ABSTRACT

Title of dissertation: THERMODYNAMICS, REVERSIBILTY AND JAYNES' APPROACH TO STATISTICAL MECHANICS

Daniel N. Parker, Doctor of Philosophy, 2006

Dissertation directed by: Professor Jeffrey Bub
Department of Philosophy

This dissertation contests David Albert's recent arguments that the proposition that the universe began in a particularly low entropy state (the "past hypothesis") is necessary and sufficient to ground the thermodynamic asymmetry against the reversibility objection, which states that the entropy of thermodynamic systems was previously larger than it is now. In turn, it argues that this undermines Albert's suggestion that the past hypothesis can underwrite other temporal asymmetries such as those of records and causation.

This thesis thus concerns the broader philosophical problem of understanding the interrelationships among the various temporal asymmetries that we find in the world, such as those of thermodynamic phenomena, causation, human agency and inference. The position argued for is that the thermodynamic asymmetry is nothing more than an inferential asymmetry, reflecting a distinction between the inferences made towards the past and the future. As such, it cannot be used to *derive* a genuine physical asymmetry. At most, an inferential asymmetry can provide *evidence* for an asymmetry not itself forthcoming from the formalism of statistical mechanics.

The approach offered here utilises an epistemic, information-theoretic interpretation of thermodynamics applied to individual “branch” systems in order to ground irreversible thermodynamic behaviour (Branch systems are thermodynamic systems quasi-isolated from their environments for short periods of time). I argue that such an interpretation solves the reversibility objection by treating thermodynamics as part of a more general theory of statistical inference supported by information theory and developed in the context of thermodynamics by E.T. Jaynes. It is maintained that by using an epistemic interpretation of probability (where the probabilities reflect one’s knowledge about a thermodynamic system rather than a property of the system itself), the reversibility objection can be disarmed by severing the link between the actual history of a thermodynamic system and its statistical mechanical description. Further, novel and independent arguments to ground the veracity of records in the face of the reversibility objection are developed. Additionally, it is argued that the information-theoretic approach offered here provides a clearer picture of the reduction of the thermodynamic entropy to its statistical mechanical basis than other extant proposals.

THERMODYNAMICS, REVERSIBILTY AND JAYNES' APPROACH TO STATISTICAL MECHANICS

by

Daniel N. Parker

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2006

Advisory Committee:

Professor Jeffrey Bub, Chair
Professor Joseph Berkovitz
Professor Mathias Frisch
Professor Norbert Hornstein
Professor Robert Rynasiewicz

PREFACE

The foundations of statistical mechanics pose a unique set of problems in the philosophy of physics. Despite a relatively unproblematic ontological structure of microphysical particles, the dynamical laws that govern their motions and a consensus that the macroscopic features characterised by the laws of thermodynamics are to be accounted for through the statistical analysis of this underlying ontology, there remains considerable dispute as to exactly *how* this fundamental description is to be linked with the macroscopic features of the world. Further, as Hagar (2005) has noted, there exists very little agreement as to what the problems faced by the foundations of statistical mechanics are, and even as to what would constitute solutions to these problems as they are often relative to the approach one endorses.

This dissertation endeavours to explore and defend an objective Bayesian approach to statistical mechanics, first championed by E.T Jaynes (1983). Instead of conceiving of statistical mechanics as a proper physical theory, Jaynes envisions statistical mechanics as being an expression of a more general theory of statistical inference based on the formalism of information theory. Hence, the probabilities appearing in statistical mechanics are to be thought of as epistemic, useful in describing the expected behaviour of thermodynamic systems, rather than conceiving of them as objective, physical features of the world.

Among the most perplexing problems in the foundations of statistical mechanics is the reversibility argument. Given that the fundamental dynamics governing the microconstituents of the universe are time-reversible, if a uniform probability distribution over all the possible ways a given non-equilibrium thermodynamic system might be fundamentally described implies that we should expect the entropy of the system to increase monotonically towards equilibrium in the future, then, symmetrically, one should expect the same to hold in the past. In other words, if we expect a presently unmelted ice cube to melt into a pool of water in the future, the reversibility argument implies that the most likely *past* history of the ice cube is one where it started out as a pool of water and spontaneously formed into the ice cube as a highly unlikely fluctuation. Despite our apparent memories to the contrary, the very same considerations that indicate that the ice cube was once more melted than it now is apply equally well to our memories: it is vastly more likely that our memories themselves arose as spontaneous fluctuations than as reliable indicators of past states of affairs.

Against this backdrop, this thesis defends 3 central claims:

1. The nature of temporal asymmetry cannot be explained by appealing to the formalism of statistical mechanics constrained by a uniform probability distribution over microstates compatible with the initial state of the universe and the dynamical laws of motion.
2. The Jaynesian account provides a conceptually respectable interpretation of statistical mechanics, accounting for the behaviour of thermodynamic systems and supplying a satisfactory account of the reductive relations between statistical mechanics and thermodynamics.
3. By conceiving of statistical mechanics as being fundamentally concerned inference, the sceptical challenge posed by the reversibility objection can be diffused without appealing to or explaining the physical origins of thermodynamic irreversibility.

The first chapter of this thesis examines Albert's (2000) recent arguments to the effect that the low entropy initial state of the universe is sufficient to solve the reversibility objection and account for the distinction between the past and the future. In it, I argue that, given the sceptical challenge posed by the reversibility argument, the fundamental laws of motion and a uniform probability distribution constrained by the initial low entropy state of the universe and the present macrostate of the universe cannot have the explanatory force that Albert claims it to have.

The second chapter introduces the maximum entropy formalism for statistical mechanics, developing the framework for an epistemic approach to statistical mechanics first championed by Jaynes (1983). Here I argue that the Jaynesian approach to statistical mechanics provides a clear and satisfactory description of thermodynamic processes. Further, I claim that this framework links the thermodynamic entropy to its statistical mechanical basis better than other extant reductive accounts of entropy.

Chapter 3 reviews some criticisms of epistemic approaches to statistical mechanics, focusing on criticisms to the effect that epistemic interpretations of probability are in principle incapable of explaining the success of the laws of thermodynamics as well as charges that Jaynesian accounts of statistical mechanics rely on assumptions and results from ergodic theory to which they are not entitled. It is argued that there exists no completely satisfactory interpretation of probability in the context of statistical mechanics and moreover that these criticisms place an explanatory burden on epistemic approaches to statistical mechanics that they deny. In regard to ergodic theory, some problems with the theory, conceived as a foundational programme in statistical mechanics are reviewed and the (ir)relevance of ergodic results to the Jaynesian programme is discussed.

The fourth chapter recounts Reichenbach's (1956) branch systems account of irreversibility, which may be seen as an attempt to ground the direction of time in the behaviour of local thermodynamic systems, in contradistinction to Albert's more global approach. Although problematic as originally presented, I argue that Reichenbach's branch systems proposal, when reconceived as an epistemic tool, serves to constrain the inferences one can make into the local past by limiting one's inferences to those systems about which one has some knowledge of its past (or future).

Chapter 5 investigates the issue of records in the face of the reversibility objection. I argue that the fact that our memories and records are apparently well-correlated with how we should expect the world to appear if they were veridical provides good evidence for the fact that they are. Insofar as statistical mechanics is conceived as a theory of inference, the claim is that in spite of the reversibility argument, the best inferences one can make are those that take our records as veridical. This approach is contrasted with Albert's account of records, which seeks a physical explanation of how it is that records could be veridical given the reversibility objection, and seeks to account for why we have records of the past but not of the future.

In sum, this thesis envisions the *sceptical* concern presented by the reversibility argument as taking precedence over the *physical* problem of accounting for the origin of irreversible thermodynamic behaviour, and looks to diffuse the argument by casting statistical mechanics as being a theory of inference rather than a physical theory. This is not to say that there does not exist any physical origin of irreversibility, only that such an origin is not forthcoming from the formalism of statistical mechanics itself and the fundamental laws of motion.

TABLE OF CONTENTS

List of Tables……viii

List of Figures……ix

Chapter 1: Reversibility and the Big Bang……1
- The Need For Explanation……1
- The Problem: Irreversibility from Reversibility……6
- The Big Bang to the Rescue?……11
- Albert's Solution Scrutinised……15
- Further Problems for Albert……29
- Who Cares What Happened Ten Minutes Ago?……35

Chapter 2: The Jaynesian Approach to Statistical Mechanics……43
- The MEP as a Statistical Mechanical Formalism……44
- Interpretations of Entropy……52
  - The Boltzmann Entropy……53
  - Bridging the Theories……61
  - Gibbs Entropy……71
  - Could Entropy Be a Measure of Ignorance?……75
- Non-Equilibrium Considerations……81

Chapter 3: Criticisms and Problems with Epistemic Approaches……86
- Interpretations of Probability in Statistical Mechanics……86
  - Objective Interpretations of Probability……88
  - Criticisms of Epistemic Interpretations……94
- Why An Epistemic Approach?……104
- Ergodic Theory……109
  - The Claims of Ergodic Theory……109
  - Problems with Ergodic Theory……112
  - Jaynes and Ergodic Theory……116
- The Reversibility Objection Reconsidered……124

Chapter 4: Branch Systems……130
- Reichenbach's Branch Systems Proposal……131
- Refining the Branch System Proposal……136
- An Epistemic Branch Systems Account……142
- Conclusion……152

Chapter 5: Records……155
- Are Records Low Entropy States?……156
- The Scope of Thermodynamics……160
- What Records Are……169
- When Should Records Be Considered Veridical?……175
- Response to Possible Objections……182

Can There Be Records of the Future?..190
Ready Conditions…………………………………………………………………195
A Note on the Direction of Time………………………………………………207

Chapter 6: Conclusion…………………………………………………………...211

References……………………………………………………………………………216

LIST OF TABLES

Table 1: Reichenbach's Matrix of Thermodynamic Systems over Time………………132

LIST OF FIGURES

Figure 1: Albert's Pinballish Device…………………………………………………………..13
Figure 2: Reichenbach's Branch Systems……………………………………………………134
Figure 3: Reichenbach's Revised Branch Systems...…………………………………….....137

Chapter 1: Reversibility and The Big Bang

1.1 The Need for Explanation

What is it that one looks for in an explanation of irreversible processes? At its most basic level, it would appear that such an explanation would demonstrate that a thermodynamic system in a non-equilibrium state, either isolated from the environment or in thermal or diffusive contact with a reservoir, is overwhelmingly likely to evolve into an equilibrium condition towards the future temporal direction and not the past temporal direction, in accord with the microscopic dynamics of the system. Obvious examples of such behaviour are furnished by the observations that

1. Unmelted ice cubes in warm glasses of water melt until the water comes to a uniform temperature, but never 'unmelt', forming an ice cube from such a glass.
2. When two gases at different temperatures are brought into thermal contact with each other, they eventually come to the same equilibrium temperature, but once at that temperature do not re-establish a temperature difference between them.
3. When milk is poured into coffee, the two substances mix but do not spontaneously separate after mixing.

What is needed for such an explanation?

First a description of what it is for a system to be either in an equilibrium or non-equilibrium state is required. While statistical mechanical definitions of these conditions differ (e.g. the macrostate for which there are the most compatible microstates consistent with the macrocondition in Boltzmann's conception or the invariance of the phase averages associated with macroscopic variables across the ensemble in the Gibbsian approach), the observable behaviour to be explained is the phenomenological behaviour of thermodynamic systems. A system is in thermodynamic equilibrium iff some preferred set of thermodynamic observables remain constant in time. Conversely, a system is in thermodynamic non-equilibrium iff it is not in equilibrium.

The above definition should suffice for now.[1] It leaves open the specific state variables that are taken as the primitive thermodynamic observables. These comprise the usual measurable quantities such as temperature, pressure and volume, though certain systems may require additional observables. A list of these observables will be generically referred to (interchangeably) as a macrocondition, macrostate or macrodescription.[2] A related notion is that of the microstate, which is understood to be the precise specification of the intrinsic properties of each component of the system, picking out the system's exact phase point in appropriate phase space.

Second, in order to link these definitions with the microscopic, statistical mechanical description of thermodynamic systems, some additional postulate regarding how the properties of the microscopic description are to be associated with the phenomenological observables of thermodynamic systems is required. Again, while the details of how this link is to be made vary according to the statistical mechanical approach one subscribes to, any proposed explanation of irreversible processes must dynamically connect the thermodynamic approach to equilibrium from non-equilibrium macrostates. Ideally, the explanation would contain the following elements:

---

[1] One might loosen the definition of equilibrium to accommodate small and fleeting changes in the values of observables (van Lith 2001).

[2] Throughout the present discussion, I will eschew talk of entropy, and speak only of equilibrium and non-equilibrium conditions, relying on an intuitive understanding of what it is, say, to be in a highly non-equilibrium state. This is done to avoid importing any particular interpretation of entropy into the discussion, either in describing irreversible processes at the macroscopic scale or any interpretation at the microscopic one.

1. A detailed description of how it is that non-equilibrium states probabilistically evolve towards equilibrium ones on the basis of the dynamics of the theories that govern the behaviour of its microscopic constituents.
2. Why it is that we observe isolated systems evolving from non-equilibrium states towards equilibrium ones in the future temporal direction only, and not towards the past.

For the purposes of this chapter, I will treat the dynamical explanation as being unproblematic. Clearly, this aspect of the explanation and theory of irreversible processes is a thorny one, and the detailed analysis required will depend on several factors. On the one hand, the form of the explanation will depend on what theory one takes to govern the microscopic dynamics of thermodynamic systems, whether it be Newtonian mechanics, quantum mechanics or some other theory. On the other, the way in which the approach to equilibrium is described will differ between schools of thought. Significant differences exist between the Gibbsian approach, which utilises ensembles of systems to describe the dynamical evolution, kinetic theories as well as master equation descriptions.[3] Additionally, the way in which the dynamics are described typically involves statistical or probabilistic assumptions, whose nature and implementation depend on the particular interpretation of probability one favours. What I shall assume for the present purposes is that whatever theory correctly describes the microscopic constituents of thermodynamic systems, the vast majority of microstates that are compatible with any macrocondition will evolve towards an equilibrium state, where the thermodynamic observables do not change with time.

[3] See Sklar (1993) for a general description of these and other approaches.

Since the following discussion is independent of the detailed dynamics of such processes, I assume, for the purposes of this chapter, that we can ignore this problem and focus on the second aspect, namely the temporal asymmetry of thermodynamic processes. For the moment, the details of how one describes the dynamical evolution and the *interpretation* of the probabilistic assumptions (as they are used to describe the likely evolution of thermodynamic systems) are to be taken as orthogonal to the problem of explaining the apparent temporal asymmetry of the thermodynamic approach to equilibrium states in the future but not the past temporal direction, given that the dynamical theory that governs the microscopic constituents of thermodynamic systems is taken to be temporally symmetric.[4] Indeed, the hallmark of most theories that are taken to plausibly describe the microdynamical evolution of physical systems is that they are temporally symmetric. [5]

A dynamical theory is temporally symmetric (or, equivalently, time-reversible) just in case, for any sequence of states allowed by the theory, the time-reversed sequence is allowed as well, so the theory is incapable of picking out a privileged temporal direction.[6] So it seems that the theory governing the microscopic behaviour of thermodynamic systems is incapable of picking out a privileged temporal direction: given a non-equilibrium state at some time t, if the theory demonstrates that a system is overwhelmingly likely to evolve towards equilibrium at some point later than t, then it is

[4] This assumption, however, will be questioned in Section 1.5 in regards to Albert's interpretation of probability.
[5] A notable exception is discussed by Albert (2000, Ch. 7), who considers the GRW theory of quantum mechanics where an explicitly time asymmetric spontaneous collapse of the wavefunction could generate the required thermodynamic time asymmetry. Of course, claiming that the probability distribution over microstates of a system samples less than the full accessible phase space associated with the macrostate can also generate a non-lawlike temporal asymmetry.
[6] In the case of Newtonian mechanics, this amounts to reversing the velocities of each particle in the system.

overwhelmingly likely to have evolved *from* an equilibrium state in the past. As a guiding question, how could it be that time-reversible dynamics at the microscopic scale give rise to temporally asymmetric behaviour at the macroscopic level? How does irreversibility arise from reversibility?

As a concrete example, suppose that there is a glass of water with a half-melted ice cube in it (suitably isolated from its environment). Given this present macrocondition, we can follow its underlying (Newtonian) dynamics to either predict or retrodict its future or past macrocondition, respectively. In each case, these probabilistically described dynamics would indicate that the system spontaneously evolved *to* its present non-equilibrium state *from* an equilibrium one, and that it will return to an equilibrium state in the future. Based solely on the uniform probability distribution over the microstates compatible with this macrocondition and the dynamics that underlie thermodynamic systems, it would appear that any non-equilibrium macrocondition one might come across popped into existence as an enormously improbable fluctuation from a past equilibrium macrocondition, and will return to an equilibrium macrocondition in the future.

Further, it would seem that we often have records of past non-equilibrium conditions: we remember the unmelted ice cube being in the glass ten minutes ago. But can our memories or records of the evolution of the ice cube be taken as veridical? Given that we take our memories and records of past events to be describable in statistical mechanical terms and thus are also governed by time-reversible dynamics, the above concerns apply equally well to our own memories. Just as, on the basis of the probabilistically described dynamics, we could retrodict that the ice cube arose as a spontaneous fluctuation from an equilibrium state, so too can we retrodict that our current memories of the ice cube most likely arose out of spontaneous fluctuations as well. In fact, taking our memories as statistical mechanical systems, it would appear that all our memories arose spontaneously from equilibrium states, and should not be taken as veridical. Here and throughout this work, this problem will be referred to as the *reversibility objection* or *reversibility argument*.

1.2 The Problem: Irreversibility from Reversibility

Something is very wrong here. The above conclusion directly contradicts our experience concerning other ice cubes melting, our memory of the half-melted ice cube once having been fully unmelted, and our intuitions about how glasses of water with ice cubes in them behave. In particular, a half-melted piece of ice in a glass of water naturally leads us to the conclusion that some time in the past, whether or not it was observed to be the case, there was a less melted ice cube in the glass of water and furthermore that it came to be the case that, at present and in line with the underlying dynamics, the ice evolved to the half-melted state that we now observe. Additionally, we expect that the system will continue to evolve towards the equilibrium condition where the ice is fully melted and the glass of water is at an equilibrium temperature. So, the question becomes: how can we reconcile our intuitions, memories and putative experiences of thermodynamic systems with the theories we take to determine the evolution of such systems, given that these theories seem to undermine the veracity of these very experiences, memories and intuitions?

The easy answer is to invoke an additional postulate to the effect that, in the past, the system was in a non-equilibrium state; that is, the ice cube was fully unmelted. By doing so, the intuitive history of the ice cube can be saved. Call the present time, where we have before us a half-melted ice cube, t_0. Let t_{-1} and t_1 be the times in the past and future, respectively (say, ten minutes before and after t_0). Given only the present state of the ice cube at t_0, the dynamics alone would have us believe that the ice cube is fully melted at both t_{-1} and t_1, but we can apparently reconcile this with our memories of there being a fully unmelted ice cube in the glass of water at t_{-1} if we posit that at t_{-1}, there was in fact a fully unmelted ice cube in the glass of water. This will restrict the probability distribution to those states that are compatible with this posit, solving the problem by cropping out those microstates that would have lead to past anti-thermodynamic behaviour. In this case, the overwhelmingly probable evolution of the system would be in line with our memories and our experiences of other ice cubes since its dynamical description would render it likely that the ice cube went from being fully unmelted at t_{-1} to being half-melted at t_0, and will almost certainly evolve towards an equilibrium, melted state at t_1. And so it would seem that we have the explanation we have been looking for. But the imposition of this initial condition at t_{-1} does not come for free, and it is to the justification of this imposed initial condition to which we now turn.

The first question that needs to be addressed is exactly what it is that needs to be justified. Imagine, again, that at present (t_0), we have before us a glass of water with a half-melted piece of ice in it, and that we remember that ten minutes ago (t_{-1}) the ice cube was fully unmelted. Now *this* present situation at t_0 (the ice cube, glass of water and our memory), taken as a composite thermodynamic system, introduces two worries. First, we need to inquire how, and particularly what aspects of, the evolution of the ice/water system need to be explained in order to complete the explanation of the irreversible process we observe. In particular, and assuming our memories or records of ice/water system's macrostate at t_{-1} can be taken to be veridical, do we need a complete account of how it is that this system first came into being, or do we even need to be at all concerned with its past history, prior to t_{-1}? Specifically, is the system's state prior to its being observed relevant to the explanation of irreversible processes (such as ice melting) that we actually observe?[7]

As the second worry, can our memories or records of the evolution of the ice cube be taken as veridical? As noted above, since our memories and records of past events are themselves statistical mechanical systems and governed by time-reversible dynamics, we can ask the same questions about our memories. As we could retrodict that the ice cube arose as a spontaneous fluctuation from an equilibrium state, so can we retrodict that our current memories of the ice cube most likely arose out of a spontaneous fluctuation as well. So, taking our memories as statistical mechanical systems, it would appear that all our memories arose spontaneously from equilibrium states, and should not be taken as veridical.[8]

Let us leave this second worry aside for the moment, and concern ourselves with the first problem. How is it that we can reconcile our records and memories of the ice being fully unmelted at t_{-1} with the time-reversible dynamics that render it overwhelmingly likely that the ice spontaneously formed from an equilibrium system in the past? The reconciliation is achieved by stipulating that the ice was fully unmelted ten minutes ago, contrary to what our retrodictions would render likely. And such posits are clearly sufficient to recover the experimental and experiential content of thermodynamics.

Up to this point, we have taken the laws of thermodynamics, and buttressed them with the claim that, in both temporal directions, the underlying dynamics of such systems

[7] In Chapter 4, I will argue that the answer to this question is 'no', though I will assume an affirmative answer for the purposes of this chapter.
[8] In Chapter 5, I will present considerations intended to deny this conclusion, while still acknowledging the *prima facie* worry posed by the reversibility objection.

drive them towards equilibrium (in a manner which will depend on exactly how one describes the underlying dynamics). The temporal asymmetry gets built into the description by stipulating *that*, at some point in the past, the system found itself in a non-equilibrium situation (say, that someone dropped an unmelted piece of ice in a glass of water). In this way, it seems that we can make sense of irreversible processes. We can imagine that a scientist drops a piece of ice in a glass of water or removes a partition allowing a gas to expand into an empty chamber. From the moment when we know a system to be in a non-equilibrium state, we can follow its evolution towards its future equilibrium state. Thus the proposed explanation consists of two elements:

1. The underlying or reducing dynamics of thermodynamic systems, which render it likely (in the appropriate sense of likely given the particular statistical description of the underlying dynamics) that a non-equilibrium thermodynamic state will evolve towards an equilibrium one in both temporal directions.
2. The imposition of an initial, non-equilibrium condition in the past of the thermodynamic system which, when evolved according to the underlying dynamics proper to such a system, renders it likely that the system will evolve as it is observed to evolve experimentally.

It is commonplace to think we need more, that the positing of an initial non-equilibrium condition in the history of the system of interest is not enough, and reference needs to be made to the thermodynamic condition of the universe itself. To see why, consider again our glass of water with a half-melted piece of ice in it at t_0, and the stipulated condition that the ice cube was fully unmelted at t_{-1}. There appear to be two problems with this explanation of the irreversible melting of the ice cube. First, while the imposition of an initial condition at t_{-1} renders it likely that the ice cube melted in the way it was observed to melt experimentally, it is in conflict with the retrodiction based upon the underlying dynamics. If we take the underlying dynamics and uniform probability distribution seriously, and believe them to accurately describe the overwhelmingly likely future evolution of the system, then, by parity of reasoning, one ought to have equal confidence in their description of the past evolution of the system. The imposed initial condition, by virtue of the retrodictions based on the underlying dynamics, would appear to be an extraordinarily unlikely event. We might ask how one could ever be justified in thinking that any present non-equilibrium state came from a highly non-equilibrium initial condition, rather than through a spontaneous fluctuation from equilibrium. As for the second concern, consider times in the past before t_{-1}. According to the time-reversible underlying dynamics, the same reasoning that led us to believe that the half-melted ice cube arose as a spontaneous fluctuation *from* an equilibrium condition at t_{-1} should lead us to expect that the initial condition that we posit *at* t_{-1} (the one that allowed us to evade this unintuitive consequence of the present macroscopic condition arising as a spontaneous and unlikely fluctuation) itself arose from an enormously unlikely fluctuation *before* the time of the stipulation of an initial condition (say, twenty minutes ago). The upshot of this is that although imposing the initial condition at t_{-1} allows us to recover the expected behaviour of the ice cube coming from a fully unmelted state to the present half-melted state, and on to fully melted state in the future, this comes at the cost of expecting that the fully unmelted ice cube in the glass of water itself arose as a spontaneous fluctuation, one that is presumably *even more* unlikely than the half-melted ice cube arising as a spontaneous fluctuation. And note that postulating an initial condition before t_{-1} (say, twenty minutes ago) won't help either, since things will still go horribly wrong before *that* initial condition.

1.3 The Big Bang to the Rescue?

Both these concerns are alleviated, on many accounts, by postulating the existence of a highly non-equilibrium condition at some point in the early universe. For unless we appeal to such a highly non-equilibrium state at some point in the distant past, at the moment (or a short time after) the big bang, it would render it overwhelmingly likely that the past non-equilibrium states that we recall, or that we posit, were preceded by states evolving in the past temporal direction closer and closer to equilibrium states. And so without the posit of a highly non-equilibrium state shortly after the big bang, it would seem highly improbable that anything that we take ourselves to know about the past, either through memories or records of past events, would be true.

And this stipulation, to the effect that the universe came into being in a highly non-equilibrium state, would in fact be necessary if the thermodynamic system whose apparent irreversible behaviour we were trying to account for was the universe itself. Indeed, it would be improbable that the universe is, and was, in the highly non-equilibrium state (which includes glasses of water with ice cubes in them) that we remember it being in yesterday, a year ago or, as our best theories suggest, a billion or ten billion years ago, unless we make such a posit. In fact, the reversibility argument seems to entail that present non-equilibrium state of the universe most likely arose as an improbable fluctuation, and nothing about the past was as we take it to have been. However, the universe is but one (large) thermodynamic system among many, and it seems intuitively irrelevant to invoke *its* thermodynamic condition in order to understand the behaviour of our glass-of-water-with-ice system.

Albert (2000) is a strong proponent of the view that the initial state of the universe can solve these problems, and recognises these issues. Albert acknowledges that this cosmological posit (in his parlance the 'past hypothesis') cannot directly entail that the ice cube was fully unmelted ten minutes ago or that it did not arise as a spontaneous fluctuation from a prior equilibrium state, but he thinks that the posit, in conjunction with the macrostate of the rest of the present universe, can guarantee the veracity of our memories of such events and furthermore can show that the apparent history of the ice cube is in some sense typical.

To explicate his view, Albert considers a "pinballish device" replicated in Figure 1. At the bottom of the device sit several glasses of warm water, some of which contain half-melted ice cubes. On the basis of the microscopic dynamics of the system in question, we would expect that the ice cubes arose from an equilibrium state in the past. However, if we add the posit that ten minutes ago the ice cubes were fully unmelted and at the top of the pinballish device when they fell, then things will come out right. Here is Albert's claim:

> It is (to begin with) certainly *not* the case that this last posit will make either the present macrocondition or the five-minutes-ago macrocondition overwhelmingly probable: this posit (as a matter of fact) will make *no* particular present or five-minutes-ago macrocondition overwhelmingly probably. What it will do (rather) is to make certain prominent thermodynamic *features* of the present and five-minutes-ago macroconditions overwhelmingly probable (their average temperatures, for example, and the degree to which what ice there is in them is *melted*, and so on), but it will clearly assign similar probabilities to a rather wide variety of quite *distinct* five-minutes-ago macroconditions (macroconditions associated with the ice cubes having landed in quite different sets of glasses, for example). What we *have*, though, in this last posit, and what we were *lacking* in the previous one, is a probability-distribution relative to which what we remember of the entirety of the last ten minutes, and what we know of the present, and what we expect of the future, is (you might say) *typical*. (Albert 2000, 84)

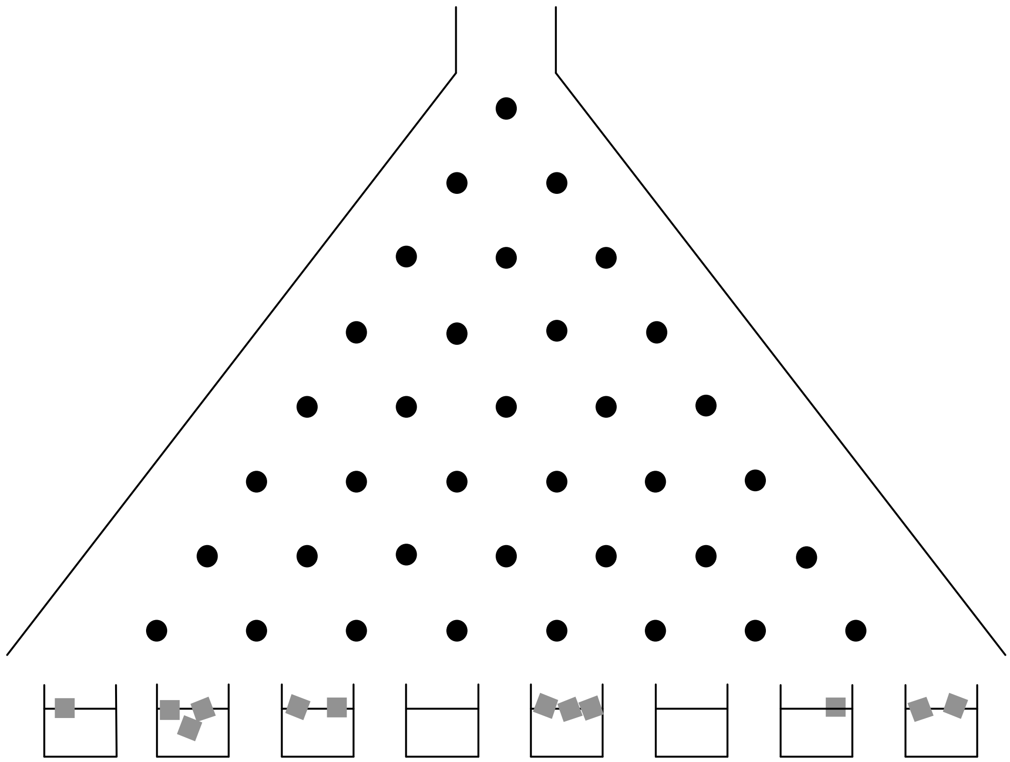

Figure 1: Albert's pinballish device. At present, there are partially melted ice cubes in some of the glasses.

Nothing here seems particularly objectionable, save for the vagueness of Albert's notion of typicality. Let us specify exactly what Albert is claiming. As he notes, based on the posit to the effect that ten minutes ago the ice cubes were all at the top of the device, we might expect to find a variety of different configurations of ice cubes now (that is, some glasses of water might be empty, some might contain more than one ice cube, etc.), only one of which is actually realised. But what we do find is that some gross thermodynamic features will be consistently attributable to the system based on the initial posit, features that will remain over many runs of the experiment.

Of course, we cannot expect that the temperatures *or* the degree to which the ice cubes are melted be the same in every run of the experiment. If many ice cubes, say, wind up in a single glass, those ice cubes will melt more slowly than cubes that are alone, each in a single glass. The temperature of the glass with numerous ice cubes in it will be lower than glasses without ice cubes or with a single ice cube. After a number of runs of this experiment, we would find that these thermodynamic features are not constant relative to the initial posit, but are described by a probability distribution describing the present values of these features. Albert's notion of typicality is thus summed up as

> A probability-distribution relative to which a certain highly restricted set of sequences of macrostates – a set which happens to include what we remember of the entirety of the last ten minutes, and what we know of the present, and what we expect of the future – is overwhelmingly more probable than any *other* [set of] such sequence[s]. (2000, 84)

So far, so good. In fact, based on the present macrostate of all the *other* glasses, one can infer whether or not any unmelted ice cubes were present ten minutes ago in any particular *remaining* glass. So it would appear that the initial posit, the macrodescription of the rest of the pinballish system, and the standard measure over the phase space allow one to trace the thermodynamic history of the ice cube in the glass. But Albert quickly makes the move to cosmological considerations. He immediately notes that the story he has just told gets everything right (regarding the typicality of the present ice cube situation), but that at times before the initial posit, things will still go horribly wrong, since it would seem that *that* situation must have arisen as a spontaneous fluctuation from equilibrium. And so one must accept the granddaddy of initial posits: the 'past hypothesis' to the effect that the universe began in highly non-equilibrium state, or at least the portion of the universe to which we have epistemic access (Albert 2000, 85).

What the past hypothesis gives us then, in accordance with the description of typicality above, is a probability distribution over some vague macrocondition, characterised by a non-equilibrium macrostate somewhere in the distant past, which is supposed to restrict the sequence of macroconditions from the distant past until now to those which make it overwhelmingly probable that the universe evolved pretty much in the same way that we take it to have evolved. Furthermore, given the present state of the universe, one should be able to determine (more or less) the thermodynamic evolution of any system of interest. The past hypothesis allows us to trust our memories, any records of the past we may have, and to validate the posit to the effect that the ice cube, half-melted and sitting in a glass of water, was fully unmelted ten minutes ago and didn't arise as a spontaneous fluctuation from an equilibrium state. Or so Albert claims.

1.4 Albert's Solution Scrutinised

This inference, if sound, would be miraculous. The informality of the argument aside, it appears downright implausible that the mere stipulation of a non-equilibrium state of the universe somewhere in the distant past could justify my memory of an unmelted ice cube ten minutes ago, somehow make it altogether improbable that the ice cube formed as a spontaneous fluctuation ten minutes ago, and testify to the veracity of any records to that effect. Extrapolating Albert's notion of typicality to the case of the universe, let us see how probable the history of our ice cube seems.

Let us say that the universe began in some well-defined macrostate, over which we assign a uniform probability distribution according to the standard measure.[9] Suppose that we can then trace the history of each universe compatible with this initial macrostate. How many of these possible universal histories will lead to an ice cube in a glass of water, fully unmelted ten minutes ago, along with a memory of the ice cube being fully unmelted ten minutes ago, and in accord with the apparent evolution of the ice cube up to

[9] Of course, this is already giving Albert too much. All we are justified in positing is that the universe (or the part to which we have epistemic access) began in some arbitrary highly non-equilibrium condition.

the present? In what sense is this sequence of events *typical*? It is certainly not typical in that we could predict, on the basis of the past hypothesis, that there would be a glass of water with an ice cube before us at present, or even that such an event would be likely relative to the probability of there not being an ice cube before us at present. Even more, it is not typical of all the possible evolutionary histories of the universe that I exist, since most histories would dictate that I was never born. In fact, one might be inclined to say that the present situation is not typical at all.

However, what Albert is claiming *is not* that the exact present situation (or the ice cube having been unmelted ten minutes ago) is overwhelmingly likely given the past hypothesis, but that it is much more likely than if the past hypothesis were not true. If the universe didn't begin in a particularly non-equilibrium state, then it would appear extraordinarily unlikely that I could exist at all, that there ever could be a glass of water with an ice cube before me. On this point, I am inclined to agree. But, as I shall argue, this is a long way from establishing that the records and the memories I have of the history of the ice cube are overwhelmingly likely to be veridical, and is a long way from establishing the general validity of the second law of thermodynamics.

Extrapolating Albert's notion of typicality to the case of the universe, let us see how probable the history of our ice cube seems. Consider as the event space *all* the microstates on the energy hypersurface of the universe, use the standard statistical measure (Albert's statistical postulate), and the following propositions:

> B (for big bang): the portion of this event space that contains all possible microstates presently compatible with the initial macrocondition of the universe (i.e. the past hypothesis).
>
> U (for unmelted): the portion of the event space compatible with an unmelted ice cube in the glass of water ten minutes ago.

H (for half-melted): the portion of the event space compatible with a half-melted ice cube in the glass of water.

M (for macro-knowledge): the portion of the event space compatible with the macrostate of the *rest* of the universe; that is, everything not including the ice cube in the glass of water.[10]

We then look to establish that

$$P(U|H\&B\&M) > 1/C \tag{1.1}$$

where *C* is a positive constant. In words, we look to show that conditionalising on the present macrocondition of the universe *and* those present states compatible with the past hypothesis is more likely than some threshold probability such that we can justifiably infer that the ice cube was indeed less melted in the past. This relation amounts to a necessary condition on Albert's proposed explanation.[11] Indeed, if the past hypothesis fails to establish the inequality above, then its explanatory value is of little or no worth.

Substituting the definition of conditional probability, we find that (1.1) can be expressed as

$$P(U\&H\&B\&M)/P(H\&B\&M) > 1/C.$$

We can simplify the above equation by noting that almost all unmelted ice cubes ten minutes ago evolve to presently half-melted ice cubes by dropping the H term from the expression that conjoins it with U (in any case this will not alter the inequality):

$$P(U\&B\&M)/P(H\&B\&M) > 1/C. \tag{1.2}$$

Resubstituting the definition of conditional probability, we can rewrite the equation as

[10] Actually, Albert claims we only need to take into account any macroscopic knowledge about the universe that we *happen* to have (2000, 96). I consider the stronger claim where the entire present macrostate of the universe is given.

[11] One might think that a natural value for *C* would be 2, indicating that it is more likely than not that the ice cube was indeed unmelted five minutes ago. However, I restrict myself to the weaker claim that C should be *at least* 2, since Albert's typicality condition requires that "a certain highly restricted set of sequences of macrostates … is overwhelmingly more probable than any *other* such sequence". For the present purposes, it suffices to let *C* be greater than 2, though it should not be too large.

$$P(B|U\&M)/P(B|H\&M) > 1/C*P(H\&M)/P(U\&M)$$

This can be simplified by noting that the terms M that appear on the right side of the equation do no work and can be dropped. This is because the macrostate of the present universe is exhaustively described by the conjunction H&M, and any *apparent* correlations between the macrostate of the *rest* of the universe and the state of the ice cube are, by the reversibility argument, almost certainly the result of a spontaneous fluctuation and not in any way correlated with the past, unmelted, state of the ice cube. Thus, we can rewrite the above as

$$P(B|U\&M)/P(B|H\&M) > 1/C*P(H)/ P(U)$$

The right side of the equation now places a strong lower bound on the inequality, since the measure associated with a half-melted ice cube on the event space is presumably orders of magnitude larger than that associated with an unmelted ice cube. Accordingly, we can drop the constant C:

$$P(B|U\&M) >> P(B|H\&M). \tag{1.3}$$

(1.1) is thus equivalent to saying that the initial, non-equilibrium state of the universe, given that there was an unmelted ice cube in the glass of water ten minutes ago along with the present macrodescription of the universe, is much more probable than its likelihood given that there is a half-melted ice cube in the glass now.

So does the past hypothesis solve our trouble? Recall the problem with which we began. It appeared that, no matter how far from equilibrium we find a thermodynamic system, the underlying dynamics dictated that in both temporal directions the system would move towards an equilibrium state. In fact, based on the underlying dynamics and a uniform statistical distribution, nothing in the present situation could ever imply that the system was, or ever will be, further from equilibrium than it is now. More to the point, there is nothing in the present state of affairs that could, in itself, ever provide any grounds for believing that the *universe* was ever further from equilibrium than it is now. Albert clearly recognises this, calling it the fundamental insight of Boltzmann and Gibbs (2000, 93). But if this is the case for our present macrocondition (the ice half-melted in the glass of water), then surely, *mutatis mutandis*, this applies to the fully unmelted ice cube ten minutes ago. *Nothing* in *that* macrocondition could ever count as evidence for the universe having been further from equilibrium than it was ten minutes ago. And so, looking at (1.3), we are forced to conclude that conditionalising on the highly non-equilibrium state of the early universe adds nothing to what we've been looking for: a reason to think that the ice cube was previously less melted than it is now. Albert's explanation seems to be dead from the start.

