Entropy **2012**, *14*, 344-369; doi:10.3390/e14020344

OPEN ACCESS

entropy

ISSN 1099-4300

www.mdpi.com/journal/entropy

Article

Interventionism in Statistical Mechanics

Stephen Leeds

Department of Philosophy, University of Wisconsin, Curtin Hall, 3243 N. Downer Avenue, Milwaukee, WI 53211, USA; E-Mail: sleeds@uwm.edu

Received: 13 January 2012; in revised form: 9 February 2012 / Accepted: 9 February 2012 / Published: 17 February 2012

Abstract: I defend the idea that the fact that no system is entirely isolated (“Interventionism”) can be used to explain the successful use of the microcanonical distribution in statistical mechanics. The argument turns on claims about what is needed for an adequate explanation of this fact: I argue in particular that various competing explanations do not meet reasonable conditions of adequacy, and that the most striking lacuna in Interventionism—its failure to explain the “arrow of time”—is no real defect.

Keywords: interventionism; Albert; arrow of time

PACS Codes: 01.70. +w, 05.20.-y

1. Introduction

In the last chapter of *Time and Chance* [1] David Albert presents his by now well-known GRW-based approach to statistical mechanics. In the course of his defense of that idea, he brings up a possible response:

“It has often been suggested…that nothing even remotely as up-to-date as quantum mechanics is going to be required here—that…the sorts of perturbations we were talking about are already *all over the place*, if one simply stops and looks, in (say) the *Newtonian* picture of the world. The idea is that since none of the … systems of which we have ever had any experience … are genuinely *isolated* ones, the perturbations in question can be seen as arising simply from the interactions of the … system with *its environment*” [2].

The view in question, which Albert goes on to reject, is often called Interventionism. It is an idea that has been in the literature at least since Boltzmann, and has had now and then notable supporters: the reader is likely to have heard of Borel’s calculation of the effect on a gas on Earth of moving a gram of matter on Sirius; more recently, Michael Redhead and Katinka Ridderbos [3] have advanced a version of Interventionism, and John Earman [4] has indicated some sympathy with the view. Overall, however, the weight of opinion has generally been negative; and indeed there are objections to Interventionism that, depending on one's expectations about what work it is supposed to do, can seem devastating.

The qualification in my last sentence is, however, important. It seems to me that virtually all the philosophical issues concerning the foundations of statistical mechanics have to do with different conceptions of what we need to explain, and what would count as explaining it; for this reason, what can seem like a fatal objection to a particular approach, given particular expectations about what we should be trying to explain, might turn out to be no objection at all, given other conceptions of our explanatory task. This is the case, I believe, with Interventionism. A large part of a defense of Interventionism—at least of the one to be offered here—is really an argument for taking seriously the kinds of why-questions that Interventionism can answer, and for not insisting on answers to the questions that it can't.

This said, the organization of the paper will come as no surprise. I will begin with the questions about statistical mechanics to which some variety of Interventionism, if the mathematical and empirical details work out favorably, would provide a satisfactory answer; I will then compare the account with some of the currently popular approaches (including Albert's own, when he puts aside the GRW approach). I will then argue for ignoring the issues about which Interventionism has nothing to say. Since among these issues is one which many people consider the heart of the subject—namely, why entropy increases in the forward time direction—this will be the most controversial part of the argument. I will conclude with a few sketchy remarks on what a worked-out interventionist account might look like.

Let me now enter the usual caveats. First, our subject is classical statistical mechanics; except for some mention of GRW (here chiefly as an example of an account with some resemblance to Interventionism) we are for the most part going to pretend that quantum mechanics is irrelevant. Next, since this is a subject in which there would be very little to say if we were to stick merely to what is known, I am going to indulge in a great deal (though, I think, no more than the usual amount) of speculative dynamics. And finally, although I speak of arguing for and against different conceptions of our explanatory task, I do not really think there is enough agreement on what counts in general as a good explanation to allow us to establish or refute any of the competing conceptions, and I do not claim to do so here. What I am hoping to do is to raise considerations which tend to tell in favor of one conception, and against the others.

2. From GRW to Interventionism

I began with Albert, not only to have a snappy summary of the position, but also because his GRW story [5] is a close cousin of the version of Interventionism that I want to discuss, and attractive for much the same reasons. Let us review the GRW account. We take as the object of interest a system S which is quasi-isolated in the sense that the magnitude of any fluctuation in external forces is extremely small compared to its internal energy. An example might be a gas in a confined space or (Albert's favored example) two bodies of different temperatures in thermal contact. We suppose that S is to be described by giving its macrostate M, which we might identify with its single-particle density function or with some more coarse-grained description of the state of the system—e.g., the values of particular local or global thermodynamic observables. There will be various well-confirmed generalizations concerning how a system initially in a given M will develop in time, and it is our goal to explain these. Thus, in the case of two bodies of different temperatures in thermal contact, one might take the explanandum to be a generalization relating the composition and temperatures of the bodies at an initial time to their temperatures 10 min later. This is the only generalization Albert explicitly addresses, but it is reasonable to set one's sights somewhat higher. There are many cases in which the time-development of the single-particle density function of a given kind of system is empirically well-approximated by, or has a particular probabilistic distribution around, the solution to an appropriate evolution equation, and a completely satisfying account will need to address these as well. I will assume, as I expect Albert does, that the general pattern of explanation that applies to the simpler cases will carry over reasonably straightforwardly to these.

Let p be a point in the phase space Γ of S, representing a microstate m, and thereby also (non-uniquely) a macrostate Albert, restricting his discussion to the case of the two bodies in thermal contact, calls p a *normal* point if the system, started out in (the state represented by) p, is determined by the dynamics of the system in the absence of interventions to conform to our explanandum generalization, *i.e.*, to be in the predicted macrostate in t = 10 min—that is, to be in the macrostate M* that our generalization predicts for systems started off in macrostate M. Other points, naturally, are *abnormal.* Then the explanation offered by the GRW account consists in three parts. The first is a claim about the structure of Γ: that the abnormal points are in some appropriate sense scarce, or scattered. The second is a claim about the perturbing influences on the system, namely that these are the GRW collapses, probabilistically distributed as specified by that theory. Finally, it is claimed that the combination of these two features—the distribution of normal points, together with the perturbations—has the desired outcome: for *every* p, the probability is nearly 1 that our system, when started out in p, will, by a combination of the unperturbed dynamics of the system and the GRW perturbations, be in the predicted macrostate at t = 10 min.

Although I think the intuitive idea behind the account is clear enough, and although my purpose here is not to criticize the account—my goal is in fact to take it over, replacing the GRW collapses with classical outside interventions—it still seems worth pointing out some ways in which the account might be improved, or anyway clarified. We can simplify our discussion by thinking of the perturbations as coming at regular very brief intervals, with the system developing freely between these. Notice first that it is not enough to require that the first collapse will almost surely set the system "on course" to arrive at the predicted macrostate 10 min later, if left unperturbed; we need to insure that later collapses don't interfere with this process. Albert's account as stated will plausibly achieve this if we see the claim about the distribution of normal and abnormal points as applying not just to the case of t = 10 min. Rather one might for each t define the normal states for t analogously to the way the normal states were defined for t = 10 min, and make the analogous claim about their distribution. This will have the consequence that the system, after having pursued the predicted course of macrostates up to t = 5, and now in macrostate M', say, is likely to be perturbed in such a way that it is now "on course" to arrive at M* in 5 min.

We are now in effect requiring that the set of points we want to count as normal meets a very strong condition: we are requiring that a system whose phase point is initially normal will, if not perturbed, evolve in accordance with the empirical predictions forever, or at least for periods of time comparable to those for which we expect to observe the system. I think it is clear that this is Albert's intention; still, given that we want to claim that normal points are quite common—*i.e.*, that the abnormal points are in some sense scarce or scattered—this is quite a strong claim, perhaps stronger than what we need. Sticking for the moment with our picture of the collapses coming at fixed discrete intervals, it looks as if we can get away with a weaker notion of a normal point, namely as one which will evolve freely along the predicted path of microscopes until the time of the next collapse. In the actual case of the GRW perturbations, the interval between collapses is probabilistically distributed, but this presumably means only that we need to define the normal points as those that will evolve "correctly" for a length of time long enough so that a collapse within that period of time is extremely probable. It is worth keeping this in mind, given that the best result we have in the general area of showing "most" points evolve in accordance with an empirically confirmed evolution equation, namely Lanford's, doesn't guarantee such behavior for more than a fraction of a second.

