption of initial randomness.

Albert (2000) has forcefully argued that not only the probability postulate but also the past hypothesis plays an important role in ensuring the reliability of records: the *past hypothesis* can at least in principle, as he puts it, play the role of the "mother of all ready conditions". Conditionalizing the occurrence of a putative record on the past hypothesis, that is, can ensure that that record is reliable. This reliability of records, according to Albert and Loewer has the consequence that the past, unlike the future, is dynamically insensitive to small changes in the present macro state: if we feed a state that differs only slightly macroscopically from the actual present state of the world into the dynamics, conditional on the past hypothesis and assuming the probability distribution induced by the initial equiprobability postulate, then we find that the macro-past with overwhelming probability would have been what it actually was and this is so, because conditionalizing putative traces of the past on the past-hypothesis ensures that these traces will with overwhelming probability have been reliable. For Albert and Loewer it is this dynamical robustness of records, when underwritten by the past hypothesis, that underlies the asymmetry of causal influence. We do not take ourselves to have any influence over the past, because the thermodynamic account entails that the past would have been what it was anyway even if we had decided to act differently from the way we actually did. Thus, while I have suggested that the asymmetry of records or traces is a special case of the asymmetry of causation, Albert and Loewer's strategy is to reverse the order of explanation. The thermodynamic account underwrites the asymmetry of records and hence of knowledge, which in turn allow us to account for the asymmetries of influence, counterfactuals, and causation.

In what follows I want to examine critically whether the past hypothesis can play the role that Albert and Loewer assign to it. I want to begin by asking more generally what the connection might be between the existence of records or traces of the past and the fact that the entropy of a closed system is overwhelmingly likely not to decrease—a fact that Albert and Loewer take to be underwritten by the past hypothesis. Albert and Loewer emphasize that many of our inferences about the past would be radically false, if these inferences were made only on the basis of the present macro-state and an equiprobability distribution over micro-states compatible with that state. Because of the time-reversibility of the micro-laws we would mistakenly conclude that entropy was overwhelmingly likely to have been higher in the past. For example, if we encountered a half-melted ice cube in a glass of water, the dynamical laws plus an equiprobability postulate would lead us to infer not that the present state of the ice cube was evidence of someone having put an unbelted ice cube into the glass earlier, but rather as evidence for the presence of a glass with water that anti-thermodynamically had spontaneously begun to freeze. To block this mistaken retrodiction, Albert and Loewer maintain, we need to postulate the past hypothesis. That is, without an assumption that can ensure thermodynamically normal behavior, many of our retrodictions would be radically mistaken.

But this argument can only establish that the past hypothesis is a necessary condition for the reliability of many of our records and does not yet show why we do not also have records of the future. After all, our inferences toward the future based on the dynamical laws face no similar 'thermodynamic obstacle' as our inferences toward the past. In the first instance, then postulating a past hypothesis merely ensures that our dynamical inferences towards the past are *as good* as our inferences toward the future. But what we are still looking for is a reason why we can know *more* about the past—that is, why our inferences toward the past are in some sense more powerful than our inferences about the future. What else, then, can we say about the connection between records and the thermodynamic arrow?

The example of the half-melted ice cube mentioned above might suggest that the connection between records and the direction of entropy increase is that the record state is one of non-maximal entropy and that the systems ready state must have been one of even lower entropy. Albert has argued, criticizing a suggestion by Reichenbach along these lines, that this is false and that a system locally in equilibrium can also function as a record (Albert 2000). I want to amplify Albert's discussion here. Since the measuring system cannot be a closed system during the measuring interaction, there are in fact no restrictions on the relative entropy of the record state and the measurement state and the entropy of the record state can be higher, lower, or the same as that of the ready state. It is easy to find examples of each kind.

First, an exposed film is an example of a system that ends up in a state of higher entropy than its ready state. An unexposed film consist of an emulsion of silver bromide molecules. During the exposure silver atoms are formed. The sum of the molar entropies of atomic silver and bromine is higher than that of silver bromide, and hence the entropy of the system constituted by the exposed film goes up, despite the fact that after being exposed the film intuitively contains a much larger amount of 'information' and might seem to be much more highly 'ordered' than the unexposed film. Thus, one has to be careful not to confuse the thermodynamic notion of entropy with a more intuitive notion of information.