One might rightly object that even though no particular non-equilibrium state can in itself increase the probability that the universe's entropy was ever lower, the existence of low entropy states like an unmelted ice cube does increase the probability of a low entropy past *relative* to higher entropy states such as a half-melted ice cube. But it's hard to see how that's going to help since the inequality is quite strong in the sense that the left side of (1.3) needs to be orders of magnitude greater than the right side. This worry can be made clearer by considering (1.3) without the conditionalisation on the macrostate of the rest of the universe (that is, $P(B|U) \gg P(B|H)$). Here one might be inclined to think that the big bang state *is* better correlated with the unmelted ice cube than the half-melted cube is, on the order of $P(H)/P(U)$.

At best, this is an unargued for conjecture, and Albert provides no substantial reason to think that this is true. In fact, one can argue that it is clearly false, for the only way that the inequality can hold is if the existence of a presently half-melted ice cube, given the past hypothesis, virtually guarantees that it was unmelted ten minutes ago. But this is false.

Imagine that I walk into an otherwise empty room with an apparently half-melted ice cube sitting in a glass of water. I need not infer, even as a matter of everyday reasoning, that it was unmelted ten minutes ago. There is virtually an infinity of other histories the ice cube could have, both in accord with the second law and those exhibiting anti-thermodynamic pasts: someone could have left the room just moments before I entered, having placed the glass of water with the *half*-melted ice cube (fresh out of the freezer) in the room. The ice cube need not have ever been in a more unmelted state.

So if (1.3) is to hold, the macrocondition of the rest of the universe must do non-trivial work in guaranteeing that the inequality is satisfied, in a manner similar to the way that the macrostates of the *other* ice cubes fixed the thermodynamic history of the *first* ice cube in the pinballish device; that is, they serve as records of the fact that the ice cube was previously unmelted. Returning to the example of the ice cube in an empty room, it is clear that this move fails as well. There is, presumably, nothing in the present macrostate of the universe that can tell me whether or not a half-melted ice cube was dropped in the room just moments before I entered. More generally, there may not be *anything* about the present observable universe that serves as a reliable indicator or record of a system's thermodynamic history either because no lasting records were formed, because any apparent records underdetermine its history, or because any such records have 'washed out', in the sense that they didn't last (came to an equilibrium state) up to the present. Appealing to the rest of the universe doesn't seem to help.

This criticism can be expanded to less exotic scenarios. Consider a case where I have a clear memory of an unmelted ice cube and a half-melted cube before me now.

Does this situation license the inference that the ice cube was unmelted ten minutes ago *because* there are records of this past state of affairs? It does not, since the entire probabilistic derivation could be run anew, this time taking H to be the phase description of the claim that the ice cube is presently half-melted *along with* my memory of the unmelted ice cube, and U to be the description of the unmelted ice cube ten minutes ago *and* the formation of my memory. But now we are in the same situation as before. Is there anything about the present state of the rest of the universe that, along with the past hypothesis, fixes the history of the ice cube/memory system by serving as a record of this new state described by U?

Even if there is, one can easily incorporate *that* into our description of the system of interest and ask if there is anything about the *rest* of the universe that fixes *that* composite system's history, and so on. If and only if this regress can continue all the way to the point where there is no more "rest of the universe" to consider is Albert's proposal successful in vitiating the reversibility objection.[12]

But the regress cannot continue in this way, because many facts about the past state of the universe are forever lost to us. Indeed, the regress halts when there is nothing about the present macrostate of the rest universe that can fix the history of the system under consideration. The present macroscopic description of the universe underdetermines its history, and there may be nothing (or at any rate, nothing left of any past macroconditions) that would serve to ground the history of the system of interest.

Perhaps it is instructive to return to Albert's "pinballish" device and identify in what ways this model differs from the full scale past hypothesis that Albert offers as analogous to this case, since in that instance it appeared that Albert's proposal had a

[12] Note that this is exactly the case when considering the pinballish device.

certain plausibility. Specifically, we can identify three elements that differ between the case of the whole universe and the pinballish device:

1. The difference in the times indexed by the past hypothesis and the present are much greater than the time required for an ice cube to melt. Conversely, the pinballish device involves time scales on the order of the time required for an ice cube to melt.
2. In the case of the pinballish device, the glasses of water cannot interact with one another, while presumably the subsystems of the universe can.
3. The past hypothesis references a vague macrostate in the early universe that is insufficient to generate specific predictions about the macrostates of present subsystems, while the pinballish device allows us to near deterministically expect that we will find, at present, half-melted ice cubes in the glasses at the bottom of the device.

These three elements are salient in explaining why the pinballish machine's "local" past hypothesis does prevent the reversibility objection from going through, and why it fails to be convincing when considering subsystems of the universe. We can again write

$$P(U\&B\&M) / P(H\&B\&M) > 1/C \qquad (1.2)$$

where we now understand B to be the initial state of pinballish device where, say, 12 ice cubes are at the top of the apparatus (call this the 'local past hypothesis'), and M is the present macrodescription of the rest of the glasses, some of which contain half-melted ice cubes. Here the three elements described above conspire to allow this inequality to be satisfied.

Since the time between the local past hypothesis and the present macrostate is short and commensurate with the time it takes an ice cube to become half-melted from a fully unmelted state, it is overwhelmingly likely that the presently half-melted ice cube in the glass evolved from one of the unmelted cubes described by the local past hypothesis rather than as a spontaneous fluctuation, and we can trace the thermodynamic history of the ice cube from the time of the local past hypothesis to the present. But there is nothing analogous in the case of the global past hypothesis: the ice cube did not begin in an unmelted state 13 billion years ago. Why should I not believe that the presently half-melted cube arose a spontaneous fluctuation?

In addition, the local past hypothesis asserts that one should expect a total of 12 half-melted ice cubes to be gathered presently in glasses at the bottom of the device. If 11 of the ice cubes are accounted for in M, the remaining glass should have the last half-melted ice cube in it, one that was previously fully unmelted. In this case, M serves to fix the history of our system of interest in a way that the macrostate of the universe cannot. As such, the H and U terms from (1.2) can be dropped, since they are both contained in the measure associated with B&M, so (1.2) reduces to

$$P(B\&M) / P(B\&M) = 1 > 1/C.$$

The equality is thus satisfied, in part because of the temporal proximity of the local past hypothesis to the present situation, but also because each ice/water system is isolated from the others, *and* because the specificity of the local past hypothesis guarantees this to be the case. In a universe such as our own, one does not expect or believe ice cubes to be causally isolated from the rest of the universe, stemming from the time of the big bang. If we see ice cubes in everyday life, we quite reasonably infer that they came into being through some interaction with the rest of the universe, even if there is no trace of the interaction.

Furthermore, the macrostate of the rest of the universe cannot fix the history of our ice cube, even if it is restricted to those microstates compatible with the past hypothesis, in the way the local past hypothesis did. The local past hypothesis reported a *very specific* macrocondition, one where we knew that all the cubes (exactly 12) would wind up, half-melted, in glasses of water at the present time, one that specified that there could be no causal interaction between the glasses, and one that precluded any of the cubes emerging from a spontaneous fluctuation. Conversely, the full-blown past hypothesis is of a very different sort: the macrocondition it refers to is, at best, a general description of a macrodescription that cannot possibly satisfy the aspects of the pinballish device that made the corresponding condition work there.[13]

The main argument for the postulation of a highly non-equilibrium condition in the early universe was supposed to rest with its ability to explain irreversible thermodynamic processes like ice cubes melting in water. Does it accomplish this? Recall the above result, labelled (1.3). We noted that the extent to which irreversible processes are explained by reference to the initial condition of the universe is sensitive to the scale of the systems being considered. As a concrete example, consider a gas found to be slightly out of equilibrium, perhaps as the result of a small fluctuation of the sort we expect to observe for such systems. Now, there are at least two distinct histories one can attribute to the system. First, the system might have come to its present state from an even lower entropy state, and it is this possibility that the past hypothesis is intended to render plausible. But the system could easily have come to its present state as the result of a past fluctuation. Even if the past hypothesis could render it not entirely unlikely that the gas was previously in an even lower entropy state, it should not eliminate the possibility that the gas came to its present state as the result of a fluctuation. If we assume that the past hypothesis does not eliminate this as a real possibility (as we should), it is virtually

[13] Specifically, the past hypothesis is the claim that "the world first came into being in whatever low-entropy highly condensed big-bang sort of macrocondition it is that the normal inferential procedures of cosmology will eventually present to us." (Albert 2000, 96)

irrelevant to explaining the non-equilibrium state of the gas. But in the case of the universe itself, it is highly relevant. What about the cases of interest to us, those everyday processes like ice cubes melting in glasses of water?

The second law of thermodynamics, understood as being a law that only holds on the average, acknowledges the fact that anti-thermodynamic behaviour will occasionally occur as the result of spontaneous fluctuations. To be sure, the relative magnitude of these fluctuations will be dependent on the size of the system being considered. Usually, a statistical explanation is used to (causally) explain the occurrence of a single event by citing its increased probability of occurrence given another event. In the case of the past hypothesis, we have a single event explaining the occurrence of a myriad of events, namely those falling under the purview of the second law. In a sense, conditionalising on the highly non-equilibrium statistical mechanical state of the early universe ought to be interpreted as a causal explanation of a higher-level law. But this appears to be an odd sort of causal explanation. Surely the degree of relevance of the proposed explanation varies considerably with the size and time scale of the events we look to explain, whether it be the thermodynamic behaviour of the universe itself or a single ice cube melting in a glass of water. Can a lawlike generalisation such as the second law of thermodynamics be subsumed by appealing to a single event when the statistical relevance of the cause to the myriad of effects varies widely?

We can investigate this claim with reference to Albert's notion of typicality. While one might be justified in thinking that the universe that we see before us is "typical" of the sort of universe that one would expect had the universe originated in a highly non-equilibrium state (i.e. we find galactic structures, clusters and superclusters of the sort astronomers observe), it does not follow that small subregions of the universe, such as a glass of water with an ice cube in it at present, are typical to even a remote degree by positing the past hypothesis. Yet the second law is statistically valid on all thermodynamic scales, but even the most generous understanding of what is typical only covers the largest of these systems, and is a long way from explaining everyday thermodynamic phenomena.

To be sure, one expects stricter agreement with the second law (i.e. less prominent fluctuation phenomena) for larger rather than smaller thermodynamic systems, and the past hypothesis roughly mirrors this feature. The past hypothesis can rule out large fluctuations of the universe as a whole (at least in our epoch) while, as noted above, it should be completely irrelevant for the sorts of fluctuations we normally see in small thermodynamic systems. Nonetheless the challenge, insofar as Albert claims that the past hypothesis is intended to "underwrite the actual content of our thermodynamic experience" (2000, 159), is to provide an account that can demonstrate that the *fluctuations* we see in thermodynamic systems, and not just the monotonic increase in entropy, are predicted by appealing to the past hypothesis.

One might defend Albert against this worry by noting that Albert conjectures that whatever the history of a thermodynamic system (i.e. whether its present state was the result of a spontaneous fluctuation or a normal thermodynamic process), the present probability distribution of the system will randomly sample the microstates accessible to the system insofar as the distribution is used to predict the future evolution of the system. Consequently, the probability distribution associated with the system conditionalised on the past hypothesis is virtually identical for the purposes of prediction to a probability distribution that includes all the microstates accessible to the system.

This defence misses the mark. The concern expressed here is not that *future* fluctuation phenomena are correctly predicted whether or not one conditionalises on the past hypothesis, but whether, upon encountering a non-equilibrium thermodynamic system, conditionalising on the past hypothesis correctly characterises the probability that the system arose *in the past* by a normal thermodynamic process and not by a fluctuation (however small that probability may be). In the absence of some direct calculation of these probabilities from the past hypothesis itself, it is but an article of faith that these probabilities will match with whatever the actual frequencies are, thus "underwrit[ing] the actual content of our thermodynamic experience".

To indicate why I think this task is impracticable, consider two points that vitiate the applicability of Albert's conjecture. First, Winsberg (2004a) notes that the way the probability distribution samples the accessible phase space of a system will depend on *when* the conditionalisation is implemented.[14] As such, even if one accepts Albert's claim that the probability distribution randomly samples the phase space and thus correctly *predicts* fluctuation phenomena over the past (say) ten minutes when conditionalised on the macrostate ten minutes ago, this need not imply that it will correctly *retrodict* fluctuation phenomena over the past ten minutes when conditionalised on the present macrostate. Second, there is no clear or straightforward conceptual relationship between the region of the phase space accessible to a system and its probability distribution.[15] When these concerns are conjoined with the fact that the actual frequencies will be dependent on the size of the system under consideration, it appears that the past

[14] This point can be made plausible by noting that the probability distribution assigned to an unmelted ice cube five minutes ago, conditionalised on the past hypothesis *and the macrocondition of the rest of the universe five minutes ago* can differ significantly from the distribution associated with the presently half-melted ice cube conditionalised on the past hypothesis *and the macrocondition of the universe now*.
[15] This point will be more fully argued in Chapter 2.

hypothesis is powerless to provide a clear statement as to when it is reasonable to think that a system came to its present state as the result of a fluctuation rather than via a normal thermodynamic process extending into the system's past.

It would seem that we are still without the solution to the problem with which we began. Recall that what we wanted was some sort of postulate that made it overwhelmingly likely, given the half-melted ice cube before us and contrary to what the dynamics alone would have us believe, that the ice cube was fully unmelted ten minutes ago. Further, we hoped to show that this unmelted ice cube did not arise as a spontaneous fluctuation from an equilibrium state. However, what we discovered was that postulating the asymmetric initial condition of the early universe as being relevant to irreversible processes was equivalent to showing that the existence of non-equilibrium states strongly supported the proposition that the universe began in a highly non-equilibrium state. This result indicated a serious problem with the claim that a highly non-equilibrium state at the time of the early universe explains the everyday thermodynamic processes we set out to explain: no present macrocondition can count as evidence that the universe was ever further from equilibrium than it is now, and this goes for any past macrostate as well. Conditionalising on the macrocondition of the early universe cannot help much. While it is true that the existence of systems not in equilibrium supports, to some degree, the existence of an initial non-equilibrium condition in the early universe, we still lack an explanation of irreversible processes that solves the problem of reconciling the reversible dynamics with the apparent actual history of thermodynamic systems, and any reason to think that our records and memories of the past are veridical.

1.5 Further Problems for Albert

Albert claims to take the problem of justifying the veracity of our memories and records seriously. As we have already seen, Albert's solution to this problem is the past hypothesis, the claim that the universe began in a highly non-equilibrium macrostate. But how does Albert justify his postulation of the past hypothesis? The preceding section in part evaluated the extent to which this claim was successful in justifying the veracity of such records (records to the effect that the ice cube was unmelted ten minutes ago), and it was found wanting. Indeed, taking our memories or any record of the past as a thermodynamic system, Albert's explanation of their veracity does no better than his putative explanation of the evolution of our half-melted ice cube. Even if Albert could appeal to the argument that the past hypothesis validated the veracity of one's memories and records (which I maintain he cannot) and thereby argued for the past hypothesis on abductive grounds, it would be hard to see how his argument would be anything but circular.

The problem, as presented above, was that the present macrocondition of the universe, the ice cube, our memories, etc. represented (if we take the time reversibility of the underlying dynamics seriously) a state that would approach thermodynamic equilibrium both towards the past and future temporal directions. This is contrary to what we presumably remember, but the same consideration entails that our memories themselves arose as spontaneous fluctuations. Thus, the apparent history of our ice cube, together with the memory of the ice cube having been less melted ten minutes ago than it is now, constitute the very *explanandum* that the past hypothesis was supposed to explain. Notwithstanding its failure to do so, the appeal to the veracity of our memories in order to justify the past hypothesis would simply beg the question.

Luckily, Albert recognises that this move would be fallacious (2000, 94). Unfortunately, his solution to the justification or corroboration of the past hypothesis falls into the same fallacy once removed. Albert's argument is as follows:

> Our grounds for believing [the past hypothesis] turn out to be more like our grounds for believing general theoretical *laws*. Our grounds (that is) are *inductive*; our grounds have to do with the fact that the proposition that the universe came into being in an enormously low-entropy macrocondition turns out to be enormously helpful in making an enormous variety of particular empirical *predictions*. (2000, 94)

If there weren't independent evidence for the existence of the big bang, this would certainly be true.[16] Nonetheless (and the gratuitous use of the word 'enormously' aside), Albert's argument fails to establish the past hypothesis, for it is question begging.

Our grounds for believing in theoretical laws are generally inductive (or abductive), and the reason why we tend to believe in most theoretical laws is, indeed, because they are enormously helpful in making empirical predictions. But, more accurately, one does not need laws in order to make predictions. I can predict that the Buffalo Bills will win this year's Superbowl, that there will be peace in the Middle East by the autumn, that the Earth will continue to revolve around the Sun, or that the gravitational constant of the universe will change tomorrow. In short, I can make a myriad of predictions, with or without the help of theoretical laws. I simply do not need their help. However, if I want my predictions to turn out *right*, if I want them to *accurately* predict future phenomena, then my predictions are best served by being buttressed by theoretical laws of one sort or another.[17]

---

[16] It is interesting to note that if this line of reasoning is valid, one could determine that the universe existed in a highly non-equilibrium state (a sort of primordial soup) without the benefit of modern cosmology. It is a testament to Boltzmann's (1898) ingenuity that he was able to see this before the advent of 20th century physics.
[17] At least in the case of predicted physical phenomena.

Fair enough. However, if these laws are inductively supported, if they are thought to be correct, it must be because they have generated predictions that turned out to be correct. And it seems to be an obvious necessary condition to corroborating any theoretical law that I must be able to check that the predictions that are made on the basis of that law do, in fact, turn out to be correct. Unfortunately, checking that a prediction turns out right relies on having some knowledge, a record, of what that prediction was. One cannot claim inductive support of a theoretical law without a prediction that is based on that law. And here's the rub: the problem, as it was posed, presumes that one cannot trust one's memories or records of such predictions and so no prediction can ever provide inductive support to any claim, past hypothesis or otherwise. In sum, the past hypothesis cannot be confirmed inductively since we have no method of confirming past predictions, and no way of trusting that we actually made any present predictions in the future.

Albert's argument for the past hypothesis presumes that we *can* verify predictions, that we can trust our memories of such predictions. But this is precisely what is at issue. Thus, it is hard to see how Albert's argument for the past hypothesis is anything but question begging.

A further issue that is raised with respect to Albert's discussion of the past hypothesis is whether his construction of the past hypothesis is consistent with his interpretation of probability. In describing the nature of the past hypothesis, Albert claims that

> ….If the project of statistical mechanics is on anything remotely like the right track, then, when all the data are in, the initial macrocondition of the universe had better turn out to be one relative to which – on the standard uniform probability distribution over microconditions – what we think we know of the history of the world, and what we expect of its future, is *typical*. (2000, 85)

Albert takes the position that the *only* time where the uniform probability distribution over all accessible microstates is the appropriate one to use is for the initial macrocondition of the universe. In fact, he explicitly rejects the view that the uniform probability distribution is the right distribution to choose for any present state of affairs (in contradistinction to almost all standard accounts of statistical mechanics).

This view (whatever its justification) is hard to reconcile with Albert's interpretation of probability. In a footnote he writes

> Here (by the way), and everywhere else in statistical mechanics, and (as a matter of fact) everywhere *simpliciter*, it seems to me of great help, it seems to me to spare one an enormous amount of confusion, to be thinking of probabilities as *supervening* in one way or another on the non-probabilistic *facts of the world*, to be thinking of them (that is) as having something or other to do, by definition, with *actual frequencies*. (2000, 81)

Despite the melange of strongly worded claims and hedging, one can distil the two salient features of Albert's interpretation. First, he sees probabilities as being non-tychistic, as being reducible to categorical properties of the system in question. Second, he embraces a strong frequency view of probability, where the probability is to be identified (somehow) with the *actual* frequency of identical events.

It is not my intention here to criticise Albert's interpretation of probability, but rather to investigate its consistency with his account of statistical mechanics.[18] What does it mean for the initial macrostate of the universe to *have* a probability distribution? According to Albert, the probability simply *is* the actual frequency of events from identically prepared systems. Yet there has only been (to the best of our knowledge) *one* universe. The *actual* frequency of microconditions is just the one that did, in fact, evolve from the original big bang. It is, with probability one, the microcondition that the universe found itself in at the time indexed by the past hypothesis. If the probabilities of which

[18] The issue of the interpretation of probability will be taken up in Chapter 3.

Albert speaks were, in some sense, tychistic, then perhaps one could make sense of what it would mean for the initial macrostate of the universe to possess a probability distribution. But since he explicitly rejects this, and sees the probabilities of systems as reducible to their categorical properties, there really is no probability distribution to speak of.

One might be tempted to interpret Albert as thinking of these "actual frequencies" as being frequencies across possible worlds sharing the same initial macrostate. While this might seem a plausible stretch, it still fails to capture what Albert could mean by the probability distribution of the initial macrostate of the universe. If we thought of these frequencies as being frequencies across possible worlds, then this interpretation would be one where the probability would be described by an ensemble of worlds, of initial macroconditions. But Albert rejects ensembles as being wrong-headed and of no relevance for statistical mechanics (2000, 67-70).

Finally, the inconsistency between Albert's past hypothesis and his interpretation of probability leaves his notion of typicality in troubled waters. Albert defined the notion of typicality in terms of the probability distribution of a certain set of macrostates (see section 1.3 above) which were thus made overwhelmingly more probable than any other sequence. In the case of the pinballish device (Figure 1), we could make sense of what the probability distribution Albert was speaking of meant. We could, in principle, construct such a device and run it an infinite amount of times, thereby generating a probability distribution of sequences of macrostates. But by pushing the typicality condition back to the initial macrostate of the universe, the notion of typicality loses all meaning, since the probability distribution is then simply the microcondition that the world found itself in at the time referred to by the past hypothesis. What the typicality condition then amounts to is that the world is just the way it is with probability one: a platitude if ever there was one. As such, it is unclear that the past hypothesis coupled with a notion of typicality can do the work Albert claims it can do.

So far, I've presented considerations to the effect that the initial non-equilibrium state of the early universe cannot have the explanatory power that it is often thought to have. The postulation of such a state cannot make it overwhelmingly likely that a half-melted ice cube was previously less melted that it is now, nor can it establish the veracity of our records and memories to that effect. While the target of these arguments have been specifically directed against Albert's recent presentation of this argument, the general point holds against all those arguments purporting to establish that the explanation of thermodynamic irreversible behaviour can be somehow explained by exclusively appealing to the early macrostate of the universe. Furthermore, I have argued that the existence of this past macrostate cannot be established by appealing to inductive grounds in the same way that one establishes theoretical laws. While it is by no means my intention to deny that the universe did, in fact, begin in a particularly non-equilibrium state, this fact alone cannot have the explanatory scope that it is often thought to have.

The upshot of this discussion is that any correct explanation of the time asymmetric behaviour of thermodynamic systems needs more. The non-equilibrium state of the early universe cannot be proven by appealing to thermodynamic systems, past or present, nor can it explain their irreversible behaviour. One must look elsewhere, say to a branch system approach (e.g. Reichenbach (1956), Davies (1974)), or to independent grounds for believing the veracity of our records (Horwich 1988), or for some other fundamental source of temporal asymmetry that "drives" thermodynamic irreversibility such as the causal asymmetry or radiative asymmetry. The big bang can't do it all by itself.

1.6 Who Cares What Happened Ten Minutes Ago?

Here I would like to present a sketch of how I will construct what appears to be a correct and cogent approach to the two problems that faced us above: providing a reasonable expectation that the entropy of thermodynamic systems (including the universe) was lower in the past and justifying the veracity of our records to that effect. The approach I favour is one that combines a branch systems approach with arguments that emphasise the primacy of the epistemological aspects of the problem over the physical. In particular, I see the solution to both the stated problems as involving a reformulation of the nature and mode of presentation of each one.

Albert applies a sort of (benign) double standard to his discussion of the explanatory scope of the past hypothesis. His central concern is to validate our memories and posits about the state of unmelted ice cubes in glasses of water ten minutes ago, despite the fact that the present state of the ice cubes, along with the statistical mechanical description of the underlying dynamics of the molecules that comprise the ice cubes, make it overwhelmingly likely that the ice cube was in fact more melted ten minutes ago than it is now. This validation can be achieved by stipulating that the ice cube was, in fact, unmelted ten minutes ago, as noted in Section 1.2. But then we would find the same concern presented to us at times before ten minutes ago, where we would be forced to infer (based on the underlying dynamics) that the ice cube came to its unmelted state from a more melted state. And so we were forced to adopt the past hypothesis, where the universe began in a highly non-equilibrium state.

What about the past hypothesis applied to the universe as a whole? Here we see that the postulation of the initial non-equilibrium state works exceptionally well. Take H to be the current state of the universe, U to be the state of the universe at the time indexed by the past hypothesis.[19] In this case,

$$P(B|U) = P(B|B) = 1 >> P(B|H), \tag{1.3'}$$

and the inequality is satisfied. We therefore find that the initial condition of the universe is highly relevant to explaining the thermodynamic history of the universe itself, from the time of the initial posit to the present: a satisfying but perhaps unremarkable result. Of more interest are the consequences of retrodicting the history of the universe before the time indexed by the past hypothesis. According to the underlying dynamics, we should expect that the initial condition of the early universe arose as the result of an amazingly and overwhelmingly unlikely spontaneous fluctuation from an equilibrium state of the universe itself.

Of course, this was Boltzmann's (1898) proposal for explaining the tendency of the (observable) universe to evolve towards an equilibrium state from the time of the initial posit, despite its consequences for times prior to it. According to Boltzmann's 'anthropic' argument, it just so happens to be the case that we live in a small region of the universe which, at some point in the past, underwent a spontaneous and highly improbable fluctuation into a non-equilibrium state, and is currently evolving back towards equilibrium. The justification or explanation offered for why such an unlikely fluctuation occurred is provided by a sort of transcendental or anthropic reasoning: the only way that living creatures such as ourselves could exist is if we found ourselves living

[19] Here there is no "rest of the universe" to consider.

in a portion of the universe that was evolving from a highly non-equilibrium state towards an equilibrium one.

Unfortunately, this argument, whatever one thinks of its plausibility, is vitiated by considerations of contemporary cosmology, for which we have independent evidence. According to standard theories of cosmology, the universe began as an initial singularity with a 'big bang', a highly non-equilibrium state. Whenever one takes to be the time of the "early universe" posit, whether it be at the time of the big bang or shortly afterwards, it follows, by the same considerations presented with regard to the ice cube, that the initial state of the universe was preceded by and arose due to a previous fluctuation from an equilibrium universe. Should we be concerned about this "prehistory" of the universe, for which we have no evidence? Apparently, we can either bite the bullet and accept this consequence of time-reversible dynamics, or we can stipulate that there is some special, time-asymmetric condition, that blocks this inference to the history of the universe before the time of the big bang.

Albert seems curiously unbothered by this retrodiction to times before the one indexed by the past hypothesis. To him, it seems outside the central concern of his programme to consider or worry about times before the past hypothesis or even areas of the universe beyond which we have local contact. Why? Albert takes as his central concern the validation and explanation of *our* experiences, *our* memories of processes that behave in accordance with the second law of thermodynamics. He is altogether unconcerned with what happened before the time indexed by the past hypothesis, and regions of the universe to which we do not have epistemic access, since this is beyond the scope of our possible experience, and beyond the concerns that Albert sets out to address.

We have already seen that the past hypothesis cannot have the explanatory import that Albert claims it to have, since it can neither render it likely that our ice cube was less melted than it is now, nor can it testify to the veracity of the memory that it was less melted than it is now. But it there appears to be (or at least it seems that Albert thinks there is) a disanalogy between the case of the thermodynamic state of the universe and our ice cube; namely that we need to worry about the past history of the ice cube before the posit to the effect that the ice cube was fully unmelted ten minutes ago whereas in the case of the claim that the universe was in a highly non-equilibrium state some time in its early history, we do not.

The most obvious source of this disanalogy is that we have records of, or at least can potentially have records related to, the state of the ice cube beyond ten minutes ago, while in the case of the universe it seems we cannot have records of its macrostate before the initial state. Thus, Albert implicitly concludes, we need not be bothered by the parallel inference that the universe itself arose as a spontaneous fluctuation because we are blocked from having any records of that fluctuation, in contradistinction to our experiences with, and memories of, ice cubes.

More generally, we can point to four ways that the histories of thermodynamic systems of the sort we experience differ from the thermodynamic description of the universe:

1. The non-equilibrium history of a thermodynamic system is but one among many histories of thermodynamic systems we wish to reconcile with the time-reversible underlying dynamics. In the case of the universe, each of these thermodynamic systems can be seen as subsystems of the universe and therefore, potentially, the postulation of a highly non-equilibrium initial macrostate of the universe is a

parsimonious explanation of the history of the smaller thermodynamic subsystems.

2. The initial, non-equilibrium macrostate of the universe is explained by some asymmetric law or fact about the universe stemming from contemporary cosmology or some other argument that makes it different from the history of everyday thermodynamic systems.
3. We have records and memories of the non-equilibrium histories of everyday thermodynamic systems, whereas we have no such records of the universe before the time indexed by the past hypothesis.
4. We have records and memories of *other* thermodynamic systems that lend inductive support to the non-equilibrium history of any particular thermodynamic system, while in the case of the universe we have no such support.

The first disanalogy is the one that Albert focuses upon as providing grounds for the belief that individual thermodynamic systems behave in accordance with the second law, and as justification for the veracity of our memories. However, the alleged validity of this disanalogy as a solution to the problem with which we began has already been debunked. The exclusive appeal to the non-equilibrium state of the early universe cannot be the *explanans* that we are looking for.

As for the second disanalogy, there exists a rather large literature that attempts to explain the early, non-equilibrium state of the universe. Sklar (1993) reviews many of these approaches, from transcendental arguments, teleological arguments, multiple universes, inflationary models of the universe and Penrose's claim that the initial singularity from which the universe emerged must have a Weyl tensor that is identically zero, and hence would describe a universe in an extremely low entropy state early in its history. Each of these proposals has a similar flavour; they all look to establish the plausibility of the low entropy state of the universe over the seemingly more probable state of the early universe as a high entropy one, and thereby attempt to block the inference that the universe itself must have arisen as a spontaneous fluctuation from equilibrium.[20]

Whatever the plausibility of these arguments, they seem to be specific to explaining the thermodynamic condition of the universe itself, and don't have any bearing on the explanation of the irreversible behaviour of common thermodynamic systems. What these arguments do (if they are successful) is provide independent grounds for the belief that the universe did begin in a highly non-equilibrium macrostate, that the past hypothesis is correct.[21] Such arguments may be needed to establish this fact, since I have argued above that one cannot infer the truth of the past hypothesis by appealing to its ability to guarantee the veracity of our memories and records, or that the present non-equilibrium macrostates of thermodynamic systems did not arise as spontaneous fluctuations from equilibrium.

These considerations relocate the problem that we posed above. The last two disanalogies between the case of the universe and common thermodynamic systems point to our records and experiences with such systems. If we didn't have the capacity to form (apparent) records of past events, there really would be no issue. We could just as well conclude that the present, half-melted ice cube had formed as a spontaneous fluctuation, just as we seem to be entitled to when considering the prehistory of the universe. In other words, if these records are in fact veridical, then we would appear to be justified in believing in the low entropy past of thermodynamic systems. However, if they are not,

[20] See Earman (forthcoming) for a criticism of these approaches.
[21] Price (1996) has offered several considerations arguing that these approaches are in some sense misguided, that these putative explanations miss the mark.

then we may as well have no records at all; we may as well believe that (say) our ice cube arose as a spontaneous fluctuation from equilibrium.

Suppose that I walk into a room that is empty with the exception of a glass of water with a half-melted ice cube in it. What ought I to infer about the history of this system? Should I suppose that the ice cube was fully unmelted ten minutes ago, and before that the water and ice cube were not in thermal contact (someone dropped the ice cube into the glass)? Or perhaps that immediately before I entered the room, someone dropped the ice cube, half-melted, into the water? Maybe the ice cube was three quarters melted five minutes ago, when someone dropped it into the water. Or again, perhaps the water was previously in equilibrium with its surroundings when it spontaneously formed a half-melted ice cube in the glass. Or perhaps the ice cube spontaneously formed, fully unmelted, in the glass ten minutes ago from previous an equilibrium state somewhere in the past.

Each of these scenarios is compatible with the dynamics governing the behaviour of the ice/water system, and solely on the basis of the evidence I have upon entering the room, none of these possibilities are overwhelmingly likely to be the correct description of the history of the system (measured over the infinite continuum of other possibilities). After all, the only knowledge I seem to have is that there is, before me, a glass of water with a half-melted piece of ice in it. Of course, we have more than that since we have experiences with other glasses of water with ice cubes in them, and we seem to know that the latter possibilities mentioned (those which exhibit past anti-thermodynamic behaviour) are not in accord with our experiences of ice cubes. More to the point, we never *see* cubes of ice spontaneously form out of warm glasses of water, and we never *observe* strong anti-thermodynamic behaviour occurring towards the future.

Consequently, we seem to have good inductive grounds to discount retrodictions in the past temporal direction that indicate that the present macrostate evolved from a macrostate closer to equilibrium than it is now, despite what retrodictions based on the underlying dynamics of thermodynamic systems would have us believe.

Now, this inductive conclusion to the effect that the ice cube did not arise as a spontaneous fluctuation depends on the veracity of our previous records and experiences with ice cubes, which is itself called into question by the problem before us. Without some independent means of ensuring the veracity of our records of past experiences with ice cubes, we are blocked from making any inference regarding the past history of the ice cube that we have encountered upon entering the room.[22] It appears that all we are left with is the present macrostate of the ice cube and our apparent records of events that may never have taken place, and nothing more. Whatever the correct solution to the problem of explaining irreversible thermodynamic processes, it must start here and build forwards.

[22] In fact, without the assurances of our records, we cannot have any confidence that the underlying dynamics dictate that the ice cube arose as a spontaneous fluctuation since our belief in the empirical adequacy of our theories also depends on the veracity of our records that have, in the past, confirmed those theories.

Chapter 2: The Jaynesian Approach to Statistical Mechanics

E. T. Jaynes (1983) considered his work on statistical mechanics to be on the one hand a major break from traditional approaches to the subject, and on the other a natural extension of Gibbs' formalism and understanding of statistical mechanics, whose thought was fettered by the philosophical *zeitgeist* of the late 19th century. In Jaynes' work, one finds an approach to statistical mechanics that utilises the tools of information theory, and a radical reconstruction of the central explananda and explanantia of the foundational problems of the theory. Here statistical mechanics is to be regarded as properly being a theory of statistical inference in the face of incomplete information rather than as a physical theory in its own right. Specifically, Jaynes' approach can be characterised by the following claims:

1. The probabilities that appear in statistical mechanics ought to be thought of as epistemic probabilities measuring the degree of ignorance of the exact microstate of a thermodynamic system that is defined by a set of macroscopic constraints.
2. The probability distribution that maximises the information-theoretic entropy for a thermodynamic system is the one that assigns equal probability (on the standard measure) to all microstates compatible with the thermodynamic macrostate.
3. Ensemble and ergodic methods are irrelevant and conceptually inadequate for explaining the phenomena of either equilibrium or non-equilibrium statistical mechanics.
4. The central project of the reduction of thermodynamics to statistical mechanics is to explain the appearance of experimentally reproducible thermodynamic processes and link the theoretical terms of the two theories, rather than reconciling the apparent paradox between the time-reversible microdynamics and irreversible macrophenomena.

5. Boltzmann's formalism and conception of statistical mechanics is an inadequate characterisation of thermodynamic phenomena, with a suitably interpreted Gibbsian approach being far superior in terms of practical use and conceptual clarity.

The following two chapters will be devoted to examining the nature of each of these claims, and in the process describing in some detail Jaynes' conception of statistical mechanics. I will then attempt to relate and expand (where necessary) this position to address mainstream problems in the foundations of statistical mechanics, indicating when a Jaynesian view can be of use in solving the central problem of deriving irreversibility from reversibility. In section 1 of this chapter, I consider how the Jaynesian approach can reproduce the formalism of equilibrium statistical mechanics, and argue in the second section for an information-theoretic conception of entropy as being superior to either the standard Gibbs entropy or the Boltzmann entropy. The final section takes up the issue of non-equilibrium statistical mechanics. In Chapter 3 I explore the issue of using epistemic probabilities in the interpretation of statistical mechanics, criticising the ability of an objective, physical interpretation to meet the explanatory goals of foundational statistical mechanics. The next section of the chapter examines ergodic theory, and discusses several objections raised against the Jaynesian approach from the perspective of ergodic theory. Finally, the reversibility objection is reformulated for an epistemic approach to statistical mechanics, and a preliminary solution is discussed.

2.1 The MEP as a Statistical Mechanical Formalism

In this section, I present and develop the Maximum Entropy Principle (MEP) and apply it to equilibrium statistical mechanics, providing the background and the formalism to further discuss the foundational and philosophical importance of the principle as it applies to statistical mechanics. The present discussion of the MEP will be restricted in two ways. First, Jaynes claims that the MEP is not specific to statistical mechanics, but is a general schema for probabilistic reasoning and the assignment of probability distributions based on constraints imposed by empirical data or knowledge. In what follows, the MEP will be discussed only in the context of statistical mechanics, and its wider and more problematic role as a rule of statistical inference and probability assignment will be bracketed. Second, the MEP formalism discussed here will be limited to a finitely partitioned event space, rather than a continuous measure. The extension to a continuous measure will be discussed further in Chapter 3, where the use of epistemic interpretations of probability in statistical mechanics will be discussed in more detail. The purpose of this section, and this chapter more generally, is to glean from the MEP approach any interpretational advantages of foundational value for statistical mechanics while leaving the technical problems of its scope aside.

Jaynes (1983) presents the MEP as being an updated principle of indifference in the spirit of Laplace and Bernoulli, incorporating the then novel results of information theory in order to solve some of the problems facing epistemic accounts of probability. In particular, Jaynes looks to solve the problem of generating the correct or preferred probability distribution for a random variable, taking on *n* distinct possible values, given some constraints by maximising the information-theoretic entropy, given by the expression

$$S_I = -\sum_{i=1}^{n} p_i \log p_i \ . \qquad (2.1)$$

The choice of this formula as the one to be maximised is intended to be the least biased measure of one's ignorance over some event space subject to a set of constraints. The function that satisfies the following *desiderata* is sufficient for (2.1) up to a constant (Shannon and Weaver 1949):

1. The entropy function is a continuous positive function of p_i.
2. If $\forall i,j$, $p_i = p_j$, then S_I is a monotonic increasing function of n.
3. $S_I(p_1, \ldots, p_n) = S_I(w_1 \ldots w_2) + w_1 S_I(p_1/w_1, \ldots, p_k/w_1) + w_2 S_I(p_{k+1}/w_2, \ldots, p_n/w_2)$. This is a composition law where the first k events are grouped together as a single event $w_1 = (p_1 + p_2 + \ldots + p_k)$, and likewise for the events from k to n. This ensures that the entropy function is unchanged by regrouping or re-labelling the probabilities in the event space.

Generally, the constraints on the probability distribution will come in the form of the average, or expectation, values of some quantity of interest. The resultant probability distribution can then be solved by the method of Lagrange multipliers. For instance, Jaynes considers the following problem: given the fact that a die, after having been (independently) tossed a number of times, has an average value of 4.5 dots face up, what is the most reasonable probability distribution to assign for the result of the next toss? There are two constraints in this case, namely that

$$\sum_{i=1}^{6} p_i = 1, \qquad \sum_{i=1}^{6} i p_i = 4.5$$

where (2.1) is to be maximised. Using the technique of Lagrange multipliers, the probabilities for each p_i are given by

$$p_i = \frac{e^{-\lambda i}}{Z(\lambda)}$$

where the partition function is

$$Z(\lambda) = \sum_{i=1}^{6} e^{-\lambda i}$$

and the constraint expressing the mean value of the rolls determines λ by

$$-\frac{d}{d\lambda}\log Z = 4.5\,.$$

Generally, if there is more than one constraint at work in determining the probability distribution, the partition function is a function of several λ_m, each of which is determined by equating the mean value of the constraint with the partial derivative with respect to λ_m of the logarithm of the partition function. The resultant probability values, as calculated by Jaynes (1978), are:

p(1) = 0.05435, p(2) = 0.07877, p(3) = 0.11416,

p(4) = 0.16545, p(5) = 0.23977, p(6) = 0.34749.