Let us turn now to the sense in which the abnormal points, however defined, are scarce or scattered. The notion of scattering required cannot be a topological one: in fact, since, by reversibility arguments, the cardinality of abnormal points is as great as that of the normal ones, and because, by a continuity argument, every normal (abnormal) point is surrounded by a neighborhood (in the usual topology) of other normal (abnormal) points, the abnormal points are, from a topological point of view, no more scattered than the normal ones. It seems inevitable, then, that the appropriate notion of scattering will be described in terms of some more or less natural measure, say the microcanonical measure conditionalized on the energy of the system (I will refer to this as the mc-measure); thus, it might be sufficient to claim that the bad points come in isolated clusters of small individual (mc) volume, the volume of all the clusters also being small. A large part of our discussion is going to turn on questions about what role the various measures and distributions that show up in statistical mechanics are playing there, and whether we are entitled to use them to play such a role. For this reason, it is worth emphasizing that the role of the mc distribution [6] in the GRW account is not in any way that of a probability: it is merely a way to describe the structure of phase space. The only probabilities in the story are those of the interventions; it is for this reason that the account requires that *every* phase point p has high (intervention-driven) probability to evolve in the right way. If we had required only that mc-*most* p so evolve, we would then need to explain the connection between this (mathematical) fact and the observation that most actual systems do go to equilibrium: we would need to say something about how the states of empirically given systems are distributed, and how this relates to the mc-distribution. One distinctive feature of the GRW account is that in it nothing needs to be said about the distribution of actual systems: they all have high probability of going to equilibrium, whatever phase point they are in initially.

I have been speaking of the GRW account, but in fact, as the quotation from Albert suggests, the entire summary applies just as well to Interventionism—at least to one version of it. The sole difference is in the origin of the interventions: for GRW they are the spontaneous localizations; for the interventionist, the perturbations are supplied from outside. The interventionist also takes the interventions to be probabilistically distributed; here, however, the relevant notion of probability is not chance, as it is in GRW, but frequency: our interventionist will claim that as a matter of empirical fact the outside interventions on any given system are distributed in a way that can be modelled by a probability distribution, or by one or another member of a family of probability distributions. It is not an automatic requirement on interventionist accounts (or for that matter on GRW- type accounts) that they refrain from assigning probabilities to the initial states of systems, but my account, like Albert's, will not do so; as will emerge in the next section, I think this is one of the attractive features of both of these approaches.

The obvious difficulty with Interventionism is that it offers no explanation for the distribution of the interventions. There is no escaping this; the best strategy for the interventionist is to argue that every theory needs to suppose a more or less definite distribution of something—whether of collapses, as in GRW, or of the actual states of systems—and that the assumptions he needs to make about the interventions will turn out to be rather weak. As to how weak an assumption will be needed, the hope is that the situation turns out to be like the familiar explanations of why it is that a well-balanced coin comes up heads half the time. Here too one has a phase space—that of the various parameters characterizing the coin's trajectory at the moment of release—and the phase space has an interesting structure which is most easily characterized in terms of a particular measure (Lebesgue measure over position × velocity space): namely as an alternating sequence of narrow bands of roughly equal measure—those destined to land heads and tails. And likewise, it is convenient to model the "choice" of initial points in the phase space, *i.e.*, the tossing of the coin, by one or another probability distribution. One might wonder what is gained by these explanations, seeing that we need to presuppose a probability distribution of the tosses in order to get a distribution for the outcomes. But of course a great deal is gained: although something needs to be presupposed about the tosses (it is not every possible distribution of tosses which will "harmonize" with the layout of the bands so as to produce heads on the average half the time) we don't need to presuppose much: any distribution of tosses that varies reasonably slowly over the bands will do the trick. Likewise, one hopes, in Interventionism. We cannot get away without making some assumption about the distribution of the incoming perturbations, but it might be that any distribution under which there was substantial probability that the system would be moved onto an abnormal point would need to be very weird indeed.

Which account is preferable, GRW or Interventionism? I have no quarrel with anyone who claims that, considered solely on their merits as explanations, and leaving aside the question what we have reason to believe, the GRW account is a better explanation. I can agree with this even while remarking that no GRW account can entirely leave out the influence of external perturbations—which of course do occur [7], and about which the GRW account must assume at least this much: that they are distributed in such a way that it seldom if ever happens that an external perturbation undoes the work of the quantum collapses by knocking the system right back into an abnormal state. Such an assumption is of course the same kind of claim about external perturbations as that which the Interventionist needs to make—namely, a claim about how, as a matter of empirical happenstance, they are distributed; however, it presumably will be weaker than what the Interventionist needs to suppose, and I am entirely in sympathy with the view that the less we attribute to such empirical happenstance, and the more to fundamental physical law (in the present case the GRW collapses), the better for our explanations. However, given that so far we have no direct evidence for the GRW collapses, an Interventionist who concedes this much to GRW is conceding very little. And in fact Albert's criticisms of Interventionism, although they appear in the GRW section of his book, have nothing to do with that rather speculative theory. Rather, they are based on his adherence (when he brackets his GRW views) to the approach, stemming originally from Boltzmann, that is most popular nowadays: that in which we explain the facts of thermodynamics by invoking, not dynamical probabilities, as does GRW, but rather probability distributions on initial states—whether on the states of the particular quasi-isolated system whose behavior we want to explain, or, in a variant that has recently gained some currency, on the entire universe at the time of the Big Bang. It seems to me that the Interventionist point of view is best appreciated when set against the more familiar initial-state approaches, and that is what I want to do in the next section.

3. The Initial States Approach

Let's begin with the kind of account closest to Boltzmann's own, in which we invoke a distribution on the states of the quasi-isolated system of interest. In Boltzmann's original version of the account, the favored distribution is the mc distribution; but let us leave open the possibility that we will want to invoke some other distribution instead, and just refer to the distribution we will want to use as μ. So according to this account—at least in its most straightforward version (we'll be looking later at some less straightforward versions)—the explanation of why a glass of ice-water in a warm room, described as being in a particular macrostate M, with a given Hamiltonian H, goes to equilibrium is that (A) μ-most states consistent with M are good, *i.e.*, destined to melt in the predicted relaxation time.

It is widely acknowledged that when μ is taken to be the mc distribution, the account runs into difficulties, and it is widely agreed that these have something to do with the fact that mc-most states consistent with M will also evolve to equilibrium if run in reverse. I agree that some version of a reversibility objection is a problem for the account, but it is important to see exactly why.

A good place to begin is by noting that our outline of the Boltzmann account is as it stands incomplete. To see this is to consider a proposed explanation whose incompleteness is more obvious: namely one consisting only of a specification of H, M, and the dynamical laws D. No one would put this forward as explaining why the ice-water melts, but why not? Well, there simply is no acceptable model of explanation that would cover this instance: the conjunction H&M&D does not logically or mathematically imply the explanandum; nor is it a claim about chances or frequencies that would allow us to assimilate our explanation to any familiar statistical model of explanation. Notice, however, that A, above, is merely a mathematical consequence of H&M&D. It follows that, for exactly the same reasons, H&M&D&A cannot explain why the ice-water melted. Our explanandum is not a logico-mathematical consequence of H&M&D&A; and if A made any claim about chances or frequencies such a claim would already be derivable from H&M&D, which it isn't.

The point I am making is in one way or another widely conceded; different writers have different ways of acknowledging that there is something more that needs to be added to the explanation, specifically something more that needs to be said about μ. In [1], Albert speaks of his preferred replacement for the mc-distribution as the *right* one to use in making inferences. In later work (so far unpublished), he interestingly refers to it as the "true" distribution. Other authors signal the special role of their preferred μ by speaking of sets of large μ-measure as *probable*—not "probable according to μ" but probable *tout court*. As I see it, these various locutions are merely metaphorical placeholders for a substantive claim about μ, and, once the need for such a claim is acknowledged (which, I am sorry to say, is not universally the case) there is a pretty natural idea about what it might be. I propose that the claim we need to make—let us call it (C)—is some version of the idea that μ be a close match to the empirical statistics. This will have the advantage of allowing our explanation to conform to a standard model: it will be a statistical explanation based on frequencies [8].

Since it is going to be (C), and (C) alone, which accounts for the distinctive role of μ in our explanations, we want it to pick μ out uniquely. For (C) to require that μ be a best possible fit to the empirical statistics would surely rule out any μ that anyone has ever heard of, but there is a standard Lewis-style fix for this: ask that μ be both quite a close match, and that it be much simpler than any better matching distribution. Putting things so generally leaves of course a lot of freedom in specifying (C); not only as concerns criteria for closeness of match and simplicity, but also as concerns the acceptable macrostates M for describing the initial conditions of the systems of interest. There are however also some natural constraints on what (C) needs to look like: in particular there are some criteria for closeness of match that seem to be built into the nature of what we are trying to do.

Here is one such. Although we are assuming that the empirical frequencies will not be an exact match to μ, we should require that the discrepancies be in a sense random: it can't be that there are easily specifiable domains in which the match holds up, and easily-specifiable domains in which there will be significant mismatch. The reason for this is that we are anticipating that our explanations end at this point: we are not expecting to show why it is the empirical frequencies happen to match μ. There is of course nothing wrong with explanations that end in a brute statistical generalization, but such a brute generalization needs to be reasonably across-the-board: one whose successes and failures conform to an easily specified pattern is one which, as we say, cries out for explanation, not one in which explanations can come to rest. To say this is, no doubt, to say something not about Nature, but about ourselves and our explanatory preferences; but explanation is, after all, a human activity, and it is appropriate here, as in analogous cases elsewhere [9] to take account of what we do and do not consider a good explanation.

I submit that this is the underlying reason that most people who pursue some version of the Boltzmann program nowadays begin by conceding that the mc-distribution is not the μ we are looking for. Had the mc-distribution been a close match to the empirical frequencies, I think there is little question, given how natural a distribution it is, that we would have been satisfied to take this as a brute statistical fact, one that itself did not require further explanation. But it is not such a match, and in fact we can point to ways in which mismatch is guaranteed: the phase points occupied by actual systems consistently fail to fall into the enormous region of phase space occupied by the time-reversals of normal points, *i.e.*, by points which evolve to equilibrium in the past time-direction.