Second, an example of records that have much lower entropy than their respective ready states are tree rings, which climate scientists use as a proxy for historic temperatures. Roughly, the size and density of tree rings are a function of the yearly temperature and therefore can be used as a record of the temperature. The chemical entropy of cellulose and the other constituents of a tree trunk is much lower than that of the CO_2 and water out of which they are formed through photosynthesis.

Finally, an arrow or other sign arranged out of pebbles found on a beach to signal the way, is a record that has the same thermodynamic entropy as the random pattern of stones out of which it was formed.

So the question remains: since the ready state of a measuring system can have higher, lower, or the same entropy as the record state, what accounts for the fact that there are records only of the past? Now, there are several further interesting and important connections between the existence of records and the second law of thermodynamics, but none of these, I think, get at the kind of connection that Albert and Loewer are after: On the one hand, very many (and perhaps most) records depend on the presence of friction and on the availability of low-entropy energy reservoirs. This suggests that the second law of thermodynamics does play an important role in the existence of records. On the other hand, the second law of thermodynamics also is a great destroyer of records. Some of what Albert and Loewer say might be taken to suggest that records of past macro-events are pervasive and that traces of virtually the entire past are somehow 'baked' into the present macro state of the world. But it is important not to exaggerate how prevalent records of the past really are. Precisely because of the increase of entropy and the evolution toward equilibrium very many past events have left no macro-traces in the present. Thus, as I have argued elsewhere (Frisch 2010), Albert and Loewer's account has the consequence that we can influence the past: every time our decision to perform a certain action are correlated with the occurrence of some past event E that has left no other macroscopic traces in the present, the decision will come out as a cause of the past event E, according to their account.

There is one argument, however, for the claim that the past hypothesis entails that we can know much more about the past than the future, which parallels Albert's discussion of the claim that the past hypothesis can function as an ultimate ready condition and which, therefore might be able to support the latter claim. This argument, which I want to call the 'constraint argument' appeals to the fact that low-entropy states correspond to much, much smaller regions of the phase space available to a system than states of higher entropy do. Imposing a low-entropy constraint on the past of the universe (or of a finite subsystem) thus puts a severe constraint on the system's past. In explicit premise conclusion form the argument can be represented as follows:

1. Low-entropy states correspond to extremely small regions of the phase space available to a system.
2. Therefore, imposing a low-entropy constraint on the past of the universe puts a severe constraint on the system's past.
3. Since the universe evolves toward states of ever higher entropy, its future evolution is much, much less constrained.
4. Therefore, we can know much more about the past of the universe than about its future.

But what this argument does not take into account is the fact that the different macro-states of a system that are accessible and of interest to us do not correspond to phase space regions of comparable size. Thus, when we know that a system is in an equilibrium state, then we know all there is to know about the system macroscopically, even though we have not been able to narrow down very much what region of its phase space the system occupies. While in some sense we know much less about a system when we know that it is in equilibrium than we know of the system that it is in a specific extremely low entropy state—we have narrowed down the region of phase space that it currently occupies much less—in another sense we know just as much about the system: in both cases we know its exact current macro-state.

In fact, if all we know of the system's past is that it was in *some* extremely low entropy state, but we know of the system that it will evolve into its equilibrium state, then macroscopically we know *more* about the system's future than its past. Take as an example the case of a body of gas that is expanding in a box and of which we know that it was compressed into 1/100 of the total volume in one of the eight corners of the box (but know nothing more about the past state). Then while the system's past micro-state is much, much more constrained than its future micro-state—we have narrowed down the phase space regions that were accessible to the system in the past much more than the regions accessible in the future—we have full knowledge of the system's future macro-state but do not know everything there is to know about the system's past state.