This dice problem presents a simplified example that is useful in order to expound the features of the MEP formalism. Unlike the case where the mean or expected value of the die face is 3.5 (which would give an equiprobable distribution), this case updates the principle of indifference to handle more complex cases where the observed frequencies deviate from equal probability and/or where there is good reason to think that the experiment is biased in some manner. Jaynes (1967) claims that the MEP serves to lay down the probability distribution based on a set of given constraints which does not eliminate any possible future outcome and provides the most "honest" probability assignment possible.

In the case of equilibrium statistical mechanics, the MEP functions in a similar way to the dice problem presented above, where the constraints take the form of the values of physical measurements performed on a system of interest, such as volume, total energy (usually in the form of temperature measurements), etc. These constraints serve to restrict the probability distribution to the (partitioned) region of the phase space that is compatible with the results of measurement, namely the macrostate. By maximising the information-theoretic entropy, the MEP assigns equal probability to the regions of phase space compatible with the constraints. In this way, the probability distribution generated by these constraints encodes the information one has about the system in such a way as to allow another person to recover the values of the constraints from the probability distribution. According to Jaynes (1978), any probability distribution other than the MEP distribution would either encode more or less (i.e. bias) information than was actually possessed by the generator of the distribution because it would privilege a certain class of microstates compatible with one's macroscopic knowledge without sufficient reason.

The notion of a macrostate is defined by listing the macroscopic constraints that are synchronically at work in a thermodynamic system. Thus, a macrostate can be thought of as a list detailing the results of measurements made on the system at some particular time, or reflecting some preparation of a system by means of controlling one or more of its macroscopic variables. It is important to note that according to this definition, each physical system can correspond to several macrostates,[23] and since the information-theoretic entropy is a function of the probability distribution generated by those constraints, it is a property of the macrostate of the system, not its microstate. In this sense, entropy is an anthropomorphic concept: it is "a measure of the degree of ignorance of a person whose sole knowledge about [the system's] microstate consists of the values of the macroscopic quantities that define its thermodynamic state." (Jaynes 1983, 238) Thus, entropy, unlike the energy of a system, is not an intrinsic property of physical systems. Rather, it is a relational measure between the physical system and one's ability to discern its microstate based on given informational constraints about the macroscopic

[23] For instance, the same physical system could have its macrostate defined by its temperature and volume, or also by its temperature, volume and net magnetisation.

properties of the system. Thus, a more "complete" description of the thermodynamic state of a system, based on the inclusion of a greater number of measurements of thermodynamic parameters, will generally lead to a different entropy assignment than a less exhaustive set of measurements would.

In what follows, I will adopt the following terminology: by a *physical* system, I mean the actual system that is characterised by its microstate, including the values of all the thermodynamic observables possessed by the system, insofar as they are determined by the actual microstate of the system. *Thermodynamic* systems, by contrast, are systems characterised by a given macrostate, which is a partitioning according to some chosen set of thermodynamic observables. Thus every particular microstate might correspond to many thermodynamic systems, in the sense that the same microstate can be described as different macrostates. *Statistical mechanical* systems are associated with probability distributions over microstates, though how such a system is to be understood may depend on one's interpretation of probability. If the probabilities are thought of as being objective, then the statistical mechanical system might be a dispositional property, an ensemble over similar systems, etc. However, if the probabilities are interpreted subjectively, then a statistical mechanical system will generally refer to ascriptions of belief regarding the actual microstate of the physical system, however the ascription of the probabilities is effected.

Let us now apply these ideas to equilibrium statistical mechanics for the canonical case. Following the dice example, we consider a set of observables [X_1, $X_2 \ldots X_m$] and the results of a sequence of measurements of these observables (say, temperature or energy, magnetisation, stress, etc…), and interpret the results of these measurements to be the expectation values or constraints operative in constructing the

MEP probability distribution. Then the Lagrange multipliers solving the problem are given by solving the equation

$$\langle X_i \rangle = -\frac{\partial}{\partial \lambda_i} \log Z, \ \sum_j p_j = 1$$

where

$$Z(\lambda_1 \ldots \lambda_m) = \sum_{j=1}^{n} e^{-(\lambda_1 X_1 + \lambda_2 X_2 + \ldots \lambda_m X_m)}$$

and the resultant probability distribution is

$$p_j = \frac{e^{-(\lambda_1 X_1 + \lambda_2 X_2 + \ldots \lambda_m X_m)}}{Z(\lambda_1 \ldots \lambda_m)}.$$

This reduces to the usual canonical distribution when the only known constraints are the energy and volume of the system. Essentially, this completes the derivation of probability distribution for a canonical system with an arbitrary number of constraints. It is easy to see that this method yields results identical to ordinary Gibbsian statistical mechanics. The usual statistical mechanical relations apply, where for instance, the pressure is

$$\langle p \rangle = \frac{1}{\beta} \frac{\partial}{\partial V} \log Z$$

and the internal energy is given by

$$\langle U \rangle = -\frac{\partial}{\partial \beta} \log Z$$

where β=1/kT and is the Lagrange multiplier for the constraint corresponding to the system's energy. In general, the information-theoretic entropy, as defined above, obeys the following relation:

$$kS_I \leq S_E \tag{2.2}$$

where S_E denotes the thermodynamic or experimental entropy as defined in conventional thermodynamics, relative to the macrostate. The equality holds if and only if the information-theoretic entropy is calculated from the usual canonical ensemble distribution, since any additional constraints over and above the present macrostate (say, some information about the past macrostate of the system) would further constrain the information-theoretic entropy. In this sense, the experimental and information-theoretic entropy represent two different concepts, the former defined by its role in ordinary thermodynamics while the latter is a general property of probability distributions, and is only given physical dimension by the Boltzmann constant.

At first, this relation may appear somewhat strange, since it seems to deny the possibility of reducing the thermodynamic entropy, which is a property of a macrostate, to some related *intrinsic* property of the microstate. Indeed, one might think that the whole project of reducing thermodynamics to statistical mechanics is predicated on providing just such an identification between the concepts of the two theories, and the reduction of thermodynamics to statistical mechanics serves as a paradigm example of such a reduction (e.g. Nagel 1961). Nonetheless, as many writers have noted (e.g. Sklar 1993, Yi 2002), the supposed reduction is extremely problematic, especially in the case of entropy, in part because its value is dependent on the way the macrostate is characterised. In the following section, I will argue that the information-theoretic approach to characterising the thermodynamic entropy holds significant advantages over other extant proposals.

2.2 Interpretations of Entropy

While the entropy of a system is well-defined for thermodynamic systems, there exists much debate over the nature of its statistical mechanical surrogate. While I do not want to present an exhaustive account of the literature on this topic, I would like to present some considerations to the effect that the information-theoretic entropy best serves as the reducing concept for the thermodynamic entropy.

The intertheoretic reduction of thermodynamics to statistical mechanics presents considerable challenges, especially when trying to link or identify the theoretical terms of each theory with the other. In the case of entropy, the attempt to identify some property of statistical mechanical systems with its thermodynamic description is highly problematic. While the problems associated with Gibbs' (1902) approach are well known (see Ridderbos and Redhead (1998), Ridderbos (2002) and Sklar (1993)), recently a number of authors have taken to defending the Boltzmann (1898) conception of entropy, such as Albert (2000), Lebowitz (1994), Callender (1999) and Goldstein (2001), among others. In this section I will motivate and argue for an alternative approach to understanding the nature of entropy associated with Jaynes (1983), where statistical mechanics is to be understood as a theory of inference about certain kinds of physical systems, and the entropy is understood to be a property of an epistemic probability distribution over the possible microstates of such systems.

The strategy of this section is as follows. Rather than directly arguing for the cogency of the Jaynesian approach to entropy, I will pose a set of problems in sections 2.2.1 and 2.2.2 that question the capacity of the Boltzmann approach to successfully reduce the thermodynamic entropy to its statistical mechanical surrogate.[24] Specifically, I will claim that the Boltzmann conception of entropy is unexplanatory and generates an unacceptable ambiguity in reducing thermodynamics to statistical mechanics, or will be unable to explain why thermodynamics is empirically successful. In section 2.2.3, both the fine-grained and coarse-grained Gibbs' entropies will be considered, reviewing a few conceptual problems associated with each. Finally, in section 2.2.4, I will demonstrate how these fundamental problems about the nature of entropy can be resolved by adopting the view that statistical mechanics (and thermodynamics) is best characterised as a theory of inference.

At the very least, a successful reduction of the thermodynamic entropy to some statistical mechanical concept should establish a reductive basis for the entropy values of the different equilibrium states of thermodynamic systems, the only states for which the entropy is defined in conventional thermodynamics. As a thermodynamic system evolves from one equilibrium state to another, say by the removal of a constraint on the system, the entropy takes on well-defined values only for the initial and final states of the system. I shall argue that the both the Boltzmann and Gibbs approaches fail to even meet this weak reductive requirement, even though determining how to extend the concept of entropy to non-equilibrium situations is considerably more difficult.

2.2.1 The Boltzmann Entropy

The Boltzmann conception of entropy (roughly) associates the entropy of a thermodynamic system with the number of microstates that are compatible with its

[24] I restrict my discussion throughout to a syntactic conception of theories, since how to characterise intertheoretic reduction on the semantic view is still an open question. See Morrison (2000) for one proposal.

macrostate. More precisely, the Boltzmann conception claims that the entropy of a system is a function of the macrostate of the system and the entropy of a *microstate* is the logarithm of the volume of the phase space associated with the macrostate to which the microstate belongs:

$$S_B = k \ln W \tag{2.3}$$

where W is the phase volume of the macrostate and *k* is Boltzmann's constant.[25] As such, the definition of the Boltzmann entropy makes no direct appeal to a probability distribution over microstates. In criticising the Gibbsian entropy that identifies the entropy as a property of a probability distribution over an ensemble, Albert claims that the "thermodynamic entropy is patently an attribute of *individual* systems. And attributes of physical systems can patently be nothing other than attributes of their *individual microconditions*" (2000, 70). Albert claims that S_B satisfies this requirement.

---

[25] There are a few reasons to reject the Boltzmann notion of entropy. First, the definition fails to correspond to the thermodynamic entropy in general, and does so only in cases where there are no interparticle forces or potential energy between internal degrees of freedom (Jaynes, 1965). Furthermore, there is considerable ambiguity in the meaning of Boltzmann's distribution function (which serves to count the possible microstates associated with a macrostate).

A more precise characterisation of the Boltzmann entropy envisions a six-dimensional μ-space where each particle in the system is represented by a point in this space, identified by its position and momentum in three spatial degrees of freedom. He then defines a distribution function f(x, p, t) that describes the system in this μ-space, coarse-grained into small, six-dimensional cells. The number of particles in any such cell, R, for a given time t is then

$$n_R = \int_R f(x, p, t) d^3x d^3p$$

Based on these considerations, for a suitably idealised kinetic model, Boltzmann shows that the quantity

$$H = \int f \log f d^3x d^3p$$

can only decrease over time until it reaches a minimum, and thus sought to define the thermodynamic entropy as

$$S \equiv -kH$$

This is the content of Boltzmann's H-theorem.

Now, this definition of entropy came under attack in Boltzmann's own time through the famous objection of Loschmidt and Zermelo, who raised the reversibility and recurrence objections that suggested that the H function could not remain stationary once it reached a minimum. As a result, Boltzmann attempted to reinterpret the meaning of the distribution function so as not to represent the *actual* distribution of particles in the various cells, but as the most *probable* number of particles in a cell, or as the *average* number of particles in a cell, or perhaps the probability that a *given* particle will be found in a particular cell. This indeterminism regarding the meaning of the distribution function obfuscates its extension to more general cases (Jaynes, 1967). In what follows I will put these worries aside.

What is lurking in the background of this quotation is the claim that any proposed reduction of thermodynamics to statistical mechanics must adhere to a supervenience condition. Any reduction of the properties or ontology of the higher-level theory should be stated exclusively in terms of the properties and ontology of the lower-level theory. Here Albert seems to be worried that the Gibbs entropy commits a category mistake: it identifies a property of many systems with those of a single system. Putting this specific worry aside, the general point holds: that the reduced theory's properties and ontology supervene on those of the reducing theory is a necessary condition on a successful reduction.[26]

Broadly, one might conceive of intertheoretic reduction as being characterised in one of two ways. One tradition is that often associated with Nagel (1961), where the reduced theory is in some sense "derived" from the reducing theory by means of bridge laws that connect the theoretical terms of the two theories. Another tradition is associated with Kemeny and Oppenheim (1956), where one thinks of a theory T_2 as being reduced to a theory T_1 if the *phenomena* explained or predicted by T_2 can also be explained by the reducing theory T_1 without essential reference to T_2, and T_1 is a more fundamental or more encompassing theory than T_2 is. On this view, those theoretical terms of T_2 that do not also appear in T_1 are in a sense dispensable or eliminable: they need not be thought of as genuinely referring terms because they play no essential role in explaining the phenomena (though one might still want to retain them).

A familiar example of this latter type of reduction is the reduction of Galileo's laws for free falling bodies to Newtonian mechanics: one can account for the uniform

[26] Depending on one's metaphysical stripe, it might be thought that a stronger relation is required, such as identity or realisation.

acceleration of freely falling bodies near the surface of the earth using the more general theory of Newtonian gravitation, demonstrating Galileo's laws to be applicable for a restricted domain of Newtonian mechanics. Viewed in this way, one reduces T_2 by showing how its range of phenomena can be predicted or explained by appealing to the reducing theory, perhaps in the process demonstrating the reduced theory to be an approximation to a more exact description furnished by T_1.

It should be noted that there are two types of reductive explanation that one can offer for the behaviour of a given thermodynamic system. Insofar as the values of the thermodynamic observables are thought to supervene on the properties of the microstate and that microstate evolves according to deterministic laws of motion, one could (in principle) explain the evolution of the system by claiming that, given the laws of motion, the microstate of the system was such that it evolved in the way that it did. But, as Batterman (2002) notes, such explanations, while obvious, often do not address the explanatory goals of a theory. In statistical mechanics, the question one really wants to ask is why such behaviour is generally to be expected; that is, why systems exhibit the *patterns* of behaviour they do, such that they evolve in accordance with the laws of thermodynamics.

In statistical mechanics, one attributes to a microstate a set of intrinsic properties or values of thermodynamic observables such as temperature, pressure, volume, etc., and explains why a thermodynamic system moves from one equilibrium state to another (due to, say, an adiabatic process or the sudden removal of a constraint) by appealing to a probability distribution over microstates and dynamically advancing these microstates in time. Although specific accounts of this explanation vary greatly, the behaviour of the system is explained by claiming that the overwhelming majority of microstates evolve to the new equilibrium state, and tend to stay there for an extended period of time.[27]

The statistical mechanical explanation does not make essential reference to entropy: the behaviour is fully explained by appealing to the probability distribution over microstates and its dynamical evolution.[28] Strictly speaking, the entropy, as defined in conventional thermodynamics, is not an observable but a function of other properties of the system, useful because it characterises an exact differential whose change in value is independent of the path according to which the system evolves. Indeed, it is only as an afterthought that one connects the final statistical mechanical equilibrium state with any notion of entropy (witness Boltzmann's famous H-theorem). One might reasonably expect that a reducing concept of entropy would be in some sense tied to this more fundamental explanation.

Callender (1999) argues that the Boltzmann entropy *can* have explanatory import. He argues, citing Railton, that

> "The stability of an outcome of a causal process in spite of significant variation in initial conditions can be informative … in the same way it is informative to learn, regarding a given causal explanation of the First World War, that a world war would have come about … even if no bomb had exploded in Sarajevo". To be sure, it would be wrong to think that the number of states [the microstate] is *not* "drives" [the microstate] toward equilibrium. But finding out about the (typical) "inevitability" of thermodynamic behaviour does carry with it modal and explanatory force. S_B quantifies this modal force. (373)

To investigate this claim more closely, consider a box divided into three chambers by removable partitions, where one mole of an ideal gas (at temperature T) is initially confined to the left-most chamber at t_0, obeying the familiar relation PV=NkT. The

[27] One takes thermodynamics as an approximation to statistical mechanics in that such evolutions are not deterministically assured, but merely statistically certain.
[28] As such, whether or not the Boltzmann entropy is in fact the correct reducing concept will be immaterial to the larger programme of explaining irreversibility by appealing to the initial probability distribution of microstate of the universe.

Boltzmann entropy for the system (whatever its actual microstate) is defined relative to these observables. Intuitively, the appropriate probability distribution to use is one where each microstate compatible with the macrostate of the system is assigned equal probability. In this case, the microstates counted in calculating the Boltzmann entropy coincides with those microstates included in the probability distribution assigned to the system.

When the partition constraining the gas to the left side of the box is removed, the gas expands to fill the accessible volume to the left of the last partition, coming to equilibrium at t_1. The Boltzmann entropy increases and the new equilibrium macrostate is successfully predicted by evolving the initial probability distribution according to the laws of motion. Insofar as the probability distribution and the Boltzmann entropy initially coincided, one might think that the Boltzmann entropy serves a useful explanatory role.

However, the measure associated with the probability distribution, as it evolves dynamically to (more or less) spread itself out evenly over the accessible volume, remains constant as a result of Liouville's theorem. As such, the microstates included in the probability distribution sample fewer microstates than are used in calculating the Boltzmann entropy. When the second partition is removed and the gas expands to fill the whole box (coming to equilibrium at t_2), the Boltzmann entropy increases yet again. *Why* the gas expanded (or should be expected to expand) can be explained by appealing to the probability distribution which samples less than the whole accessible phase space at t_1: although the Boltzmann entropy and the probability distribution matched up before the first partition was removed, this was not so afterwards. From this perspective, it appears that the Boltzmann entropy fails to meet our explanatory goals.

In order to recover an explanation, proponents of the Boltzmann entropy appeal to the notion of typicality. Intuitively, we know that for a system with a large number of degrees of freedom, such as the ideal gas described above, 'most' of the microstates compatible with initial state of the gas are such that upon removal of the partition, the gas will expand to fill the accessible region of the box. Furthermore, as the number of particles (N) constituting the gas approaches infinity, the set of abnormal microstates (where the gas does not expand) rapidly approaches measure zero (Goldstein and Lebowitz 2004).[29] Given the large number of molecules constituting the gas, it is reasonable to expect that the gas will expand to fill the accessible volume of the box and remain in equilibrium afterwards. As Goldstein and Lebowitz write, "the fact that [W] essentially coincides for large N with the whole energy surface … also explains the evolution towards and the persistence of equilibrium in an isolated macroscopic system" (2004, 57).

Two points are in order regarding this proposed explanation. First, as the gas expands from t_0 to t_1, we know that the probability distribution samples less than the full accessible phase space W at t_1, and *mutatis mutandis* for t_2. Hence, a dynamical assumption is needed to argue that as the gas expands, the probability distribution more or less randomly samples the microstates contained in the accessible phase volume over the relevant time scales so that the actual microstate of the gas remains a typical one. Otherwise, the expansion of the gas from t_1 to t_2 would fail to be explained by appealing to the typical, expected behaviour of the system from t_0. Such a dynamical assumption turns out to be very difficult to prove rigorously, especially for any realistic thermodynamic system. However, this assumption can be made plausible by appealing to

[29] It is not clear what the relation between this notion of typicality and the one described by Albert is.

computer simulations and models such as the Kac ring (Bricmont 2001, Garrido, Goldstein and Lebowitz 2004).[30]

Second, the appeal to the typical behaviour of thermodynamic systems purports to be the result of the large disparity in the number of the degrees of freedom associated with the macroscopic and microscopic scales. As Lebowitz remarks,

> S_B typically increases in a way which *explains* and describes qualitatively the evolution towards equilibrium of macroscopic systems. This behaviour of S_B is due the separation between microscopic and macroscopic scales, i.e. the very large number of degrees of freedom in the specification of macroscopic properties. It is this separation of scales which enables us to make definite predictions about the evolution of a *typical individual realisation* of a macroscopic system where, after all, we actually observe irreversible behaviour. (Lebowitz 1995, p. 2 emphasis original)

Intuitively, one might think of a series of flips of a fair coin, where the typical long run behaviour is 50% heads and 50% tails. In order to determine the result of any particular flip, one would need to specify the relevant values of each degree of freedom of the toss, namely the angular momentum of the coin as it is tossed, how high the coin is tossed, the air currents in the room, the bulk modulus of the table where it bounces when it lands, etc. Lebowitz's claim is that even if *some* of the degrees of freedom are fixed (or their values known) across many tosses, one should still expect typical behaviour from the coin as long as the fixed degrees of freedom fall far short of the full specification of the coin's (and the room's) physical state; that is, even if we knew (say) how high the coin was flipped, it would still not be enough to alter one's expectations of the typical long run behaviour of the coin due to the vast number of other degrees of freedom operative.

[30] How strong this assumption needs to be is a matter of some controversy. Earman (forthcoming) argues that the system must possess a property stronger than that of mixing to demonstrate that the microstate will remain typical, while Callender (1999) for instance argues that only a property weaker than ergodicity is required.

In the case of thermodynamic systems, the number of degrees of freedom for an ordinary gas is so vast that specifying another constraint (say, the net magnetisation) on the gas over and above a small set of pre-existing macroscopic constraints such as the total energy and volume of the system should not, if Lebowitz's claim is correct, alter the expected typical behaviour of the system. However the system is described, one should expect typical behaviour to ensue as long as the unconstrained degrees of freedom remain large relative to those that are specified. In the next subsection, a counterexample will be presented that undermines this claim.

2.2.2 Bridging the Theories

Alternatively, one might think of intertheoretic reduction as being a relation between two theories, where the theoretical terms of T_2 are identified with those of T_1 via a set of bridge laws. Here one seeks an identification or some nomological relation between the properties of the two theories, thus explaining the reduced theory by linking the reduced theory's ontology with that of the reducing theory. In this section I argue that no matter how this reduction is conceived, the Boltzmann entropy cannot do the job.

It is important to realise that the entropy of a thermodynamic system is only defined relative to a macrostate; that is, relative to the macrodescription one can or chooses to give it. There is nothing privileged about any particular macrodescription (expect for perhaps a full characterisation of the microstate) and thus to any entropy assignment, unless one adheres to a strict distinction between the observable and unobservable properties of a system. It is precisely this relational aspect that leads to trouble for the Boltzmann entropy.

As many have noted, the entropy of the system is *not* an intrinsic property of the microstate of the system, as Albert would like, but rather makes essential reference to the macrostate of the system; that is, a microstate cannot uniquely define a corresponding macrostate, and the entropy of a system is dependent on the description of the macrostate. As an extreme case, imagine a Laplacian demon that knows the exact microstate of a system. For such a demon, the macrostate simply *is* the microstate and the entropy of the system is zero, though for any non-ideal agent the entropy would be positive.

Callender (1999) takes a different approach to defending the adequacy of the Boltzmann definition. Instead of arguing that the entropy is an intrinsic property of physical systems, he acknowledges that the entropy is indeed relative to the characterisation of the macrostate, but contends that this need not compromise its objectivity. He claims that the relative nature of entropy

> does not imply, as Jaynes thought, that entropy is anthropocentric in nature. It implies merely that Boltzmann's entropy, to its credit, reproduces an ambiguity [in the definition of the entropy] already existing in thermodynamics. How many microstates correspond to a particular thermodynamic description is still an objective matter, even if a system admits more than one such description. (371)

There is a sense that we want more out of an intertheoretic reduction than a reproduction of ambiguities. In thermodynamics, the values of observables such as temperature are defined for equilibrium situations relative to the known constraints on the system (i.e. the macrostate):

$$\left.\frac{\partial S}{\partial E}\right)_{X_i} = \frac{1}{T} \qquad (2.4)$$

where the X_i are the operative constraints, to be held constant during the partial differentiation. For differently defined macrostates of a physical system, the entropy S(E, X_i) takes on different forms. Although there is no glaring inconsistency here, one might think that the ambiguity Callender describes as existing in thermodynamics ought to be clarified and reduced to an unequivocal theory of statistical mechanics, not merely reproduced at the reducing level. As noted in section 2.2.1, there is often an expectation that a reducing theory should provide a more fundamental and correct description of the world, resolving any conceptual tensions or empirical inadequacies that may exist at the level of the reduced theory. Usually this observation is stated as a challenge for accounts of intertheoretic reduction that naïvely require the establishment of a strict identity or nomological relation between the theoretical terms of the two theories. However, in this case the identity is almost too good: one could reasonably hope that statistical mechanics would resolve this ambiguity rather than recast it in its own theoretical terms. Below I will argue that, as long as entropy is understood in the Boltzmann sense, statistical mechanics does not have the conceptual resources to resolve this ambiguity on its own terms.

Returning to the quotation, Callender's claim that the number of microstates associated with a macrodescription is correct, but these considerations fail to absolve the Boltzmann entropy of its faults. Jaynes, who thinks of the entropy as a measure of one's ignorance as to the exact microstate of a system, would agree with Callender that the entropy in this sense is a fully objective. He writes that "this is a completely 'objective' quantity, in the sense that it is a function only of the [macroscopic quantities], and does not depend on anybody's personality. There is then no reason why it cannot be measured in the laboratory" (Jaynes 1983, 238). Jaynes' claim that the entropy is anthropocentric is just the claim that it is dependent on *which description* (i.e. which list of macroscopic quantities) one chooses to use in order to define the thermodynamic state of the system. In this regard there seems to be no disagreement between Callender and Jaynes.

However, if Callender wants to claim that the entropy is an *objective* relational property of the *system*, then here Callender and Jaynes differ. For Jaynes, the fact that the entropy is relative to a *description* is precisely the reason why it cannot be an 'objective' property. This is not to say that a relational property cannot be objective (as, say, the relativistic distance between two space-time events is), but that the fact that the relation is dependent on the description one *chooses* to or *can* give it compromises the possibility of its being objective, in the sense of it being a property of a physical system.[31]

But even if one allows this relational concept of entropy to be objective, I believe there is good reason to reject Callender's argument. It is Callender's contention that the entropy, though relative, is still an objective property of the microstate of the system. In the case of Jaynes' approach to defining entropy, a thermodynamic system has but one entropy: the information-theoretic entropy generated by the method described above, and it is this entropy that figures in the laws of thermodynamics. Of course, the value of the entropy will vary depending on the description that one chooses, but that is the entropy to be employed in making predictions or inferences. Unlike Jaynes' approach, Callender's view implies that *each* relativised entropy assignment is an objective property of the system. Thus, a particular physical system will have many different entropies corresponding to each possible description of its thermodynamic state, representing theoretical terms (call them $S^{(1)}$, $S^{(2)}$…, $S^{(n)}$) that will refer to some physical property of the system. This raises the question: to which entropy do the laws of thermodynamics refer?

---

[31] Here we can distinguish two senses of objectivity. First, there is the notion of the entropy being objective as not "depending on anyone's personality", in the sense that once the macrodescription is fixed, the entropy has a unique 'objective' value that does not depend on the beliefs or caprice of any individual, unlike, say, personalistic probabilities. The second, stronger sense of objectivity that Callender appeals to takes the entropy to be a *property* of a physical system.

It is not true that this multiplicity of entropies is innocuous because it merely shadows an ambiguity that already exists in thermodynamics. Such ambiguities as to the scope and domain of the theory seem to constitute a weakness in its conceptual rigour. More specifically, we must believe that there is no single entropy that corresponds to *the* laws of thermodynamics, but we are asked to treat each entropy as equally real and equally capable of reducing the entropy of thermodynamics. The only way I can see this as being possible is if we begin to think of thermodynamics not as a single physical theory, but as a class of theories ($T^{(1)}$, $T^{(2)}$…, $T^{(n)}$) each with its proper entropy figuring as a central theoretical term.[32] To be sure, these theories have a common or at least comparable mathematical formalism, and the theoretical terms that arise in these theories appear, superficially, to be similar. Nonetheless, these theories ascribe to physical systems different properties, and therefore are about different theoretical objects.

In and of itself, there is nothing wrong with this, as long as each of these sets of laws correctly describes thermodynamic processes from their own vantage point. The problem is that, in certain situations, one might demonstrate reproducible violations of the second law. As Grad points out,

> We come now to a basic question, how to choose an entropy in a given situation. We claim that the interests of the individual are paramount … we turn to aerodynamics. The existence of diffusion between oxygen and nitrogen somewhere in a wind tunnel will usually be of no interest. Therefore the aerodynamicist uses an entropy which does not recognise the separate existence of the two elements but only that of "air". In other circumstances, the possibility of diffusion between elements with a much smaller mass ratio (e.g. 238/235) may be considered quite relevant. (1961, 325)

The problem arises because some of these sets of laws will incorrectly characterise physical processes since they will leave out properties that are crucial to understanding

[32] One might appeal to a semantic conception of theories or a framework approach (Winsberg 2004b) to avoid this consequence.

the behaviour of physical systems. As an extreme example, Maxwell's demon is able to control the exact microstate of a system: this demon could generate microstates that, from the perspective of any less detailed macrodescription, would produce violations of the second law appropriate to this set of thermodynamic laws.[33]

More realistically, Jaynes (1992) provides a more general scenario based on discussions of Gibbs' Paradox of Mixing. In Gibbs' Paradox, one considers the same physical process relative to two different descriptions. If (say) equal amounts of the same ideal gas in a box at equilibrium are separated by a partition and the partition is removed, there is no net change in the entropy of the system, since the initial macroscopic state of the gas can be recovered by simply reinserting the partition. However, if the gases on either side of the partition are different, then upon removal of the partition, the gases mix irreversibly, and there is a net increase of the total entropy of the system:

$$\Delta S = nR\log 2$$

where n is the total number of moles of gas.

Now, the sensitivity of the entropy to the manner in which the system is described would be innocuous from the vantage of the S_B explanation if one could expect the system to behave in a typical manner and in accordance with the 2nd law *no matter how* the system is described, but this is not so. Drawing on the example provided by Grad, imagine one has a box of gas with oxygen on the left side of the partition and nitrogen on the right, kept at constant temperature. As before, if the partition is removed, the entropy of the system as it mixes will increase if one is sensitive to the fact that the

[33] Putting aside the question as to whether such demons are in fact possible. In any case, such a demon would violate the typicality condition because it can control all the degrees of freedom associated with the system.

gases on either side of the box are of different species, but will remain unchanged if the system is simply described as "air".

Now, instead of removing a partition, two semi-permeable membranes with pistons attached are placed at the centre of the box, one that is transparent to oxygen but opaque to nitrogen and another that is transparent to nitrogen but opaque to oxygen. When the pistons are slowly and isothermally pulled away from the centre of the box in opposite directions, work can be done by the system ($W=T\Delta S$). From the more fine-grained perspective, there is no mystery here: the work has been extracted from the system at the cost of an increase in entropy. Yet, from the more coarse-grained description, this is a violation of the 2nd law: work has been extracted from the system with no corresponding increase in entropy, since if the gas is simply described as "air" the initial and final states of the gas are the same.[34]

But whether or not the two gases are identical species or different is often a matter of the specificity of a description. As van Kampen notes, "the question is not whether they are identical in the eyes of God, but merely in the eye of the beholder" (1984, 309). Of course, the two gases need not be different elements or molecules for this sort of example to go through: they could be isotopes of the same element or even the same elements electrically polarised in different directions.[35] The point is that as long as one describes the elements of the gas as distinguishable and has some physical means of *exploiting* this distinguishability, one can reliably and reproducibly generate what, from a

[34] Here I interpret the 2nd law to be the statement that no net work can be extracted from a thermodynamic system operating in a cycle. Although no change in the system has been observed from the more coarse-grained perspective, this is not true from the fine-grained perspective: in order to complete the cycle, the pistons would have to be pushed back to the centre of the box, reestablishing the original state of the gas at the expense of work.
[35] Examples of this kind include isotope effects, where isotopes (or molecules containing isotopes like heavy water) may behave differently as a function of pressure or temperature.

more coarse-grained description, is *atypical* behaviour leading to violations of the 2nd law. Here, the explanation proposed by appealing to typical behaviour fails categorically, *irrespective* of the dynamical assumptions about the evolution of the probability distribution and *in spite* of the fact that the microscopic degrees of freedom far outnumber the macroscopic ones.

Callender puts this sort of problem aside, for the appropriate laws of thermodynamics relative to a description *will* work for 'typical' microstates, but not necessarily those specially arranged microstates created by Maxwell's demon, or for that matter a more fine-grained macroscopic description than the one in our possession. He writes that

> the worry is essentially asking for an independent justification of the imposition of our "natural" probability metric on [the phase space], which is something we do not have. But we can only try to solve one problem at a time, and anyway, the problem of justifying the "natural" probability metric is a very large one common to all of the different approaches to SM (1999, 371-2).

I read this passage as asserting that the problem of justifying the use of a particular probability distribution for a statistical mechanical system is completely orthogonal to defining the entropy (in the Boltzmann sense) and explaining why a particular physical system evolves in the way that it does, and he is surely right about this.[36] But to divorce the question of how two physical theories are reducible to one another from the *properties* or *behaviour* of physical systems themselves effectively abandons the project of reduction itself. Rather, what we seem to be looking for is a reducing *explanation* of why one physical theory is operationally or instrumentally successful given a more fundamental theory and certain background assumptions (which may or may not be

[36] As long as one doesn't conceive of the entropy as being related to this probability distribution.

justified), not a proper reduction of one theory to another. Yet the explanation itself is unsuccessful since atypical behaviour can be reproducibly generated.

If this is the case, then for the Boltzmann entropy the ambiguity is here to stay, since there seems to be nothing at the statistical mechanical level that could be used to resolve it. The opacity of the entropy is essentially tied to its relational character. It depends on a macroscopic characterisation, a characterisation that does not appear at the statistical mechanical level where the ontology is a probability distribution over microstates (however it is to be justified). As noted above, in statistical mechanics the explanation of thermodynamic processes is achieved by appealing to the properties of this probability distribution. Unfortunately, the probability distribution is conceptually unconnected to the entropy and to the explanation of thermodynamic processes according to the Boltzmann definition. It is impossible for statistical mechanics or even for some theory that might ultimately reduce statistical mechanics (if there be one) to work out this ambiguity. The entropy does not supervene on the statistical mechanical probability distribution, nor does it supervene on the actual microstate of the system. But that the theoretical terms of the reduced theory supervene on the terms of the reducing theory is a natural necessary condition on a successful reduction.

This has the general character of a reference class problem. Such problems are not uncommon when dealing with statistical theories or statistical explanations, where there may be no obvious privileged reference class on the basis of which one forms explanations or predictions, and there is a long literature regarding these issues in various sub-disciplines in philosophy. Concerning issues in the philosophy of biology, problems arise in stating evolutionary transition probabilities, which may vary in value depending on which causal factors are taken into account in the statement of the probabilistic law.

Here we seem to have an analogy with the case of thermodynamics: the reference class chosen will influence the formulation of the theory's laws. In discussions of evolutionary theory, responses to this problem vary from suggestions that the probabilities be thought of as epistemic (Rosenberg 1994), to their values being interpreted instrumentally (Giere 1976), to the specific instantiations of the laws one uses being a matter of pragmatics (Sober 1984).

Popper, in developing the propensity interpretation of probability, writes that "[propensities] are not properties inherent in the die, or in the penny, but in something a little more abstract, even though physically real: they are relational properties of the experimental arrangement – of the conditions we intend to keep constant during repetition" (1959, 37). Whether it be in thermodynamics, evolutionary theory or the propensity interpretation of probability, the point *seems* to be the same: even if the theoretical terms are essentially relative to our interests or epistemic abilities in the sense that they provide pragmatic, operational or instrumental characterisations of the phenomena we seek to describe, this need not rob the theories of their lawlikeness or explanatory power.

I say the point 'seems' to be the same for each of these cases because the question of reduction of thermodynamics to statistical mechanics throws a wrench in the problem. From the reductive perspective, if entropy is not needed to explain the phenomena, and if thermodynamics is characterised as being an instrumental, pragmatic or operational theory, then what point is there to seeking a reduction of the theory to statistical mechanics via bridge laws? Only if the reference class dependence were the *same* at both the reduced and reducing levels could one reasonably argue that the theory has been reduced. On the Boltzmann reading of entropy, the reference class dependence is not the same.

To put the point slightly differently, the project of intertheoretic reduction is predicated on the establishment of some sort of connection between the theoretical terms of each theory, whether this connection be logically necessary (as an identity claim) or nomological (see Sklar 1967). But no such connection exists as long as one conceives of the entropy as being relative to a chosen description, as not supervening on the underlying probability distribution (even if it is itself relative). Indeed, if the picture described above is correct, thermodynamics is neither explanatory nor lawlike from the perspective of statistical mechanics. As a reducing concept the Boltzmann entropy is unsuccessful.

2.2.3 Gibbs Entropy

I have argued that a necessary condition on intertheoretic reduction is that the terms and ontology of the reduced theory must supervene on those of the reducing theory, and the Boltzmann entropy fails this requirement. What we have to work with in statistical mechanics is a probability distribution over microstates. However this probability distribution is interpreted, I think that the entropy must supervene on it. The Gibbsian conception of entropy does supervene on the probability distribution. In this subsection, I will present and discuss both the fine-grained and coarse-grained versions of the Gibbs entropy.

The fine-grained Gibbs entropy is defined as

$$S_{fg} = -k\int \rho(x)\ln(\rho(x))d\Gamma \tag{2.5}$$

where the function $\rho(x)$ is a probability density function over the phase points x of the phase space Γ of dimension $2fn$, where n is the number of particles and f is the degrees of freedom of each particle. The probability density function is usually interpreted as an objective distribution of microstates of an infinite ensemble of macroscopically similar systems to the one under consideration.

Return to Albert's concern that the Gibbs entropy commits a category mistake because it confuses the properties of a collection of systems with those of an individual system, and the entropy is a property of individual systems. Albert's worry is a good one, though perhaps not damning. One could conceive of thermodynamics as being a theory not about individual systems but, like its statistical mechanical counterpart, as a theory about ensembles of systems. This would evade the charge of the proposed reduction being a category mistake. Nonetheless, it is evident that we *do* apply the laws of thermodynamics to individual systems and, furthermore, it is the behaviour of these individual systems (in accordance with the laws of thermodynamics) that we are ultimately seeking to explain.[37] Also, the fine-grained entropy does not address the problem posed by Liouville's theorem, which demonstrates that the measure of the probability distribution on the phase space (and hence the entropy) is invariant as it evolves dynamically in time. This is a serious problem as it is seemingly evident that the entropy of thermodynamic systems *does change*, and was apparent to Gibbs (1902).

An obvious way to avoid the issue is to deny that the probability distribution always evolves dynamically. Thus, Gibbs suggested the method of coarse-graining the phase space into small but finite regions, usually justified as objective by appealing to our

[37] A general worry on this point is that the notion of equilibrium on the Gibbsian interpretation is that of a stationary probability distribution whose thermodynamic observables never change. But individual thermodynamic systems fluctuate in and out of equilibrium all the time!

epistemic limitations or the resolution of measuring instruments. The coarse-grained probability density function is defined as

$$\rho_{cg}(x) = (1/V_{cg})\int \rho(x)\,d\Gamma \tag{2.6}$$

where V_{cg} is the phase space volume of the coarse-grained cell and the integration is performed over the whole cell containing the point x. The coarse-grained Gibbs entropy is then defined as

$$S_{cg} = -k\int \rho_{cg}(x)\ln(\rho_{cg}(x))\,d\Gamma \tag{2.7}$$

Unlike the fine-grained Gibbs entropy, the value of the coarse-grained version can change over time, since the probability distribution does not always evolve dynamically, but changes as the result of the coarse-graining procedure.[38]

Both Ridderbos and Readhead (1998) and Ridderbos (2002) argue that the coarse-grained Gibbs entropy cannot be the appropriate surrogate for the thermodynamic entropy through consideration of the spin-echo experiment. Briefly described, the experiment is conducted on a set of nuclear spins that are aligned via an external magnetic field (say in the x direction). A radio pulse is then applied to the spins in order to tilt them perpendicular to the magnetic field in the yz plane (a low S_{cg} state), where they begin to precess about the direction of the magnetic field. As the spins' precessions decay at different rates due to imperfections in the magnetic field, they become unaligned (a high S_{cg} state). Then a second radio pulse is applied, flipping the spins in the yz plane

[38] Exactly how the entropy changes will depend on how the coarse-graining is implemented (Lavis 2004).

and the decay is reversed so that the spins evolve *back into an aligned state* (a low S_{cg} state).[39]

Ridderbos and Redhead (1998) press on the fact that the apparent decrease in entropy by an isolated system after the second pulse is *prima facie* in violation of the 2nd law, interpreted as the claim that the entropy of an isolated system never decreases. Their conclusion is that if the second law is not violated in the spin-echo experiment, then the fine-grained Gibbs entropy must be the appropriate reducing concept, because the entropy does not decrease at any point in the experiment, but remains constant.[40]

Ridderbos (2002) offers a related argument against the coarse-grained Gibbs entropy. She claims that after the second pulse, the fine and coarse-grained probability distributions will behave very differently. If one uses the coarse-grained probability distribution, which ignores the correlations between the precession rates and the magnetic field, one should not expect the spins to realign. Conversely, the fine-grained entropy, which retains the 'information' about the correlations, *does* predict that the spins will realign.