I intend to rely heavily in the sequel on the condition just stated for a given distribution μ to be an acceptable one to use in explaining the facts of statistical mechanics. I have no argument to offer for it besides the claim that it expresses our usual practice, but it may be helpful, by way of showing how I intend to use the condition, to point to an approach that I think violates it. This is one reading of the idea, put forward by Huw Price [10] that the reversibility issues do not undercut the ability of the mc-distribution to explain the approach to equilibrium in the time-forward direction, since there is an excuse for its failure when we try to apply it backwards. This excuse is that the initial state of the universe plays the role of a boundary condition, which "overrides" the mc-distribution when we try to apply it in reverse. Now it seems to me that there are two ways to read Price here. He might be saying that there is a distribution that *does* match the empirical frequencies, even in the regions of phase space that are of interest in the reversibility arguments: namely the mc-distribution in some way conditionalized on the macroscopically described initial conditions—what I will from now on call the mc* distribution. (This is the line taken by Albert, which I discuss below). If so, then there is no conflict with what I am proposing, so long as we agree that it is then not high mc-probability but high mc* -probability that is doing the explanatory work. But it is not clear to me that Price does take his position to stand or fall with the claim that the mc*-distribution is an across-the board good match to the empirical statistics [11]. His view might be, and I think the word "override" suggests, that it is the mc-distribution that is doing the explanatory work; so long as we can explain away the failure of the mc-distribution to apply backwards, we do not need to find a distribution which can be used backward as well as forward. If so, then his view violates the condition I am proposing. By way of response to it, it seems to me that talk of "overriding" the mc-distribution makes sense only if one thinks of the mc-distribution as in some sense operating behind the scenes, despite not matching the statistics. This might be the case, for example, if it was legitimate to think of systems as having a propensity to conform to the mc-distribution, since a propensity can be real despite not being realized. But talk of propensities is surely ruled out here, and I see no other way to make sense of the idea that the mc-distribution is overridden, or that it is in some way the "right" distribution, despite not matching the statistics.

Returning to the more common view, that reversibility problems do indeed disqualify the mc-distribution, the current most popular fix is that of (among others) Albert in [1]: we take μ to be the mc*-distribution, that is, the mc-distribution conditionalized on macroscopic initial conditions; this is usually combined, as it is by Albert, with a specific hypothesis about what the initial conditions were like, e.g., that they were a state of extremely low entropy. (Albert calls this the Past Hypothesis (PH)). The advantage of this over the mc-distribution is that, it is claimed, it does not retrodict that our glass of ice-water was previously less melted.

I have been speaking as if it were unambiguous to speak of conditionalizing the mc-distribution on macroscopic initial conditions, but in fact, there are two distributions that can be considered to be the mc-distribution conditionalized over PH. One possibility is to begin with the mc-distribution over the phase space whose points represent states of the entire universe; then the mc*-distribution at time t after the Big Bang (in some standard reference frame) would be the result of conditionalizing the mc- distribution over the part of phase space which represents points whose backward evolution over time t leads to a state satisfying the PH. The mc*-distribution thus becomes a distribution over the phase space of the entire universe. The other possibility, more in keeping with the original project of explaining why the glass of ice-water melts by finding an appropriate distribution over the phase-space of the ice-water, is to think of the mc* distribution for the ice-water as the result of beginning with the mc-distribution over the phase space of the ice-water and appropriately conditionalizing *that*. What subset R of the phase space of the ice water are we supposed to conditionalize over? Here I can think of nothing better than what Eric Winsberg proposes in [12]: R (better called R_t, since it depends on how long the ice-water is after the Big Bang) should consist of those phase points p which are compatible, not only with the system's macrostate, but also with the PH in the following sense: p is the restriction to the phase space of the ice water of a state of the entire world which, traced back in time t seconds, fulfills the PH. Let us call such states of the world t-PH states. The mc*-distribution—for the glass of ice-water, at t—is the conditionalization of mc over R_t.

Will the mc* distribution work both forward and backwards? Winsberg offers a good reason for thinking not. If one wants to show it mc*-improbable that a glass of ice-water in a warm room was earlier even more melted than it is now, it must be the case that R differs substantially from the subset of the system's phase space picked out by its macrostate. But it is by no means clear that there is *any* state of a glass of ice-water which cannot be embedded into a t-PH state of the entire world. Grant that it would require extremely delicate correlations for the world to be so set up that, starting from the current macrostate containing a glass of ice-water and looking backwards, we first see the ice melting, and then everything arranging itself to conform to the PH. At best, what such arguments show is that certain states of the world are *improbable* according to one or another measure. They are not arguments that show the dynamical *impossibility* of anything: in particular, they do not tend to show what we need to show: that there are states of our glass of ice-water which cannot in any way whatever be embedded in a t-PH state of the world. Indeed, so far as one has hunches about this sort of thing at all, they are all on the side of the thought that, for any given state of the glass, there is *some* way to arrange the rest of the world so as to allow things to evolve (in the past direction) to make the PH hold [13]. We have thus been given no reason to suppose that R is any smaller than the entire phase space for our glass of water. If R is just the original phase space, or close to it, then conditionalizing over R accomplishes nothing: the mc*-distribution just is the mc-distribution, and we are still stuck with a reversibility problem.

This leaves us with the first of our two options for what we should take as the mc*-distribution, namely as applying to the entire universe. The mc* distribution will now be the (evolving) distribution over the phase space of the entire universe (supposed to be permanently fixed—let us pretend for now there is no creation or annihilation of particles), which begins as the uniform distribution over the set of phase points in which the world is in the PH state, and which then evolves by Hamiltonian dynamics. If we use this distribution to assign probabilities, there is no reason to expect problems analogous to Winsberg's: if we want to know the probability, given that the world contains a glass of ice-water at certain coordinates at t, that it contained a more melted glass of ice-water at the same coordinates a minute ago, it won't matter if every possible state of the ice-water itself can be matched with *some* state of the rest of the world to produce a t-PH state of the whole world: what we care about now is the mc*- measure of states of the entire world, and we might hope to argue that the measure of such states in which the ice-water was previously more melted will turn out very small.

Does this mean that the mc*-distribution matches the empirical frequencies well enough to conform to a reasonable version of our condition (C)? It is clear that the usual claims about the statistics of individual systems—e.g., that mc- most microstates of a glass of water in macrostate M are normal—can be reformulated as claims about states of the world, e.g., that mc*-most states of the world in which there is a glass of ice-water centered at (x,y,z) (and in which no strong outside forces are destined to impinge) are also states in which there is a glass of somewhat more melted ice at (x,y,z) a few seconds later. Suppose these claims are in fact correct. In what sense, however, can we say that we have found a match between the mc*-distribution and the empirical statistics? "There being a glass of ice-water centered at (x,y,z)" is indeed a macrostate of the entire world, in the sense of selecting a well-defined part M of the phase space of the world ; but clearly we don't want to claim that the empirical distribution of the phase-points occupied by the world on those occasions when it happens to be in M is much like the mc*-distribution conditionalized over M: the great mc*-majority of phase-points corresponding to M represent macrostates in which there are no life forms anywhere at all. In what sense, then, can the empirical frequencies be said to match the mc*-distribution? Things seem to work out well so long as we only ask whether the world is in subsets M' of M which themselves can be described in terms of the same glass of ice-water—e.g., the set of points in M which are destined to evolve into a world in which a more melted glass of ice-water is present at x,y,z. But this extent of match between distribution and empirical frequencies does not come close to meeting our requirement that the empirical statistics resemble the right μ in a reasonably across-the-board way, not simply in particular specifiable respects.

This might suggest that in specifying the descriptions of the world relative to which we can expect the mc*-distribution to resemble the empirical statistics, we must not use "partial" descriptions like M: what we need is detailed macrodescriptions of the entire world. I am not sure whether such a restriction meets the requirement that the mc*-distribution matches the empirical statistics in a reasonably across-the-board way, but there is in any case a more serious problem, namely that the match between mc*-distribution and empirical statistics now becomes completely trivial.

For any "full" macroscopic description M* of the entire world, the world will occupy phase points in M* at most once (or perhaps twice if the expansion of the world is fated to reverse itself); it goes without saying that the distribution of these one or two points over the subsets of M* will not be seriously discrepant with the mc*-distribution, or indeed any other [14]. This however is not the kind of non-trivial resemblance between the mc*-distribution and the empirical statistics that we need if we are to justify the explanatory role of the mc*-distribution on the grounds that it is a best fit to the empirical statistics.

I conclude there is reason to be sceptical of those versions of the initial states approach in which we justify the use of a particular distribution in our explanations by claiming that it matches the empirical frequencies. There is however a related idea which Albert sometimes endorses, one which sidesteps the issue of finding a distribution that matches the empirical statistics. This is the idea that all we need to show is that the starting states of the world that lead to what we might call "good histories" are typical for the initial macrostate M (alternatively, for the initial part PH of phase space occupied by points that satisfy the Past Hypothesis) in the sense of occupying an mc- large proportion of the part of phase space occupied by M (or PH). What is a "good" history? Since we are trying to validate the use of the mc-distribution in statistical mechanics, a good history must at least be one in which forward-looking calculations using the mc-distribution are predictively successful; one might also ask, as Albert does, that such histories would validate use of the mc*-distribution [15] in postdiction, or even in the rough calculations of conditional probabilities that guide our lives outside the usual domain of Statistical Mechanics.