Imagine that we were playing the following game: we are each given a box with a gas in some randomly prepared macro-state and our goal is to acquire as much information as possible about the system's macro-state. Whoever can determine more of the macroscopically available information about the gas in the box wins the game. Imagine now that I received a box in which the gas is in equilibrium and evenly spread throughout the container and I am able to determine this (along with the values of all the thermodynamic parameters characterizing the state of the gas); and that you received a box with a gas very far from equilibrium compressed into a small volume in one of the corners of the box and with the partitions constraining the gas just removed. Let us assume that you were able to determine that the gas was compressed into one of the corners, but perhaps because the box is rotating and you were not able to keep track of its rotations you were not able to determine into which corner the gas was compressed. Then it seems clear that I won the game: I have been able to determine all the information about the gas that is macroscopically accessible, while you have not been able to do so. And this is so, even though if our goal were to be able to constrain the possible micro-states of the system as much as possible you would have won—and it would not even have been close: you would have won that contest by many, many orders of magnitude and the fact that you failed by a factor of eight to maximally constrain the available region of phase space would make no difference to that other contest. Thus, even though you handily would have won the game of narrowing down the phase space region which the system occupies, you lost of the game of determining the system's macro-state as precisely as possible.

Thus, the constraint argument fails and, therefore, cannot successfully be adopted to support the claim that the past-hypothesis can function as ultimate ready state. In fact, we can now see that analogous considerations also call into doubt the claim that the past hypothesis provides a constraint sufficient to ensure the reliability of records. Here is an explicit version of that argument—the ready-state argument:

1. Inferences based on records presuppose assumptions about an earlier ready state.
2. The past hypothesis can function as the mother of all ready-states.
3. Therefore putative records of the past are reliable conditional on the past-hypothesis: If R is a record of a past state S, then $Pr(S/R\&\boldsymbol{PH})\approx 1$.
4. Therefore small counterfactual changes to the present are associated with changes to the future but not to the past.

Both the ready-state argument and the constraint argument rely crucially on the idea that the past hypothesis puts a severe constraint on the past evolution of the universe. But like in the case of constraint argument, it is unclear whether positing that constraint can on its own carry the burden in the argument that it would need to carry. As our discussion above has shown, positing that there is *a* low-entropy constraint is not enough to ensure a unique past evolution. It is compatible with the assumption of an extremely low-entropy initial state that a system that evolved into a macro-state slightly different from its actual macro-state began its life in a different, non-actual low-entropy state. Thus, if the macro state of the gas in our example, when it is close to equilibrium, had been slightly different, it might have evolved from a low entropy state that had the gas constrained into a different corner than the one in which it was actually located. Recall that the phase space regions corresponding to the forward evolutions of macroscopically different low entropy states are highly fibrillated—they have 'fingers' spread throughout the entire phase space volume that corresponds to the later high-entropy state. This has the consequence that if we move the system's present phase space point away from its actual trajectory to one corresponding to a slightly different macro-state, we are as likely to land on a trajectory that has evolved from a macroscopically different low-entropy past than on a trajectory that evolved from the actual low-entropy past. (Of course, the overwhelming majority of trajectories one which we could land evolved from a high-entropy past. But these trajectories are excluded by positing the past hypothesis.)

Now Albert may respond (and in fact has responded) to this kind of argument by insisting that the past-hypothesis contains much more than the claim that the universe began its life in an extremely low entropy state: it states that the universe began its life in whatever "Big Bang-ish sort of state" cosmology eventually presents us with. Yet it is unclear how much any additional cosmological constraints might help, since any such constraint will be much more coarse-grained than the kind of small-scale differences in the states of medium-sized macroscopic objects involved in common sense causal claims. No matter what the constraints on the initial state of the universe cosmology will eventually present to us, these constraints will be compatible with many different histories of 'medium-sized dry goods.' That is, even Albert's richer past hypothesis appears not be sufficient to ensure the reliability of records.