Two points emerge from this analysis. First, the fine and coarse-grained probability distributions are empirically distinguishable and second, we can explain why the spins realigned using the fine-grained distribution but we cannot using the coarse-

[39] A detailed physical analysis can be found in Ridderbos and Redhead (1998). It is worth noting that the Boltzmann entropy will behave similarly (Lavis 2004) and be subject to the same objections levelled against the coarse-grained Gibbs entropy.
[40] The appropriateness of S_{fg} is defended against the charge that it fails to account for the thermodynamic increase in entropy by appealing to the intervention of the environment for everyday thermodynamic systems. They claim that the spin-echo experiment furnishes a rare example where the system is effectively isolated from the environment. Whether or not this justification works, S_{fg} still succumbs to Albert's worry that the Gibbsian approach commits a category mistake.

grained one.[41] This second point reinforces the conclusion of the three-chambered box example from Section 2.2.1 in that the entropy, if it is to figure in lawful generalisations, must be tied to whatever explains the observable phenomena at the statistical mechanical level. But this is not the case for the coarse-grained Gibbs entropy. In addition, it succumbs to the worry expressed in Section 2.2.2 that the coarse-graining (however it is to be justified) fails to match up with the reference class relative to which we have a probability distribution that explains the spin-echo behaviour.

2.2.4. Could Entropy Be a Measure of Ignorance?

It appears that neither the Gibbs nor Boltzmann entropy can do the job of reduction as long as one maintains the following desiderata for a reducing concept of entropy:

1. It should be a property of *individual* thermodynamic systems, even if it is relative to a description.
2. It should help to e*xplain why* thermodynamic systems (should be expected to) obey the (statistically certain) laws of thermodynamics by virtue of it being a function of a probability distribution over microstates.
3. In cases where the laws of thermodynamics are *prima facie* violated (as in the spin-echo experiment), it should be able to account for and explain the anomalous behaviour on its own terms.
4. It should make contact with the concept of entropy as it is understood in thermodynamics.

Both S_{fg} and S_{cg} fail to satisfy 1, with S_{fg} further failing 4 and S_{cg} failing 3. S_B does not satisfy conditions 2 and 3. The fourth point expresses the requirement that however entropy is to be understood in statistical mechanics, it should be clearly related to the

[41] Lavis (2004) denies the force of the first point, since he does not take the coarse-graining to be justified on the basis of our epistemic limitations.

explanatory role entropy serves in the description of ordinary thermodynamic processes. In extraordinary cases like the spin-echo experiment, the force of this point can be made intelligible by citing Sklar's discussion of the spin-echo experiment:

> But doesn't the [spin-echo] experiment convince us that even when the fine-grained entropy is conserved, we should still expect an asymmetric increase in the kind of entropy familiar to us from thermodynamics? To be sure, the information about the original order of the system in question can't vanish from the system as a whole without something like outside intervention to allow it to dissipate into the outside environment. But what the spin-echo experiment shows us is that there is loss of information to be accounted for even when the full information has spread itself out into correlations among the micro-components of the system without truly disappearing altogether. (1993, 253)

So what notion of entropy can do the job? In this subsection, I want to suggest that conceiving of entropy as a function of an epistemic probability distribution can satisfy these requirements.

If this is right, one can diagnose the problem with the Boltzmann and Gibbs interpretations of entropy as lying in the conception of thermodynamics as being a proper physical theory. In contradistinction to this admittedly rather intuitive view, Jaynes (1983) offers an alternative way to think about thermodynamics and statistical mechanics: one that sees the theories as being an application of certain rules of inference about the values of thermodynamic observables based on incomplete information regarding the exact microstate of the system; that is, statistical mechanics is not a physical theory proper, but a theory of inference to be applied to certain kinds of physical systems. The laws of thermodynamics are then best conceived as a set of inference tickets, and the intertheoretic reduction of thermodynamics to statistical mechanics is best effected by connecting one's epistemic probabilities regarding the microstate of the system to how one ought to expect the system to behave as it evolves over time. Addressing the first requirement laid out above, it is clear that an epistemic probability distribution over the microstates of a physical system is a probability distribution intended to describe the one's knowledge of or information about the system itself, not an ensemble of similar systems. In this sense it avoids one of the conceptual problems associated with the Gibbsian approaches to entropy.

In Jaynes' approach to statistical mechanics, an epistemic probability distribution over microstates at some time t is generated by assigning equal probability (on the standard measure) to all microstates compatible with the knowledge one has about the system t. The entropy is calculated as in (2.5), though it is to be understood in an information-theoretic sense; that is, as a measure of one's uncertainty regarding the exact microstate of the system. This knowledge is codified by the constraints operative on the system, both in the form of the known synchronic values of thermodynamic observables at t and by dynamically advancing the probability distribution generated by the values of the system's known thermodynamic observables at times other than t (Jaynes 1983, 1985). Dynamically advancing this probability distribution and calculating the expected value of the observables at t_x can then generate predictions as to the values of thermodynamic observables at times other than t. Conceived as an approach to making inferences about the sort of physical systems thermodynamics describes, Jaynes' interpretation of statistical mechanics is rather plausible, though it may fall short of meeting the explanatory goals of those who insist on conceiving of the probability distribution as a property of physical systems.[42] But the present concern is to describe

[42] It should be noted that insofar as statistical mechanics is to be thought of as a theory of inference, it should not be thought to *explain* why a particular thermodynamic system evolves in the way that it does. Rather, one should read the results of statistical mechanical predictions as the best inferences one can make as to the expected state of the system, given the knowledge at hand.

how this scheme can satisfy the *desiderata* spelled out above in a clear and straightforward way.

Statistical mechanics, interpreted as a theory of inference, provides a looser framework for assigning entropy values to thermodynamic systems because it is defined relative to one's epistemic situation, rather than being interpreted as a property of physical systems themselves or of their objective probability distributions. It is the flexibility afforded by such a conception that allows it to satisfy the latter three *desiderata*, while retaining a clear and straightforward reduction of entropy to its statistical mechanical counterpart.

The essential feature of Jaynes' approach is that one conceives of the entropy as being a measure of one's *actual* uncertainty regarding the exact microstate of the system. In this respect, one's actual uncertainty may be greater than the general scheme indicates. In the case of the expansion of an ideal gas described in Section 2.2.1, it is clear that one can ascertain the values of the relevant thermodynamic observables at equilibrium, but one is hardly able to follow the dynamical evolution of the probability distribution as the gas expands to fill the box. In some sense, the best one can do is to offer vague descriptions (as in Section 2.2.1) to the effect that the vast majority of microstates contained in the original probability distribution evolve to the new equilibrium state, and furthermore that the distribution more or less evenly samples the accessible phase space specified by the new values of the equilibrium observables. In this case, one's actual uncertainty regarding the exact microstate of the system is virtually defined by these values: knowing that the system evolved from a "lower" entropy state does not, in a practical sense, help to delimit the possible microstates the expanded gas might be in at t_1. As such, it is perfectly appropriate to say that the entropy of the system has increased upon expansion: the actual microstate of the system is more uncertain than it previously was. Further, given the assumption that at t_1 the dynamically advanced probability distribution randomly samples the macrostates accessible microstates, we can, for the purposes of making further inferences (say about the state of the gas at t_2), justifiably neglect the past known history of the system. In this way, one can speak of the entropy of ordinary thermodynamic systems increasing.[43] Here the proposal satisfies *desiderata* 2 and 4.

What is to be done if a system reproducibly behaves contrary to expectations? This is an indication that the wrong reference class was chosen to characterise the probability distribution, or that closer attention must be paid to the details of the dynamical evolution; that is, the incorrect inference was the result of insufficient information about the constraints operative on the system or the result of ignoring the dynamical implications of already known constraints. In the former case, one can readily admit that the (incorrect) inferences were the best ones that could be made given the information at hand, but additional information was needed to correctly constrain the probability distribution so as to generate a correct set of thermodynamic relations for the system. Unlike S_B, there is no worry that the entropy is calculated on the basis of a reference class other than the one used to generate the probabilities.

An example of the latter case is furnished by the spin-echo experiment. As per the passage by Sklar cited above, there is a perfectly ordinary sense in which the entropy of the spin systems does increase before the second pulse is applied, as described for the

[43] One might also argue that the effect of performing measurements on thermodynamic systems typically will "randomise" the probability distribution, in the sense that even if one could follow the exact evolution of the probability distribution, the actual microstate of the system would be just as uncertain as if nothing were known of the previous macrostate.

case of the ideal gas. Before the second pulse, the entropy makes contact with the usual thermodynamic concept when the probability distribution is generated solely on the basis of synchronic values of the thermodynamic observables. Still, someone who applied such a distribution would be utterly mystified by the observation that the entropy seems to spontaneously decrease after the application of the second pulse. But when the information about the spin system's history is included and one attends to the dynamical details, a simple and correct description (paralleling the S_{fg} description) results, thus explaining why one would have been wrong to ignore these details. Of course, one should understand in applying statistical mechanics that not all physical systems will be like the ideal gas, where various details of the dynamical evolution and of the past states of the system can be liberally ignored when inferring the system's expected evolution. The greatest advantage to understanding entropy in this way is that it can make use of this additional information to make more rigorous inferences in a conceptually coherent way. Unlike S_{cg}, we are not faced with the problem of picking the *right* reference class and yet having a prescribed method for generating the probability distribution that fails to characterise the phenomena. In this way, the information-theoretic interpretation of entropy satisfies requirements 2, 3 and 4.

I would argue that this approach is foisted upon us by the inadequacies of the standard reductive accounts of entropy. The central thesis of this section has been a criticism of the claim that the Boltzmann conception of entropy can successfully reduce the thermodynamic entropy to a statistical mechanical surrogate, and the same point has been made for the Gibbs entropy, whether in its fine-grained or coarse-grained incarnation. I have argued that the reduction of the thermodynamic entropy according to the Boltzmann interpretation fails to explain the success of thermodynamics, or results in an unacceptable ambiguity in statistical mechanics that precludes the possibility of a successful reduction. In the case of the Gibbs entropies, it is difficult to understand the entropy as a property of the individual systems that the laws of thermodynamics are taken to describe. Further, Gibbs' fine-grained incarnation of entropy fails to make contact with the laws of thermodynamics due to its constancy upon dynamical evolution, while the coarse-grained entropy and Boltzmann entropy fail to explain apparent anti-thermodynamic phenomena.

As long as one insists that the probabilities appearing in statistical mechanics are objective *and* that the entropy is an objective property of thermodynamic systems, it appears that any attempt (thus far) to reduce the thermodynamic entropy to some statistical mechanical surrogate will leave many thermodynamic processes unexplained, and can generate cases where the laws of thermodynamics are violated. The solution offered here is to re-characterise the entropy as a measure of one's ignorance and to understand statistical mechanics and thermodynamics as theories of inference.

2.3 Non-Equilibrium Considerations

In these important senses, the Jaynesian notion of entropy provides substantial advantages to more traditional definitions. Beyond agreement with the Gibbs formulation of the probability distribution for statistical mechanical systems, Jaynes (1965) also offers an "almost unbelievably short" derivation of the second law of thermodynamics. He considers a thermodynamic system in equilibrium at t=0, where the only known constraints are those associated with the equilibrium ensemble. As shown above, the thermodynamic (S_e) and information-theoretic (kS_I) will be equal. We then allow the system to undergo an adiabatic change. Since the probability distribution obeys

Liouville's theorem, the information-theoretic entropy S_I remains unchanged throughout the evolution (provided that one can follow the evolution of the probability distribution).

At some later time *t'*, the system will find itself in equilibrium once more, with a new thermodynamic entropy S_e'. However, the inequality (2.2) still holds, and we see that

$$kS_I \leq S_e'$$

and since $kS_I = S_e$, we have

$$S_e \leq S_e'$$

which states that the experimental entropy cannot decrease, a rough statement of the second law.[44]

A few remarks are in order regarding this proof. First, there is clearly no requirement that the time *t'* be later than *t*. *t'* could have been chosen earlier than *t* and the proof would have gone through just as well. Of course, the formulation of the second law need not make explicit reference to time, referring only to cyclical processes where thermodynamic state variables are allowed to change. Nonetheless, this proof fails to explain why the entropy of thermodynamic systems always increases towards the future. Jaynes (1978) argues that the reason for this is that statistical mechanics only seeks to characterise *experimentally reproducible* phenomena, and phenomena where the entropy increases towards the past temporal direction are not reproducible in this way. I will postpone the discussion of this notion to the next chapter.

[44] This proof can be given intuitive content by thinking of the probability distribution on the energy hypersurface compatible with the initial state *e*. Under the adiabatic change described, the phase volume of the distribution remains constant as required by Liouville's theorem. If this distribution is "complete" in the sense that there are no controllable macroscopic constraints not taken into account that will induce anti-thermodynamic behaviour (as in the spin echo experiments), then the distribution defined solely by the final equilibrium state (*e'*) of the system must contain almost all the trajectories from the initial distribution.

Another point worth noting about this proof is that it fails to demonstrate that the thermodynamic entropy increases monotonically in time, only that the entropy at times after t=0 must be greater or equal to the entropy at t=0 (Sklar 1993). However, it is not clear that this is even a desirable or necessary property of a successful reduction of the second law. Callender (1999) argues that, given the Poincaré recurrence theorem of classical mechanics, it is guaranteed that the thermodynamic entropy of closed systems will, in a large but finite time, decrease and come arbitrarily close to its initial value. If the statistical mechanical entropy (whichever interpretation is employed) demonstrated monotonic increase over time, it could never properly characterise the behaviour of thermodynamic systems.

Still, while Callender's argument holds some sway, all it really demonstrates is that the entropy cannot monotonically increase for *all* times. What we are interested in here, however, is whether or not it is possible to derive a monotonic increase of the entropy for more local thermodynamic changes, namely a process where a thermodynamic system moves from one equilibrium state to another one of higher entropy. As Sklar (1993) complains, Jaynes' proof fails to provide any details of the dynamics of the approach to equilibrium, such as the relaxation times and transport coefficients for thermodynamic processes. Nonetheless, both these objections miss the mark: this proof of the second law is not intended to give a detailed description of the dynamics of irreversible thermodynamic systems, but to demonstrate the MEP's ability to reduce the second law to its statistical mechanical basis. Of course, this is but one aim of statistical mechanics, and we would like to get more out of the theory than just this demonstration. From a foundational perspective, we would like to show that the Jaynesian approach is sufficient in principle to predict the dynamical features of thermodynamic systems and account for the monotonic increase in the entropy of irreversible processes. But in order to answer these questions, it is necessary to turn to a more general formalism that is capable of treating non-equilibrium situations. For irreversible processes, a more general notion of entropy is required since the usual entropy is only defined for equilibrium states in thermodynamics.

Unlike equilibrium, descriptions of irreversible processes lack temporal and spatial uniformity; that is, we expect irreversible phenomena to be time-dependent and demonstrate spatial variation in their description. Luckily, however, the principle of maximum entropy can easily accommodate these complexities. As before, the description of a macrostate is thought of as a set of constraints on the system, though in this case such constraints can demonstrate temporal and spatial variation. In general, we have in our possession a set of temporally and spatially localised measurements of thermodynamic observables [$X_1(x,t)$, $X_2(x,t)$,…$X_n(x,t)$]. As before, the method of Lagrange multipliers furnishes a probability distribution

$$\rho = \frac{1}{Z}\exp\left[\sum_{i=1}^{n}\int_{R_i} -\lambda_i(x,t)X_i(x,t)d^3xdt\right] \qquad (2.8)$$

where R_i refers to the space-time region associated with the measurement. The partition function is

$$Z(\lambda_1 \ldots \lambda_n) = \sum_{i=1}^{n}\exp\left[\sum_{i=1}^{n}\int_{R_i} -\lambda_i(x,t)X_i(x,t)d^3xdt\right] \qquad (2.9)$$

with the Lagrange multipliers determined by

$$\langle X_i(x,t)\rangle = -\frac{\partial}{\partial\lambda_i(x,t)}\log Z \qquad \text{for x, t in } R_i. \qquad (2.10)$$

Intuitively, this formalism codifies the way in which information constricts the probability distribution. Each datum is incorporated by cropping any microstates that are incompatible with the data out of the probability distribution and renormalising it. Jaynes (1985) refers to this distribution as a "calibre" because it measures the cross-section of a tube of world lines in phase space that are compatible with the information. The information-theoretic entropy thus depends on the entire *known* macroscopic history of the process.[45]

Predictions of the values of macroscopic variables of the system can then be determined by dynamically advancing the probability distribution according to the equations of motion and evaluating the phase averages of the variables at any place and time. Of course to do so is a highly nontrivial task, but foundationally the concept is clear. One can then reproduce the thermodynamic entropy values for equilibrium situations and extend it to non-equilibrium situations. This can be done by generating a probability distribution where the only constraints on the system are the predicted or known synchronic values of thermodynamic observables.

In sum, the MEP method provides a clear account of how to describe both equilibrium and non-equilibrium processes, as well as furnishes a satisfactory reductive relation between the thermodynamic entropy and the formalism of statistical mechanics. Despite these advantages, the MEP method has been subjected to intense criticism, and exploring some of these arguments is the purpose of the next chapter.

---

[45] This effectively formalises Albert's (2000) proposal for generating the probability distribution associated with a given thermodynamic system, except the low entropy state of the universe shortly after the big bang is not included among the data. As argued in Ch. 1, how one would incorporate the "past hypothesis" into the probability distribution is at best unclear. Without some means of calculating how the past hypothesis (or any other constraint) constrains the present probability distribution, such information is useless. This stands in contrast to what might be termed an objective account of the probabilities where the distribution is constrained by our knowledge irrespective of whether or not we know how to implement such knowledge in attributing a probability distribution to a system.

Chapter 3: Criticisms and Problems with Epistemic Approaches

The formalism of statistical mechanics presented in the previous chapter captures the essence of how one can update a probability distribution on the basis of new information and supplies a rule for generating predictions of macroscopic state variables based on the dynamical evolution of a probability distribution. In this way, Sklar's worries that the MEP cannot describe non-equilibrium processes are answered. On the one hand, the Maximum Entropy method provides an elegant and clear approach to dealing with thermodynamic phenomena. Conversely, however, Jaynes' approach requires one to interpret the probabilities involved in statistical mechanics as epistemic, a view that many philosophers find objectionable since they often look to ground the understanding of thermodynamic processes in *real*, *objective* properties of statistical mechanical systems (e.g. Albert 2000, Loewer 2001, Goldstein 2001).

3.1 Interpretations of Probability in Statistical Mechanics

In this section, I will not endeavour to settle deep and difficult issues in the philosophy of probability. Rather, I will discuss the applicability of various popular approaches to the interpretation of probability that are of particular importance from the perspective of statistical mechanics, assessing the relative strengths and weaknesses of different interpretations of probability in their capacity to provide consistent and cogent explanations of thermodynamic phenomena. In particular, I look to advance considerations that demonstrate (or at least render plausible) that 'objective' conceptions of probability fail in important respects to capture or explain, in physical terms, thermodynamic phenomena.[46] I then turn to defending the use of epistemic probabilities against arguments designed to demonstrate the insufficiency or conceptual inadequacy of such a view applied to statistical mechanics. Finally, I suggest in a preliminary way how some of the putatively negative features of Jaynes' approach can be of considerable use in solving some foundational issues in statistical mechanics.

The broad debate between those who maintain objective and epistemic views of probability in statistical mechanics reflects a strong difference in intuition. On the objective side, there is a sense that the probabilities *must* be an objective matter for there must be some fact (or collection of facts) that explains the regularity of thermodynamic phenomena. Otherwise, how could such phenomena be accounted for unless they supervene on the probabilistic facts of statistical mechanical distributions? Merely appealing to one's ignorance cannot *make it the case* that systems are statistically distributed as they actually are, and cannot have any explanatory import as to why the phenomena occur in the way that they do.

On the other side, there is a strong intuition that the probabilities could be *nothing but* measures of one's ignorance. After all, there is but one microstate that fully and completely describes any physical system, and the laws that determine the trajectory of that microstate are completely deterministic. Thus a Laplacian demon, given the microstate and the laws of motion, could determine the whole past and future evolution of the system where no reference is ever made to probability. Indeed, the only room for probability seems to stem from our own inability to follow the trajectory of individual phase points and our epistemic limitations in determining the exact microstate of a

[46] The qualifier "in physical terms" is important. This is because it is the supposed inability of epistemic probabilities to provide a physical basis for explaining thermodynamic phenomena that supplies the primary objection to interpreting the probabilities in this way.

physical system. Alternatively put, the appearance of probabilities seems to be the result only of our own limited capacities, and are in no way related to the actual properties of physical systems.

The strategy here is to debunk the former intuition, namely that construing the probabilities as objective potentially gives us an explanation of why statistical mechanics works. I argue that the various major proposals for interpreting the probabilities of statistical mechanics either fail to provide a coherent account of those probabilities as they relate to physical situations, or even in cases where they can, cannot provide a justification for using the probability distribution that they do. In other words, while there is widespread agreement that the uniform, equiprobable distribution is the right one to use, there is no objective rationale for using it (beyond an application of the principle of indifference). Here I briefly consider four different objective interpretations of probability: finite frequentism, hypothetical or relative limit frequentism, propensity accounts and Loewer's best systems account.

3.1.1 Objective Interpretations of Probability

(i) Finite frequentism: Finite frequentism, where the probability of an event is identified with the relative frequency of the event over a set of independent actual trials, is fraught with difficulties (see Hajek 1999). Nonetheless, the view still has some supporters as an interpretation of probability in statistical mechanics. For instance, Albert writes

> It seems to me to spare one an enormous amount of confusion, to be thinking of probabilities as *supervening* in one way or another on the *non*-probabilistic *facts of the world*, to be thinking of them (that is) as having something or other to do, by definition, with *actual frequencies*. (2000, 81)

Recall that what we are looking for is some construal of probability that has the capacity to ground our understanding of statistical mechanics and furthermore can be used to justify the microcanonical distribution that one uses in statistical mechanics.[47] With regard to this first issue, a finite frequency view fails completely to capture the essential features of statistical mechanics. Since the proportions are finite, the probabilities can only exist as rational numbers. But as statistical mechanics describes a continuous probability distribution, it is hard to see how any finite frequentist view could accommodate this fact. Coarse graining the event space is surely of no help here because the trajectories of most phase points are highly unstable with regard to their macroscopic evolution: coarse graining obscures the very features that the probability distribution looks to describe. In addition, the finiteness of the event space implies that many sets of microstates (those not of measure zero) will fail to be realised, and thus will be attributed zero probability, though the equiprobable distribution would assign non-zero probability.

Further problems arise when one considers how the event space is defined. What grounds the fact that the probability distribution of, say, an ice cube in a glass of warm water is the equiprobable one according to the standard measure? Since the goal of this endeavour is to ground a *general* rule for the assignment of probabilities to statistical mechanical systems, let us take the event space to be all those statistical mechanical systems that have existed, or ever will exist.[48] What do the frequencies of *those* microstates have to do with the probability distribution associated with this very glass of water? More worrisome is the fact that many statistical mechanical systems are

[47] The microcanonical ensemble is the Gibbsian ensemble appropriate for isolated systems, where the energy of the system is conserved.
[48] It would seem folly to think that we could just look at the frequency of microstates among glasses of water *just like* the one we are considering at present. Surely very few such glasses have existed, or will exist, and there remains the difficulty of spelling out exactly what one means by “just like”.

subsystems of each other, or are correlated in some way. Even though, say, two systems individually might have equiprobable distributions, the composite system might not have such a distribution. Given each of these reasons, it seems appropriate to question the existence of a privileged set of events that could serve to ground the probability.

Setting aside these metaphysical worries and assuming there were some way of defining the probabilities in such a way that they supervened on the non-probabilistic facts of the actual world, there would still be the epistemological problem of justifying the equiprobable distribution as the one that corresponds to the actual frequency. Surely there is no evidence, on the books at least, that any given physical system has ever exhibited extreme anti-thermodynamic behaviour, though one believes that such a chance exists. Based on our limited experience, we have no reason to believe that all microstates are equally probable as long as we identify the probability with actual frequencies. On this point Albert argues that the equiprobable distribution "*does* seem to be some sort of a *fact* – or at any rate it seems to yield correct *predictions* to *suppose* that it is some sort of a fact … And so the sort of fact that it is must be an *empirical* one, a *contingent* one, a *scientific* one." (2000, 65)

This doesn't seem to get us out of the bind. Just because *assuming* the *actual* frequency of microstates to be associated with an equiprobable distribution over the points on the phase space compatible with the macrostate gives us the right predictions does not demonstrate that this is the unique, true distribution, or that it is warranted as a choice over other possible distributions. Further, it seems impossible, on Albert's definition of probability, to give meaning to his past hypothesis, since as we are given the world but once there is no actual frequency of microstates of the initial state of the universe.[49]

(ii) Hypothetical or Ensemble frequency: The hypothetical frequency or ensemble view of probability identifies probability with the frequency of events relative to an infinite number of independent, counterfactual trials. While this view eschews many of the standard problems of the finite frequentist account of probability, it is not clear exactly what can ground the frequencies of such counterfactual events. Given a box of gas in some macrostate, what could make it the case that the relative frequency of counterfactual microstates is the equiprobable distribution, other than an application of the principle of indifference? Unlike quantum mechanics, where the frequencies are grounded by the Born rule and thus by the wavefunction, classical statistical mechanical systems are thought to be deterministic and each phase point lies on a unique trajectory. There seem to be no facts about the fundamental ontology of classical mechanics that could generate such frequencies. And even if one could find something to ground counterfactual frequencies in statistical mechanics, why should we take these frequencies as a guide to what will happen in the actual world (Loewer 2001)?

While this interpretation of probability can be employed in many ways to generate an explanation of how (or why) statistical mechanics works, it is most often used in the context of ergodic theory, where the phase averages of the probabilities are equated to the infinite time averages. A more detailed critique and exploration of ergodic theory will be presented in section 3.3.

---

[49] Of course, this will only be a problem for Albert or anyone else who simultaneously holds an actual frequentist view of probability yet wants to appeal to the probability distribution of the initial universe.

(iii) Propensity or Dispositional interpretation: Propensity accounts of probability associate probabilities with dispositional properties that measure a system's tendency to be in one or another microstate. This interpretation seems to be of little or no use in the case of statistical mechanics: if these dispositional properties are reducible to other, more fundamental properties (as the dispositional property of fragility is putatively reducible to the molecular structure of the bearer of the property), then it is hard to see how the probability could be reducible to anything other than one's ignorance of the actual microstate of the system (Clark 2001). Conversely, if the probabilities are thought to be tychistic, then we are required to deny either the fact that a statistical mechanical system has a unique microstate or that the trajectories are deterministic; that is, we deny the classical nature of the system. But giving up these fundamental aspects of the classical world in order to uphold an interpretation of probability that leaves the nature of probability metaphysically mysterious does not seem like a viable move (van Lith 2001).

(iv) Best Systems/Theoretical interpretation: An alternative account of probability is offered by treating probability as a theoretical term defined in the context of a theory, including an approach recently put forward by Loewer (2001) that draws on Lewis' best systems account of laws. In this approach, probability terms in the theory acquire meaning by virtue of the role they play in a theory; that is, through their interrelationships with the other theoretical terms in a given theory. Here I will only consider Loewer's proposal, as I take it to be the most sophisticated approach of this sort.

Loewer argues that even though frequentist and dispositional accounts of probability fail, one can still salvage the notion of objective probability by *stipulating* that the equiprobable probability distribution over the initial macrostate of the universe is a *law* in Lewis' (1994) sense, in that it is part of our strongest, most informative and simple systematisation of the totality of facts about the world. As a law the probabilities obtain objective status and, Loewer argues, these probabilities over the initial state trickle down, as it were, to individual subsystems of the universe so as to imbue ordinary statistical mechanical systems with meaningful and objective probability assignments.

Putting aside objections to Lewis' best systems account of laws[50] and accepting that the probabilities do in fact "trickle" down to individual statistical mechanical systems, I still see two problems facing Loewer's attempt at justifying the objectivity of the probabilities in statistical mechanics. First, even if we accept Lewis' view of laws, it isn't clear that the equiprobable distribution over the initial condition of the universe is one of those laws. While Albert (2000) argues for the lawlike status of this initial condition, I have presented arguments in Chapter 1 that cast doubt on the lawfulness of this proposition by questioning its explanatory scope (i.e. strength).[51] If the uniform distribution over the initial state loses its place as part of the best systemisation of the world, then Loewer loses his rationale for considering the probability objective.

Second, the nature of the probabilities that Loewer endorses is completely obscure. In other interpretations of probability we can identify what the probabilities *are*, whether they are dispositions, actual frequencies, ensemble frequencies or measures of degrees of belief. Loewer's argument, by contrast, fails to provide a positive account of to what the concept of probability in statistical mechanics refers. Indeed, Loewer carefully considers, but ultimately rejects, the various objective interpretations of probability presented above before suggesting his 'best systems' account. Yet this account seems different in kind from the others in that it serves as an umbrella argument for the

---

[50] For criticisms see, for example, Armstrong (1983) and van Fraassen (1989).
[51] Note that it is the lawfulness of the *probability distribution* of the initial state of the universe that is in question, not the initial state of the universe itself.

objectivity of probability rather than constituting a viable interpretation in its own right. In fact, Lewis (1994) views his argument for the objectivity of probability as remaining neutral between various interpretations, favouring whichever one works best in the context of the theory in which it appears. But Loewer argues (and justly in my opinion) that *none* of these interpretations can adequately characterise the notion of probability as it appears in statistical mechanics! In this sense, it is not clear what Loewer is proposing.

Alternatively, one could interpret these probabilities (as suggested above) as acquiring meaning through their interrelationships with other theoretical terms within the theory; that is, as a throwback to post-positivist attempts at imbuing the theoretical terms with meaning. Such 'meaning holism' strikes me as a questionable way to approach the explanation of statistical mechanical phenomena, since it is often the case that the interpretation of probability one employs in statistical mechanics will determine what role the probabilities play in the theory. If an ensemble or hypothetical frequency approach is taken, then the probabilities are usually employed via ergodic theory through the use of phase averages over the probability distribution. Or in the case of Jaynes' approach, the probabilities are measures of one's degree of ignorance and not tied down to the physical system in question. Insofar as all these approaches work (to some extent) or at least appear as viable interpretations, it's hard to see how the probabilities could acquire meaning within the context of the theory since their role cannot be specified without their interpretation.

3.1.2 Criticisms of Epistemic Interpretations

In published criticisms of Jaynes' work on statistical mechanics, and more generally of those approaches that attempt to use some flavour of the principle of indifference, one finds roughly three attacks aimed at the cogency of viewing the probabilities as degrees of belief or as measures of ignorance, namely:

1. The principle of indifference cannot be applied *a priori* without a justification of the measure used, and is therefore at best arbitrary, or at worst incoherent.
2. One's *ignorance* of the exact microstate of a physical system cannot (causally) account for or explain the regularity and apparent lawlikeness of thermodynamic phenomena.
3. There are possible worlds where probability assignments generated according to the principle of indifference will fail to reproduce or correctly characterise thermodynamic processes (such as a world where the global entropy curve is decreasing).

The first criticism of Jaynes' view has a long and storied history, generally falling under the rubric of Bertrand's paradox. Roughly, the paradox rests on the fact that the partitioning or parameterisation of an event space can be accomplished in any number of ways. Thus, the rule will assign equal probabilities to each elementary event in the event space, though different parameterisations of the space will generally lead to vastly different probability assignments for a single trial. For instance (in a simple finite case), a roll of 'six' on a fair die might be accorded probabilities of 1/2 or 1/6, depending on whether the event space is defined as 'six'/'not six' or by the number of faces on the die, respectively.

Things become more difficult in moving to a continuous event space. In the case of the Maximum Entropy Principle, this paradox demonstrates that the naïve extension of the entropy expression to the continuum,

$$H(p) = -\int p(x)\log p(x)dx \tag{3.1}$$

cannot define a unique value for the entropy since x can be re-parameterised in any number of different ways. The issue can be resolved by introducing a background measure, m, to the equation and relativising the entropy to this new measure:

$$H(p,m) = -\int p(x)\log\frac{p(x)}{m(x)}dx\,. \tag{3.2}$$

The entropy now becomes invariant under re-parameterisation, and thus not vulnerable to Bertrand's paradox. In the case where the background measure is the usual (but now invariant under re-parameterisation) Lebesgue measure, one recovers the naïve entropy expression in (3.1).

Of course, this move only pushes the question back to one about the apparent arbitrariness or justification of the choice of background measure. For Jaynes (1973), the justification of the measure to be used in a given experiment stems from the physical symmetries and invariances at work in the problem. Here Jaynes argues, following Jeffreys (1967), that the expected results of an experiment should not change under re-parameterisation[52] (that is, an alternate description of the experiment) and this is sufficient, for any well-posed problem, to pick out the appropriate background measure.

It should be noted that there is an alternate tradition to the problem of justifying the choice of background measures where background measure is interpreted as representing a prior probability distribution to be conditionalised upon given a new constraint (See Uffink 1996a for a detailed review and criticism of this tradition). Although this approach seems impracticable for dealing with the problem at hand in that it does not supply a rule for assigning a probability distribution to a thermodynamic state,

[52] For instance, a change in the zero point chosen for a quantity of interest.

it is this interpretation of the background measure that has attracted the most critical attention.[53]

On Jaynes' view, the symmetries that go into the choice of background measure, while they may be justified physically, still represent symmetries in one's knowledge rather than symmetries that exist in the physical world. In this sense, the application of the Maximum Entropy Principle is not entirely justified *a priori*; that is, empirical considerations can enter in one's choice of the background measure and thus empirical considerations come into play as part of the explanation of why it is that the principle works. However, there is nothing perniciously question begging about this. In the case of statistical mechanics the sorts of symmetries involved, those that generate conservation of energy and the incompressibility of the phase volume on a phase space (through Liouville's theorem), are of vital importance to describing the properties of statistical mechanical systems and the dynamical behaviour of such systems. In fact it would seem odd to require (but not without historical precedent) that, say, the conservation of energy must be established *a priori*. Why require that any feature of the dynamics needs to be established *a priori*?

Nonetheless, even if the justification of the standard measure as the background measure is accepted, it still remains obscure why it successfully captures thermodynamic phenomena. More precisely, the justification of the background measure, though based on physical symmetries, still remains a characterisation of one's knowledge and the resulting probability distribution thus does nothing to explain why the microstates of

[53] For instance, Guttmann (1999) argues that Jaynes' approach is significantly weakened by problems facing the justification of the background measure in this interpretation.

thermodynamic systems are actually distributed in the way that they are. As Albert puts it,

> Suppose that there were some unique and natural and well-defined way of expressing, by means of a distribution function, the fact that "nothing in out epistemic situation favours any particular one of the microconditions compatible with [a given macrostate] over any other particular one of them." *So what?* Can anybody seriously think that that would somehow *explain* the fact that the *actual microscopic conditions* of *actual thermodynamic systems* are *statistically distributed in the way that they are?* Can anybody seriously think that it is somehow *necessary*, that it is somehow *a priori*, that the particles that make up the material world must arrange themselves in accord with *what we know*, with *what we happen to have looked into?* Can anybody seriously think that our merely being *ignorant* of the exact microconditions of thermodynamic systems plays some part in *bringing it about*, in *making it the case*, that (say) *milk dissolves in coffee?* (2000, 64)

This argument, or something like it, is often taken to be the main or primary reason why the probabilities involved in statistical mechanics cannot be interpreted as measures of one's ignorance. Let us begin by stating a few truisms:

1. Determinism: What *makes* it the case that in some particular instance milk dissolves in a cup of coffee is the fact that this dissolved state lies on the trajectory in phase space of the actual microstate of the system.
2. Statistical Explanation: The reason that milk almost always dissolves in coffee is that almost all the initial microstates of actual instances of milk and coffee systems lie on trajectories where the milk and coffee mix (appeal to typical behaviour).
3. Induction: The fact that (almost) all past milk and coffee systems have mixed renders it reasonable to think that the next milk and coffee system will mix, but does not *make* it the case that it will mix.
4. Triviality: The fact that (almost) all actual milk and coffee systems mix makes it the case that (almost) all actual milk and coffee systems mix, but is no explanation of that fact.

An important distinction latent in these claims is that there is a mismatch between the *explanation* of the statistical regularity of thermodynamic phenomena and a *causal* account of such phenomena, seemingly conflated in the passage by Albert quoted above.

What *makes* it the case, what *causes* it to be the case that a particular milk/coffee system mixes is the fact that the system's microstate lies on a deterministic trajectory that passes through a set of phase points corresponding to a mixed coffee and milk macrostate. Now, the sense that this *does* provide a complete causal description of the fact that a given milk/coffee system mixes is what drives what Loewer (2001) calls the "paradox of deterministic probabilities". Since the system is deterministic, the probabilities can be nothing other than a measure of our ignorance.

The intuition can be put more poignantly by considering what would have to be the case if the probabilities were thought of as somehow contributing, somehow *making it* the case, that a given milk/coffee system mixes. Attributing *causal* powers to probability distributions raises the spectre of overdetermination, since the deterministic trajectory of the microstate is already sufficient for deciding the dynamical evolution of the system. Perhaps even more serious or more fundamental, one might worry how it could be that the microstates that the system is *not* in would have a causal influence on what actually happens. Surely then, one should not think that the probabilities associated with a statistical mechanical system play any causal role in the evolution of that system.

Albert's criticism of the ignorance interpretation's inability to provide a causal explanation of why any particular coffee and milk system mixes falls flat. It is clearly not true that his own interpretation of statistical mechanical probability provides a causal account of this fact. In this sense, no interpretation of probability can make any claim to explaining what *makes* it the case that a given coffee and milk system mixes and, as argued earlier in section 2.2.1, Callender's appeal to typical behaviour fails to characterise the modal and explanatory force of the probabilistic explanation.

Alternatively, one could read Albert as arguing that the *modal force* of such an explanation cannot have anything to do with what we happen to have looked into, what we happen to know, rather than seeing one's knowledge as playing a causal role in the evolution of thermodynamic systems towards equilibrium. But this is simply false: the probability distribution associated with a statistical mechanical system is clearly (at least in part) a function of our knowledge, and Albert says as much when he states that, as one of his fundamental laws, the probability distribution is to be constrained by whatever "information—either in the form of *laws* or in the form of *contingent empirical facts*—we happen to have" (2000, 96). I still see no reason to think that the probabilities need to be interpreted in an objective vein in order to explain why a given thermodynamic system behaves in the way that it does.

Even if single-case probabilities do not demand objectivity, one might still argue (or read Albert as arguing) that the *regularity* of thermodynamic processes demands an objective interpretation, that nothing but an objective interpretation of probability can explain why it is that we see the regularity that we do. Again, however, it is not clear that Albert's own account can do this as I would maintain that offering an actual frequency account of probability that identifies probability with proportion is in no way an explanation of the proportions that one actually sees. Similarly, offering an ensemble frequency or propensity account of probability provides no explanation, since it is at best unclear what could make the probability distribution take on the values it does other than an application of the principle of insufficient reason.

Still, an argument can be offered to show the *explanatory* poverty of any non-objective interpretation of probability in statistical mechanics by appealing to counterfactual situations where the uniform probability distribution over all compatible microstates *fails* to properly characterise the evolution of thermodynamic systems. Indeed, we could imagine possible worlds populated by microstates whose trajectories lead to anti-thermodynamic behaviour (Sklar 1993) or a universe with a globally decreasing entropy curve (Loewer 2001). In such cases, it is argued, the fact that the maximum entropy principle assigns probability distributions that predict macroscopic phenomena that are radically at odds with what happens (in these counterfactual situations) demonstrates its explanatory poverty; that the principle does work turns out to be a contingent fact about our world.