From the point of view we have been taking so far, this suggestion is a non-starter. There is no sense in which the one and only actual starting state of the world can be said to be empirically distributed in a way that approximates the mc-distribution (within its macrostate), and therefore no way to see the mc-distribution as singled out by the empirical frequencies as the "right" distribution to use in a statistical explanation of the fact (at least we hope it is a fact) that the initial state of the actual world is a good one—*i.e.*, one that will lead to a good history. But I take it that the idea is not to offer a statistical explanation for the goodness of the initial state. Rather, the proposal offers a new sufficient condition for finding a feature of the history of the world (in this case, its goodness) as unsurprising, as needing no explanation: namely that most initial states (within M, or PH let's from now on take this qualification as understood) lead to histories with this feature, in some sense of "most" that we find intuitively natural.

I have no quarrel with this idea; indeed, given that the view I am defending also requires us to consider some feature of the world—the distribution of interventions—as requiring no explanation, I am hardly in a position to quarrel with it. On the other hand, I do find its current popularity a little surprising, given that the mathematical claim on which the idea rests—namely that mc*-most initial states are good—is one that not only has not been proved, but is one that, so far as I can tell, is quite unlike anything anyone has ever been able to prove. It seems to me that much of the popularity of the approach rests on two ideas that I believe to be wrong. One is that the claim that mc-most initial states are good is simply a more general form of a conjecture we already have reason to believe. The other is that postulating some feature of the initial state of the universe has a kind of explanatory legitimacy which postulating statistical features of later states of the universe lacks. I will say something about each of these.

Beginning with the idea that to claim mc-most initial states are good is simply to generalize from what we already know or believe about ordinary macroscopic systems to the case where the macroscopic system is the whole world, the fact that the world as a whole is isolated, whereas ordinary macroscopic systems are not, makes it hard to give such an argument without blatantly begging questions against the interventionist. To see this, let us first be clear about what the argument requires. What is needed is a generalization G about ordinary macroscopic systems which when applied to the entire world says or implies that mc-most initial states are good. This suggests that we will want G to say something like the following about ordinary macroscopic systems S: for any initial macrostate M of S, mc-most states compatible with M will, if allowed to evolve in isolation, lead to histories which are the analogue for S of what in the case of the entire world we have been calling a good history. Now since a good history is at minimum one in which use of the mc-distribution is predictive in the usual realm of statistical mechanics, and since so much of this realm—arguably all of it—involves quasi-isolated subsystems of the world, it seems clear what the analogue for S of a good history needs to be: it needs to be a history in which, supposing S evolves into a set of quasi-isolated subsystems, we can successfully use the mc-distribution to predict the behavior of *these*.

Now of course we do expect that if a macroscopic system S evolves into quasi-isolated subsystems X, Y, Z, then we can use the mc-distribution to predict the thermodynamic behavior of X, Y, Z. But this is because we expect to be able to use the mc-distribution predictively for *any* quasi-isolated system. The argument needs however to give us reason to believe G, which is a claim about mc-most states of the original system S. And this is a claim for which we have no evidence at all. When we are interested in predicting the behavior of a given system X in a given macrostate, we *never* go back to the original S from which X arose and try to calculate whether G holds of S—*i.e.*, whether mc-most of the states of S that evolve in isolation so as to produce an X-like system will validate the use of the mc-distribution on X. It would be a task greatly beyond what we are able to do, and also pointless: we will already know that S did not evolve in isolation, and we are already confident that whatever the origins of X, we can use the mc-distribution on it. Someone in the grip of the overall picture might insist that G *must* hold of S, even if we have never checked any particular case—how else could it be the case that, however S evolves into later subsystems, we can use the mc-distribution on these? But interventionism gives us one alternative (and there may be others): it is consistent with everything we know that, for an isolated system, the mc-distribution will correctly predict its future only for a short time [16].

Let us turn now to the support the present view derives from the idea that we are required to derive any statistical claims from initial conditions. This is in fact Albert's objection to Interventionism: the interventionist has no right to end his explanation with a brute-fact claim about the distribution of interventions; such a distribution must be derived from initial conditions. This seems to me to be simply a prejudice, and indeed one of pretty recent vintage. Introducing a distribution on the initial state of the world is indeed *one* way to go about showing that some feature of the world is to be expected, or at least unsurprising; but why think it is the only way? Currently we believe the world had a beginning in time, but can one seriously believe that if we were to abandon this view, it would never be possible to claim that any statistical feature of the world is unsurprising enough to require no separate explanation? (not even the fact that not all the babies born in the next hour will be girls?) It seems to me particularly ironic that this view is often presented as a development of Boltzmann's ideas, given that Boltzmann's original idea was that what could be treated as unsurprising, as an unexplained explainer, was not a fact about the initial state of the universe, but rather one that plays itself out over history, namely that the empirical distribution of systems is approximately the microcanonical distribution, conditionalized on the macrostate. He was indeed wrong about his specific candidate for unexplained explainer, but this is no argument in principle against the guiding idea [17]. Surely, if we resist the thought that the universe was somehow "selected" at the beginning, as if God flipped a coin, there is no a priori reason to insist on tracing back to initial conditions. My suggestion is a statistical claim about individual systems or their environments, supposing it to be a reasonably natural one, gains no further support by being derived from a distribution on the initial state of the universe: a distribution whose only recommendation can be that it too strikes us as natural.

Against this, a possible response is that if the distribution we invoke over individual systems or their environments is time asymmetric—and in the particular case of interventionism it has been thought they need to be so—then there is much to be said for tracing this asymmetry back to the beginning. This is the topic of the next part of the paper; before taking it up, I want to summarize, in a preliminary way, what I take to be attractive about the interventionist picture.

It is that the probabilities it invokes will meet, or at least can be reasonably hoped to meet an appropriate version of the condition I have been calling (C): modulo a bit of idealization (I will say more about this in the last part of the paper) they are supposed to reflect the actual frequencies of interventions. For an interventionist there is no problem of invoking probabilities that frankly don't match the frequencies (Price, in one reading), or probabilities that one has no good reason to think match the frequencies (Albert, applying the mc*-distribution to individual systems), or probabilities that *can't* match the frequencies, since they are invoked only once (also Albert, this time applying the mc-distribution to the whole world at the Big Bang). The interventionist indeed offers no explanation of why the interventions are distributed the way they are. This would be fatal if it turns out that their distribution needs to be in some way special or interesting or unexpected; but if all we require is something pretty humdrum, then I think the position is worthy to stand against positions that invoke the naturalness of a distribution over initial states [18].

4. What about the Arrow of Time?

I have postponed until now a set of issues that many people take to be central to the whole subject—namely those that have to do with the striking time-asymmetry of the phenomena we are trying to explain. It is often held that any adequate explanation for the approach to equilibrium in the forward time direction must itself be in some way time-asymmetric; otherwise we will be in the embarrassing situation of being equally able to "explain" the (non-existent) approach to equilibrium in the past direction. And in fact, interventionists are sometimes at pains to point out exactly where their accounts are time-asymmetric. Thus, Ridderbos and Redhead maintain that their version "cannot be applied in the reverse time direction to argue that equilibrium will be approached into the past; in the 'ordinary' time direction the 'incoming' influences are the influences from the environment on the system, and these are uncorrelated, but in the reversed time direction the 'incoming' influences are the influences the system exerts on its environment, and these will be correlated" [19].

Price cites this passage in the course of an extended attack [20] on Interventionism; he seems to be reading Ridderbos and Redhead as proposing a time asymmetry in interactions which does not itself rest on some asymmetry in background conditions. It is not plausible that they are suggesting anything so radical, but it must be admitted that the exact bearing of their remark is rather unclear. Is their point that although a system is likely to go to equilibrium in time direction d when the interventions (*i.e.*, what we count as interventions when we count d as the future time direction) are uncorrelated, this is no longer the case when we stipulate that the interventions are correlated? But surely it depends on what sorts of correlations we are talking about: if the correlated outside influences are you and me banging in unison on a glass of ice-water, one would still expect the ice to melt in the future time direction. On the face of it, this is true for any correlations in influences that one can describe in terms other than by explicitly saying that they are correlations precisely geared, given the actual state of the system, to lead it towards lower entropy. The asymmetry that Ridderbos and Redhead are pointing to is, then, not that interventions are correlated in the past time-direction and not in the future; it is that they are precisely geared, in the past time-direction, to lead the system in the direction of lower entropy, but not so in the future direction. This of course says no more than that the ice melted "forward".

Any attempt to use an asymmetry of incoming and outgoing influences to explain why systems go to equilibrium in one time direction and not the other will I think run into the kind of difficulty just sketched. There is no reasonable candidate for an asymmetric kind of intervention, and no reasonable notion of probability, such that, looking in the time-reversed direction, the interventions acting on a given system—given only the initial (*i.e.*, time-reversed final) state of that system—will probably be such as to drive the system towards lower entropy. So I shall not be trying to set up our interventions to be time-asymmetric: they may be, and our formulation will allow for this possibility, but we shall not rely on this to do any explanatory work.