There is one further reply one can give to my criticism of the ready-state argument. Once we focus not merely on the history of a single system, like the body of gas or the entire universe taken as a whole, but consider a complex system with many different interacting subsystems, it seems plausible that we would find that otherwise reliable records or traces of the past of a system could only be misleading or unreliable in a close counterfactual situation, if the micro-state of the system in question exhibited correlations of the kind that the initial randomness assumption, which we discussed in the previous section, is intended to exclude. Imagine, for example, that we took a picture of the body of gas as it was expanding and the picture clearly indicates that the gas was compressed into a corner marked with a red dot. Then in the actual situation this picture serves as a record of the gases past state. If we now consider a counterfactual change to the macro-state of the gas such that the resulting counterfactual macro-state is overwhelmingly probable to have evolved from a non-actual low-entropy past, then the resulting 'world' would have to have contained strong correlations among the variables characterizing its initial micro-state. For it would have to be the case *either* that the photograph of the initial state of the gas was unreliable and there were strong correlations among the incoming light waves that mimicked the light waves reflected by the gas at its actual past location and masked that the light was reflected at the gas's counterfactual past location; *or* that there were strong initial correlations among the constituent atoms of the gas that resulted in improbable fluctuations in the gas's macro-history—or perhaps both. In any case, the counterfactual 'world' would have to violate the assumption of initial randomness.

Thus Albert and Loewer' premises do after all seem to entail that traces or records of the past are overall reliable, and hence that there are certain counterfactuals—those associated with small macroscopic changes to the present—that are temporally asymmetric, and hence, perhaps, that there is a causal asymmetry. But if what I have just argued is right, then the entire burden of the argument is carried by the probability postulate or the assumption of initial micro-randomness. The past-hypothesis does no work—no work, that is, aside from providing a *necessary condition* for the reliability of records by preventing an overwhelmingly anti-thermodynamic past and thereby ensuring that inferences to the past do not go radically wrong.

4. Conclusion

Where does this leave the debate about the two strategies I distinguished above? The initial randomness assumption, I have argued, is quite naturally thought of as a causal assumption that fits well with our explanatory practices of trying to explain correlations between distant events in terms of earlier common causes. We would expect initial states to be randomly distributed in the absence of an even earlier common cause. Thus, I think the burden of proof lies with defenders of a reductive strategy to show that the interlocking pieces of our causal representations of phenomena—our time-asymmetric explanatory practices, the positing of time-asymmetric causal relations, and the assumption of initial randomness—can be reduced to a set of non-causal assumptions. I am skeptical that Albert and Loewer's thermodynamic account can discharge this burden. If, as I have argued, the entire work in their account is done by the equiprobability or randomness assumption, they have not succeed in deriving the causal and explanatory asymmetries from what are clearly acausal assumptions.

References

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

Arntzenius, Frank. 2010. "Reichenbach's Common Cause Principle", *The Stanford Encyclopedia of Philosophy (Fall 2010 Edition)*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2010/entries/physics-Rpcc/>

Earman, John. 2011. Sharpening the Electromagnetic Arrow(s) of Time. in Craig Callender (ed.), *The Oxford Handbook of Philosophy of Time*. Oxford: Oxford University Press.

Elga, Adam. 2000. Statistical Mechanics and the Asymmetry of Counterfactual Dependence, *Philosophy of Science* suppl. 68: 313-24.

Frisch, Mathias. 2005. *Inconsistency, Asymmetry and Non-locality: A Philosophical Investigation of Classical Electrodynamics*. New York: Oxford University Press.

———. 2006a. Causal Asymmetry, Counterfactual Decisions and Entropy. *Philosophy of Science* 72 (5):739-750.

———. 2007. Causation, Counterfactuals and the Past-Hypothesis. In *Russell's Republic: The Place of Causation in the Constitution of Reality*, edited by H. Price and R. Corry. Oxford: Oxford University Press.

———. 2010 "Counterfactuals and Entropy," in Andreas Huettemann and Gerhard Ernst (eds.) *Time, Chance, and Reduction*. Cambridge University Press. 13-33.

Lewis, David. 1986. Time's Arrow. In *Philosophical Papers*, vol. II. Oxford: Oxford University Press.

Loewer, Barry. 2007. Counterfactuals and the Second Law. In *Causality, Physics, and the Constitution of Reality: Russell's Republic Revisited*, edited by H. Price and R. Corry. Oxford: Oxford University Press.

Sober, Elliott. 1984. Common Cause Explanation. *Philosophy of Science* 51: 212-241.