The source of this objection is easy to identify, for it seems that the maximum entropy principle assigns probabilities that are insensitive to these counterfactually occurring proportions. In a sense, this is true: the principle works by assigning probabilities based on given constraints, and if there are no identifiable constraints at work in generating the proportions that differ from the maximum entropy assignment, then the method will indeed make false predictions. However, this is by no means damning since it was never claimed that statistical mechanics, viewed as an application of a more general approach to *statistical inference*, need always or even usually make correct predictions. Rather, as Jaynes repeatedly claims, the method provides the best predictions possible *given the information*, and in order to make better predictions, we would need more information. For instance, Jaynes claims that "in calling a probability subjective, we mean that it is not a physical property of the system, but only a means of describing our information about the system; therefore it is meaningless to speak of verifying it empirically" (1983, 23).

Before discussing this claim more carefully, consider Loewer's worry that the maximum entropy principle would work out terribly wrong if the universe were an entropy-decreasing one. I believe there are two reasons to think that Loewer's concern is unjustified, or at least inconclusive. It is unclear why or how the fact that this counterfactual universe exhibits a global entropy decrease immediately implies that its *subsystems* demonstrate anti-thermodynamic behaviour, or behaviour not predicted by the maximum entropy principle. Even in such a universe, would it not be the case that a gas confined to one side of a box by a partition would expand to its equilibrium state once the partition was removed?[54] Is it even true that the laws of thermodynamics would no longer hold, and be 'explained' by the maximum entropy principle, in a universe with globally decreasing entropy?

The point can be rephrased by considering whether or not the maximum entropy principle could serve to describe thermodynamic processes in such a universe. In a universe whose entropy exhibited a monotonic decrease, the entropy curve would likely look much like the curve associated with retrodictions based upon the present macrostate of the universe alone, in that it would appear that the present universe evolved to its present state as a highly improbable fluctuation from an abnormal microstate. If such an inference is licensed by the maximum entropy formalism (and by most other naïve accounts of statistical mechanics), then why should we think that our *interpretation* of probability has failed us in some way?[55]

Now, even in naïve approaches to statistical mechanics (i.e. ones that can't cope with the reversibility objection) one would not be inclined to think that just because the universe has decreased in entropy in the past, this provides a reason to think that it will

[54] If Mars is treated as a blackbody in equilibrium, then it is an entropy-decreasing environment, since the temperature of incoming radiation from the Sun is higher than the temperature of the outgoing radiation. But surely a gas would mix in such an environment!

[55] Of course, in the actual universe the reversibility objection appears in the foundational literature on statistical mechanics as a problem to be solved, whereas in the universe Loewer envisions such behaviour is by no means problematic.

continue to do so in the future. In fact, it is exactly the content of the reversibility objection that the present macrostate of the universe represents a trough in its entropy curve. If this curve were the entropic path of the universe, it would seem that the maximum entropy formalism (along with others) would capture this apparently strange thermodynamic behaviour quite easily. Again, even in such a universe, one would quite reasonably expect the entropy to subsequently increase in the future.

Stretching this objection, it is possible that Loewer's objection to an epistemic understanding of probability envisions an abnormal universe where there is a continued decrease in entropy into the future. Nonetheless, it is not clear why this sort of objection should be damning. Such a world would effectively be the mirror image of our own, where the reversibility objection would loom large, just in the *opposite* temporal direction.[56] Insofar as Loewer might think this universe to be an objection to the use of epistemic probabilities assigned according to the principle of indifference, it amounts to the claim that such an interpretation cannot solve the reversibility problem, in either temporal direction. But this is not an objection to the conceptual inadequacy of epistemic interpretations of statistical mechanical probabilities in general. Rather, it is the claim that such an interpretation cannot solve a particular foundational problem. However, it is among the central contentions of this dissertation that such an interpretation *can* be of use in solving this problem, and Loewer's argument provides no reason to think otherwise.

A second and different problem with Loewer's objection is that there may be some way that the entropy-decreasing nature of the universe might be amenable to being formulated as an operative constraint under the maximum entropy formalism, in the same

[56] I don't here want to take a stand as to whether the 'direction' of time would be flipped in such a world. More specifically, I won't assume that the thermodynamic arrow of time is the fundamental or definitive one.

way that any fact about the world that is relevant to the behaviour of thermodynamic systems constrains the possible microstates of the system. But it remains obscure how this fact about the universe could be translated into a constraint that would alter the predictions or retrodictions generated by a probability distribution, in the same way that I argued that it is opaque how Albert's past hypothesis actually constrains the probability distributions of local thermodynamic systems (See Ch. 1). If it is unclear how a probability distribution over the initial condition of the universe can be practicably applied to such local, individual subsystems, then it is equally unclear how constraining the probability distribution of the universe to entropy-decreasing trajectories affects the predicted behaviour of individual thermodynamic systems, such as the gas expanding into a box after a partition has been removed.

3.2 Why an Epistemic Approach?

Having discussed the major conceptual objections to applying some version of the principle of indifference to generate statistical mechanical probabilities, I take it that none of these objections presents insurmountable difficulties. In particular, it seems possible to avoid the apparent arbitrariness of the choice of measure by considering the physical symmetries and dynamics involved in the systems we look to describe. Nor does the possibility that the probabilities assigned according to the maximum entropy formalism might differ from actual proportions seem conceptually problematic; this is exactly the same fact that we face in trying to rescue statistical mechanics from the reversibility objection. That this is a hard problem (on any view of statistical mechanical probability) doesn't appear to vitiate the MEP's *conceptual* adequacy. Finally, I have charged that critics of the principle of indifference have placed explanatory demands on statistical mechanical probability distributions that no interpretation can meet, and thus in no way constitutes an objection to the use of probability conceived of as a measure of one's ignorance.

Up to this point, I have presented considerations to the effect that there exist significant conceptual problems facing so-called objective interpretations of probability in statistical mechanics, as well as responded to charges that the maximum entropy principle, viewed as an extension of the principle of insufficient reason, is conceptually unfit to serve in that role. But little has been said in terms of what advantages there are to thinking about the probabilities of statistical mechanics in this way beyond the claim that it provides a clear and effective reductive link between statistical mechanics and thermodynamics. In essence, I believe that some of the very aspects of epistemic probabilities that many critics have denounced are exactly those aspects that provide significant advantages to solving foundational problems in statistical mechanics. Specifically, the worry that motivated two of the aforementioned objections, the problem that the probability distribution was not tied to some actual, intrinsic feature or property of thermodynamic systems (or perhaps the actual microstates of such systems), strikes me as the interpretation's greatest advantage for two reasons. First, it identifies statistical mechanics as part of a more general theory of *statistical inference*, delimited by certain features of physical (dynamical) systems, in order to generate physical predictions. Second, the fact that one doesn't think of the probability distribution as being linked to an objective property of a system indicates that the reversibility problem encountered in the foundations of statistical mechanics is not as dire as one might *prima facie* think.

With regard to the first of these claims, I have largely eschewed a detailed discussion of the thorny issue of the maximum entropy principle's status as a general theory of statistical inference. However, I have attempted to motivate the idea that we should have good confidence in the claim that the maximum entropy formalism provides a correct and non-arbitrary account of the predictive success of statistical mechanics. Furthermore, there is a certain naturalness to the idea that statistical mechanics, a theory that itself introduces no new *physical laws* or ontology, a theory that has an outstanding range of applicability across differently constituted and described systems, should be part of a more general theory about the nature of statistical inference rather than a theory grounded in the nature of reality.

Now, thinking of statistical mechanics in this way is tantamount redefining the explanatory goals of the theory, and with good reason. We have seen how Albert and Loewer, for instance, discount the idea that an epistemic interpretation of probability can account for the manifest temporal asymmetry of thermodynamic systems, and furthermore that it cannot explain the regularity or apparent lawlikeness of thermodynamic phenomena. What they fail to mention is that no objective interpretation of probability can do these things either. So what should the explanatory goals of statistical mechanics be, and how might statistical mechanics, conceived as a theory of inference, meet them?

At the very least, we are looking for a theory that (usually) generates correct predictions based on present macrostates, and it would appear as if almost any naïve approach to statistical mechanics can manage this. Yet, paradoxically, the same considerations that led to those predictions also lead to retrodictions that we take to be radically at odds with what we recall or of which we have apparent records. Even worse, it appears that these same considerations should lead us to believe that these records are themselves spurious, in that they too are more likely to have arisen out of some sort of molecular chaos than as the result of some veridical reflection of past states of affairs. In sum, the bare bones version of statistical mechanics leads to the reversibility objection, where a sceptical disaster looms in that it appears that almost everything we believe about the past is most likely false.

The problem can be posed as a dilemma: either all of our beliefs about the past are false, or they are not. In the case of the latter (which we hope is the correct option), it should be the explanatory goal of a philosophical account of statistical mechanics to construct a true theory of how it is that these beliefs are largely correct, or at least that we are justified on the basis of the theory itself (or some suitable extension of the theory) in believing that these beliefs are correct in spite of the *prima facie* considerations to the contrary. If *this* is the explanatory goal that we set for ourselves, then it is patently appropriate that we view statistical mechanics as a theory of inference regarding the sorts of systems that can be described by thermodynamics.

Indeed, we understand that, in calling probability epistemic, the physical properties of thermodynamic systems are not directly tied to, or definitionally related to, the probability distribution assigned to its statistical mechanical description. It follows from this that just because the probability distribution one might *assign* to the retrodicted macrostate of a system based solely on its present macrostate might be of higher entropy than its present macrostate, this does not in itself make it the case that the physical system was *actually* in a higher entropy state in the past. Thus, conceiving of the probabilities as degrees of ignorance offers two paths to solving the reversibility objection, both of which we will have occasion to use:

a) Retrodicted probability distributions, unlike predicted ones, are inappropriate in some circumstances for determining past macrostates.

b) Retrodicted probability distributions need to be brought into line with (i.e. further constrained by) justified beliefs about the past in order to correctly describe the past macrostates of thermodynamic systems.

In some cases we shall see that it is appropriate to block retrodictions of anti-thermodynamic behaviour because the system did not exist in the past, or was ill-defined or ill-suited to be described as a statistical mechanical system. In such cases we can speak meaningfully of the 'creation' of a statistical mechanical system; that is, of the creation of a probability distribution that accurately reflects one's justified beliefs as to the past thermodynamic state of the system. Roughly, this allows us to think of thermodynamic systems as being branch systems, though of a different sort than the branch systems described by Reichenbach (1956), Grünbaum (1963), Davies (1974) and others. The discussion of this notion of branch systems will be further expounded in Chapter 4.

But it is evident that the reversibility worry is manifest in present statistical mechanical systems in the sense that the probability distribution generated by using only the present values of thermodynamic observables as constraints will lead one to retrodict anti-thermodynamic behaviour in the past. However, if the probability distribution also depends on past constraints (i.e. past values of thermodynamic observables), then the retrodicted macrostates of thermodynamic systems avoid the reversibility worry. Of course, this merely pushes the question back to whether or not one can justifiably think that those past constraints were actually operative. In this way, the nature and reliability of records (which may themselves be thought of as thermodynamic systems) becomes a pressing issue, and will be discussed in Chapter 5.

I take it that a successful explication of these aspects of statistical mechanical inference constitutes a solution to the reversibility objection in the sense described above:

it grounds and explains our beliefs regarding the past of individual thermodynamic systems, rendering it reasonable to maintain that such systems have histories that accord with the laws of thermodynamics. To ask for anything more would be to place a heavy explanatory burden on statistical mechanics that it cannot possibly bear.

3.3 Ergodic Theory

The maximum entropy formalism provides a potential justification and explanation of the success of Gibbsian statistical mechanics. Historically, however, the conceptual justification of Gibbs' methods has travelled a different route, that of ergodic theory. In this section I will briefly present the structure of ergodic theory and how it purports to answer some key foundational questions. Ultimately, I argue that ergodic theory falls short of grounding thermodynamic phenomena in any meaningful way. In the final subsection, I consider charges that the use of the maximum entropy formalism tacitly presupposes ergodic results, a charge that Jaynes always denied.

3.3.1 The Claims of Ergodic Theory

Ergodic theory takes as its primary object an ensemble of systems similar to an actual system of interest in the sense that it comprises the set of systems whose microstates are compatible with the macrostate of the system, assigning them the probabilities associated with the Gibbsian ensemble appropriate to the system. It then attempts to identify the phase averages of the thermodynamic observables with the actual values of the thermodynamic observables associated with the system of interest, thereby linking the properties of the ensemble with the properties of the actual system. As such, ergodic theory looks to justify the use of the Gibbsian measure across ensembles as being relevant to the thermodynamic variables of individual systems.

The worry that immediately arises in this approach is to understand how the *phase* averages associated with a probability distribution are to be connected with the actual observed values of a system, since the system is actually in only one microstate. How does appealing to the properties of a collection of microstates *other* than the one that is actualised bear on the actual properties of the system?

The ergodic problem has its origins in the work of Boltzmann and Maxwell, who were able to demonstrate that a stationary, equilibrium probability distribution could be demonstrated for simple models such as an ideal gas. Here the thought was that since the equilibrium macrostate was temporally invariant, its statistical surrogate ought to be as well. In a rough way, this appears to link a stationary probability distribution with the actual properties of equilibrium systems. However, it was felt that something more was needed, namely a demonstration that the derived stationary distribution was *unique*, rather than one among many: for if there are other stationary distributions where the phase averages of thermodynamic observables do not change over time, then it would be possible for equilibrium states to exist for a given system other than the one commonly associated with thermodynamic equilibrium (Sklar 1993).

Boltzmann proposed the ergodic hypothesis, which conjectured that there exists only one phase trajectory for a system and that over time, this trajectory eventually passes through every point on the hypersurface associated with the system. If this were the case, then over infinite time the trajectory of the microstate would pass through every phase point, independent of its initial state. Since the ensemble includes all these phase points, the phase averages of the ensemble should be equal to the time averages of thermodynamic observables, thus establishing the link between the ensemble and the actual system. Furthermore, since the equilibrium macrostate represents the vast majority of the accessible phase region, the overwhelming majority of the system's life is spent at equilibrium. In such a case, because there exists only one trajectory, this stationary equilibrium distribution is unique, and the stationary nature of the values of equilibrium thermodynamic observables is explained.

Unfortunately, Boltzmann's ergodic hypothesis is clearly false. First, it is not clear that the infinite time limit of such systems exists. But more importantly, because the phase trajectory is a one-dimensional line on the phase hypersurface it is impossible, from a measure theoretic view, for it to pass through every point on the accessible phase space (the measure of a lower dimensional subset on a higher dimensional space is always zero (Sklar 1993)).[57]

Contemporary ergodic theory approaches these problems from an explicitly measure theoretic framework. Here, we consider a measure preserving dynamical system, given by a measure space $<\Omega, B, \mu>$ and a dynamical time translation operator T_t that maps sets of the space onto others while preserving their measure, as required by Liouville's theorem on the standard Lebesgue measure. For such a system, Birkhoff and Von Neumann proved that the infinite time limit for phase functions defined on Ω exists, except for possibly sets of measure zero (See Kinchin (1949) for a proof).

We call a dynamical system *metrically transitive* or *metrically indecomposable* if and only if all invariant sets on Ω have a measure of either zero or one (invariance is the property that all dynamical translations on a set map the set onto itself). If a system is

[57] There were attempts to prove a weaker conjecture that proved intractable. Rather than the strong claim that the trajectory passes through every point, one could try to prove that the trajectory passes arbitrarily *close* to every point. This is known as the quasi-ergodic hypothesis.

metrically transitive, then it is impossible to decompose an invariant set of positive measure into two or more invariant sets of positive measure. More intuitively, the condition of metrical transitivity asserts that a phase point has access to the whole of the accessible phase space, and its trajectory is not restricted to some smaller portion of it.

Given that a set is metrically transitive, one can show that the infinite time limit of an observable is equal to the phase average of that observable (if the limit exists). In such a case, it appears that ergodic theory has achieved its goal of equating phase averages with time averages, and in doing so linking the properties of the ensemble with those of individual systems.

3.3.2 Problems with Ergodic Theory

Before moving on to consider how successful the ergodic approach is when a system is ergodic, it is necessary to note a couple of formal problems facing the applicability of ergodic theory to statistical mechanical systems. First, there is the issue of dealing with sets of measure zero, whose infinite time limits may not exist. Secondly, one must prove that real statistical mechanical systems are in fact metrically transitive, which is necessary and sufficient for demonstrating ergodicity; that is, the equality of phase averages and infinite time averages.

Measure Zero: Recall that an objective of ergodic theory is to justify the use of the standard measure over the Gibbsian ensemble as the appropriate one for attributing probability distributions to statistical mechanical systems. However, it is apparent that ergodicity will fail for some sets of measure zero, in the sense that their infinite time averages may not exist. Thus, some rationale is required for discounting such sets as being negligible or unimportant for such systems.

We can identify two problems generated by such sets. First, there is the worry that *even if* one can identify sets of measure zero with events of probability zero, such probabilities are not to be identified with impossibility: the probability of obtaining an infinite series of heads on a fair coin is zero but is surely not impossible.[58] Indeed, the actual microstate of a physical system is assigned measure zero, but one would hardly want to discount the actual microstate of a system as being unimportant or negligible. So on what basis can we discount such sets as being irrelevant to the properties of statistical mechanical systems?

Even if one can come to terms with the meaning of attributions of zero probability to events, a further problem arises in establishing the link between sets of measure zero and events of probability zero. Why should one assign a measure zero set zero probability? Part of what ergodic theory seeks to establish is the appropriateness of the standard measure for assigning probabilities to statistical mechanical systems. To justify neglecting sets of measure zero *because* they are associated with events of probability zero assumes the very link that the ergodic programme seeks to establish. Such bootstrapping indicates the need to appeal to external physical considerations to demonstrate that measure zero sets can be neglected.[59]

Metrical Transitivity: A second, perhaps more substantive worry regarding the ergodic approach involves establishing the metrical transitivity of statistical mechanical systems. Indeed, the claim that the phase averages can be equated with the time averages fails if a trajectory can be restricted to less than the full accessible phase space. Although *everything* turns on this property of dynamical systems, it has proved to be incredibly

[58] How one deals with events of probability zero will depend on one's interpretation of probability.
[59] This has turned out to be a difficult problem. See Malament & Zabell (1980), Sklar (1993), Vranas (1998) and van Lith (2001) for detailed discussions of this problem.

difficult to demonstrate that any sufficiently interesting system is, in fact, metrically transitive.

Sinai claimed that metrical transitivity could be proven for the case of hard, elastic spheres in a box (roughly an ideal gas). However, the status of this 'proof' remains in question, as it was never published (Uffink 1996b). In any case, there appears to be good reason to think that the ergodic behaviour of statistical mechanical systems may be the exception rather than the rule, as indicated by KAM theory. This approach, initiated by Kolmogorov, Arnold and Moser, considers time-independent perturbations applied to quasi-periodic Hamiltonian systems, including those with more realistic assumptions such as the existence of interparticle forces. Even under such perturbations, the orbits of the trajectories remain quasi-periodic; that is, the trajectories are invariant and of positive measure, yet not metrically transitive. The apparent stability of such systems under perturbations seems to throw a wrench in the hopes that statistical mechanical systems would, in general, prove to be ergodic (Sklar 1993).

Metrical transitivity has turned out to be exceptionally difficult to demonstrate for anything approaching a sufficiently complicated system, and there appears to be good reason to think that many statistical mechanical systems are not ergodic: at best we can hope that some are. But this doesn't seem good enough for the foundational project ergodic theory looks to develop. For those systems that are not metrically indecomposable, we still lack an explanation of why the usual statistical mechanical algorithm for making predictions works. And even if a system is ergodic (in which case we would apparently have such an explanation), shouldn't we think that whatever it is that explains the success of statistical mechanics for metrically *decomposable* systems should also explain the success of statistical mechanics for genuinely ergodic systems (Earman and Redei 1996)?

The above discussion has centred on technical issues surrounding the demonstration of ergodic behaviour for statistical mechanical systems. Yet a foundational worry remains as to whether the ergodic approach has the necessary explanatory force required to justify the empirical success of statistical mechanics. Note that ergodicity attempts to link the properties of statistical mechanical ensembles with the properties of individual physical systems by equating to phase averages of the ensemble with the *infinite* time average of an individual system, thereby rationalising the use of the usual Gibbsian ensemble. However, this may not be enough, for one still needs to explain how the infinite time limit is related to the actual values of thermodynamic observables that are the results of measurement. How are the infinite time averages that the ergodic theorem refers to related to the experimentally determined values of thermodynamic observables?

A common argument proposed to solve this worry is to appeal to the fact that the characteristic time required to make a measurement of an observable is long (i.e. infinite) in comparison to the characteristic time of microscopic processes. Thus, the results of measurements are to be thought of as infinite time limits, and the use of phase averages is thereby justified. However, if this rationale were true, then it would have to be the case that, as a matter of experience, we could not measure the values of non-equilibrium observables, since the measurements would take infinite time compared to the length of the equilibrating process. Since we can, in practice, measure non-equilibrium values, it is clear that this attempt to link the infinite time average to observed thermodynamic observables fails. As a result, it is altogether obscure how these time averages are to be connected to the actual results of measurement. Furthermore, ergodic theory's reliance on infinite time averages to account for equilibrium ensembles makes the prospect of expanding the foundational explanation to include non-equilibrium behaviour seem intractable.

3.3.3 Jaynes and Ergodic Theory

For these reasons, Jaynes felt that the ergodic programme could not have the foundational significance it is commonly thought to have. Between the technical and, more importantly, the conceptual problems facing the approach, Jaynes rejected the possibility of basing the foundational programme of statistical mechanics on the ergodic approach and furthermore saw it as altogether unnecessary for the purpose of justifying the use of the Gibbsian ensemble.[60] Rather, as we have already seen, Jaynes saw the justification of the Gibbsian ensemble as stemming from the fact that it maximises the information-theoretic entropy, where the probabilities are to be thought of as epistemic instead of intrinsic or objective properties of the system in question.

There exists a set of criticisms of the Jaynesian approach that argues that Jaynes' rejection of ergodic results is only skin deep, and that ultimately Jaynes requires the use of ergodicity to back up his programme. Sklar (1993), for instance, asserts that although one knows that the Gibbsian equilibrium ensemble is temporally invariant, part of what ergodic theory seeks to establish is the uniqueness of this temporally invariant ensemble.

[60] Jaynes (1978) further claimed that Gibbs himself saw ergodicity as unimportant to statistical mechanics, though the exegetical quality of Jaynes' reading of Gibbs may be suspect.

To this, Sklar claims, the Jaynesian has no response: could there not be other distributions where the values of thermodynamic observables do not change over time?[61]

This criticism seems to be misplaced from the Jaynesian perspective. Again, what *justifies* the use of the usual Gibbsian ensemble is that it maximises the information-theoretic entropy, not that it is the only temporally invariant distribution. There may well be other temporally invariant distributions, but that there exists a single, correct probability distribution to use only seems to be a concern for those who conceive of the probability distribution as being an objective and uniquely characterised property of a physical system. What motivates Sklar's objection is antithetical to the whole Jaynesian project.

Sklar fleshes out this concern by considering cases where there may be more than one temporally invariant distribution because the system is metrically intransitive due to an unknown constraint operative on the system.[62] In such a case, the distribution associated with the Jaynesian algorithm may lead to incorrect values of equilibrium thermodynamic observables. Here Sklar rightly observes that this should lead one to look for this unknown constraint, thus bringing the ensemble into line with the values of the observables. Sklar proceeds to argue that

> our initial probability assignment was *wrong*, suggesting that there is an objective rightness or wrongness about a priori probability distributions. But from a subjectivist point of view, even of the [Jaynesian] sort, our initial probability distribution was *right*, being the uniform probability assignment having the appropriate invariance characteristics relative to the knowledge we had. (1993, 193)

There appears to be no paradox here. If one conceives of statistical mechanics as being a theory of inference in the face of incomplete information, one makes the best

[61] This will be the case if the system is not metrically transitive.

[62] This would have the effect of restricting the accessible phase region of the system to a lower dimensional hypersurface of the full phase space, thus assigning positive measure to less than the full space.

predictions one can based on whatever information is at hand. To be sure, there will be instances where these predictions will turn out to be wrong, since the known constraints will be insufficient to correctly characterise the equilibrium state, perhaps because the system is not metrically transitive. But to say that the probabilistic inference was wrong or unjustified *because* it led to a false conclusion is fallacious; if, despite overwhelming odds, I win the lottery, does this mean that I was ‘wrong’ to make the original inference to the effect that I would not win, even if there was some unknown, hidden, conspiratorial fact that guaranteed that I would win?

Of course, the worry here is more substantive, for in the case that the system is not metrically transitive one’s predictions could turn out to be consistently wrong. If one cannot find some constraint that leads to correct predictions, then it would appear that there is a strong sense in which the problem is rather trenchant. However, it should be realised that this does not reflect a latent reliance on ergodic theory: whether a system is or is not metrically transitive is a feature of the dynamical system itself, independent of ergodic theory. Although metrical transitivity is a necessary and sufficient condition for a system to be ergodic (presuming the infinite time limit exists), explaining the failure of the MEP method by appealing to a lack of metrical transitivity does not reflect any hidden dependence on ergodic methods. In the ergodic approach, one appeals to metrical transitivity in order to establish the equality of the phase and time averages, thereby justifying the use of usual Gibbsian ensemble. Jaynes never appeals to this crucial aspect of ergodic theory in order to justify the maximum entropy formalism. Indeed, Jaynes eschews any reliance on ergodic methods, and even metrical transitivity in the justification of the method:

> *Even if we had a clear proof that a system is not metrically transitive, we would still have no rational basis for excluding any region of phase space that is allowed by the information available to us.* … This shows the great practical convenience of the subjective point of view. If we were attempting to establish the probabilities of different states in the objective sense, questions of metrical transitivity would be crucial, and unless it could be shown that the system was metrically transitive, we would not be able to find any solution at all… The only place where subjective statistical mechanics makes contact with the laws of physics is in the enumeration of the different possible, mutually exclusive states in which the system might be. Unless a new advance in knowledge affects this enumeration, it cannot alter the equations which we use for inference (1983, 10-11, original italics)

The unimportance of metrical transitivity can be emphasised in a different way by reflecting back on Earman and Redei's observation that statistical mechanics works even in cases where the system is not metrically transitive. The point there was that there appears to be good evidence that metrical transitivity is *irrelevant* to explaining the success of statistical mechanics in making predictions of the values of thermodynamic observables. Although explaining *why* metrical transitivity is irrelevant to the success of statistical mechanics is a project of foundational interest, there is no reason to think that, whatever the correct explanation, the Jaynesian approach will not be entitled to it as long as it is a property of the dynamical system itself and does not essentially rely on ergodic assumptions.

A more nuanced criticism of Jaynes' rejection of the importance of metrical transitivity for his programme appears in Guttmann (1999). Guttmann appeals to a technical objection first raised by Friedman and Shimony (1971) and reformulated by Seidenfeld (1986), alleging a conflict between the maximum entropy method and Bayesian conditionalisation. He argues this objection is of particular importance when assessing Jaynes' claim that his approach does not tacitly rely on ergodic results.

Return to the example in Chapter 2 where a probability distribution is to be assigned for a die. Simply given the constraint that the sum of the probabilities of each event equals 1, we find that the expected value of a toss is 3.5, since each outcome is assigned an equal probability of 1/6. Now, consider the additional information that the roll is further constrained to give an odd outcome of 1, 3 or 5. Here, we find a new constraint at work restricting the probabilities according to $p(1) + p(3) + p(5) = 1$. According to Bayesian conditionalisation, this ought to lead to an equally weighted new distribution where $p^*(1) = p^*(3) = p^*(5) = 1/3$. The MEP, simply using this new constraint, will agree with the Bayesian method, but if this new constraint is taken together with the previous one (that the expectation value of the roll be 3.5), the resultant probability distribution will look quite different:

$p^*(1) = 0.21624$, $p^*(2) = 0.31752$, $p^*(3) = 0.46624$

There thus appears to be an ambiguity between different possible applications of the Maximum Entropy method. Which algorithm should one adopt? Should one retain the previous constraint that leads to a conflict with Bayesianism or should one forget about any previous constraints? It should be noted, however, that this disagreement between possible algorithms will not occur when the evidence itself is not probabilistic, but singles out a specific event or set of events with certainty on the prior distribution.

Guttmann picks up on just this point in arguing that this technical objection is particularly salient in statistical mechanics when a system is not metrically transitive. In such a case, the constraints do not pick out a unique, invariant portion of the phase space of positive measure, and the alleged tension with Bayesianism emerges. Therefore, he argues, the proof of the metrical transitivity of statistical mechanical systems becomes important. Based on these considerations, Guttmann concludes:

> [T]hose who accept the ergodic approach may try to resolve the conflict by stating that the MEP is justified because there are objective physical reasons as to why nonergodic systems are not likely to be found. However, because Jaynes objects to the ergodic approach, what makes it reasonable for him to suppose that we shall never encounter nonergodic systems? (60)

This worry seems altogether misplaced, for several reasons. First, Jaynes rejects the *ergodic approach*, by which he means the claim that the establishment of the equality of phase and time averages is of any importance for the foundations of statistical mechanics. However, whether or not systems are metrically transitive is not part of the ergodic approach: it is a property of the dynamical system itself. Jaynes would not deny that proofs of metrical transitivity do indeed characterise objective properties of dynamical systems, and given such proof, that metrically decomposable systems are not likely to be found. What Jaynes does deny is that the ergodic approach constitutes a potential justification of the Gibbsian ensemble, irrespective of whether or not systems are metrically transitive. Surely he could accept the fact that there is no conflict with Bayesianism, should a general proof of metrical transitivity be found.

Even if this potential conflict is shown to exist because systems are not metrically transitive, this need not be troubling from the MEP perspective. The objection relies on a technical objection to the MEP insofar as it potentially conflicts with ordinary Bayesian conditionalisation, but this potential conflict is not a closed issue. It is not clear that objection poses a serious or unassailable worry from the perspective of the MEP theorist (Uffink 1995). Furthermore, as argued in section 3.1.2, this worry only arises if one views the MEP as a rule for conditionalising on new evidence, rather than as a method of assigning probability distributions given a set of constraints; that is, if one takes the background measure as a prior probability distribution upon which to be conditionalised rather than a measure fixed by the symmetries or inherent properties of the event space being considered.[63] Jaynes clearly preferred this latter perspective. If we don't view the MEP algorithm as being a rule to generate posterior probabilities based on priors in the face of new evidence, then why should any conflict with any proposed rule of conditionalisation be worrisome?

This point can be clarified in the case of statistical mechanics by returning to the die described above. In that example, the prior distribution led to a prediction of expectation value 3.5 for the roll.[64] Upon learning that the roll is odd, should one keep this prior calculated expectation value of the die roll as a constraint? This surely depends on how one interprets the expectation value as being a constraint on the system. In the case that the expectation value is merely a prediction based on incomplete information (one functionally dependent on the prior distribution), this expectation value surely does not serve as a constraint on the posterior distribution, since the updating is not done conditionally on the prior distribution, but on the set of constraints actually operative. Hence, the set of constraints to be used are the following:

$$\sum_{i=1}^{6} p_i = 1, \qquad \sum_{i=0}^{2} p_{2i+1} = 1.$$

This set assigns equal probability to the odd numbered rolls, as one would expect. The conflict with Bayesianism arises when one maintains that the original expectation value of a roll should remain 3.5. But there will be instances where it is natural or one would be required to retain previous constraints, such as when the constraint is still operative in the face of new information and not merely a prediction based on previous constraints. For instance, if rigorous testing established that the average value of previous rolls of the die

---

[63] In the case of statistical mechanics, this corresponds to the standard measure over the phase space.
[64] Strictly, one should not speak of prior and posterior distributions based on newly acquired evidence, since this presupposes that one considers the MEP as a method of conditionalisation.

was 3.5 and one was given the additional information that all the previous rolls were odd, then it would be appropriate to generate a probability distribution according to Seidenfeld's calculation, since this would constitute evidence that the die is severely biased.

Applying the MEP to statistical mechanics, it is clear that when one learns a new value of a thermodynamic observable, one is not conditionalising on previous predictions of the values of observables, but adding to a list of constraints operating on the system from which one can re-derive an appropriate probability distribution based on the (new) known set of constraints (see section 2.3). The acquisition of new knowledge about the system surely does not obviate previously known constraints; learning the net magnetisation of a thermodynamic system, say, does not change the already known values of the temperature or volume of the system. These constraints should continue to be employed in generating a novel probability distribution. Guttmann's concerns do not seem to hold any water.

Metrical transitivity, when it can be demonstrated, is clearly a useful property of statistical mechanical systems from the Jaynesian perspective since it guarantees that no additional constraints are needed to correctly characterise the equilibrium distribution. But it by no means necessary for the success of the MEP method. As quoted above, even if a system is not metrically transitive, this does not alter the fact that the MEP method gives one the best predictions possible given the information at hand.

Conversely, even though metrical transitivity is both necessary and sufficient for a system to be ergodic, there remain serious concerns facing the claim that ergodicity provides a genuine solution to the foundational problem of identifying a unique distribution sufficient to characterise the equilibrium state, thereby relating the properties of the ensemble to those of the actual system.

3.4 The Reversibility Objection Reconsidered

Many of the standard interpretive issues in the foundations of statistical mechanics have been addressed, ranging from the interpretation of the probabilities appearing in the theory to the justification of the use of those probabilities to predicting and accounting for the values of thermodynamic observables. The preceding two chapters have been devoted to arguing for the success and cogency of the maximum entropy method. It also defended it against criticisms averring the inadequacy of any epistemic approach and against accusations that the Jaynesian approach is either incoherent or relies on objective features of physical systems to which it is not entitled. Generally, these objections stemmed from a misinterpretation of the goals, claims and motivations of the Maximum Entropy interpretation, usually through smuggling the expectations and demands of ‘objective’ interpretations into the criticisms themselves. However, one foundational problem that has yet to be addressed is the reversibility objection. In this section, the issue will be discussed in a preliminary way, to be further explored in later chapters.

Assume that the maximum entropy method can successfully interpret the success of statistical mechanics and successfully relate the phase averages to thermodynamic observables by making the best predictions possible based upon available evidence. The foundational worry still remains: could we not equally well apply these same predictions as retrodictions, inferring that the entropy of thermodynamic systems was previously greater than it is at present, and contrary to what the second law would have us believe?[65]

Jaynes' solution to the reversibility problem is an appeal to what he calls "experimentally reproducible processes". In essence, Jaynes looks to restrict the application of the MEP method to those phenomena that exhibit some form of statistical regularity. Indeed, he writes that "it is not the business of statistical mechanics to predict everything that can be observed in nature; only what can be observed reproducibly" (1983, 297). Thus, given a present thermodynamic system, one is only licensed in making predictions because the future evolution of such systems is, to a high degree of statistical certainty, uniform. Conversely, there are many ways that the system could have come to its present state from the past. Jaynes argues that this fact indicates that there is a fundamental asymmetry of inference between the past and the future, and it is inappropriate to apply statistical mechanical reasoning in order to retrodict past thermodynamic states, thus avoiding the reversibility objection.

There is a grain of truth to this claim, but it is clearly insufficient as it stands insofar as it is supposed to solve the reversibility objection. Sklar (1993) takes Jaynes to task on exactly this point:

> …processes in one time direction are experimentally reproducible. The same non-equilibrium state always follows a statistically lawlike evolution to equilibrium. But this is not so in the other time direction, because many distinct routes from many distinct initial non-equilibrium states all lead to the same final equilibrium. This is certainly true. But it is just that fact that there is this parallelism in time of systems – that distinct systems show experimental reproducibility in one time direction and not the other and that it is the same time direction for all systems – that we want explained when we are trying to *account* for temporal asymmetry. From this perspective it is hard to see Jaynes' argument as anything but question-begging. (258)

[65] Assume that the only known constraints are present ones. Presumably, any previously known constraints would be cast into doubt by the reversibility objection as discussed in Chapter 1.

For what sort of temporal asymmetry are we looking to account? What is this parallelism of which Sklar is speaking? Presumably, Sklar is concerned with the monotonic increase over time of the thermodynamic entropy for all such systems, and that these systems all demonstrate this increase towards the future, rather than the past. In other words, Sklar is looking for an account that grounds the past thermodynamic properties of physical systems, an account that draws on the properties or features of the probability distribution associated with its statistical mechanical counterpart.

I have already questioned the possibility of such an account on two fronts. First it is unclear that any 'objective' notion of probability is meaningfully attributable to statistical mechanical systems. Yet one could only expect that the actual macroscopic thermodynamic properties of systems would supervene on the underlying probability distribution if the distribution is an intrinsic, 'objective' property of such systems. Second, it has been argued that any attempt to link the statistical mechanical properties of a system to a measure of entropy, whether in the Boltzmannian or standard Gibbsian sense, is fraught with difficulty.

As such, Sklar's demands seem unreasonable. These explanatory aims look for too much. Rather, the suggestion of the last two chapters has been that the probabilities appearing in the theory are to be thought of not as inherent properties of a system, but as measures of one's ignorance regarding the exact microstate of a physical system, past, present or future. Otherwise stated, we seek from statistical mechanics a method of making correct predictions or retrodictions of the thermodynamic properties of physical systems.

On this view, if there is a temporal asymmetry between past and future statistical mechanical states, this is to be accounted for as an asymmetry between the way one forms opinions or inferences regarding past and future states of affairs, one that is not determined by (though related to) the actual past and future macrostates of physical systems. One can get a handle on this difference by considering the reversibility objection, applied to an ordinary non-equilibrium thermodynamic system. In such a case, the use of the probability distribution applied to the present system leads one to retrodict that it arose as a spontaneous fluctuation from a past equilibrium state. But notice that under the interpretation of probability being entertained here, this probability distribution is not directly tied to the *actual* past values of thermodynamic observables; that is, the essential link between the two has been severed. The fact that one's retrodictions indicate a past equilibrium state need not demand that the system *actually* arose as a spontaneous fluctuation.

Now the reversibility objection takes on a different form. The problem one must address is *not* the worry that the entropy of any given non-equilibrium thermodynamic system was most likely higher in the past, but one of aligning one's retrodictions with the actual past of physical systems. The asymmetry to be accounted for is one of inferential practices: how and why are inferences to the past differently constrained from those looking towards the future? In what circumstances might one's retrodictions be blocked or constrained so as to move them in line with the actual history of thermodynamic systems?

Let us return to the passage by Jaynes quoted in Section 3.3.2. It is certainly true that, looking at statistical mechanics in this way, it is not necessary for statistical mechanics to predict or retrodict everything that occurs in nature. It suffices that the predictions or retrodictions form the best inferences one can make given the available evidence, even if this underdetermines a thermodynamic system's past or leads one to infer a past state of higher entropy. This merely amounts to a restatement of the reversibility objection in epistemic terms. However, what justification can one offer to restrict the application of the MEP to experimentally reproducible processes?

Clearly, as Sklar notes, simply avowing that the MEP can only be applied to experimentally reproducible processes begs the question. In what is perhaps a lighter moment, Jaynes offers a justification for this restriction by appealing to the fact that scientists are, in practice, only interested in experimentally reproducible processes (1983, 297). Terming this a "sociological convention," he takes this as a justification for ignoring retrodictions of past states of affairs. Surely this will not do. If the MEP project is on the right track, the ultimate response to the reversibility objection should not appeal to the pragmatic interests of scientists, but to some feature of the approach itself that solves the problem on its own terms; that is, by accounting for this temporal asymmetry as one reflecting an asymmetry of inferential practices.

The solution that I offer is twofold, to be further elucidated in the following chapters, and briefly introduced in section 3.2. First, we shall look to block inferences to the past that retrodict anti-thermodynamic behaviour when one lacks sufficient information in the form of *past* thermodynamic constraints to adequately characterise the macrostates of such systems. The idea here is to pick out an appropriate range of times during a given system's past and future where it can be considered a well-defined statistical mechanical system. Such a suggestion is tantamount to adopting a branch systems proposal (Reichenbach 1956), where thermodynamic systems are thought of as 'branching' off from some larger system and only exist as quasi-isolated systems for a finite period before rejoining the larger system. Adapted to an epistemic framework, this implies that one can identify the moment that a statistical mechanical system came into being; that is, when the initial knowledge of the system's macrostate was acquired.