What then *are* we going to say about the arrow of time, or anyway the arrow of processes in time? One familiar approach is to take the relevant probabilities to be symmetric in time—systems tend to be in states that will move to equilibrium when run forwards or backwards—and to locate the source of the asymmetry of the universe in the low-entropy initial condition. There may well be a version of Interventionism that follows this line, deriving the relevant time-symmetric probabilities from (time-symmetric) interventions, and then invoking a low-entropy initial condition. This paper would be shorter, and probably less controversial too, if its author had been able to subscribe to such a point of view. However, since I find it completely obscure what work the initial low entropy of the world is supposed to be doing here, I propose to argue for a form of Interventionism that resolutely refuses to explain the arrow of time.

To see the bare bones of the view I want to propose, it will be helpful to look at an extremely simple model, one in which a given system has no dynamics of its own, but is driven from state to state by probabilistically distributed interventions. To make things definite, let us suppose that the interventions consist in repeated equiprobable selections of a number from 1 through 6—think of these as done by rolling a die. The system whose motion is determined by these rolls will be one with a countable infinity of possible states. For each state p of the system, there will be 6 nearest neighbor states, and these will be at 6 designated "positions" with respect to p—the 1-position, the 2-position, *etc*. In this model, time is discretely ordered and infinite; it is convenient to suppose it is labelled by the numbers ....−1, −(1/2), 0, 1/2, 1. The jumps of the system from state to state will take place at integral times; the rolls of the die will take place at the half-seconds in between, and they will relate to the jumps in the obvious way: if the system is at state p at (t=) n, and the roll of the die at n + (1/2) gives say a 4, then at n + 1 the system is in the (unique) state q which is at p's 4-position.

I want there to be something analogous to entropy levels in our system. Let us suppose the points in the phase space come labeled with a level-number, a positive or negative integer. And let us suppose that for each state q, five of its six neighbors are at a level one higher than that of q, while one neighbor is at a level one lower. I want the setup to have as little interesting structure as possible consistent with all this; the easiest way to achieve this is to have countably many points at each level, of which countably many have a neighbor one level down at their 1-position, countably many at the 2 position, *etc*. A consequence of there being countably many points at each level is that there is no distinguished or salient normalized probability over the phase points and no distinguished or salient sense in which one entropy level is more probable than another. The only probabilities in the story are those that have to do with the rolling of the die; I want to think of this as a process which occurs frequently, in contexts not involving our system, and whose statistics conform extremely well to that of independent equiprobable Bernoulli trials.

We cannot speak either of phase points or trajectories as probable, but we *can* speak of the probability of a trajectory, say one of length n, given a specification of an initial point—this probability will be $(1/6)^{n-1}$. And so it makes sense to assert, and presumably we can prove, that for any p, a trajectory that has p as an initial point will probably have mostly increasing entropy in the "future" time direction—the direction corresponding to increasing values of t.

Suppose there is only one run, and that in this run the entropy fairly steadily rises (in the future time direction) from the initial point p until the process ends (the system is shut down) at a point q. Can we explain this "increase" by pointing out that, given p, the probabilities were in favor of this happening? I expect everyone to agree that we can. Many explanations aim at showing why the explanandum was to be expected, or at least is not surprising; when the explanandum is itself a series of events, such explanations often begin with an initial condition or event which itself is left unexplained, and then proceed by showing that the successive events were each probable, given the preceding ones, according to one or another objective notion of probability. This is the sort of explanation we have here.

I now want to add one additional feature to the model—better, to reveal one that has been there all along—namely a kind of symmetry in the neighbor relation. It will now be the case that, for all p and q, if q is p's neighbor at say position 4, then p is likewise q's neighbor at position 4; similarly for all positions. It should be clear that we can consistently add this feature to our model. It has the consequence that, to the extent we see the roll of the die as "causing" the jumps of the system, we can see it doing so in either time direction: if the system is in state p at t = 3 and in state q at t = 4 and the die shows 6 at t = 3½ then q is p's neighbor at position 6 and p is q's neighbor at position 6; we can think of the roll of the die as "driving" the system from p to q in one time direction, or from q to p in the other. Even better, we can describe the "law of motion" of the system from a point of view that recognizes neither direction of time as privileged: taking a history as an assignment of rolls of the die and of states of the system to some stretch of times, a history is "allowed" only when, if the die shows n at t + (1/2), the states occupied at t and t + 1 are each in the n-position with the respect to the other. Notice that now that neither direction of time is privileged in describing the law of motion of the system, there is no point at which any feature of the model needs a privileged direction of time; despite the time-directed suggestions of my talk about rolling a die, the "rolls" are just a kind of event that takes place on the half-seconds, and exhibits a certain kind of distribution.

Go back now to our single run, rising in entropy in a particular time direction—the direction we have conventionally labelled with increasing values of t—from an initial point p (perhaps better called a "boundary" point). Suppose the boundary point on the other end is q. Then once again, it seems to me that we can explain why the trajectory increased in entropy in the p to q "time" direction exactly as we did before, and that the explanation works in just the same way. Although our previous talk of p as an "initial" point may well have suggested that we were thinking of it as temporally earliest, the explanation required only that p be initial in the sense that it is the point on the trajectory that we are not trying to account for, that we are taking as, for the purposes of our explanation, just given. Given p, the rest of the explanation consisted in showing that the rise in entropy away from p was probable, and of course we can still show this.

What we cannot do, of course, is to take q as our "initial" point, and explain the decrease in entropy in the q to p time direction by showing this was probable, since in fact it is, given q, improbable. Earlier, before we described the law of motion time-symmetrically, my last sentence would have made no sense, and it may be felt that it is precisely this difference that makes our explanation of the p-to-q entropy increase problematic now in a way in which its earlier counterpart was not. But just why does our explanation now become problematic? It is not because we are now unable to explain something we could explain before, namely why it is that, taking q as given, entropy decreased from it in the p-direction: no explanation of this type was available in the earlier version, since there we had no probabilities at all governing the q to p direction. It seems to me that, in the second version of our story, our explanation has exactly the virtues it had before, and should be equally acceptable, provided we continue to think there is no better description available of the probabilities governing the die. This last proviso is not of course trivial. We have given unconditional probabilities for the die, and we may suppose the frequencies bear these out as correct; still, one might also wonder whether there are conditional probabilities that are relevant here. It might be that the probability of the die rolling, say 3, conditional on the state of the system immediately "before" the roll is well defined, likewise the conditional probability on the state of the system immediately "after", and one or both of these might differ from the unconditional probability in such a way as to make it likely that entropy is likely to increase in the "forward" direction and decrease in the "backward" direction, given any given initial state of the system [21]. I want to suppose, however, that we have reason to reject such options. Perhaps we have reason to reject any time-asymmetry in the probabilities governing the die; perhaps we think of the rolls of die as causally independent of the state of the system, and we think this feature is best expressed by treating the probability of the die rolling n as conditionally independent of the state of the system. In that case, my suggestion is that we have met the *only* challenge raised by the improbability of the decrease from q to p: if we continue to think our original description of the probabilities governing the die was correct, then our explanation of the increase from p to q also survives unchanged.

I want to claim that the interventionist can tell a similar story about statistical mechanics. Of course there are differences in the two cases. The systems treated by statistical mechanics evolve both by interventions and freely, and interventions come perhaps continuously, perhaps at variable time intervals (perhaps both). A more serious difference is that although it may be realistic to think of the interventions at work on a given system as modelled by a probability distribution, it does not seem realistic to suppose that one and the same such distribution will fit the actual pattern of interventions for every system at all times. I will later be suggesting that we would do better to suppose that each system S is during different stretches of time exposed to interventions whose pattern is best modelled by some distribution in a broad family. For now, let us imagine a simple—oversimple—version of Interventionism, in which one and the same distribution D of interventions applies to all systems at all times. We might or might not be thinking of the interventions as time-oriented. An example of taking the interventions as time-oriented is to take D as assigning probabilities to the "incoming influences"; given a spatial region R at time t, in which there may be (but need not be) an actual system to be intervened on, D will assign probabilities to possible distributions of particles and forces outside R that are so located, and moving in such a way, as to reach R within say the next second after t (in some once-and-for-all chosen rest-frame). One non-time-oriented alternative might be to take D to assign probabilities to everything that is going on in a fixed (spatial) region surrounding R during a fixed period of time before and after t. The second alternative is closer to our example of the die, since here it is one and the same distribution that models the "influences" from both temporal points of view. But the first alternative is not essentially different, and perhaps a little easier to think about. Here, the most direct way to avoid postulating an unexplained time-asymmetry in the interventions is to claim that "incoming influences from the future"—let's call these IIF's'—are governed by a distribution D* which is just the time-reversal of D [22]. I will assume the dynamics works out so that we can show that, given any state of our system at t, it will be probable according to D (or D*) that the system evolves towards increased entropy in the forward (backward) time direction. And here, as with the example of the die, I want to claim that looking only at future-facing probabilities alone—that is, at D—we are able to explain the entropy-increasing histories of individual systems, and thereby of the entire world: we take as given the initial state of the world, and at each t then and thereafter we point to the fact that the individual systems that comprise the world at t interact via interventions on each other; on each system, the actual pattern of interventions is well represented by D, so it is in this sense likely that the system, given its state at t, will follow an entropy-increasing path. If the relevant probabilities are as large as we expect, this shows that it was likely, given the state of all the systems that make up the world at t, that the world followed a trajectory at t in which virtually all the individual systems increased entropy; repeating this explanation at fixed small time intervals (or pointing out that it could be repeated at fixed intervals, however small the interval) explains the entire trajectory.