The second component to the proposed solution involves arguing for the veracity of records of past macrostates. Since the reversibility objection is taken to apply equally well to one's records and memories (also construed as statistical mechanical systems), an account needs to be developed that will assure that possessed information regarding past states of affairs are veridical. Given that records (as a matter of fact) exist of the past, the retrodictions based on MEP inferences are differently constrained from predictions, thereby generating the desired asymmetry.

The important thing to note is that we are not attempting to ground the fact that thermodynamic systems, past, present and future, always monotonically increase in entropy. What is being endeavoured here is a scheme that grounds the validity of our inferences to past states of affairs while still correctly predicting future ones on the basis of statistical mechanical reasoning through the MEP method. The following two chapters provide an account of how this may be done.

Chapter 4. Branch Systems

In Chapter 1, the idea that the initial low entropy state of the universe was sufficient to ground the irreversible behaviour of thermodynamic systems was found wanting. Further, it was argued that this sort of global explanation could, at best, provide a consistent model of a universe where the laws of thermodynamics were approximately true, although it would still be entirely improbable that this model represented an accurate account of the entropic history of the universe's subsystems even if it could, in principle, recover the veracity of one's beliefs and memories of the past. The upshot of this argument was that such global approaches to explaining or grounding our thermodynamic experiences fail to provide any reason to think that the entropy of the universe was ever lower than it presently is. As such, one should look for a more local approach; one that begins with the thermodynamic properties of individual thermodynamic systems that we have experiences and contact with, first establishing that these local systems obey the laws of thermodynamics, and then extending this to progressively larger and larger systems further and further into the past. In effect, this is an inversion of the sort of project that Albert (2000) envisions, where instead the low entropy past of *local subsystems* of the universe grounds the inference to the low entropy past of the universe as a whole. After all, what purpose does it serve to provide a transcendental account of how our beliefs about the past *could* be veridical without giving any reason to think that they are?

An influential line of thought that adopts what might be termed a 'local' approach to explaining the irreversible behaviour of thermodynamic systems is that of branch systems, championed by Reichenbach (1956) and with variants proposed by Grünbaum (1963), Davies (1974) and Winsberg (2004b), among others. Briefly stated, the branch systems proposal sees the origin of irreversible behaviour as stemming from the fact that individual thermodynamic systems come into being by finding themselves as quasi-isolated systems that 'branch' off from the rest of the universe. If these systems branch off (or form) in low entropy states, then they are overwhelmingly likely to evolve to higher entropy states before coming back into contact with the rest of the universe. According to Reichenbach, it is the *parallelism* of these branch systems, the fact that they all evolve in the *same* temporal direction, that accords time a global direction (at least in our region of the universe), and in turn grounds the inference that our region of the universe began in a low entropy state.

This chapter will be structured as follows: In the first section, I will briefly summarise Reichenbach's basic branch systems proposal. The second section will discuss Reichenbach's more refined proposal along with the variants of Grünbaum and Davies, and investigate the various criticisms of branch systems approaches levelled by Sklar and Albert. This will be followed by a reconstruction of the branch systems account from an explicitly epistemic perspective, demonstrating how the usual criticisms of branch system analyses can by evaded or answered if one conceives of statistical mechanics as being a theory of statistical inference applied to thermodynamic systems.

4.1 Reichenbach's Branch Systems Proposal

Reichenbach motivates the branch systems argument by considering an infinite ensemble of identical thermodynamic systems all beginning in identical low entropy thermodynamic states and permanently isolated from the environment. These states can be arranged in a matrix:

Table 1: Reichenbach's matrix of thermodynamic systems over time

y^{11}	y^{12}	y^{13}	y^{14}	…	…	…
y^{21}	y^{22}					
y^{31}						
y^{41}						
y^{51}						
…						
…						

The first superscript indexes the system while the second indexes the time, assumed to be given in discrete intervals. Further, Reichenbach defines the transition probabilities as long run frequencies either across times or across systems. For some arbitrary entropy values A and B, the forward transition probabilities (from entropy A to entropy B) are given (in Reichenbach's notation) as either

$$P(A^{ki}, B^{k, i+1})^{i}$$

or

$$P(A^{ki}, B^{k, i+1})^{k}$$

where the superscripts outside the parentheses indicate whether the long run frequencies are to be evaluated relative to the long term behaviour of a single system (the time ensemble indexed by *i*) or relative to the long run behaviour across many systems at a single time (the space ensemble indexed by *k*). This distinction between the long run probabilities of the space and time ensembles contains the seed of Reichenbach's attempt to ground the direction of time in the behaviour of many systems, rather than the time-directed behaviour of a single system (including the universe as a whole) because the reversibility objection only applies to single systems. In Reichenbach's notation, the reversible property of statistical mechanical systems can be stated as

$$P(A^{ki}, B^{k, i+m})^{i} = P(A^{ki}, B^{k, i-m})^{i}$$

Where *m* denotes an arbitrary number of time steps. The reversibility property is captured by the formalism by stating that the transition probabilities of B from A are equal over long (infinite) times in either temporal direction.

Conversely, a similar relation may not hold in the case of the space ensemble. The claim that

$$P(A^{ki}, B^{k, i+m})^{k} = P(A^{ki}, B^{k, i-m})^{k}$$

will depend on how a collection of systems is arranged in the column *k* for which the long run frequencies are determined. For instance, if one looks in the early columns of the array where *i* is not too large, and the array is arranged such that each system has a very low entropy in the first column, then is it most probable that a system having a low entropy at time *i* will have higher entropy in later columns but lower entropy in previous columns because the array was explicitly constructed in this way. Thus, the collective features of the space ensemble will generally depend on the way in which the matrix is constructed, whereas the long-term behaviour of a single system will be symmetric over time.

Reichenbach employs this distinction to generate an asymmetry in the behaviour of isolated systems. To establish this asymmetry, he makes two assumptions designed to link (and delimit) the properties of the space and time ensembles. The first assumption is that of *row invariance*, and amounts to the assumption that rows/systems are independent of each other; that is, conditionalisations on the entropic state of any other system in the array at any time do not affect the transition probabilities of the system of interest.

Second, Reichenbach posits the property of *lattice invariance*: this condition explicitly links the properties of the space and time ensembles by stating that the long run transition frequencies are equal whether one counts across a given row or a given column. Reichenbach terms an array satisfying these conditions a *lattice of mixture*. Based on these assumptions, he is able to show that no matter how the states of the initial column are arranged, they will approach an equilibrium distribution in the later columns.

Reichenbach gives these posits physical meaning by considering some small section of the universe's entropic curve, with various thermodynamic systems 'branching off' from this main system and remaining in isolation afterwards, as depicted in figure 2. No matter how (or in what entropic states) the systems branch off from the main system, they will move towards an equilibrium (canonical) distribution at later times. Furthermore, if the entropies of the systems when they branch off from the main system

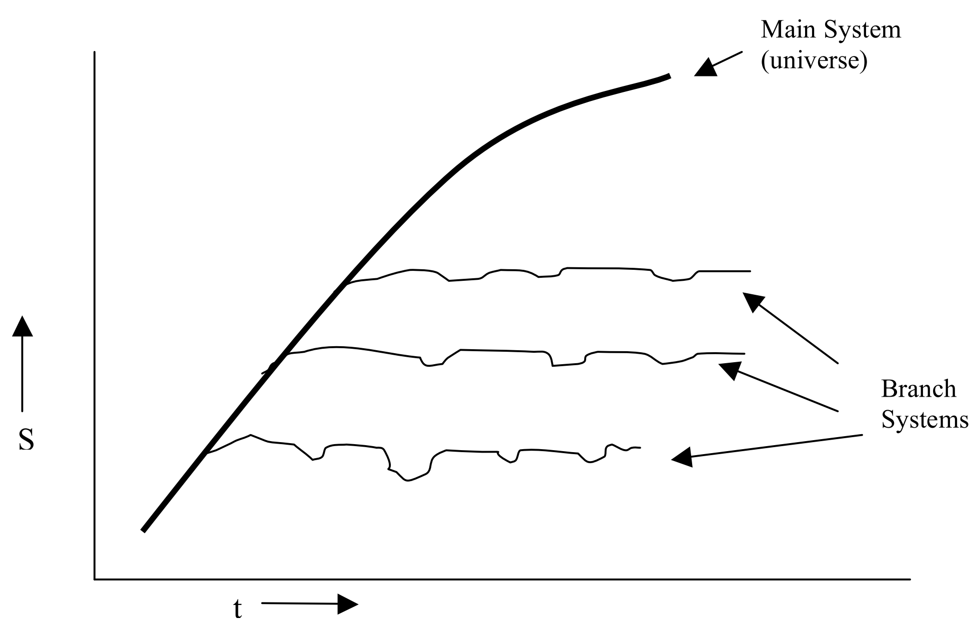

Figure 2: Reichenbach's Branch Systems. The universe is on a long upgrade in entropy and branch systems, once formed, remain isolated and subject to spontaneous fluctuations.

differ from the equilibrium distribution in that a disproportionate number of these systems branch off in states of non-maximal entropy, then these systems will exhibit time asymmetric behaviour early in their lifetimes, i.e.

$$P(A^{ki}, B^{k, i+m})^k > P(A^{ki}, B^{k, i-m})^k \text{ for } 0 < m < i$$

for small *i* because the systems are improbably distributed with respect to their entropies in the first column.

Reichenbach goes on to *define* the direction of time as the direction in which most branch systems increase in entropy. Thus, if Reichenbach can show that the array of real thermodynamic branch systems is one where the systems are more likely to be formed out of an equilibrium distribution, this can serve to provide a rationalisation of the direction of time through the behaviour of thermodynamic systems.

While this picture represents an over-simplified description of the behaviour of real thermodynamic systems (as Reichenbach readily realises), several important features of his approach can be identified. First, Reichenbach assumes that these branch systems split off from a universe that is on a long upgrade of its entropy curve. This assumption seems suspect, even if it is true, as it is part and parcel of the reversibility objection that one ought to think of any physical system, including the universe, as having greater entropy in both its past and future.[66]

Nonetheless, Reichenbach is right to point out that the entropic curve of the universe, whatever it may be, is insufficient to pick out a direction of time because it is susceptible to the reversibility objection. Rather, Reichenbach argues that the solution to this problem lies in the statistical properties of the *collective* time-directed behaviour of the various quasi-isolated branch systems of the universe. Thus, appealing to the initial

[66] Reichenbach offers a few cryptic remarks as to why this assumption is reasonable (116).

non-equilibrium distribution of thermodynamic systems when they branch off from the main system generates the temporal asymmetry. Effectively, the asymmetry is introduced by asserting that prior to these branch-off points, the systems did not exist (as well-defined thermodynamic systems), and therefore did not arise from a higher entropy past as the reversibility argument suggests. Claiming that before these branch-off points, the systems simply did not have a history generates the asymmetry.

4.2 Refining The Branch System Proposal

It is my contention that Reichenbach was right to focus on the collective behaviour of individual isolated systems as supplying the germ to the solution of the reversibility objection, though he went about it in the wrong way. The cogency of Reichenbach's project rests on his ability to demonstrate that individual branch systems do in fact form in non-equilibrium states (or collectively in a non-equilibrium distribution), and that this assertion can somehow be made reasonable by appealing to the local entropy grade of the universe.

To see why this claim is problematic, consider Reichenbach's more refined account of branch systems (depicted in figure 3), where the universe undergoes changes in the direction of its entropic curve (as it must in the long run) and where branch systems are not permanently isolated from their environments, but rejoin the main system after some time.

These amendments to the model introduce new complications. First, the finite lifetime of branch systems indicates that the long run time ensemble frequencies cannot be meaningfully defined for such systems since the probabilities (on Reichenbach's view) are only defined for infinite or long finite sequences of thermodynamic states, much longer than the lifetime of a typical branch system. Nonetheless, if one accepts the postulate of lattice invariance, this difficulty can be easily overcome because it links the probabilities defined for the space ensemble to those of the time ensemble. Since there does exist a large number of branch systems, the time ensemble probabilities can be meaningfully defined, though this indicates that the lattice of mixture assumption may now be doing more work than it was originally intended to do.

A second point stressed by Reichenbach is that the periodic nature of the universe's entropic curve implies that the direction of the entropic increase of branch systems will be different depending from what part of the universe's entropic curve the branch systems originate. Since Reichenbach defines the direction of time as the one for which the majority of branch systems increase in entropy, this has the consequence that it is meaningless to speak of the direction of time for the universe as a whole, but only for certain subsections of the universe.

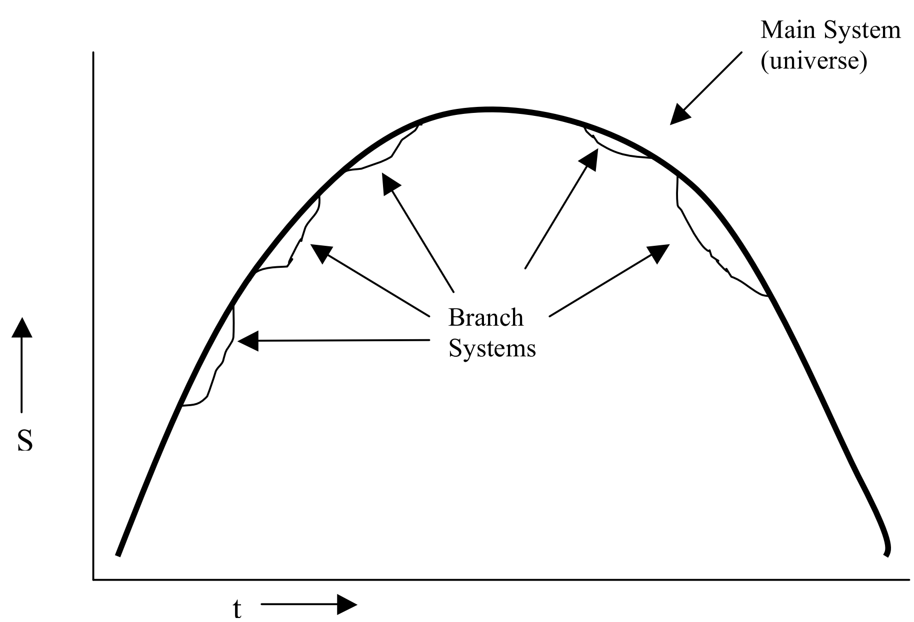

Figure 3: The universe is on a long upgrade in entropy followed by a downgrade and branch systems, once formed, remain isolated until they recombine with the main system.

Reichenbach (1956, 136) summarises his completed proposal through five assumptions:

1. The entropy of the universe is at present low and is situated on a slope of the entropy curve.[67]
2. There are many branch systems, which are isolated from the main system for a certain period, but which are connected with the main system at their two ends.
3. The lattice of branch systems is a lattice of mixture.
4. In the vast majority of branch systems, one end is a low point, the other a high point.
5. In the vast majority of branch systems, the directions towards higher entropy are parallel to one another and to that of the main system.

Sklar (1993) identifies Assumptions 2 and 4 as being fairly innocuous. Assumption 2 merely asserts the existence of branch systems; that is, systems that are isolated for a finite period of time from the rest of the universe. Assumption 4 can be read as delimiting those branch systems that are of interest to us, discounting those systems that do not experience a net entropy change over the course of their lifetimes such as systems that branch off in equilibrium states and remain there until they rejoin the larger system.

Reichenbach asserts that Assumption 1 is insufficient to derive the remaining assumptions. He notes, in particular, that Assumption 5 is not deducible from assumption 1 since there is nothing contrary to the laws of physics that prevents a branch system's entropic curve from being counter-directed to the rest of the universe. He suggests that it might be possible to show that assumptions 2-5 are in some sense probable given Assumption 1, but denies the meaningfulness of this claim since the probabilities would only acquire meaning over long temporal spans of the universe. Whether or not one

[67] Grünbaum denies the necessity of this assumption in his account of branch systems.

endorses Reichenbach's interpretation of probability, I would maintain that the low entropy past of the universe is not sufficient to (even probabilistically) ground the temporal anisotropy expressed by the second law, as argued in Chapter 1.

However, Reichenbach claims to show that Assumption 5 can be derived from Assumptions 3 and 4 as follows: consider an ensemble of systems in low entropy states in the initial columns of the array. Such an ensemble will approach an equilibrium distribution in the later columns given the assumption that it is a lattice of mixture. Similarly, an ensemble of systems with their low entropy ends in the last column (as we are now dealing with finite arrays) will approach an equilibrium distribution in the initial column given that it too is a lattice of mixture. However, if one combines the two arrays just described into a single array, where half the systems have their low entropy ends on the left side of the array and the other half on the right, then, Reichenbach claims, the resultant array is not a lattice of mixture because such arrays always approach an equilibrium distribution on one end of the array. Hence, one can show the parallelism of branch systems given Assumptions 3 and 4.

Sklar points out that the proof is flawed. The result that Reichenbach derived, that for a lattice of mixture one would find an equilibrium distribution in the later columns, depended on the assumption that the initial column was in a non-equilibrium distribution and that the final column (in a finite array) was not constrained. In this combined array, this condition is not satisfied, yet the array still satisfies the lattice of mixture condition.[68] As such, Assumption 5 remains an independent assumption, not derivable from Assumptions 1-4. Insofar as it is the parallelism of branch systems that is supposed to

[68] Sklar seems to identify the lattice of mixture assumption as the culprit in claiming that it is "not as innocent as it first appears" (323). However, it is not the lattice of mixture assumption that is at fault, but the insistence that arrays of branch systems are constrained in the initial column but not in the final one.

supply a direction to time, Assumption 5 merely asserts this directionality without explaining why it holds. What we are looking for is some explanation as to why the majority of branch systems evolve in parallel, and why they evolve in the same direction as the universe as a whole: Reichenbach has given us neither.

An alternative approach is suggested by Davies (1974) and Grünbaum (1963), who rely on Schrödinger's formulation of the 2nd law. Consider two systems (A and B) in non-equilibrium states at t_1. As per the reversibility worry, at times both earlier and later that t_1, one should expect the entropy of these systems to increase. Hence, at some time t_2 not too far from t_1, the following relation will hold:

$$\left(S_A^2 - S_A^1\right)\left(S_B^2 - S_B^1\right) > 0$$

irrespective of whether t_2 is earlier or later than t_1. Indeed, this relation will hold if the entropies of *both* systems were lower in the past *or* the future. Since no assumption was made regarding the nature of the systems A and B, if one system is taken to be an arbitrary branch system and the other the whole universe, one can expect the entropy of the universe to evolve parallel to any given branch system.

Nonetheless, the inequality is false if the systems evolve in anti-parallel directions. As such, the inequality merely states that systems should evolve in parallel, without giving any justification for why it should hold. It fails to explain why one should think that systems always increase (decrease) in entropy in the same temporal direction.[69]

A further worry can be introduced with respect to this argument because it fails to establish that the entropy was lower in the past. At most, the relation demonstrates that systems evolve in parallel, not that the entropy is higher or lower in either temporal

---

[69] One might also worry as to *why* the differences in entropies are to be multiplied together, rather than (say) added. There seems to be no obvious rationale choosing this inequality other than the fact that it correctly reflects the parallelism of thermodynamic systems.

direction. The entropic asymmetry is introduced by stipulating that branch systems are often formed in low entropy states and before their formation, did not have a history. But if t_1 is chosen at some point other than at the beginning of the system's history, then the usual reversibility objection applies: one should infer that the systems' entropies are higher in both temporal directions, including the entropy of the whole universe. In order to establish the monotonic increase in entropy, one must find some principled way to insist that the uniform probability distribution obtains *only* when the branch system is formed, and at no other time.

Albert (2000) asserts that the branch systems proposal is "sheer madness". On the issue just raised, Albert asks

> How is it (to begin with) that we are to decide at *exactly* what moment it was that the glass of water with ice in it first came into being? And even if we *could* decide that, what then? How is it (exactly) that the medium-sized system we decided to focus on was the glass of water with the ice in it and not (say) the *room* in which that glass is currently *located*, which also contains the table on which the glass is currently sitting, and the freezer from which the ice was previously removed, and the person who first got it into his head to do the removing? The uniform probability-distribution over the possible microconditions of the macrocondition of *that* system, at the moment when *it* came into being, will (after all) differ quite radically – even insofar as the glass of ice water *itself* is concerned – from the one we have just been talking about! And why not the *building* the room is in? And why not the *city* the *building* is in? And even if all *that* could be decided, very serious questions would remain as to the *logical consistency* of all these statistical-hypotheses-applied-to-individual-branch-systems with *one another*, and with the earlier histories of the branch systems to branch systems branched off *from*. (89)

It is not clear, as Winsberg (2004b) notes, that there is any obvious problem or logical inconsistency lurking in considering the glass of ice water to be a branch system with respect to the room, or the room to be a branch system with respect to the building, and so on (assuming that one can meaningfully talk about each of these systems as being at least quasi-isolated from its environment). I take it that an epistemic interpretation of branch systems can answer these worries, for it seems that these concerns only have bite if one adopts an objective interpretation of statistical mechanical probabilities. As Winsberg writes, Albert's concern seems to be "how does the universe know the precise moment at which a branch system comes into being so that *it* can know to apply the [uniform probability distribution]?" (2004b, 715)[70]

In any case, if one interprets the probabilities as epistemic, such a question doesn't make any sense, because it is agents that apply probability distributions to macrostates, not the universe. But the question remains: why should one apply a uniform probability distribution only when the branch system is formed, and at no other time? The solution to this problem lies in reinterpreting the notion of what a branch system is, and how it is related to the thermodynamic state of the rest of the universe.

4.3 An Epistemic Branch Systems Account

Imagine that I walk into an otherwise empty room with a half-melted ice cube sitting in a glass of water. If I were to apply a uniform probability distribution to the ice-water system at this moment I would, in a familiar way, both predict and retrodict that the ice cube was, and will be, more melted than it is now. On the usual branch systems proposal, one assumes that, in the past, the ice/water system was formed at a particular time by some ordinary thermodynamic process, say by someone previously entering the room and dropping an unmelted ice cube (fresh out of the freezer) into the glass of water. As we have seen, this amounts to postulating that such systems have low entropy pasts which, though intuitively true, merely asserts this fact without explaining it, and blocks

[70] If these are taken to be serious problems, I suspect that it will generate severe difficulties for Albert's own proposal. If one can't meaningfully talk about the moment of creation of an isolated glass of ice water, then exactly how does the past hypothesis guarantee that it was previously in a state of lower entropy?

any anti-thermodynamic behaviour by asserting that such systems had no history before the moment of their formation where they became effectively isolated from the environment.

Although these claims seem reasonable, they are left unjustified by branch systems theorists such as Reichenbach, Grünbaum and Davies. Further, it is at best unclear why an agent who just walked into the room with the ice water in it should just *assume* that the ice cube was previously less melted than it currently is. Conversely, it is also unclear why an agent should just *assume* that the ice cube was previously *more* melted then it currently is. Insofar as one conceives of statistical mechanics as being a theory of inference regarding thermodynamic systems, it is how such an agent ought to reason regarding the past and future states of our ice/water system that is crucial.

Consider a proffered bet as to the entropic state of the ice cube five minutes ago, given that I just walked into the room and found a half-melted ice cube before me. If I am wise, I should refuse to bet, for I can imagine three types of processes that are compatible with the present state of affairs:

Type 1: The system has been isolated for the past five minutes and the presently half-melted ice cube arose as a spontaneous fluctuation from some state of even higher entropy.

Type 2: The system has been isolated for the past five minutes and the ice cube was less melted five minutes ago, either itself arising from a spontaneous fluctuation or through some previous interaction with the environment.

Type 3: The system was formed through an interaction with the environment within the last five minutes; say by someone dropping a less melted ice cube into the glass of water two minutes ago.

Considering this last case, it is clear that, unless I can be *assured* that the system has been isolated for the previous five minutes, I should not bet as to its state five minutes ago. If the system was not formed then as an isolated system, its properties (its energy, for instance) at that time are completely unknown. The upshot is that without any further information about the history of the system, one should not consider any proffered bet. The reason for this is that any bet as to the past entropic state of the system operates on a completely undefined event space. Insofar as one regards the probability distribution as a reflection of one's ignorance as to the microstate of the system, the probability distribution only takes on definite values when an agent acquires some definite knowledge of the state of the system; that is, at the moment I enter the room. Any retrodiction operating before this moment is completely unreliable. In this way, we can speak of the time of the creation of a branch system as being the time at which one is willing to bet as to the past entropic state of the system: at the moment I enter the room.[71]

This account of what one means by the creation of a branch system differs significantly from traditional accounts in two ways. First, the previous accounts of Reichenbach, Davies and Grünbaum look (or need) to establish when it is that a branch system comes into being as an objective matter, and furthermore why it is that branch systems are formed in low entropy states in the past and not the future. In contrast, here one takes the formation of a branch system as not being the moment where the quasi-isolated system actually came into being, but the moment at which one can treat the system as a well-defined statistical mechanical system; specifically at the moment when one can assign a well-defined probability distribution over possible microstates of the system.[72]

---

[71] This notion of the formation of a branch system will be further clarified and discussed below.
[72] Understood in this way, it seems as if Albert's worries regarding the branch systems proposal are unfounded.

Second, the non-existence of the branch system before I enter the room should not be misconstrued as an assertion that the ice/water system was in no definite state five minutes ago.[73] Herein lies the advantage of the severing the link between the probability distribution and the actual thermodynamic state of the physical system. Whatever microstate obtained completely determined the values of the system's thermodynamic observables, but the non-existence of a well-defined probability distribution only implies that an agent should not bet on the values of those observables. A statistical mechanical system is a description of one's knowledge of the physical system, not a description of the physical system itself (see Chapter 2).

One can offer a more general description of branch systems on this approach. The lifetime of a branch system is the duration of a physical system for which one can meaningfully apply a probability distribution over the system's possible microstates. The intuition here is the one should restrict one's inferences to the times that one knows (or suspects) the system to be quasi-isolated from its environment, and at no other times. This definition may appear overly restrictive, since we do normally make inferences beyond those applicable under this description, but it will suffice for the present.

Suppose, given this characterisation of branch systems, that I can be assured that the ice/water system has been isolated for the past five minutes (say I've been standing outside the only door to the room for the last few minutes). Such information would rule out an event of type 3, where the system was brought into being in the very recent past. Upon entering the room, what history should I infer for the system? Should I be willing to bet that the cube arose as a spontaneous fluctuation from equilibrium?

---

[73] Assuming the ice was actually in the water.

Perhaps. Knowing that the system has been isolated for the past five minutes does not guarantee that it was always an isolated system, since it still leaves open the possibility that the ice cube was *even less* melted ten minutes ago than it was five minutes ago (as a type 2 event). Furthermore, I know that the universe has the entropic resources (whether or not the universe's entropy was higher or lower in the past) to create an ice/water system with very low entropy ten minutes ago. Naturally, this does not imply that the present state of the system did not arise as a spontaneous fluctuation from equilibrium, but it does mitigate the inference that, solely based on my present knowledge of the system and the rest of the universe, I should immediately conclude that the entropy of the ice/water system was previously greater than it is now.

It is exactly these same considerations that fail to eliminate the possibility that the system will find itself in a lower entropy state five minutes from now, even if it remains isolated over the *next* five minutes. It could be the case that the system is constrained to be in a lower entropy state beyond the time it is known to be isolated, and thus to behave anti-thermodynamically in the future. But it is a curious fact that we virtually never see such behaviour.

There are several related questions one might ask at this point regarding how one ought to bet as to the entropic state of the ice cube at different times, given the knowledge that it is isolated for the previous and next five minutes:

1. How should one bet on the state of the system five minutes from now?
2. How should one bet on the state of the system ten minutes from now?
3. How should one bet on the state of the system five minutes ago?
4. How should one bet on the state of the system ten minutes ago?

It should be emphasised that the concern here is to underwrite the *inferences* to the past and future states of the ice cube.

The central difference between the inferences we make between the past and the future lies in the fact that we *believe* that events in the past are differently constrained than events in the future, but it is the nature of this difference over which people differ. Rather than locating this constraint somewhere in the distant past (as the past hypothesis), the branch system approach seeks to provide a more local constraint (or set of constraints), namely how individual branch systems are constrained in their local pasts or, more precisely, *how* and *why* we *believe* such systems to be differently constrained in their pasts from their futures. And this is because we seem to have an abundance of records of the past, and virtually none of the future.

With this in hand, return to the question of what one's inferences ought to be regarding a presently half-melted ice cube and the knowledge that the ice cube has been isolated for the past five minutes. It seems that the inferences towards the future are fairly straightforward: on the condition that the ice/water system will be isolated for the next five (ten) minutes and is not constrained in some way in the future, one should expect the system to continue to melt over the next five (ten) minutes, reaching equilibrium and staying there (modulo expected fluctuations) for a long time. If the system is not energetically isolated over the next five (ten) minutes, then all bets are off.

Note that this does not rule out the possibility that the system is constrained in the future to behave in a way contrary to what one should predict. An example of such behaviour is furnished by the spin-echo experiments, or perhaps a Maxwell's demon that has manipulated the microstate of the ice/water system so as to recreate an unmelted ice cube five minutes from now, without altering the energy of the system. To be sure, such cases lead to apparent anti-thermodynamic behaviour in the future by virtue of the fact that the system is constrained to be in a particular (set of) microstate(s) further down the line.[74]

Despite the seemingly strange behaviour of spin-echo or demonically influenced systems, there is nothing generally wrong with the inferences we do in fact make towards the future, and any systematic failure of these inferences can be plausibly attributed to a lack of information as to how the system is constrained (past, present or future). I take it that this is a natural consequence of an epistemic approach to statistical mechanics.

The lesson holds, time-symmetrically, with respect to retrodictive inferences. Given the knowledge that the ice cube is presently half-melted and has been in isolation for the past five minutes, one should retrodict that the ice cube was *more* melted five minutes ago, and arose as a spontaneous fluctuation unless there is some known constraint in its past or future that would indicate otherwise, say by some knowledge about its state ten minutes ago. In a sense, the bullet is bitten: rather than endorse the claim that one ought to regress to the point of the early universe (as Albert does) in order to explain the fact that we should believe that the ice cube was previously less melted, the suggestion here is that one should:

1. Restrict one's inferences to times when the system is, or is presumed to be, isolated.
2. Trust those inferences as the best ones that one can make under the circumstances, only to be altered in the case that there is good reason to believe that some discernable constraint about the system's past or future is missing from the description.

[74] These constraints are, in an important sense, time-symmetric. One could describe this as either a future constraint on the state of the system, or as a *past* one arising through, say, the demon's previous interaction with the system. I take it that whether one interprets this system as being constrained in the past or in the future is merely a matter of description. The same goes for a spin-echo system where one is only aware of the system's instantaneous present state and perceives anti-thermodynamic behaviour.

This proposal seems methodologically plausible.[75] But it bears the unfortunate consequence that in cases where the system is known to have been isolated, *and* there is no missing constraint, one should be prepared to admit that the best inference one can make is that the system evolved to its present state as the result of a highly improbable spontaneous fluctuation. Of course, this conflicts with commonsense and experience: it conflicts with experience because we never witness such anti-thermodynamic behaviour, and with commonsense because we normally, in everyday life (as well as in scientific pursuits), presume that thermodynamic systems do not evolve into low entropy states from higher entropy ones.

These conflicts are only skin deep. At first glance, experience tells us that only normal thermodynamic processes occur (assuming we can trust our memories), and the current proposal seems to demarcate between the observed and unobserved evolutions of thermodynamic systems: those that are observed follow the usual laws of thermodynamics, while those that are not observed do not. But surely, one might argue, whether a system is or is not observed makes no difference to its actual evolution. So doesn't it follow that this proposal generates an absurdity?

There are several problems with this potential objection. First, while it is true that (barring quantum mechanical worries) whether or not a system is observed does not affect its evolution, it is equally true that the observations one makes do affect the *inferences* to a system's past or future state. And insofar as the concern here is regarding the inferences

[75] I take it that this is how statistical mechanics is usually applied to physical systems.

one should make regarding the past or future state of the branch system, it is entirely appropriate to distinguish between observed and unobserved processes.[76]

A second but related sense in which this argument fails is that it presumes that the inductive inference from previously observed thermodynamic processes to unobserved ones is a good one:[77] this is what drives the commonsense intuition that only normal thermodynamic processes occur. However, it is precisely this commonsense intuition that is called into question by the reversibility objection; it is a case where commonsense is *prima facie* undermined by one of our best theoretical descriptions of the world. And so it becomes a central worry in the foundations of statistical mechanics to reconcile our commonsense with statistical mechanics, vindicating the inductive inference that thermodynamic systems (almost) always follow the laws of thermodynamics as they are standardly interpreted. I want to argue that this can be achieved, but only in a restricted way.

A straightforward move, one that was resisted in Chapter 1, is to posit that a low entropy state of the early universe can ground this inductive inference, and this missing constraint in the past that, once accepted, serves to explain how we can trust our records and justifiably believe that the entropy of local thermodynamic systems was lower in the past. I argued that this posit cannot be justified as long as one takes the reversibility objection to undermine our belief in the veracity of records of the past and, even if it is assumed, cannot do the work it claims to do.

[76] Of course, it is a natural consequence of this proposal that one should never *expect* to witness anti-thermodynamic behaviour unless there is some unaccounted for future constraint on the state of the system, as is the case in the spin-echo experiments or Maxwell's demon.
[77] And perhaps it is a good inference.

The alternative is to look for constraints in the more local past that can do the requisite work. In the case of the ice cube in a glass of water, isolated for the past five minutes, a local constraint to the effect that the ice cube was less melted five minutes ago, or that someone dropped a large hunk of ice into the glass ten minutes ago, will do the job admirably. And before the time of such a constraint, the branch system simply did not exist, in the sense that there is no reasonable way in which an agent could generate a probability distribution to describe this physical system.

So the proposal is this: in cases where a thermodynamic system is known (or suspected) to be (at least) quasi-isolated from its immediate environment for some period in the past (future), one should retrodict (predict) that the system was previously (will be) in a state of higher entropy than at present, unless some constraint can be found to indicate otherwise. Now, as a matter of everyday practice, we do not believe that ice cubes fluctuate into existence (though perhaps we should), and given sufficient interest, we look for records that indicate a lower entropy past for the systems we encounter. Furthermore, if we look hard enough, we often find such records.[78]

It is the conspicuous existence of such records that ground the inductive inference that the entropies of *all* thermodynamic systems were lower in the past, so that even in the absence of such records (even once we've looked), we presume that all systems evolve thermodynamically throughout their existence even if the inference is undercut by our best theories.[79] Even so, often facts about the less local past or wider spatial environs (though still more recent than the early universe) are cited as reasons to believe that the

[78] Or we simply presume that such records exist. It is the business of many scientific disciplines that concern themselves with reconstructing the past to discover such constraints (e.g. cosmology, reconstructions of evolutionary history, planetary science, history, palaeontology).
[79] As Hume famously argued, such beliefs may merely be habitual.

ice cube was previously less melted, such as the fact that the room in which the glass of ice water sits is known to have been isolated for the past five minutes, or the wealth of evidence we have for the claim that the entropy of the solar system was previously lower, or the electrical grid that powers the freezers that usually generate ice cubes is a dissipative system. Notwithstanding all this, as was the case with the past hypothesis, the challenge is to systematically account for how this knowledge[80] (at least probabilistically) constrains the history of our now isolated ice/water system without any specific information as to how the ice cube got into (or evolved from) the glass of water in the first place. As was argued in chapter 1, it appears that no such account can be given in a straightforward way, and we are stuck, as the best inference we can draw, believing that the ice cube was previously more melted than it is now.

4.4 Conclusion

The view just outlined makes no obvious connection between the behaviour of thermodynamic systems and the direction of time. Although the proposal just offered differs significantly from those of previous branch systems theorists, the way in which the subject of the direction of time is broached is quite similar. As Reichenbach argues, the apparent direction of time is furnished by the fact that thermodynamic branch systems are (usually) constrained to be in lower entropy states at one temporal end of their lifetime than at the other. While this is true, the problem faced by Reichenbach is to explain how this assumption or obvious empirical fact can be justified in the face of the reversibility objection. Indeed, as Sklar notes, in the end Reichenbach's view basically posits this

[80] If it is knowledge and not all the result of a spontaneous fluctuation.

parallelism as a bald assumption, not grounded in any underlying feature of the physical world.

The solution proffered here also appeals to (in Reichenbach's terminology) the space ensemble's propensity for its component systems to be formed in low entropy states in (what we call) the past rather than (what we call) the future. But unlike Reichenbach, the justification for this claim of parallelism is sought in the asymmetry of the epistemic or inferential access we have to the past and the future, rather than in some feature of the physical systems themselves. Specifically, our records and experiences point to lower entropy pasts for ice cubes in glass of warm water, the solar system and the universe as a whole. The mistake committed by Reichenbach, Albert and others is to think that, even in the face of the reversibility objection that putatively undermines these records, there is some feature of the physical world, describable in statistical mechanical terms, that can make sense of why we should think that thermodynamic systems always increase in entropy towards the future and not the past.

My position, by contrast, is that it isn't at all evident that there is any such feature, and that statistical mechanics (and by extension thermodynamics) is properly characterised as a theory of statistical inference. It just so happens that the best inferences we can make point to a lower entropy past and a higher entropy future for most systems, by virtue of the fact that we have many more records of the past being constrained than we do of the future. In short, the entropic asymmetry can be nothing other than an inferential asymmetry. Indeed, if we possessed just as many records of the future as we do of the past (including records of our own actions, thoughts and lives), we might not be able to easily distinguish between the past and the future.[81]

The records we possess point to a lower entropy past at various levels of inquiry, ranging from cosmology to the local records of my coffee mixing with the cream that was added to it. Far from undermining these records, I take it that the fact that my memory of an unmelted ice cube is veridical can be justified without appealing to the past hypothesis, or any other feature of the past beyond the time that the record references; that is, the record that the ice cube was previously less melted can be justified by appealing to the properties of my own brain, the ice/water system itself and the reliability of my sense perceptions. How this is to be accomplished is the subject of the next chapter.

[81] Unless one claims that there is a 'moving now' in the sense that time possesses some sort of intrinsic directionality or there is some fundamentally time asymmetric law.

Chapter 5 Records

This chapter examines the nature of records, and the conditions under which such records may be taken to be veridical in the face of the reversibility objection. Albert (2000), for example, states that our memories, whatever their physiological makeup, ought to be considered as statistical mechanical systems, and are hence subject to the same statistical mechanical laws and problems that we associate with the everyday statistical mechanical systems that we experience. Therefore, we should think of our memories as overwhelmingly likely to have formed as spontaneous fluctuations from equilibrium memory states. But what sort of statistical mechanical systems are our memories? Are we really justified in thinking that our memories cannot be taken as veridical based on the present state of affairs alone?

It would seem that we only have memories and records of the past, but not of the future. Is this a promising avenue of accounting for the apparent asymmetry of time? In what way, given that we are thinking of memories and records as statistical mechanical systems, are the thermodynamic asymmetry and this epistemological asymmetry related to one another? There exists an extensive literature, dating back to at least Reichenbach (1956), that looks to explore and develop this relationship (see also Sklar (1993), Horwich (1988), Albert (2000) and Earman (1974)), as well as a formidable literature on this topic that appears in discussions of Maxwell's Demon (see Leff and Rex (1990), Shenker (2000), Earman and Norton (1998, 1999)). In this chapter, I look to explore this cluster of issues, with an eye towards establishing what memories and records are, how such records and memories are established, the veracity of a certain class of memories, and how this is related to the intuitive temporal asymmetry that is among the central explanantia of this work.

5.1 Are Records Low Entropy States?

Discussions of records in the context of the foundations of statistical mechanics usually draw on the idea that records (of all imaginable sorts) ought to be thought of as low entropy states, and trace back to Reichenbach (1956), with further developments by Grünbaum (1963) among others. In order to establish this thesis, Reichenbach develops a notion of macro-entropy, an entropy measure that can be assigned to systems comprising macroscopic components. To do this, he draws on an intuitive but vaguely defined notion of macroscopic ordering where low macro-entropy states are those where the constitutive components of the system are distributed such that their arrangement can be stated by means of a simple rule.[82] Reichenbach uses the example of a deck of cards: an arrangement where all the red cards are stacked together and the black cards are together constitutes a simply stated rule for describing the system, and hence is a low macro-entropy state. Conversely, a random distribution of cards does not follow any such simple rule, and is therefore a high macro-entropy state.

To illustrate the claim that records are low macro-entropy states, Reichenbach appeals to his famous example of footprints in the sand. According to Reichenbach, the existence of a footprint found on a sandy beach corresponds to a state of low macro-entropy, and thus serves as a record of someone having walked on the beach. The reason for this is that if it were not the case that the footprint arose as the result of a past causal

[82] To be sure, not all records need to draw on this notion of macro-entropy, since many record-bearing states exist at the microscopic level.

interaction, then it would have had to arise as a spontaneous fluctuation from a high macro-entropy state where the sand particles were more or less evenly distributed on the beach. Reichenbach then argues that this is a general feature of records, namely that they are low (macro)-entropy states, and that they provide evidence for the occurrence of a past event.

Two important objections have been made against this claim by Earman (1974). First the idea of using a 'simple' rule to define macro-entropy is certainly insufficient to establish the claim. Such a rule requires some privileged way of partitioning the macroscopic components of the putative record-bearing states to form a probability distribution over their macroconditions. While the justification of using the standard measure to assign probabilities to thermodynamic systems is problematic, there appears to be no disagreement as to the fact that it is the right measure to use. However, in the case of systems described by macro-entropy, there is no obvious right way to partition the states accessible to the system. Second, it is simply untrue that all (macroscopic) records exist as low macro-entropy states. Consider the remains of a city after it has been bombed: under any plausible partitioning of the "phase space" of the city, the crumbled buildings and debris describe a high macro-entropy state. Nonetheless, it seems right to think of the remains of the city as a record of its having been bombed.

Both these objections appear to me to be on the mark and have been widely disseminated throughout discussions of the nature of records in the philosophical literature (see, for example, Sklar (1993) and Albert (2000)). However, there are several points worth considering in defence of the claim that generally records are low entropy states. First, the problem of generating the right partition of the probability space in order to infer a past event is a messy business, one that will often depend on what background knowledge one has and what causal intuitions are at work in drawing such inferences. As Earman notes, finding a footprint in the sand is, under normal circumstances, rightly construed as a record of someone having walked on the beach, but this will not be true in general. If you find the footprint on the beach and are stranded on a desert island, then it seems utterly implausible that the footprint came into being via an interaction with a foot (assuming you are certain that it was not you who made the imprint). In such a case, the assigned probability that the footprint arose as a spontaneous fluctuation would be considerably higher than it would otherwise be. Also, the inference that the footprint serves as a record or trace of someone having walked on the beach does not immediately follow from the existence of the footprint itself, since the imprint could have arisen as he result of a Martian spacecraft, with landing gear shaped in the form of human feet, coming to rest on the beach. These considerations demonstrate that background knowledge and causal intuitions can and do play important roles in establishing how one generates the appropriate probability distribution. Still, this argument does not show that there is no privileged probability distribution to use, only that establishing which one the right one is a tricky and perhaps impossible one to spell out explicitly.

Earman's argument that records need not be low entropy states seems definitive. But again, the scope of this argument is questionable. First, an important lesson can be learned from the fact that there exist simple counterexamples to the thesis that all records exist as low entropy states. Most of the cases of records that Reichenbach considers are those that function as records of other events: they are written records, computer memories or outputs of measuring devices such as barometers. They are different physical systems than the systems they are purportedly records of. Conversely, the counterexamples provided to demonstrate that records are not necessarily low entropy states are cases where the record and the system recorded are one and the same (e.g. the record of the bombed city is the bombed city). Unlike such cases that appeal to causal intuitions and other background knowledge, it appears that measuring instruments only require the knowledge that they are built to reliably generate and record data. And the data that such systems record (which include computer memories and presumably human memories) *will* exist as low entropy states according to some appropriate measure.[83]

Second, the fact that we take the remains of a city to be a record of its having been bombed relies on numerous background assumptions about what we normally expect unbombed cities to look like. While the fact that this "record" is a high macro-entropy state according to any "simple" rule for assigning macro-entropy values, it is not at all obvious that it will remain so when our background assumptions and casual intuitions are taken into account. While I do not have any general scheme to offer in place of Reichenbach's, the lesson to be learned from Earman's objections is that a facile definition of macro-entropy is not to be had and a sophisticated and correct one is not easily found since it will generally not be solely an intrinsic property of the traces themselves.

What this suggests is that a completely general account of the nature and justification of the veracity of records is not possible strictly within the context of the foundations of statistical mechanics. There exist significant problems in translating statistical mechanical concepts such as entropy and justifying the appropriate measure into macroscopic phenomena, problems perhaps too insurmountable to be treated in any

[83] This follows from the fact that measuring instruments must record events from a disjunction of possible event outcomes, which must be discernable if it is to function as a record. Some cases amount to taking one potential record-bearing state as an "equilibrium" condition (such as the remains of a bombed city), and the other as a "non-equilibrium" condition (an unbombed city). However, it is this discernability condition that picks out the appropriate measure for assigning entropy values to physical systems that are interpreted as records by identifying the "information-bearing degree of freedom" (See the following section).

detail. Nonetheless, it does seem appropriate to discuss records in the context of statistical mechanics in cases where the records themselves are statistical mechanical systems, such as human or computer memories, and possibly in discussions of the records produced by measuring instruments where there exists a obvious and natural method to identify a privileged measure. The hope is that whatever the right account of records in the statistical mechanical context turns out to be, it will be instructive in extensions to the hard cases where the records are macroscopic.

5.2 The Scope of Thermodynamics

It is clear that there are significant differences between the notions of entropy as applied to thermodynamic systems and to macroscopic objects. This raises the question as to the scope of thermodynamic concepts applied to systems such as a deck of cards. In this section, the possible relations between the thermodynamic and macroscopic notions of entropy will be explored, without the intention of resolving these complicated issues. Instead, I motivate the claim that that the account of records to be offered below can be coherently applied (in principle) whether or not one can meaningfully apply thermodynamic notions to systems that are not obviously amenable to such descriptions, although the focus will be on assuming that they are applicable.

In the first place, one might think that the extension of entropic concepts beyond their original domain of application is a mistake. As Reichenbach and others readily recognise, gases possess a "natural shuffling mechanism" through the interactions of their constituent particles, whereas say, a deck of cards become shuffled by means of artificial or intentional shuffling processes. At first glance, it is not clear that thermodynamic concepts, such as entropy and temperature, can be meaningfully assigned to most macroscopic systems.

If this is the case, then many of the problems arising from the reversibility argument lose their import. If most of the systems we encounter *do not* have an appropriate representation in the formalism of statistical mechanics, then clearly the reversibility objection does not hold much bite in terms of raising any serious sceptical worry regarding the past state of most systems of the universe previously having been in states of higher entropy than they presently are. One might still claim that the reversibility objection has bite for those states that do admit a statistical mechanical description, but one can argue that the inapplicability of statistical mechanical descriptions to most systems that we, as humans, encounter mitigates the reversibility argument considerably.

The world around us comprises both thermodynamic systems and macroscopically stable systems. While, when taken alone, thermodynamic systems are subject to the reversibility argument, it is not clear that when they are considered *together* with macroscopically stable systems that this is still the case. We believe that most everyday thermodynamic systems are formed through macroscopic interactions with their environments, whether by human intentions or through some natural process. If the belief in and records of such processes are not themselves subject to the reversibility argument (say if Cartesian Dualism is true), then we seem to have good evidence that the entropy of most thermodynamic systems was lower in the past but not the future, and no serious sceptical worry is raised. To be sure, there would remain a standard set of problems for the foundations of statistical mechanics, such as how to interpret the probabilities appearing in statistical mechanics, as well as problems concerning the reduction of thermodynamics to statistical mechanics. Further, one could still maintain, in spite of the inapplicability of thermodynamic concepts to most physical systems, that the asymmetry between past and future is somehow grounded in the entropic asymmetry.

If there does exist an insurmountable chasm between these domains of discourse along the lines just described, then I submit that the only way that there can be any sort of continuity between the macroscopic world and the formalism of statistical mechanics is if one interprets statistical mechanics along the lines espoused in this work; that is, if one understands the probabilities appearing in statistical mechanics as epistemic. Consider a joint system comprising one thermodynamic system and one macroscopic system (that, by hypothesis, does not admit a thermodynamic description) such as a deck of cards.[84] One can meaningfully assign a probability to the event that the first card turned is the two of diamonds (1/52) and that the thermodynamic system occupies some particular microstate (with value *x*). Assuming these two probabilities to be independent, it is easy to generate a joint probability value for the two events, namely $1/52*x$. If the concept of entropy is understood in the information-theoretic sense as being a general property of a probability distribution rather than a property of thermodynamic systems (and only of thermodynamic systems), then one can assign an entropy to this joint system without worrying that some conceptual error has been committed. As long as one can generate probability assignments for macroscopic events, one can still use the statistical mechanical formalism to draw inferences about the composite system.[85] Furthermore, if the MEP is understood to be a completely general formalism for drawing statistical inferences, then this would appear to be a sound project. While demonstrating the truth of

[84] The deck of cards surely has a representation as a thermodynamic system. However, here the focus here is on the relevant macroscopic properties that make a particular card the two of diamonds rather than the five of hearts.
[85] Of course, the event space, dynamics and the measure associated with the deck of cards will look very different from the ones used to describe the thermodynamic system.

the antecedent is beyond the scope of this dissertation, this possibility should not be dismissed out of hand.

Alternatively, if one is prepared to accept that even macroscopic objects admit a thermodynamic description that completely specifies all of a system's macroscopic properties,[86] then it seems that the reversibility argument, applied to any (at least mesoscopic) physical system holds some water. In this case, one might rightfully be worried that all one's beliefs and records arose as spontaneous fluctuations, and that the present universe as a whole sits at a trough in its entropy curve, having come to its present state as the result of a spontaneous fluctuation as Albert does. It is the task of this chapter to defend the veracity of our records against *this* construal of the reversibility objection.

This worry requires more than the assumption that records are thermodynamic systems. That the properties of macroscopic systems supervene on the underlying, fundamental physical description is not sufficient to get the reversibility worry, as it applies to records, off the ground. Rather, what is needed is a further assumption that (most) records, insofar as they are recognised as records *of* some particular event, are records solely in *virtue* of their statistical mechanical properties. But this is false: as Merleau-Ponty notes, "This table bears traces of my past life, I have inscribed my initials my initials on it, I have left ink stains. But these traces themselves are not past: they are present; and, if I find in them traces of some "past" event, it is because I acquire my sense of the past from elsewhere, and I carry within myself this significance" (1945, 472, my translation). What makes a physical system a record isn't an intrinsic feature of the system itself, but rather that some representational content is assigned to the system,

[86] Including, say, whether a card is the two of diamonds or the five of hearts.

however it is physically described. Although the reversibility objection implies that any macroscopic system one encounters most likely arose as a spontaneous fluctuation, it falls silent on how, or why, the system counts (to us) as a record of the past.

Consider, for instance, a model of a computer memory comprising four bits. If each of these bits is readable and distinguishable by the computer, then the entropy at the information-bearing level is zero, since each individual arrangement of the possible 16 is discernable upon inspection. However, if all one can read is the total number of bits in the 1 and 0 states, then one can assign an entropy value to the set of bits, which is maximal when two bits are in the 1 state and two bits are in the 0 state. What is important for the set of bits to function as a memory is that one can identify the information-bearing degrees of freedom: if all the bits are distinguishable and readable, then there are many more possible record-bearing states than is the case if all that can be read is the total number of bits in (say) the 1 state.[87]

Each of these bits will presumably have some appropriate thermodynamic description, but this thermodynamic description must be distinguished from its representational or intentional content, which is not part of that description. Even if one can determine a bit to be in the one state, this should not be mistaken for the claim that the entropy (in the usual thermodynamic sense) of the system is zero, since many microstates can correspond to the (relatively) macroscopic bit state of the memory.

Even if there exists an appropriate statistical mechanical description for a memory state (one that is susceptible to the reversibility objection), it is not the case that the reversibility objection can be straightforwardly applied to the *contents* of the memory,

[87] See Shenker (2000) and Bennett (2003) for a discussion of these notions. Shenker notes that this treatment of memory doesn't take a stand on controversial theses such as the claim that "information is physical".

characterised as whatever the bits represent. If the reversibility objection applies to a memory state, then something like the following must be the case: given that our set of four bits is currently in the 1111 state, then it is overwhelmingly probable that this state arose as a spontaneous fluctuation from an equilibrium state where the bits were completely thermalised. So, the 1111 state and the 1010 state have equal values of thermodynamic entropy, though their representational contents may be quite different.[88] Insofar as the reversibility objection only argues that these states most likely arose as a spontaneous fluctuation and each of these states is equally improbable, it falls silent as to why the memories happen to have the representational contents they do.[89]

Note that what is meant here by thermalisation is not that the memory was in a high entropy (equilibrium) state at the *information-bearing level*, where two bits are in the one state and two bits are in the zero state. Instead, the thermalised state must be one where there is no set of distinguishable bit states at all; that is, where there is no discernable memory content. At the level of the information content, there is no entropic distinction between the 1111 state and the 1010 state: these states are equally probable *a priori* and equally distinguishable, capable of representing distinct states of affairs and are both of equally low entropy at the thermodynamic level of description.

To illustrate the importance of this claim, consider the four-bit memory cell as representing the results of a sequence of four coin flips. If the cells are perfectly correlated with the results of the flips, then the mutual information between the cells and the coins is maximal. However, if the content of the cells and the results of the flips are

[88] This does not preclude a completely thermalised state from having some representational content. There is no *a priori* requirement that records need to be low entropy states, as Earman (1974) forcefully argues.
[89] Of course, the fact that the physical systems that comprise our own memories or a computer memory have any representational content at all depends on various complex organisational features of our brains and the computer that are not themselves accounted for by merely assigning to these systems a thermodynamic entropy.

(on the average) without any correlation, then there exists no mutual information between them. Thus, given two random variables, X and Y, if the value of X is always perfectly correlated with the value of Y, then assigning a probability distribution to X is sufficient to determine the probability distribution associated with Y, and $p(x)p(y|x)=p(x)$, and the information-theoretic entropy is $H(X)+H(Y|X)=H(X)$. However, in the absence of any correlation between X and Y, $p(x)P(y|x)=p(x)p(y)$, and the information-theoretic entropy is additive: $H(X)+H(X|Y)=H(X)+H(Y)$.

If the contents of the memory cell are perfectly correlated with its environment (the results of the coin flips), then the entropy associated with the joint system is given solely by the results of the coin flips (four bits). But in the absence of any such correlation, the entropy for the coin flips and the cells is eight bits.[90] Thus, given the state of a string of memory cells, the existence of *any* correlation (on the average) between it and its environment implies that the number of bits needed to describe the joint system can be compressed. But this is not so in the absence of correlations. If the contents of the memory and the environment are uncorrelated, then no compression is possible.

As argued above, the contents of the memory are not described by assigning the memory a particular value of thermodynamic entropy, since each of the possible disjoint states of the memory possess equal thermodynamic entropy. Thus whether or not the contents of the memory are in fact correlated with the environment is orthogonal to the assignment of a thermodynamic entropy to the memory. Yet the existence of such correlations between the contents our own memories and the environment, upon inspection of the present state of the universe, is ubiquitous. If one thought that the

[90] This may also measure the physical complexity associated with the string of memory cells (Adami and Cerf 1999).

apparent correlation between the results of the coin flips and the memory cells arose randomly, then assigning a particular sequence of flips a probability of 1/16 and likewise for a particular sequence of memory cells, the chances of perfect correlation would be 1/256. So in the absence of any assumption of causal connection between the memory and the results of the flips, the correlation between them appears highly unlikely.

Now, whatever statistical mechanical probabilities are assigned for, say, the existence of some non-equilibrium system in my environment and the probability of my memory existing in a non-equilibrium, unthermalised state, this is altogether independent of the probability distribution assigned to the representational content of the memory. If both my memory and the ice cube before are to be thought of as arising independently as spontaneous fluctuations as per the reversibility objection, then one ought to expect the contents of the memory to be uncorrelated with its environment. But then the fact that our recollections *do* seem to be correlated with environment cries out for explanation.

These features are essential to the defence of the veracity of records to be given below. First, one must appreciate that the reversibility objection, as it is applied to such memory states, operates on the thermodynamic level, and not at the information-bearing level. Second, *each* definite memory state is of equally low thermodynamic entropy, though they remain distinct memory states with distinct representational contents that are not included in their physical descriptions; so even if a particular memory state did form as a spontaneous fluctuation, we should expect there to be no reason why it would come to one particular memory state rather than another. The upshot of this last observation is that one can generate a probability distribution for distinct memory states that is *independent* of statistical mechanical considerations even if the memory states, *qua* statistical mechanical systems, are still susceptible to the reversibility objection. Why is it that my memories form a largely consistent set of beliefs about the past (or a consistent set of beliefs at all), when the reversibility objection implies that the contents of my memories should be virtually random?

It is this feature of memories, combined with an examination of the statistical mechanical features of records, that forms the groundwork for demonstrating the veracity of memories based on the present macrostate of the universe alone, discussed in more detail below. Given the memory of, say, seeing a brown table before me five minutes ago and the presence of a brown table before me now, there seem to be two possibilities to account for the present state of affairs: either the record is veridical and thus the present situation is easily explicable (even if the past state of affairs recorded is of lower entropy than the present one, or the present situation arose as a spontaneous fluctuation from some higher entropy state). In the latter case, there is a very mysterious correlation to be explained: why is it that my memory reflects a record of a *brown* table, rather than a table of another colour, or even a table *at all*, if it is not causally connected to the table I now see before me? If statistical mechanical considerations indicate that my memory came to its present low entropy state as the result of a spontaneous fluctuation, isn't it highly improbable that my memory reflects a past state of affairs consistent with the present state of affairs since the *contents* of my memory ought to be completely random? Thus, there seems to be two competing probabilistic claims at work: on the one hand the sceptical worries presented by the reversibility argument, while on the other the unlikelihood of our memories, if they did arise as fluctuations, having the coherent contents they do and their apparent correlation with the outside world.

5.3 What Records Are

Albert (2000) has presented arguments to the effect that the veracity of records turns on the truth of the past hypothesis. In effect, he claims that the past hypothesis is both necessary and sufficient to ground the veracity of our records. Before turning to criticize Albert's approach and (in so doing) construct an alternative account, it is worth examining exactly what it is Albert is claiming.

Let us begin by considering Albert's central worry as to the veracity of our records, which he claims is solved by the proposition that the universe began in an exceptionally low entropy macrocondition. For Albert,

> *What [the reversibility of the underlying dynamics] entails* is not only that almost the *entirety* of what we take ourselves to know of the past (that the entropy was lower, that certain eggs splattered in certain particular shapes, that the Roman empire existed, and so on) *fails* to follow from the world's present macrocondition + the uniform microdistribution over that macrocondition + the laws of motion, but (rather) that it follows from all this that almost all of what we take ourselves to know about the past is certainly false! What will follow… is that any book describing the Roman empire is *far* more likely to have fluctuated out of molecular chaos than to have arisen as some sort of causal consequence of the *existence* of that empire; and no amount of *redundancy* among various books, or among such books and archeological artifacts and whatever else you may be able to come up with, will change that one iota. Period. (Albert 2000, 115)

Insofar as records *are* statistical mechanical systems, Albert claims, our retrodictions about the past states of records will always force us to infer that they were in higher entropy states in the past, and therefore not veridical.

Albert provides a sketch of what records are. He tells us that there are two sorts of inferential procedures that allow one to make claims about times other than the present: those that strictly rely on using the present situation, the laws of motion and the Lebesgue measure over microconditions to predict or retrodict states of affairs and those that don't. The latter he calls records. There are two points worth noting about this definition:

1. An inference to a past (or future) state of affairs need not be exclusively by means of records or retrodiction: we can possess a record of an event that could just as easily been retrodicted. What makes something a record of the past is the fact that we do not or cannot (as a practical matter) establish its veracity by means of retrodiction, or equivalently that the record is established with incomplete knowledge of the system in question.
2. Nothing in this definition restricts records to being records of the past. Any inferential procedure towards the future that is not predictive (in this strict sense) is also a record.

The question Albert wants to ask is how we could ever have such knowledge that isn't by means of retrodiction or prediction. His answer is that records are the result of measuring instruments, which are devices that undergo an interaction that produces a record at one temporal end of the interaction, provided that the device is in its ready condition at the other temporal end. This definition, at first sight, seems rather unintuitive, but the claim can be made clear by means of an example.

Consider a barometer that prints out records of local pressure measurements. This record is a physical system that allows us to make inferences about the pressure at past times (inferences that we presumably could not make purely by applying the equations of motion over the present macrocondition of the world backwards). However, the only way that these records can be considered veridical is if the barometer was in its ready condition *before* the measurement interaction took place; that is, if the barometer was set up to make such measurements (i.e. was working properly, had plenty of ink in the cartridge, was plugged in, etc.). Thus, a record is not *just* the result of an interaction, but the result of an interaction where the measuring apparatus was (or will) be in the appropriate ready condition to produce the record.

So, a record will be veridical just in case the measuring instrument found (or will find) itself in its ready condition at the opposite temporal end of the interaction. But, Albert claims, the veracity of the record presumes that we can establish that the ready condition obtains at this opposite temporal end. And for that, we'll need a record to that effect, which in turn will require another ready condition, and so on. The regress stops when we find a ready condition that is itself not a record: the initial low entropy state of the universe (the mother of all ready conditions). So it is the past hypothesis that guarantees the veracity of our records and memories. Further, the past hypothesis is precisely the law that makes it true that we can have records of the past but not of the future, since we cannot have knowledge of ready conditions in the future by virtue of the fact that there is no "future hypothesis".

There is much to question in this account. To begin, let's look at the claim that all our memories and records should be considered false based solely on the world's present macrocondition, the laws of motion and the statistical rule. In the case (yet again) of our ice cube, presently before us and half-melted, are we justified in thinking that our memory of it being fully unmelted five minutes ago is veridical? We would like to answer 'yes'. Intuitively, what are the chances of my forming a memory *of* an unmelted ice cube as a spontaneous fluctuation that began its journey from an equilibrium condition long before the presently half-melted ice cube began to form, *and* that this ice cube, which is the subject of this randomly formed memory, happens to present itself to me now?

To be sure, the reversibility objection entails that it is overwhelmingly likely that the present macrostate of any system, thermodynamically described, finds itself at a trough in its entropy curve, and that any putative record of the past most likely arose as a spontaneous fluctuation. But this does not take into account the fact that memories are generally stable over time and apparently (though perhaps spuriously) correlated with the events they supposedly record, even in the face of the reversibility objection: our memories and records will persist well into the future and the present coheres well with the way the world ought to be if these records were veridical.

To motivate this, consider Albert's example of a textbook asserting the existence of the Roman Empire. Albert's worry was that without his past hypothesis, which asserted that the universe began in a highly non-equilibrium state, we would be forced to conclude, on the basis of the time-reversible underlying dynamics, that the book, and all the claims within it, had appeared as a spontaneous fluctuation. The basis of this claim is that whatever the statistical considerations dictate will happen in the future, we can equally well apply to the past. Yet the usual thermodynamic reasoning allows us to expect that the book will still be in existence, roughly in its same condition, five minutes from *now*. By parity of reasoning, we should expect the book to have been in existence five minutes *ago*.

What about my recollections of what I read on an earlier page five minutes ago? The following considerations provide *prima facie* evidence for the claim that my memory is veridical:

1. The book existed five minutes ago. It seems unlikely that the memory of having been reading the book five minutes ago is spurious, given that I am presently reading the very same book.
2. I am following the historical narrative of the book. If the worry that my memory of reading those sentences is all the result of a spontaneous fluctuation, then the historical account I am reading right now would most

> likely not make much sense, since they would surely be completely unconnected to the *contents* (as I remember them) of the previous pages.

Surely, whether our memories are reliable or not, whether they formed as spontaneous fluctuations, we are able to identify the contents of the memories themselves and these memories generally cohere well with each other and the present world around us. So does the reversibility objection really imply that our records are unreliable, given that it gives no reason to privilege one set of memories over another?

If anything is to function as a memory or a recording device, it must be stable in the non-equilibrium state that stores the information. Of course, no memory state can be perfectly stable, but the reliability of such records does not depend on this. Shenker (2000) discusses some necessary conditions on what it is for a physical system to act as a memory.[91] Restricting ourselves to a one-bit memory, comprising two possible and disjoint memory states, Shenker argues that for any physical system to serve as a memory, its states

1. Must be distinguishable. We must be able to determine the contents of a memory.
2. Must be amenable to manipulation. We must be able to control the content of the memories and be able to use them in reasoning and logical tasks.
3. Must be stable and reliable. The storage device must not be susceptible to thermalisation, where the entire system evolves to an equilibrium state, rendering any information that was stored irretrievable.
4. Must be non-inter-accessible. The state space region associated with one memory state must be physically inaccessible from the other, disjoint memory state.

[91] Shenker takes as her paradigm example a computer memory. While the physical implementation of memories in humans will obviously be different from a computer's, the following necessary conditions should apply equally well to human memories as to computer ones.

In the case of our own memories, the discussion is simplified by the fact that we appear to have almost immediate and unproblematic access to their contents. For measuring instruments, additional problems are raised regarding the proper conditions under which a record may be taken as veridical (i.e. the functioning of the recording device and the need for the instrument to be in its ready condition). In still other cases, such as say, textual or photographic records and physical records of events not associated with any particular measuring device[92] (say footprints in the sand, the charred remains of a house), there are additional worries regarding how the records are generated and what is to be understood as the ready conditions for the system in question. I will not try to treat each of these cases (among other possibilities) individually, though the hope is that the account given below will be sufficiently general to indicate how this can be done.

The importance of the distinguishability condition has already been discussed. Surely, whether our memories are reliable or not, whether they formed as spontaneous fluctuations or not, we are able to identify memories of the ice cube being fully unmelted five minutes ago. Similarly, condition 2 indicates that we must be able, if the need arises, to use the contents of our memories in order to reason. The last two conditions are also salient to the problem of establishing the veracity of our memories. Both of these conditions are *physical* restrictions on what it is to be a memory, rather than justifications of how or under what conditions our memories can be taken to be veridical.

The fact that memories are taken to be both stable over periods of time and are unlikely to 'switch' into other memory states indicates a strong difference between how the memory functions as a store of representational content and its status as a

[92] Of course, these can be interpreted as measurements of a certain sort. The difference I am pointing to here is that these records are not generated by any physical system specifically designed to produce records of past events.

thermodynamics system. An isolated non-equilibrium thermodynamic system is not stable, and if the system is metrically indecomposable (or suitably close), permits access to every point on the accessible phase space. Hence, whatever dynamical description is afforded to a thermodynamic system will be inapplicable to the memory insofar as it refers to its representational contents. Hence the reversibility argument cannot be straightforwardly applied to the contents of the memory.

I have argued that the issue of accounting for the veracity of memory is incorrectly posed by merely stipulating that memories (or records) are thermodynamic systems and therefore, given only their present non-equilibrium macrostate, are most likely to have formed as a spontaneous fluctuation rather than from a past interaction with a system whose properties these putative records are supposedly records of. Instead, the question should be a comparative one: is it more likely that both the record and the present state of the recorded system both arose as spontaneous fluctuations, or from a common origin that speaks to the veracity of the record? The next section offers a preliminary framework for establishing the veracity of records in the face of the reversibility argument.

5.4 When Should A Record Be Considered Veridical?

The most obvious way we can establish the veracity of our records is by demonstrating them to be in line with any retrodictions we are able to make solely on the basis of the present state of affairs, the laws of nature and any statistical postulations we are entitled to. For example, Albert (2000) considers the situation where we have several billiard balls moving around on an isolated frictionless table, and ball #5 is currently at rest. Suppose we have a record of this ball having been in motion ten seconds ago, and we want to know whether or not the ball had been involved in any collisions over the past ten seconds. Since the table is frictionless, we can answer in the affirmative as long as our record is veridical. But, and as Albert readily admits, we can establish without question that the record is veridical since we could also have come to the same conclusion by cataloguing the present positions and velocities of all the other billiard balls on the table and then applying the laws of Newtonian mechanics to determine whether ball #5 interacted with any other balls over the past ten seconds. So in such a case we need not be worried that the record arose as a spontaneous fluctuation.[93]

However, Albert's definition of a record rightly emphasises the epistemology of records, not its metaphysics. The issue in question is how we can *know*, applying our ordinary linguistic concepts, that our records are veridical, not whether they are or are not. This is most easily seen from the fact that taking the exact present microstate of the universe and applying the equations of motion to it in order to determine its status can unequivocally establish the truth or falsity of any record.[94] So, we must ask what *justifies* our belief in the veracity of records? Albert's proposal is the past hypothesis.

I think this conclusion lacks nuance. Recall Albert's definition of what a record is. A record is an inference (to the past) that is not established by means of retrodiction. He claims that the record to the effect that the billiard ball was moving ten seconds ago *cannot* be established by retrodiction. This is because it is part and parcel of what a record is that its veracity must be established on the basis of an incomplete catalogue of the

[93] The fact that the record is veridical does not entail that it did not arise as a spontaneous fluctuation. However, the probability that we possess veridical records that did arise from random fluctuations is presumably so small that we can treat them as being coextensive.
[94] I take this to be at least part of Albert's point where he writes, on this topic, that "if the complete dynamical theory of the world happens to be in accord with the premises of *Liouville's* theorem, then there is a perfectly straightforward sense in which anything that constrains the condition of the world in the past necessarily also constrains the condition of the world in the future *to exactly the same degree*." (121)

present configuration of the billiard balls on the table, for if this were not so, “whatever other information that information could subsequently be *parlayed into* would necessarily *also* be of the predictive/retrodictive sort” (2000, 118). But whether or not one *actively* justifies the belief in the veracity of the record in this way is immaterial to whether or not the veracity of the record *can* be underwritten by retrodictions. If I were to ask the physicist in the street why I should believe my record that ball #5 was moving ten seconds ago, presumably she would tell me that given the setup of the physical situation and the laws of motion, the present state of affairs would guarantee it.

This consideration is parallel to the explanation Albert offers: that the past hypothesis can ground the veracity of records. Surely no one *actively* appeals to the truth of the past hypothesis in order to justify their belief that ball #5 was moving ten seconds ago, just as no one *actively* appeals to the laws of motion as warranting the belief in the veracity of the record. Yet appealing to the laws of motion is evidently a valid way of establishing the record’s veracity, while appealing to the past hypothesis an operationally false method at best. So why should we appeal to the past hypothesis rather than the equations of motion? Answering that records (*for them to be records at all*) need to be justified by other records, which need to be justified by other records and so on until we are prepared to make an assumption about a distant state of affairs for which we have no record, tries to get more out of the definition of a record than is warranted.[95]

In the example of the billiard ball, the present state of the ball served as the record-bearing condition for the inference that the ball was involved in a collision in the local past. However, most records are not taken to be indicators of their own pasts, but are

[95] If one wants to read Albert as asking how one can reliably *acquire* records without a complete catalogue of the positions and momenta of the components of the recording and recorded system, see below.

thought to indicate the past state of affairs of some other system. More specifically, we expect records to indicate a past situation just in case the putative record-bearing system, at some point in its past, *interacted* with the system of interest in such a way as to make it possible to infer the past state of this system on the basis of the record alone. This amounts to a establishing a *recognisable* correlation between the state variables of the record-bearing system and the system it records. In a great many cases, retrodicting such a correlation is largely unproblematic since it will follow from the dynamics alone in the manner described above.[96] Just on the basis of these considerations, I do not think that Albert's general claim, namely that (absent the past hypothesis) all records most likely came into being out of molecular chaos, comes out right. However, if we think of cases of records as being low entropy states, additional worries are introduced, for such low entropy states (as per the reversibility objection) seem to have most likely arisen as spontaneous fluctuations from the past, rather than as the product of a suitable past interaction. But the fact that *some* records we possess are veridical shows that not all our records arose as spontaneous fluctuations rather than as the result of some reliable causal process.

So what about records that are low entropy states? Consider the case where I have an apparent memory of some set of physical objects (a table in a room with a glass of water and an unmelted ice cube in it five minutes ago), and access to its present state (the same table or a presently half-melted ice cube). If the reversibility objection does not imply that this memory is overwhelmingly likely to have fluctuated into existence, then it must be more likely that given the present state of affairs, the contents of my memory

---

[96] We can think of billiard ball #5's current state as also being a record of a past event of another system, namely that some billiard ball (other than #5) was involved in a collision over the past ten seconds.

most likely indicate the previous existence of the table or of the unmelted ice cube. To formalise this claim, the following notation is adopted:

H: Present state of the system (a half-melted ice cube)

U: Past state of the system as indicated by memory (an unmelted ice cube)

M: Present state of my memory

We look to show that

$$P(U|H\&M)>P(\sim U|H\&M) \tag{5.1}$$

where the probabilities for U and H are defined as the phase space volume of microstates compatible with the macrostate of the system relative to the energy hypersurface associated with the system. In the case of M, the probability must be defined for both its phase space volume *and* the representational content of the memory against the space of possible memories, as discussed in section 5.2. The expression can be rearranged to yield

$$P(H\&U\&M)>P(\sim U)P(H\&M|\sim U). \tag{5.2}$$

On the right side, we note that the unconditional probability of there not having been an unmelted ice cube five minutes ago is near unity. Further, given the fact that there was no unmelted ice cube five minutes ago, we would expect (as per the reversibility objection) that the both presently half-melted ice cube and the memory fluctuated into existence from an equilibrium state independently of each other. Hence nothing is gained by conditionalising on the fact that there was no ice cube five minutes ago. Thus (5.2) can be rewritten as

$$P(H\&U\&M)>P(H)P(M)$$

or more conveniently as

$$P(U\&M|H)>P(M). \tag{5.3}$$

If the probability of my having the memory of an unmelted ice cube and there actually having been an unmelted ice cube before me *given* the present state of the ice cube is greater than the probability of my memory having arisen as a spontaneous fluctuation, then it would seem that I am better off inferring that the ice cube was previously unmelted, and that my memory is veridical.

It might at first appear that we have solved nothing, since one might argue that it is obvious that the right side of (5.3) is always greater than the left. The relevant intuition is this: because the probability of there having been an unmelted ice cube before me five minutes ago is so tiny (even conditionalised on the presently half-melted ice cube) and the present state of the ice cube does nothing to indicate why I have the memory I do (since both likely arose as spontaneous fluctuations), the inequality could *never* be satisfied. But this is not so: consider my memory of the table having been in the same place and having the same macroscopic features five minutes ago as it does right now. Even in the face of the reversibility argument, one ought to expect these features of the table to remain stable over long periods of time, so that $P(U|H)\approx 1$.[97] Then (5.3) becomes

$$P(M|H)>P(M).$$

But surely the probability of my recollection of a table having been right in front of me five minutes ago given that there is one in front of me now is far greater than the unconditional probability that I would have such a memory, even in the absence of any causal intuitions as to how I acquired that memory. Again, the reason for this is that the *contents* of my memory are related to the present situation: if I had a memory of a pink elephant in front of me five minutes ago rather than that very same table, I would be very much inclined to think that my memories were not veridical.

[97] U is now the past macrostate of the table and H its present macrostate.

This suggests that the thermodynamic description of a record, in terms of its susceptibility to the reversibility objection, is largely irrelevant to the sceptical challenge posed by the reversibility objection itself. If the two components of the memory state are independent in the sense that the *thermodynamic* state of the memory is unrelated to its representational content, then the probabilities will be separable into two components (M_T & M_R respectively) and (5.3) becomes

$$P(U\& \ M_T \ \& \ M_R|H)>P(M_T \ \& \ M_R).$$

Further, in the absence of any presumption of a correlation between the record's thermodynamic state on the one hand, and its representational content *and* the past or present state of the system putatively recorded on the other, (5.3) becomes

$$P(M_T)P(U\&M_R|H)>P(M_T)P(M_R)$$

and can be reduced to

$$P(U\&M_R|H)>P(M_R). \tag{5.4}$$

So it would appear that the thermodynamic description of the record is irrelevant to establishing or undermining its veracity.[98]

Returning to the case of the ice cube, the above analysis indicates that the veracity of my memory to the effect that the ice cube was previously fully unmelted is a contingent claim, insofar as the relevant question to ask is whether the unconditional probability that my memories have the contents they do is more or less probable than the probability that, *given* that the ice cube is presently half-melted, it was previously unmelted and that my memory indicates that state of affairs. In order to evaluate the

[98] This cannot generally be true. Clearly some records have representational content in part due to their thermodynamic description. Furthermore, one would like to take into consideration the concern that the record ought to be stable enough over time (in a low entropy state) to outdate the past state of the system recorded.

probabilities involving M_R, one would need a catalogue of all the possible memory states I might possibly have, as well as a detailed physical description of how such memory states are physically realised.[99] However, in the case my memory of the table, it is straightforward to show that it is veridical, while in the case of the ice cube it is less than obvious. But the difference is only a matter of degree, and Albert's unqualified claim that all one's memories should be thought should be thought to have arisen out of thin air, as it were, is false.

5.5 Response to Possible Objections

At this point, I will pause to consider several potential objections to the present approach. Here I will discuss four counterarguments:

1. The reversibility objection still holds: if it is more likely that the ice cube was unmelted five minutes ago, it must be more likely that it will be unmelted five minutes from now.
2. The unmelted ice cube five minutes ago + memory must be a lower entropy state than the present one, since it will most likely evolve to the present state of affairs. Therefore, the left side of (5.1) must be less probable than the right side.
3. The left side of (5.1), even if greater than the right, will be in most realistic cases be so small that in most cases it cannot ground the veracity of records.
4. There will be apparent cases of records where the probability of the record arising as a spontaneous fluctuation will be greater than the probability that the records came into being as the result of the appropriate interactions.

I shall consider these in turn.

[99] This is necessary since the physical constitution of different memories contents might be differently realised, and thus not all equally probable.

At first glance, it might appear that the usual reversibility considerations that lead to symmetric predictions and retrodictions can be applied to this case as well. If it is more likely that the ice cube was fully unmelted five minutes ago, then it should also be more likely that it be unmelted five minutes from now. Conversely, if one thinks the most likely future evolution of the presently half-melted ice cube and my memory of an unmelted ice cube is a fully melted ice cube and a slightly dimmer memory of an unmelted ice cube, then it must follow that this should have been the case five minutes ago as well.[100]

This application of the reversibility objection is too fast. The intuition underlying this reversibility objection is that the two systems under consideration (the ice cube and the memory state) can be considered either as two independent systems, or as a single system on composite event space. While in the first case each system is vulnerable to the reversibility objection, this is not so in the second. This naïve application of the reversibility argument is misplaced, since it fails to appreciate that the *content* of the memory is correlated with the existence of the ice cube; that is, this objection makes no distinction between a memory of an unmelted ice cube and a memory of an elephant having been in the room, where the present state of the ice cube does nothing to establish the veracity of my memory of an elephant. Furthermore, the dynamics (however they might be described) that apply to the memory at *the representational level* are not susceptible to the reversibility argument because the content of the memory is not part of its intrinsic physical description, as described in section 5.2. The present objection is insensitive to this fact. So even if it is more likely that the ice cube was unmelted five minutes ago, I am not forced to the conclusion that I ought to predict that it will be

[100] Albert presented this objection to me.

unmelted again five minutes from now: if the memory was formed in the past and afterwards the evolution of the memory and the ice cube are independent, then one can reasonably infer that the ice cube will be melted five minutes from now and the memory of the unmelted ice cube will endure. While the conditional stated above is presumably true on anyone's account, the worry is about establishing the truth of the antecedent. If the analysis above is correct in that it does establish the veracity of many of one's memories, then the consequent goes through and the objection is nullified.