What about the fact that we cannot hope to explain the same trajectory by beginning with the current state of the world, and using D* to show that its state a few moments earlier was probable? As with our die-example, I want to claim that the only doubt that this casts on the explanation we *can* give is to the extent it suggests that we might be able to find a better description of the probabilities of interventions, one that allowed us to see the history of the world as probable in both time directions. I want to claim flatly that, so long as we confine ourselves to candidate distributions defined over the space of distributions—as opposed to the joint phase space of systems and interventions—there simply will be no such description: there is no distribution of interventions that makes it likely that, whatever kind of system might be exposed to them, and whatever the state of that system initially, that system will be driven (in either time direction) towards a state of lower entropy. The interventions required to send a glass of water in state p towards lower entropy are necessarily quite delicately geared to p; and there is no reason to think that they will be as well geared to some other q, let alone all other q [23].

Suppose now that things work out as the interventionist hopes: a pretty natural distribution of interventions, or family of such, matches the empirical frequencies and predicts increasing entropy for any system, in any state, in either time direction, and no remotely natural competitor predicts increasing entropy only in one time direction. Then I have been arguing that we have met the one *clear* challenge posed to his account by its failure to work backwards: namely, the suspicion that he has misdescribed the probabilities of the interventions. If one cannot fault the account on the grounds of misdescribing the probabilities, then the whole explanation—which is, after all, an absolutely standard, garden-variety example of Hempelian statistical explanation—is, so far as I can see, beyond criticism.

Still, one might feel that something has been left out. I can imagine someone saying, "Don't we all know that the original Boltzmann program failed because of its inability to explain the decrease of entropy in the past direction, and how is the present account any different?" I hope it has been clear that I disagree with the presupposition: as I see it, the reason the Boltzmann program fails is that it posits a probability distribution over the states of quasi-isolated systems which turns out grossly to depart from the actual frequencies—as it happens, the mismatch is most evident in the probabilities the distribution assigns to regions that have to do with past-directed evolution, but that is not, as I see it, an essential part of the story. The interventionist has no such problem. His probability distributions are not over the phase space of quasi-isolated systems, but over incoming influences from the past and/or future, and the facts about the behavior of systems in interaction with these have no bearing at all on whether these distributions accurately represent the actual frequencies: indeed, "the set of IFFs coming into spatial region A at t which are so arranged as to drive the system which is actually in A at t from its actual state towards a lower-entropy past" is not even a well-defined subset of the domain over which the interventionist's distributions are defined. There is of course a joint domain of systems and IFFs, but our account has assigned no probabilities to regions in *this* domain, nor is there any particular reason to think one could, or should.

In explaining why entropy increased toward the future from the initial state of the world, the account explains all entailments of this fact; it therefore does explain why entropy decreases towards the past from each later state. I can imagine someone wanting more than this: someone might ask why it is that we are able to give such an explanation if we begin with the initial state, but not if we begin with the final state (pretend there is one; the objection can be given in a more complicated form without this assumption) and look backwards; where setting up the question as contrastive in form is intended to signal an expectation that the answer will point to some illuminating contrast between the two states. Such a question cannot be answered by just specifying the initial microstate (or the microstate at any time) and then deducing the entropy-increasing history from this: such an answer fails to locate a contrast between the initial and final states, let alone an illuminating one. Likewise it is no answer to the question to point out that the initial state is in the region of phase space that is determined to evolve in an entropy-increasing way in the future direction of time toward the present moment, and the final state is not in the corresponding region; this gives a contrast, but not an illuminating one: the two regions of phase space are well-defined, but we can not define them simply and intrinsically, as opposed to defining them in terms of their effect.

So what can answer the question? My view is that to frame a question in contrastive form is implicitly to put forward a theory about the subject matter of the question—namely that it allows an appropriately contrastive answer—and that, like any theory, this one can be wrong. One might *ask* why dogs wag their tails and cats don't, or why mirrors reverse left-right but not up-down, but our inability to come up with the kind of answer sought for is no sign that there is something about cats and dogs or mirrors that we don't know. The analogous point about statistical mechanics is a little hard to accept, because it resembles a kind of pessimism that, if taken seriously, would have strangled the field in its infancy: one might have argued against Boltzmann that since the initial microstate together with dynamics entails the 2nd law, there is nothing here we don't already know "in principle", and no point in looking for another level of explanation. But it does not seem to me irrational to celebrate the explanations that statistical mechanics can give, while rejecting the demand that we give a particular kind of account of why it can give these explanations; I will end this section by elaborating on why I think this is the reasonable attitude to take.

To begin with, I take it that no one is expecting to find a feature that the initial state has, and the final state lacks, which *guarantees* the increase of entropy in the appropriate time direction: we have rejected giving the entire microscopic descriptions of the two states, and although it is not logically ruled out that some reasonably natural but less detailed description of the initial state would be enough to entail entropy increase, no one believes it—that's why we do *statistical* mechanics. What we are looking for are probabilities: presumably, we are after some natural probability P defined over the phase space of the world, and descriptions D_i and D_f of the initial and final states—descriptions on the same level of detail—such that, conditional on the world's being in D_i forward entropy-increase is probable, whereas, conditional on its being in D_f, backward entropy increase is not probable.

The trouble with this idea—always, as a way of answering our contrastive question—can readily be seen by looking at the two obvious candidates, the mc distribution and the mc*-distribution. Taking first the mc- distribution, what is the right level of description of the world to serve as our D_f? If D_f is a description in terms of local thermodynamic parameters, the mc-distribution conditionalized on D_f will surely lead us to expect that entropy will be no lower in the past. If we build in some sort of sub-thermodynamic feature—the usual suggestion is some sort of "implicit order"—then we confront a problem that came up earlier in connection with Ridderbos and Redhead. Mere correlations in a state aren't enough to make it probable that it will evolve in a given time-direction in an entropy-decreasing way, else it would be probable that the maximally correlated heat death state would evolve in the forward time direction to decrease its entropy. Only a very particular kind of implicit order in the final state will make probable entropy-decrease in the backward direction, and what reason do we have to think that we can find any way to specify it short of directly saying that it is the kind of order that arises in the final microstate of an entropy-increasing history? [24].

As for the mc*-distribution, recall once again that this is an evolving distribution: the mc*-distribution at t = 0 is the mc distribution conditionalized over the initial state; the mc*-distribution at later t is the update of this. So it is entirely plausible to say that, as Albert emphasizes, the mc*-distribution predicts that the initial state will probably increase its entropy, and that the final state probably came from a previous state with more entropy: at the very least, the mc*-distribution tells us with certainty that the state at t evolved from a state t seconds earlier which satisfied the PH. I criticized Albert's view earlier on the grounds that it failed to specify an adequate sense in which the mc*-distribution matches the empirical frequencies; I want to stress at this point, however, that if we can imagine putting this objection aside—perhaps by arguing that the requirement that the "right" distribution match the frequencies is too stringent—then Albert's view seems to me a completely acceptable explanation of the increase of entropy. Where it fails is as an answer to our contrastive question. Specifically, let the mc+ distribution be defined at time t = T (where T is the time of the final state) as the mc-distribution conditionalized over the final macrostate, and at earlier times by updating in the backward time direction. Then anyone gripped by the contrastive question will surely want to ask why we cannot use this distribution, related as it is to the final state symmetrically to the way in which the mc*-distribution relates to the initial state, to postdict that the world will be at high entropy at t = 0. My view—and here I expect Albert would agree—is that the question is unanswerable; and it is no business of Interventionism (or, for that matter, of Albert's approach) to try to answer it.

5. Some Miscellaneous Remarks

I have been arguing that an interventionist does not need to explain why the interventions, looked at backward, do not drive the final state backwards towards higher entropy. If this is right, it frees things up enormously as to exactly what kind of account we might propose. I myself am far from being able to set out any precise version of the account; instead, I will end with some miscellaneous, and pretty tentative, remarks about what such an account might look like

To begin with, as I mentioned earlier, it is not believable that the empirical frequencies of interventions are well matched by a single probability distribution, applicable to all systems in all situations. Suppose, rather, that there is a family of natural distributions which meet our requirements: namely that, for any system, and any point p in the phase space of that system, interventions so distributed will probably take the system, if it begins at p, to equilibrium. Then it will be enough to claim that any actual system, at any time, is bombarded by a pattern of perturbations which, at least for a while, fits well with some member of this family. (Of course, we will also be making a similar claim about the IIF's). Speaking here of "natural" distributions is reminiscent of part of the standard justification in the Boltzmann tradition for using the mc distribution. But, although I find much to object to in this tradition, I have no quarrel with this particular part of it: my complaint was that, although natural enough, the mc-distribution did not fit the frequencies. Our distributions are required to fit the frequencies.

This is evidently an account that does not encourage talk of the "true" or "right" distribution: it seems to fit better with the thought—although it doesn't absolutely require it—that it might be a matter of choice which distribution we take to represent any particular situation. To the extent this is so, one might begin to wonder whether the probability distributions are essential to the account. It is essential that systems go to equilibrium because the interventions are empirically distributed in the way they are, but surely there are all sorts of ways to speak of empirical distributions without imposing the full structure of probability distributions: e.g., an empirical distribution can be described in terms of sample means or higher sample moments. Seen this way, our talk of probability distributions might seem to be an unnecessary theoretical layer interposed between the real explanans (the empirically given frequencies) and our explanandum. Perhaps so, but I see no way to avoid this extra layer. The reason is that what we say about the interventions will not entail that systems go to equilibrium; all it can do is show this to be probable, but for this we need probabilities.

What will the distributions be distributions over? If the interventions consist in the bombarding of the ice-water by particles, we can take the distributions to be over the number, type, and velocity of the particles crossing a spatial boundary surrounding the glass: of course our distributions should apply equally well to any closed surface in space, whether or not it happens to contain a glass of ice-water. If we want to distinguish between "incoming" interventions and IIF's, we will want two families of distributions, time reversals of each other, one for the particles whose velocity is headed inwards, one for those whose velocity is headed outwards. But, at least in this case, the simplest approach seems to be the time-symmetric one involving a distribution over all velocities. It might be wondered whether either approach does sufficient justice to the fact that IIF's are correlated in ways in which the forward directed interventions are not. But I see no reason why our favored family of distributions should specifically exclude highly correlated distributions of interventions, either forward or backward. Even a highly correlated pattern of interventions will presumably drive a glass of ice-water to equilibrium—to do otherwise, the distribution needs to be correlated with the goings-on within the ice-water, which is quite a different matter. Such correlated distributions will seldom be a best match to the incoming interventions, but it might nonetheless turn out that the most natural characterization of a family of distributions doesn't exclude them [25]. The interventionist will hope to use the time-asymmetric facts of statistical mechanics to explain why the IIF's are correlated; but this comes much later in the order of explanation—his account of the foundations of statistical mechanics does not presuppose this.

Of course bombardment by particles is not the only kind of intervention. We all believe (I wonder if we have justification for this?) that the ice-water would still go to equilibrium if encased in lead; it seems that a fully general account needs to speak about interventions against which there is no shielding, and this means gravity. One way in which including gravity makes it more difficult to say just what our distributions should be distributions of is that, as I will explain in the next paragraph, there is some reason to want the interventions tied to the motions of matter. The only approximation to the general relativistic truth that I know of where something like this is the case is in the linearized theory, where we might take the distribution to be over the (matter part of the) stress-energy tensor in the space (in some convenient frame) near the glass—or that part of space close enough to make a difference. I do not know however, whether, for a glass of ice-water deep within the gravitational field of the earth, the influence of the butterfly on Mars can be treated by this approximation.

My reason for wanting the gravitational influences to be tied to the motions of matter has to do with a question that came up earlier in connection with coin-tossing: why it is that the interventions "harmonize" in the way they do with the mc- distribution on our systems' phase spaces. There are measures on the phase space of our ice-water that make the abnormal points occupy the largest volume of phase space; a distribution of interventions that bore the same relation to such a measure that the actual distribution does to the mc measure would not at all serve our purposes. The "fit" between the actual distribution of interventions and the mc-measure is thus not trivial; what explains it? The analogy of coin-tossing is suggestive. In that case the analog to the interventions are the environmental factors that set the initial state of the coin—how I move my hand when I toss it—and here the question is why the distribution of these factors leads to a distribution over initial states of the coin which varies slowly in terms of the measure we have in mind when we describe the initial states of the coin in terms of rapidly alternating thin bands of heads and tails—namely Lebesgue measure over the position and velocity space of the coin. A crucial part of the answer to this question is that the motion of my hand—say, at the moment the coin is released—is also in a position × velocity space. Of course this isn't the whole story: the scale on which the motions of my hand vary has to be of the right size to produce a close-enough-to-uniform distribution of the initial states of the coin. But the fact that the motions of my hand can be parameterized by positions and velocities is a first important step that sets up the possibility that a reasonably smooth distribution over the one may translate to a reasonably smooth distribution over the other. It is for the same reason that I think it would be desirable if we could see the interventions on the ice-water as consisting in, or at least closely tied to, the motions of matter: smooth natural distributions over the latter may translate to smooth (enough) distributions over the phase points of the ice water. In the case where we think of the interventions as actual kicks from inrushing particles, this seems reasonable enough. My previous remark about using the stress-energy tensor of matter was an attempt to stay as close to this picture as possible in the case where it is gravitation that does the intervening, the idea being that natural distributions of the stress-energy tensor may produce distributions of the gravitational interventions whose effect on the system is likewise smoothly distributed.

A few final remarks. First, a lacuna in both the GRW and the Interventionist accounts. Both just take it as given that the "normal" states of single systems are appropriately distributed: this may be right, but we won't know it or understand why it is so until we have a better picture of what is going on that shows us why it is so easy to knock a system into a normal state. Thus, it would increase our understanding if we could show that the effect of the interventions is to knock the system into a state in which the Stosszahlansatz holds. Against this, however, it seems clear that there are states in which not only does the Ansatz fail, but in which it fails for a reasonably wide neighborhood of the state—wide enough that it is hard to imagine a single intervention fixing things. If all the fast molecules are aimed only at each other, contradicting the Ansatz, it is hard to see how a single light tap from outside could do much to alter this. A larger objection to this idea (which also applies I think to the idea that what the interventions do is to destroy correlations) is that it is not clear that a point satisfying the Stosszahlansatz is necessarily a normal point: no doubt it probably is so, by some natural measure, but we are trying to invoke no other probabilities than that of the interventions. So I think that Interventionism, at least in the form I am proposing, does not bypass the kind of dynamical arguments of which Lanford's theorem is the best-known example; indeed, in requiring a specific structure of phase space, it requires more than Lanford gives. (On the other hand, if interventions come frequently enough, the problem about sustaining the march to equilibrium may not be as severe as it is in Lanford's work.)

Second, a topic I have not discussed is whether there are in fact enough interventions around to explain the phenomena, e.g., for isolated stars. Of course even these are bombarded by photons; perhaps we need not insist that our account imply that they would probably go to equilibrium if encased in lead. In one place, facing the worry that isolated diffuse systems might never experience collapses, Albert suggests that it is enough to suppose such systems experienced collapses at some point in their past when they were not so isolated or diffuse; taking a similar line here would mean that we were supposing—contrary to the suggestion at the end of the last paragraph—a profusion of what you might call "supernormal" points—phase points whose future free evolution guarantees the approach to equilibrium over the long term. Another idea, one which would need an extended discussion, is that of regarding any system as composed of several parts intervening on each other. I do not think this is obviously wrong-headed, but on the face of it, it does present some imposing difficulties. Such an approach may not be appropriate to call interventionist, but it is very much in the spirit of the general approach pursued here, which has been to ground our explanations in distributions that are actually realized, in combination with features of the phase space of our systems which are most easily described in terms of the mc distribution.

I have not discussed a familiar kind of objection to theories that, like this one, do not show it probable that this piece of ice was previously less melted, namely that such theories cannot justify our beliefs about the past. One cannot discuss everything; I would refer the interested reader to some good discussion by John Earman, in [4], and also to a previous paper of mine [26].

References and Notes

1. Albert, D.Z. *Time and Chance*; Harvard University Press: Cambridge, MA, USA, 2000.
2. See [1], p. 132. The text reads "macroscopic two-body systems", reflecting Albert's having chosen to discuss the approach to equilibrium in terms of the equalization of temperature between two bodies in thermal contact. His point applies to systems in general.
3. Ridderbos, T.M.; Redhead, M. The spin-echo experiments and the 2nd law of thermodynamics. *Found. Phys.* **1998**, *28*, 1237–1270.
4. Earman, J. The "Past Hypothesis": Not even false. *Stud. Hist. Philos. Sci.* **2006**, *37*, 399–430.
5. Since Albert is about to appear in connection with a quite different picture, I will refer to the other one, when we get to it, as Albert's account, and call this one the GRW account. In fact, Albert's name really belongs on the GRW account, since (at least so far as I know) no one before him had the nice idea of using GRW to explain the approach to equilibrium; the account I will be calling Albert's is in fact one that has been proposed by several people.
6. In what follows, I switch often, and without notice, between talking about measures and talking about distributions. In general, if I refer to a certain measure μ* as X-measure (e.g., mc-measure), then I am always supposing that μ* can be given by a density function f with respect to Lebesgue measure; I will then refer to f as the X-distribution.

7. It is worth noting that at one point Albert himself invokes external perturbations. See [1] p. 158.
8. If we take μ to be the microcanonical distribution, then μ has the important merit of being an extremely natural distribution—though we might recall that it was not Boltzmann's first choice. Important as it is, this fact cannot be the (C) we seek: it is just one more purely logical consequence of H&M&D. Where the naturalness of some proposed μ might play a role is as a reason to think that that distribution is a match to the empirical frequencies—though, as we are about to see, there are difficulties in making this claim about the microcanonical distribution.

   Since we are looking for a (C) that plays the role which, in quantum mechanics, is usually attributed to the claim that the quantum state gives the chances of measurement outcomes, why not take (C) to be precisely the claim that μ gives the chances of our system being in one or another classical state? A view of this kind has been put forward by Barry Loewer in [27] My argument can allow this so long as we take (as Loewer himself does) the claim that μ gives chances as equivalent to some version of the idea that μ is a best match to the empirical statistics.What I do want to resist in this context is any view of chance according to which the "real" chances are not required to conform well to the actual statistics—perhaps because they haven't been given sufficient opportunity to display themselves in our world. I take this to be a genuine possibility according to the conception of chance we find in quantum mechanics, but I expect general agreement that such fundamental chances have no place in classical statistical mechanics (or indeed in the analogous place in quantum statistical mechanics). A quite different idea, discussed in a recent paper by Kevin Davey [28], is in effect to take (C) as an a priori claim about the connection between the mc-distribution and how we ought to reason: namely that we should use the mc-distribution in fixing our degrees of belief. Davey argues against this suggestion. It is worth noticing that his objection, which rests on showing cases in which we have some extra information which would make it irrational to allow our degrees of belief to conform to the mc-distribution, doesn't apply to the Boltzmann programme as I am reading it: if we know that there is a close match between the mc-distribution and the empirical frequencies, the tradition would indeed instruct us to use the mc-distribution to set our degrees of belief, but, as is always the case when we use frequencies to set degrees of belief (*i.e.*, in "direct inference", as it used to be called), this can be overridden by extra information.
9. Here are some other cases where we would reject a proposed explanation for similar reasons. "All particles that are not neutrinos obey rule X" is unacceptable as a fundamental law, not because it is not universal—it can obviously be written as a universally quantified conditional—but because it immediately invites the question "Why not neutrinos?" Similarly, a proposal that the chance of some process is x can be cast into serious doubt by showing that in a specific kind of situation, the predicted statistics are invariably violated, so long as there are a reasonably large number—say 500—of instances of such situations. This is the case even if overall the statistics are an excellent match to the predictions—which of course can happen if there are many more than 500 cases overall.
10. Price, H. Boltzmann's time bomb. *Br. J. Philos. Sci.* **2002**, *53*, 83–119.
11. I have been able to find one passage, but only one, in which Price *might* be interpreted as subscribing to the view that the mc*-distribution matches the statistics, namely in Section 3.2 of [20].

12. Winsberg, E. Can conditioning on the "Past Hypothesis" militate against the reversibility objections? *Philos. Sci.* **2004**, *71*, 489–504.
13. This is what I take to be the core of Winsberg's argument. Winsberg's own statement places a great deal of emphasis on the claim that contact with the outside world effectively "rerandomizes" the state of the glass of water. Such a claim is no part of the argument as I present it.
14. Or can we say that, given that the region of phase space in which the ice-water is fated to melt is enormous, it would be discrepant for the world to occupy a phase point outside this region? But then the empirical distribution is discrepant with the mc* distribution no matter what, since every phase point lies outside an uncountable infinity of large regions.
15. On the phase spaces of individual macroscopic systems—so this idea requires one to reject Winsberg's argument that the mc*-distribution is just the mc-distribution.
16. There is another, vague, argument for the claim which I have not seen in print, but which may be at the back of people's minds. This is the idea that since the mc*-distribution is just a conditionalized mc-distribution, we have reason to suppose that when we update it and conditionalize over the macrostate of the world today, what we get is more or less a product of mc-distributions over the current macrostates of the various systems that make up our world. If this were so, then since mc-most glasses of ice-water will melt at the right rate, it will also be the case that mc*-most initial states that involve a glass of ice-water here and now will be states in which the ice melts at the right rate. But why believe this? The current distribution doesn't factor into a product; indeed it exhibits strong correlations among the components: run backward, a glass of water will merge with all the other water in the city reservoir and so on back in just the way needed to bring about the low-entropy initial macrostate. The marginals of the current distribution on the subsystems may or may not resemble the mc-distribution in the ways we need, but I don't know how one could argue for this without delving more deeply—much more deeply—into the dynamical details
17. Likewise, ergodic approaches to foundations certainly have their problems, but it seems unreasonable to object to the modern, post-Birkhoff version of such approaches on the grounds that they do not derive their central statistical claim (that the phase point on a system can be expected to be on an ergodic trajectory) from initial conditions.
18. Note that in addition there is a sense in which Interventionism invokes less explanatory apparatus than do appeals to distributions over the initial state. The latter connect the favored distribution with the phenomena only via the assumption that the one and only initial state of the world falls into a particular typical subset (of course it cannot fall into all typical subsets); Interventionism requires no such intermediate step.
19. See [3], p. 1267.
20. Price, H. *Burbury's Last Case: The Mystery of the Entropic Arrow*; Callender, C., Ed.; Cambridge University Press: Cambridge, UK, 2002; pp. 19–56.
21. Call a state p an m-type state, for m = 1,…6, if p's lower entropy neighbor is at position m. Then one might make the conditional probability of the die rolling m, given that it was followed by an m-type state, to be 5/6, while allowing the conditional probability of rolling m to be 1/6, given any preceding state.

22. I should perhaps repeat that our assumption that there is a single distribution D governing the interventions is an over-simplification. Given a fixed characterization of the system on which the interventions and IIF's are acting, to propose a single D for the interventions such that the time reversal D* of D governs the IIFs requires substantial assumptions either about symmetry properties of D, or of the distribution of states of the system, or both. I want to avoid both kinds of assumptions (especially the second). It would be even more to ask that our D and D* be appropriately related, whatever system they are acting on. This is one reason why I think we should postulate not a single distribution but a family of such: we achieve neutrality about time-direction by requiring that the family of distributions governing the IIF's consists of the time reversals of those governing the forward interventions. I expect this requirement will be easy to meet if we are willing to accept a rather broad family of appropriate distributions. It might be that the distribution Y of IIF's on a particular system might be quite different from the time reversal of the distribution X of interventions on that systems, and yet each conform quite well to our intuitive idea of "random influences from outside". It might be clearer to read D and D* in the text as X and Y, and drop the idea that one is the time-reversal of the other.
23. What about the idea that if we move to the joint phase space of systems and IIF's, we can set things up so that the conditional probability is high, for a system in state p, that the IIF's will be such as to drive it backwards towards lower entropy? I distinguish this from the undoubted fact that the actual IIF's do in fact generally turn out to be exactly such as to drive the systems on which they intervene backward toward lower entropy. Of course this is true (and interventionism offers an explanation of why it is true); what the present idea requires, however, is that we accept the existence of correlations between systems and IIF's as the brute statistical fact that is going to explain all the rest. I have been urging that the only statistical facts qualified to play this role have to be in some way natural; it is hard however to see how any reasonably natural joint distribution over systems and IIF's could make probable the correlations that are needed. For p a phase point of a system, let A_p be the set of IIF's guaranteed to send a system in state p towards lower entropy, in the reverse time direction, for the next small interval of time. With P as our joint probability, notice that for any fixed p*, P(A_p*|p) must be high when p = p* and fall off rapidly as p moves away from p*: this follows from the fact that A_p and A_{p*} are quite different, even when p is quite close to p*, together with the assumption, needed for the correlations, that most of the weight of P(|p) is concentrated on A_p. Likewise, and for the same reasons, P(|p*) varies, by any natural measure, rapidly over intervention phase space. It is hard to see what kind of natural distribution could meet this requirement.

    In any case, doesn't the idea smack a bit much of preestablished harmony? Here is one way to help convince yourself that it is the wrong way to go. Suppose we bang on a glass of ice-water, using a random-number table to dictate the size and timings of the bangs. Then, taking as given both the final state of the system and all the interventions in the last 5 min other than our bangings, it will turn out that the inputs of the random-number table were *exactly* what was needed to drive the system backward towards a state of less entropy. To explain this amazing (?) fact, are you tempted to think that it would be illuminating to speak of a correlated joint distribution of the glass and the random number table?

24. Even if some version of implicit order gave us what we wanted, there would still be the following difficulty. For any real case of implicit order I can think of, e.g., the correlations between what is going on on your TV and mine during the World Cup, we expect to be able to explain it in terms of dynamical and statistical mechanical principles: more generally, it is reasonable to suppose that we can explain why implicit order arises, in terms of the initial macrostate evolving along a path that is probable according to mc. But if our entire explanation for the fact that entropy decreases away from D_f is that the world started in D_i and followed a P-probable path, which produced implicit order in D_f, which made backwards decrease in entropy from D_f improbable, then the explanation seems redundant: we've already explained why entropy decreased away from D_f before we mentioned implicit order.
25. Thus, if our distributions are of the form p(v, x,y,z), giving the probability that a particle of velocity v will appear at x,y,z in the next second, then they will neither demand nor forbid correlations between the velocities or locations of distinct particles.
26. Leeds, S. Foundations of statistical mechanics: Two approaches. *Philosophy Sci.* **2003,** *70*, 126–144.
27. Loewer, B. Determinism and chance. *Stud. Hist. Philos. Mod. Phys.* **2001**, *32B*, 609–620.
28. Davey, K. The justification of probability measures in statistical mechanics. *Philos. Sci.* **2008**, *75*, 25–48.