Yet some problems still remain, for the intuition driving this objection is that the reversibility of the underlying dynamics implies that any reasoning applied towards the past should equally well apply towards the future. But the asymmetry that remains to be explained is not whether one should make symmetric inferences to the past and future, but why records are formed in the past but not in the future. Insofar as the concern here is to underwrite the veracity of records we do have, and not to *explain why* we have records of the past but not the future,[101] this concern is mitigated by the fact that many records explicitly index the time of their putative formation. When I have a memory of an ice cube being unmelted five minutes *ago*, it is explicitly built into the content of the record that it is a record of the past because the memory specifically indexes that time.[102]

Moving on to other possible objections, it might be thought that in no instance could the left side of (5.1) be greater than the right, since the present situation (comprising a half melted ice cube and a memory of its being unmelted five minutes ago

[101] Or more generally why the records we possess are usually *of* low entropy states.
[102] Many record-bearing states do not explicitly index their time of creation. Even if one is prepared to point to a footprint in the sand as a genuine indicator of a stroller on the beach, it seems *prima facie* that there is no reason to prefer an inference to a past stroller rather than a future one. However, the inference to a stroller is not generated on the basis of strict statistical mechanical reasoning alone, but depends crucially on a vast network of background assumptions (Earman 1974), as will be discussed below. If these background assumptions about the world are themselves justifiable on the basis of records that *do* differentiate the past and future, then the inference to a past stroller can be made.

on the appropriate phase space description) would, with overwhelming likelihood, have evolved from the state of affairs five minutes ago (comprising the formation of the veridical memory of the ice cube's state and the fact that it was unmelted). If the past situation would likely evolve into the present one, this is tantamount to the claim that the probability of the past situation described by the left side of (5.1) *must* be smaller than the right, and hence less likely.[103]

In a similar fashion to the previous objection, this one rests on considering the system on a composite phase space that does not account for the representational contents of the record, where we can attribute and compare entropy valuations to each of the possible state of the system. This can surely be done, and it is true that the left side of (5.1) will describe a past system with lower thermodynamic entropy than the right side. But this, in itself, does not render the past state of affairs less likely than the present one, as the objection maintains: while it is true that the *thermodynamic entropy* of the joint system must be greater at present than it was five minutes ago, the probabilities appearing in (5.1) must be interpreted as incorporating the likelihood of the *contents* of the memory appearing as a spontaneous fluctuation, which are not described by the underlying dynamics. When this is taken into account in the form of correlations that exist between the recorded state and the present state of the world, the left side of (5.1) can still be greater than the right even though the statistical mechanical probabilities alone associated with the states U and H implies that the past state was of lower thermodynamic entropy than the present one.[104] The point here is that the memory and recorded systems are

[103] This objection was suggested to me by Tim Maudlin.
[104] One should also note that no assumption was made as to the state of the recording instrument (i.e. my sensory apparatus and brain states) five minutes ago. These systems are quite stable, in the sense that if one believes that they will be in a suitable entropic state to reliably record events five minutes from now, then

correlated and thus not independent of each other, so their individual information-theoretic entropies (which are the right entropies to use for making inferences) are not additive, although the thermodynamic entropies are.[105]

Third, as pointed out above, given the present situation, the probability that the ice cube was fully unmelted five minutes ago is incredibly tiny. Thus, it would seem that neither of these scenarios, either that the ice cube was unmelted five minutes ago *or* that my memory of that state of affairs is erroneous, is very likely according to the standard probability measure and the underlying dynamics. However, these two scenarios nearly exhaust the possible ways by which we could have arrived at the present situation where we have a record of the unmelted ice cube and a half-melted ice cube before us.

Finally, there will be instances where the inequality is not obviously satisfied. Cases where we have records of past non-equilibrium macroconditions that are very far from equilibrium initially seem to present additional difficulties. Here we are confronted with instances where it appears more likely that the record itself would have arisen as a spontaneous fluctuation than the system under consideration (for example, the probability that the Roman Empire existed versus the probability that the textbook I learned this "fact" from arose as a spontaneous fluctuation). Clearly, such situations represent the most serious challenge to the claim that we can take our records and memories at face value.

---

they were most likely in such a state five minutes ago. The importance of such "ready conditions" obtaining will be discussed below.

It should also be noted that in many cases the record will be formed by a process where the recording system comes to its record-bearing state by means of a normal thermodynamic process. If the recording device's ready condition is sufficiently stable, it may well be the case that the low entropy state of the record is easily offset by an increase in the entropy of the environment, whether or not the entropy in the past was higher or lower.

[105] This point is discussed in section 5.2.

Even if the inequality is not obviously satisfied for the ice cube or textbook, there is a case to be made that as long as it *is* satisfied for a large class of the memories and records we do possess, an inductive argument can furnished to demonstrate that our memories and records are, in general, veridical in spite of the reversibility objection. If my memories of the room, the table and the glass of water before me are all veridical according to the above analysis, then it would be odd to surmise that my recollection of the ice cube itself is spurious. Instead, the fact that I can be assured that I do have a large class of veridical memories provides strong *evidence* for the claim that my memories, generally, do not arise as spontaneous fluctuations but rather as the result of some reliable causal process. Further, this background knowledge can in principle be incorporated into the inequalities developed above thus both vitiating the claim of the reversibility objection (that my memories and everything around me arose as highly improbable fluctuations) and attesting to the correlation between the contents of my memory and the past state of the rest of the universe.

Now, it is clear that not all (perhaps not even most) of the records we possess are like our memories in that we have immediate access to their representational contents, nor is it the case that one always has immediate access to the present state of the recorded system. In Reichenbach's (1956) example of the stroller on the beach, we do not have the stroller herself before us as we do the remains of the ice cube, nor is it obvious that it is a human footprint, rather than, say, the shape of the landing gear of some Martian spacecraft, as Earman (1974) notes. But once we have access to a large body of background knowledge about human intentions and more generally the past states of the universe, a bootstrapping procedure can be put in place to draw inferences about a past stroller.

As a more concrete example, let us return to Albert's worry concerning our history textbook. First, records can be made to be arbitrarily redundant, and error-correcting procedures can be put in place. Perhaps it is true that the there is a higher probability (according to some appropriate measure) that the letters on the pages of my history textbook arose as a spontaneous fluctuation compared to the probability that the events described in that text actually took place, but the fact that there are thousands of tokens of these textbooks mitigates against this worry. What is the probability that *all* these texts, making *identical claims* (and hence having the same representational content), arose as spontaneous fluctuations compared to the probability that the Roman Empire actually existed? If it is still not large enough, we can always print more textbooks.[106]

This claim should not be interpreted facetiously. Of course, the right way to infer that the Roman Empire existed is not by counting the number of textbooks that claim historical accuracy. The point is that if one is prepared accept, given their identical physical composition and in virtue of them all making the *same historical claims* (rather than meaningless scribbles of ink on the pages) that it is more probable that all these textbooks arose from a common source, namely the printing press of a publishing company, than that *all* these textbooks came about as spontaneous fluctuations, then one can reasonably treat the existence of these textbooks as a record of the publishing process. This in itself does not establish the existence of the Roman Empire, but does function as a record of the intentions and actions of the publishers and authors of the textbook. Once we have such a record, the question turns to whether or not *these* actions and intentions

[106] Again, the relevant point here is that the apparent correlation between the *contents* of all the textbooks is to be considered highly unlikely if they all arose independently of each other as fluctuations.

themselves resulted from some appropriate network of justifiable records and intentions that, directly or indirectly, point to the existence of the Roman Empire.

Beyond textbooks, we have many other physical traces of the Roman Empire's existence: the coliseum in Rome, aqueducts scattered throughout Europe, Hadrian's Wall, a host of Roman coins, etc. Each of these, in some sense, functions as a record of the Roman Empire, and each of these testify to its past existence and mutually *reinforce* each other's status as *bona fide* records. In each of these cases, we see the same pattern: their persistence through the past two thousand years testifies to their origin, and by virtue of (5.3) and the large set of background knowledge we possess, it is more likely that these structures, coins etc. came into being at the time of the Roman Empire as part of a large causal network than as *independent* spontaneous fluctuations.

What are we to do in cases where the existence of a record is more likely to have arisen out of molecular chaos than the recorded system itself? The natural conclusion is that the putative record is not to be believed. But such cases are exceedingly rare. If I see a gas spontaneously unmix, I am right to question my memory, or to seek an alternative explanation to what I witnessed, than to think that I possess a record of an exceptionally rare spontaneous fluctuation.

So where does this leave the status of records? First, it is patently not the case that we should always believe that any putative records of the past we may possess most likely arose as spontaneous fluctuations. In many cases, our retrodictions, based solely on the present macrostate of the universe (including the representational content of our records), the standard statistical distribution over that macrostate and the laws of motion speak to the veracity of our memories. As long as the probability of the record arising as a random fluctuation is less than the probability of the past state of affairs recorded *and* the record obtaining given the present state of the system as per (5.3), it is more probable that our record is veridical rather than that it arose as a random fluctuation. Here I've presented the following considerations to the effect that at least most of our records are of this sort:

1. Many records can be authenticated by direct and unmediated appeal to retrodictions based on deterministic laws of motion.
2. Records of stable systems (like tables) are easily shown to be veridical, and provide a large set of background knowledge against which we can judge the veracity of records of non-stable systems by establishing the existence of a reliable causal process for the acquisition of records.
3. Given such background knowledge, one can adduce the veracity of records of non-stable thermodynamic systems that are subject to the reversibility objection. Furthermore, this background knowledge can indicate a temporal asymmetry in the records we do possess (but not explain it).

A framework has been put in place whereby one can evaluate the likelihood that any record described as a statistical mechanical system should be thought of as veridical. This section has claimed that a blanket statement to the effect that records, being statistical mechanical systems, most likely arose as a spontaneous fluctuation is insufficient to create a genuine problem. The fact that the present state of the universe coheres so well we the way we believe it to have been in the past serves to underwrite the veracity of our records, even if the face of the reversibility objection.

5.6 Can There Be Records of the Future?

The above discussion avoided discussing how records can be reliably generated. Albert argues that, based on the laws of mechanics and the present situation alone, our records more than likely emerged out of thin air, as it were, and not as the product of an interaction between some form of measuring device and the system it purportedly measured. Albert's solution to this is the past hypothesis, the proposition that the universe began in a highly low entropy state, which is both necessary and sufficient for the possibility of veridical records. In Chapter 1, I denied the sufficiency of the past hypothesis to ground the low entropy past of thermodynamic systems, which include records (insofar as they are thought to be states of low entropy). Here I would like to question the necessity of the past hypothesis to ground records of past events yet deny the existence of future ones (in the absence of a future hypothesis), and further to present considerations against the sufficiency of the past hypothesis that are specific to the issue of the generation of records.

Albert's argument for the past hypothesis serving this role works in two steps, designed to demonstrate that records can exist just in case the past hypothesis is true:

1. Records are the product of measuring instruments, which, in order to reliably record measurements, need to be in a "ready condition" at the opposite temporal end of the measurement process.
2. Ultimately, the only way that we can know that a given measuring instrument was in such a ready condition is if the past hypothesis is true.

Before examining the cogency of these claims, it appears that this definition fails to dovetail with the alternative definition of records that Albert presents, namely that records are inferences to times other than the present that are not achieved by means of prediction or retrodiction. Smart (1967) argues that a substantive explanation of why we are only in possession of records of the past, but not of the future, is necessary, and that simply *defining* records as necessarily being records of the past is insufficient to explain this asymmetry. Albert's solution is that such ready conditions can only obtain in the past, since such ready conditions can ultimately only be established by means of a distant ready condition, namely the past hypothesis. If there were a "future hypothesis", then we could have records of the future, but since there is no such future hypothesis we are restricted to records of the past.

In fact, Albert's approach fails to rule out records of the future. Consider a case, presented by Sklar (1993), where a radar operator sees a blip on a radar screen, due to an incoming Scud missile, and thereby infers that in the future, an explosion will take place. Clearly, in this instance, we see an inference towards the future that is not predictive in the sense that it does not follow from the current macrocondition of the radar screen, the laws of motion and the statistical postulate. Now, Albert's claim is that such an inference must be grounded by a ready condition. In the present case, the blip on the radar screen functions as the record of a future interaction, namely the explosion, so there must exist a ready condition *after* the explosion that guarantees the veracity of this record. What is this ready condition? How can we establish that it will obtain?

This example poses a clutch of problems for Albert's account. First, if this is a genuine record of the future, it calls into question the necessity of ready conditions for generating records. Second, even if ready conditions are required, it would appear that the regress argument that Albert employs to trace the existence of record back to the past hypothesis is superfluous, since if Albert's account of measuring devices is correct (if there is such a ready condition), we can only establish the fact that it obtains by appealing to the existence of another record (unless there is also a future hypothesis).[107]

[107] These considerations are especially troubling since Albert explicitly cites Sklar's discussion as being misguided and full of "whining" about what records are (113).

Alternatively, one might think that Albert's definition of a measuring device does not exhaust the means by which records can be acquired.

There are reasons to believe that these are all serious problems for Albert's account. Let us continue with the present case of the radar blip. What options are open to Albert? Presumably, he can either admit that this is a record since it is an inference that is not by means of prediction, or he can deny its status as a record, and thereby claim that it is a prediction. The first option is clearly a non-starter for Albert's account, since admitting that the blip functions as a record of the future would entail that there will be some ready condition that obtains *further* into the future, which needs to be justified by some other ready condition even further into the future and ends with the adoption of a future hypothesis. So what about the latter option?

Grünbaum (1963) presents a framework for denying the existence of records of the future, which he terms pre-records (as opposed to records of the past, called post-records). He discusses two cases where we seem to have *prima facie* instances of pre-records, but denies that these are, in an appropriate sense, genuine records.[108] The first of these are cases where the putative pre-record is the result of a prediction by means of a computer or theory-using agent, as when a physicist performs a calculation by means of a theory and then "records" the prediction, say by writing it down on a piece of paper. Such a marking is surely an advance indicator of what will happen (assuming the physicist's theory is correct and the deduction is valid), but this pre-record is surely parasitic on prediction, and thus does not constitute a genuine record. A second case Grünbaum

[108] I think this claim is best interpreted as admitting the existence of records of the future, but noting that recognising them as records depends crucially on the existence of certain records of the past, whereas records of the past do not depend on the recognition of records of the future. The implications of this observation in accounting for the apparent temporal asymmetry of records will be discussed below.

discusses is that of a record that is an effect of a common cause, whose existence can be used to infer another effect of the common cause. An example of such a pre-record is a barometer recording a drop in pressure that may be interpreted as a pre-record of an upcoming storm.

Though I agree with Grünbaum's analysis of these cases, I fail to see how the existence of such pre-records jells with Albert's account of records. It is clear that, insofar as records are *epistemic* tools of inference, they serve to indicate future states of affairs without explicitly referencing the manner by which they were generated. To be sure, the physicist's prediction on a piece of paper might be the result of a detailed calculation, but it is not through a direct knowledge of the calculation through retrodiction or prediction that I come to believe the information contained on that piece of paper. Instead, it is through my belief in the competence of the physicist and my belief that the physicist had the intention of calculating and recording the correct value on that piece of paper that I come to take it as a record of a future state of affairs. It is irrelevant that these beliefs regarding the competence and intentions of the physicist in calculating the prediction are, in some sense, records of a past prediction. What matters is the pattern of inference *I use* to formulate my belief that the physicist's prediction will come to pass, and this is completely independent of the actual method the physicist used to formulate the prediction. Insofar as *I* haven't established this record as a mere retrodiction or prediction, it is a record of the future on Albert's account.

Similar concerns exist in the case of the radar blip. Of course, there is an underlying physical description as to how the radar system works in generating the blip, and there is some feature of the blip that allows the operator to recognise it as a Scud missile, and some physical features of the missile that indicate that when the missile lands, its payload (on the assumption that it has one, which in turn depends on recognising the intentions of those who launched it) will be detonated, generating an explosion. There is a vast amount of details that the radar operator must presuppose in order to infer that an explosion will take place as a matter of straightforward prediction, but it is unlikely that the operator knows all these facts. It is hard to see how one would characterise the operator's inference that an explosion will take place as a prediction rather than as a record.

5.7 Ready Conditions

Even if the blip on the radar screen can be parlayed into predictive talk, Albert's account of records seems inadequate, both in the sense that he fails to eliminate traces of the future, *and* because the past hypothesis is in many cases neither necessary nor sufficient for the production of records, nor does it guarantee their veracity.

Let's begin by looking at Albert's description of the nature and production of records in more detail. For Albert, a record is an inference to times other than the present that involves less than the equations of motion, the present macrostate and the standard statistical measure over compatible microstates; that is, it is an inference based on incomplete information, something less than would enable an inference of the predictive or retrodictive variety. Such is the case in his paradigmatic example of the billiard ball. Albert says that reading the current state of billiard ball #5 (at rest) as a record of its having collided with another ball over the past ten seconds *depends* on the fact that it does not come by means of retrodiction, i.e. that it is an inference based on less than complete information regarding a closed dynamical system. Equivalently, records may be seen as (suitably interpreted) states of open systems, whose status as records depend on the recording system having (or not having) undergone an interaction with some outside system. For these systems to function as records, Albert claims, we must have some knowledge as to the state of the system on the other temporal end of the interaction, termed the ready state.

But Albert's regressive argument to the past hypothesis doesn't seem to accomplish what it purports to accomplish. In particular, I will argue three theses that are contrary to Albert's account of records:

1. The past hypothesis is not sufficient for guaranteeing the veridical production of records.
2. The past hypothesis is not necessary for guaranteeing the veridical production of records.
3. The past hypothesis, in the absence of a future hypothesis, cannot rule out records of the future.

Before proceeding to argue these claims, let's review exactly what the past hypothesis is, and what it supposedly explains. Albert claims that the past hypothesis is the proposition that the "world first came into being in whatever particular low-entropy highly condensed big-bang sort of macrocondition it is that the normal inferential procedures of cosmology will eventually present to us" (2000, 96). Further, it is the central claim of Albert's book that the past hypothesis solves the reversibility objection that is one of the central problems in the foundations of statistical mechanics; more precisely, the past hypothesis, the current macrocondition of the universe, the uniform statistical distribution over that macrocondition, along with the laws of motion, are sufficient to demonstrate that the second law of thermodynamics will typically hold in that it describes the (near) monotonic increase of entropy of thermodynamic systems from the past towards the future.

But what does this have to do with ready conditions? Consider the ready condition in Albert's paradigmatic case of the billiard balls. Albert claims that the currently stationary ball #5 functions as a record of a collision over the past ten seconds just in case (1) the collision was not inferred by means of retrodicting the history of the complete, closed dynamical system and (2) the ready condition (that the ball was moving ten seconds ago) obtained. Condition (1) ensures that the ball's status as a record is not *just* the result of a predictive or retrodictive inference (as per Grünbaum's requirement that pre-records that are generated on the basis of prediction do not constitute genuine records), while (2) ensures that the change (or lack thereof) of the state of the recording system can be taken as a reliable indication of the recorded event. Now, Albert claims that the only way that we can know that such a ready condition obtained is if we have a record of *its* obtaining, which in turn requires another ready condition, and so on *ad infinitum*, until we are left with the past hypothesis, the ready condition for which we need no record. He writes:

> It must be because we have a *record* of that other condition! But how is it that the ready condition of this *second* device (that is, the one whose *present* condition is the *record* of that *first* device's ready condition) is established? And so on (obviously) ad infinitum. There must (in order to get all this off the ground) be something we can be in a position to *assume* about some other time – something of which we have no record; something which cannot be inferred from the present by means of prediction/retrodiction – the mother (as it were) of all ready conditions. And this mother must be *prior in time* to everything of which we can potentially ever *have* a record, which is to say it can be nothing other than the initial macrocondition of the universe as a whole. (118)

The soundness of this argument does not immediately follow from the statement of the past hypothesis itself, and nowhere else are we given a clue as to how the past hypothesis could or would serve as the mother of all ready conditions. Is there something special about this (or any) low entropy state that makes it amenable or especially suited to functioning as a ready condition? Or does it gain its status as the mother of all ready conditions simply because we've already assumed its truth in solving the reversibility objection, and choosing another mother of all ready conditions would be unparsimonious?

Frisch (forthcoming) has offered considerations to the effect that the past hypothesis is not sufficient to ground the production and veracity of records. To demonstrate this, he considers an amended version of Albert's billiard ball example, where the balls are subject to weak frictional forces and ball #5 is presently moving and was stationary 10 seconds ago, rather than the other way around. Surely, Frisch argues, just because the state of the ball 10 seconds ago can function as a ready condition does not imply that the low entropy state of the balls 10 seconds ago can also serve as a ready condition. The low entropy past of the system can tell us that the ball did not begin moving as the result of a spontaneous fluctuation, but this is a far cry from giving sufficient reason to believe that the ball did, in fact, undergo a collision in the past ten seconds: we still need to know (somehow) that it was stationary 10 seconds ago.

This example demonstrates that the low entropy past of physical systems alone does not suffice to guarantee the veracity of our records, nor does it guarantee the reliable production of records. So what is it about the past hypothesis that allows it to serve as the mother of all ready conditions? Presumably, it is the fact that the time indexed by the past hypothesis is prior to the time for which we have any records, and therefore can stop the regress. Indeed, Albert writes that if one takes the mother of all ready conditions to be some time *after* the time indexed by the past hypothesis (call it t_{-x}), there can be records from times before t_{-x} whose ready conditions cannot be established, and furthermore such records could in fact trump those records whose veridical production *are* guaranteed by the ready condition at t_{-x} (2000, 118). But now the essential characteristic that allows the past hypothesis to serve as the mother of all ready conditions is not its claim that the universe began in a *low entropy* state, but that the universe began *simpliciter*. It appears that any claim about the distant past, whether it be that the universe began in a high entropy state or that God made the universe, could equally well serve as the mother of all ready conditions. If this is right, then the only rationale for choosing the past hypothesis over some other claim about the distant past is that we already have the past hypothesis on the books, and using another proposition would not be a parsimonious choice.

If the past hypothesis is to function as the mother of all ready conditions by virtue of the physical claim it makes rather than the time it indexes, then Albert owes us an argument. In what sense is the past hypothesis a more appropriate choice for the mother of all ready conditions than, say, the proposition that God created the universe? How does the past hypothesis' entropic claim serve to guarantee the veridical production of records? Without such an explanation, there could easily exist records of the future. All we would need is a proposition (not necessarily a low entropy one) about some time in the distant future to serve as a future ready condition. It does not take much imagination to come up with a few candidates: the heat death of the universe, perhaps a big crunch, or Armageddon.[109]

One might think that given the close link between the low entropy past of thermodynamic systems and the large number of recording states that are low entropy systems, the past hypothesis would be necessary for ensuring the veridical production of records. This is not so. Return to the question of whether or not a memory of a previously unmelted ice cube should be taken to be veridical, given the presence of a presently half-

[109] Again, it is not essential that these "future hypotheses" not be confirmed by means of prediction (indeed, we don't, at the moment, know what future hypothesis is true). What is required is that there be some claim about a time in the distant future that can serve as a ready condition for records of the future.

melted ice cube. Earlier in this chapter I argued that I should be prepared to accept a large class of my memories as being veridical. One might think that this assertion comes with an important proviso, namely that it had to be the case that I was in the appropriate ready condition to construct such a memory five minutes ago.

Presumably, the argument to the effect that I shouldn't be prepared to believe that I was in the appropriate ready condition to store my sensory impressions comes from the reversibility argument itself. Isn't it vastly more likely that no such ready condition obtained, and my belief that I was in such a ready condition is itself most likely the result of a spontaneous fluctuation; that is, isn't it the case that, as Albert argues, unless one is prepared to make an assumption about some sort of ready condition obtaining, aren't we stuck?

Things aren't nearly as bad as they would at first appear, because the argument goes by far too quickly and also fails to take into account the obvious intentional aspects of records, and also fails to investigate the nature of ready conditions, as was discussed above. In the case of Reichenbach's footprint on the beach, the appropriate ready condition is a relatively smooth beach. As such, one would accord a high *a priori* probability for such a ready condition to obtain.[110] While solely on the basis of the present imprint one should retrodict that it arose as a spontaneous fluctuation, one should not discount the possibility that a stroller was responsible for the imprint. As described in section 5.4, the right question to ask is whether it is more likely that the footprint arose as a spontaneous fluctuation or that a stroller was previously walking on the beach. The probability for this latter claim will surely depend on a vast amount of background

[110] The fact that ready conditions often take the form of *high* entropy states puts pressure on the claim that the past hypothesis can serve as the mother of all ready conditions *in virtue* of it's being a low entropy state.

knowledge (as Earman 1974 notes), but the point here is that one does not require some esoteric argumentation to make plausible the claim that a highly improbable ready condition obtained, it being a high entropy state.

In other cases a ready condition might be a relatively low entropy state: a string of 0 bits, manipulated reversibly to come to an information-bearing state, is an example of a case where the ready condition is in a state with the same entropic value as the record-bearing state. If the record-bearing condition is stable, then the record-bearing condition is just as probable as the ready state, and there is no entropic difference between the record-bearing state arising as a spontaneous fluctuation and the ready state arising as a fluctuation, and then subsequently entering the appropriate record-bearing state through an interaction with the system recorded.

Furthermore, when I have a memory of *this* presently half-melted ice cube before me being previously unmelted, rather than the memory of a pink elephant, I am not *merely* making an assumption that I was in some appropriate ready condition to create a memory of an unmelted ice cube five minutes ago. Rather, I want to claim that the best explanation of why I have *this particular memory* instead of some other recollection is the fact that the memory is veridical, and this in turn speaks to the right ready condition having obtained. Obviously, such an intuition is difficult to formalise, because one would have to generate some probability value for the memory being a memory *of* the ice cube, sitting in the very same glass of water on the very same table in the very same room against the space of all possible memories one could have.

As a second example, take the words and ideas presented on the previous page. If the worry that your memory of reading those sentences is all the result of a spontaneous fluctuation, then the words you are reading right now would most likely not make much sense, since they would almost surely be completely unconnected to the earlier discussion. Again, this fact immediately leads you to infer the veracity of your memories, and is not vitiated by any worries about whether or not you were in the right ready condition *before* beginning to read this section.

The account offered so far has been restricted to records of the fairly recent past. Some indications have been given as to how records of the more distant past are to be treated, as is the case when assessing the reasonableness of inferences that are not directly tied to memories such as when the number of (roughly) identical textbooks present in a classroom (or all over the earth) make it more likely that these textbooks arose from some common cause, namely a printing press. To be sure, such an inference draws on a wealth of background knowledge we have regarding the intention and actions of both the authors of such texts and the publishing company itself. Further, one might worry that *this* background knowledge might itself be the result of a spontaneous fluctuation rather than an accurate representation of the past intentions of human beings. Note that these are separate questions: first, should one infer *a* common cause for all these textbooks and second, exactly *what* common cause ought to be attributed to these textbooks? I take it the first can be answered with a straightforward 'yes', while the second is to be contemplated with the first answer already in hand, along with all the other justified records we possess.

A second example of inferences to the distant past can be furnished by looking at scientific data. Take, for instance, the COBE satellite data pointing to the thermalisation of the early universe via the cosmic background radiation. First, it appears obvious that the satellite itself and its delicate measuring instruments are too intricate to have arisen though pure chance. Even more unlikely is the possibility that the satellite spontaneously formed in space, where it recorded the CMB data, in such a way that it reliably transferred the data to earth, where it was processed to indicate the existence of the CMB.[111] Once these points are accepted, one can reasonably accept the data from the satellite as indicating the existence of the CMB.

The existence of the CMB is *not*, in itself, a record of the thermalisation of the early universe: the background radiation might reasonably be thought to come from some other source,[112] or itself as a spontaneous fluctuation. What *makes* it a *record* of the early universe's state is the fact that there exists a background theory of cosmology that permits us to recognise the CMB as a record of the state of the early universe, which is itself justified by independent theorizing on the basis of other empirical considerations.

The upshot of this point is that the memories and records we possess form an elaborate network of interconnected propositions that, taken by themselves, might each be thought to be the result of a spontaneous fluctuation. But on the whole, the fact that almost all these putative records seem to accord with what we should expect to be the case *presently* if they were veridical, and also form a consistent set of beliefs with our background knowledge of the world, provides a strong *prima facie* case for why we should think of our records as typically being veridical, and the recording instruments having been in their appropriate ready conditions. This point leaves open the possibility that, should we recall some event that fails to dovetail with everything we take ourselves to know about the past (or future), it ought not to be accepted as a justified memory (such

---

[111] One might worry that the reversibility objection can be applied to any stage of this reasoning. For example, one might think that the data itself, thought to be the result of measurements taken by COBE, are themselves the result of a spontaneous fluctuation, and perhaps (even) that no such satellite exists. Such a worry would have to be squared with the fact that many engineers remember building the satellite, that the funding for the project was part of the NSF budget, that numerous media sources reported on the satellite, and so on.

[112] Such as pigeon droppings, as Penzias and Wilson originally thought upon discovering the CMB.

as the recollection of a large system spontaneously decreasing in entropy). To sum up, we can offer the following indicators for a putative record to be justified as a veridical record, either of the past or of the future:

1. The recording system and the system recorded possess the correct intentional/representational relationship; that is, we recognise the correlation between the two systems as most likely being the result of some previous or future interaction.
2. It is more probable that both the present state of the record and the recorded system arose from some event common to their histories (either in the past or in the future) that explains the correlation than independently as the result of spontaneous fluctuations.
3. The putative record 'fits' with the web of other justified beliefs (or records) we currently have about singular events in the world, either in the past or the future.

Of these three indicators, the only one that is necessary is the first, since I take it to be part and parcel of what a record is that it be recognised as indicating some state of affairs at some time other than the present. The other two indicators can vie for priority. It can sometimes be the case that an event that fails the second condition will be rendered very likely when considered in the context of our background knowledge. Conversely, sometimes a record can be justified according to the second condition but not fit in with the set of other beliefs we hold, as in the case of Robinson Crusoe finding a mysterious footprint on the beach. When such cases occur and given sufficient interest, it becomes a matter of adjusting either one's background beliefs in the light of such evidence so as to bring them into line with observed events, either by revising these beliefs or searching for some additional information that brings these two indicators into line with each other. The existence of a strange footprint might lead Robinson Crusoe to revise his belief that he is alone on the island, or to reconsider the possibility that he is responsible for the creation of the footprint. Upon observing an apparent spontaneous decrease in entropy, as in the spin-echo experiments, one might reasonably think that such strange behaviour is explicable by some past constraint on the system's probability distribution, presently unknown to the observer. As indicated in the previous chapter, it is only in the absence of such (justified) revisions that one should believe the reversibility objection's conclusion, namely that these states of affairs arose as highly improbable spontaneous fluctuations.

The importance of ready conditions appears to be vastly overrated. In section 5.4, an argument was presented to the effect that in many cases, the veracity of one's records can be established without directly appealing to some ready condition. This was achieved by developing a probabilistic argument that eschewed considering exactly *how* the records were formed, instead focusing on what circumstances must obtain in order for a putative record to be considered to be veridical. If the record should be considered veridical, then it is appropriate to infer that the relevant ready condition obtained: one does not need independent grounds for believing it.[113]

It is a conspicuous feature of the classical world that its present macrostate forms an incredibly complex network of beliefs, apparent records of the past and presently correlated events that point to both future and past events to form a (more or less) logically consistent whole. If it were the case that the present macrostate of the universe

---

[113] This can be demonstrated along the lines of the derivation given in Section 5.4. Here (5.1) is amended to include the probability that the ready condition (RC) obtained given the present situation. Then a record is more likely than not to be veridical if

$$P(RC\&U|H\&M) > P(\sim U|H\&M) \qquad (5.1`)$$

This leads to the following inequality

$$P(M\&U|H)\, P(RC|M\&U\&H) > P(M).$$

The second term on the left side, introduced by explicitly considering the ready condition, should also be approximately equal to unity since one should be able to infer that the ready condition (whatever its form) obtained conditional on the record being veridical (M&U). We are then left, as before, with

$$P(U\&M|H) > P(M). \qquad (5.3)$$

formed as a spontaneous fluctuation, this would be utterly mysterious, for one would expect such a universe to be populated by completely random memories of events that are completely disconnected from the universe's present state: I shouldn't expect my textbook to be exactly where I remember leaving it, nor should I expect the contents or layout of the textbook to be exactly as I remember them.[114] Indeed, if my memories radically failed to jell with everything I observe in my environment, I would be willing to consider seriously the possibility that these memories, and my environment, were the products of spontaneous fluctuations.

The present macrostate of the world around us testifies to the veracity of the records and beliefs we have about the past, and does so irrespective of any assumption one makes about the entropic state of the universe in the distant past. Albert would presumably agree with the claims made in the previous paragraph, but holds that the only way that this can be true is if the past hypothesis obtains; that is, these obvious features of the world can only obtain if the past hypothesis is true. On this point, he writes that

> There might be a temptation to think that the mother of all ready conditions – the ready condition (that is) about which we are prepared to make an *assumption*, the ready condition from which all the others are thereafter *inferred* by means of *records* – must be whatever condition of *our own brains* it is that ensures the reliability of our *sense perceptions*. But a little reflection will show that this can't possibly be right. The evidence of our senses can (after all) be *overridden*, on occasion, by other sorts of records – and there are (after all) records of events which occurred well before we were born! (118)

First, there is a trivial sense in which Albert's point is correct. Of course, we do think that we possess records of events that occurred well before human history: the data from the COBE satellite, which is (indirectly) taken to be a record of the low entropy past of the

[114] I don't take it as a serious objection to this view that there is a wealth of evidence that demonstrates memories to be largely reconstructive. After all, this seems to presume we have genuine memories to begin with and furthermore, the extent to which our memories would have to be revised to form a logically coherent set of beliefs when in fact (according to the reversibility argument) there almost surely is no such coherence to be found in the world boggles the imagination.

universe, seems to be a case in point. Surely such a record can be veridical only if the universe was in a very low entropy state in the very distant past; in other words, if the past hypothesis is true. But this trivial observation does not establish that *unless* one is prepared to make an assumption about the past hypothesis being true, there could be no way that one could reasonably infer by some other means that the past hypothesis obtains.

What is fundamentally at issue here is whether or not any substantive *assumption* needs to be made in order to establish the veracity of such records. Albert correctly argues that one's records can only be considered veridical if some ready condition obtains. Yet the point that Albert fails to appreciate is that the establishment of such a ready condition does not require some further ready condition, as he argues. Rather, the fact that such ready conditions obtain can usually be reliably inferred from the present macrostate of the universe itself, for it would be almost miraculous if they did not obtain.

5.8 A Note on the Direction of Time

The view outlined above fails to make any strong claims regarding the temporal asymmetry of records. I have argued that there exists, in a reasonable sense, both records of the future and records of the past. When I receive a wedding invitation set for some future date, I see no reason to think that the invitation should not be categorised as a record of a future wedding. These constitute pre-records (in Grünbaum's terminology), and are generally explicable in terms of a past common cause (namely the intentions of the wedding planners) and thus in a sense derivative on past events, but this should not be seen to vitiate their status as *bona fide* records.

It is expressly not the case that we have records of the past but not the future. If there is a genuine epistemological asymmetry at work, it is that there is a difference between the methods according to which we come to *recognise* records of the future as opposed to records of the past. As Grünbaum claims and many of the examples presented in this chapter illustrate, it is apparent that the records we have of the future generally depend on our beliefs about and records of the past, whereas our beliefs about and records of the past do not depend on future records in the same way. An obvious candidate for elucidating and clarifying this asymmetry is through the principle of common cause: we see common causes operating in the past to generate correlated present/future events but do not see future causes at work in the world.

This is a strikingly plausible claim, and is adopted in various guises by Reichenbach, Grünbaum and Horwich, among others. The problem is that this solution then turns on establishing *why* this causal asymmetry holds in the world. If the causal temporal asymmetry is fundamental and irreducible, then the proposed explanation doesn't play by the rules, in the sense that we were looking for an explanation of the epistemological asymmetry on entropic grounds, based on the fundamental assumption of time-reversible laws. It is altogether too easy to generate asymmetries by sticking them in by hand, as Sklar writes:

> We think of records as being the causal result of what they are a record, and we think of causes as going from past to future. The trouble is that we don't want to presuppose any of that when we are trying to ground the very notion of causal asymmetry on that of the epistemological asymmetry, and we are trying to ground that latter only on some appropriate relation among events itself grounded on the entropic asymmetry. (1993, 399)

Alternatively, one can attempt to get at the causal asymmetry directly from the entropic asymmetry, and then turn to explaining the epistemological asymmetry. This is the path taken by Horwich (1988) and may also characterise Albert and Loewer's attempts to generate the causal asymmetry on entropic grounds.[115]

Savitt (1990) argues against Horwich, claiming that the causal asymmetry is insufficient to ground the asymmetry of records. Consider once again the situation where I have a memory of an ice cube in a glass of water being unmelted five minutes ago, and there is a half-melted ice cube before me now. A natural common causal explanation for the present situation is that there was, in fact, an unmelted ice cube an unmelted ice cube in the glass of water five minutes ago. But Savitt worries that if it is genuine possibility that my memory mis-recorded the situation five minutes ago, then knowledge of the past is impossible, in the same way that one does not know that a particular lottery ticket will lose in a fair lottery. Savitt concludes that the common cause principle cannot do the work that Horwich wants it to do, because to eliminate the possibility of mis-recordings, a strictly *nomic* notion of common cause is needed to explicate how one can reliably infer that recording devices always work properly. But if the common cause principle is derivative on a statistical claim about (say) the early universe's probability distribution, then no such temporally asymmetric nomic relation can hold on the basis on time symmetric laws.[116]

If Savitt's argument is correct, one should not look to *explain* the apparent epistemological asymmetry by appealing to common causes or to the statistical truth of the laws of thermodynamics. Rather, Savitt claims that the common cause formalism provides a more "evidential or methodological" (1990, 322) framework for making inferences, both towards the future and the past. In other words, we appeal to common

[115] See Frisch (forthcoming) for criticism of Albert and Loewer's approach.
[116] Indeed, Savitt affirms the existence of genuine records of the future.

causes to as a method to explain singular instances of correlated events, rather than as a universal nomic relation to account for all instances of correlated events.

Savitt's suggestion is in line with the approach developed in this dissertation. If the arguments presented in Chapter 2 to the effect that statistical mechanics is best characterised as a theory of inference are sound, then any asymmetry generated on entropic grounds can be nothing more than inferential, and any attempt to *derive* a genuine physical asymmetry on the basis of entropic considerations is misplaced. This chapter has been devoted to explaining how one can have confidence, given the way the world is at present and in the face of the reversibility objection, that the past is by and large the way we believe it to have been, and (less problematically) that the future will present itself to be in accordance with what we should expect to happen, to wit systems moving towards higher entropy states.

This point should not be misconstrued as the assertion that there is no genuine physical temporal asymmetry at work in the world. Indeed, if the arguments presented above are correct, then we have a wealth of justified claims about both past and future states of affairs that can serve as *evidence* for just such an asymmetry, whether it be causal, the result of human agency, contingent (as in the claim that the universe began in a highly constrained state but will not end in one), or something else. Yet it would be a mistake to think that some real physical asymmetry can be derived through statistical mechanical considerations.

6. Conclusion

This dissertation has endeavoured to demonstrate how conceiving of statistical mechanics as a theory of statistical inference concerning thermodynamic systems, where the probabilities are to be thought of as epistemic rather than objective features of the world, can dissolve some of the most historically trenchant problems in the foundations of statistical mechanics. These problems range from accounting for the asymmetric nature of time on the basis of the entropic asymmetry to the resolution of the reversibility objection to providing a satisfactory reduction of the macroscopic laws of thermodynamics to its underlying microphysical statistical basis.

If the central contention of this work is correct, then the project of attempting to ground the distinction between the past and future in the truth of the second law of thermodynamics is wrongheaded. Irrespective of any connections that may exist between, say, the nature of records and thermodynamics, if the character of temporal asymmetry cannot be explained or derived from statistical mechanical considerations, then it cannot provide a complete account of why we have more records of the past than we do of the future, nor can it specify what fundamentally distinguishes the past from the future.

In the preceding chapters, two accounts claiming to ground the direction of time in statistical mechanics have been found wanting. While the problems associated with Reichenbach's branch systems proposal are well documented, novel arguments have been presented against Albert's more recent thesis that all that is required to ground the truth of the second law are the fundamental dynamical laws, the standard statistical measure and the past hypothesis. In Chapter 1, I argued that these assumptions are not sufficient to achieve their aim and that further, more dubious assumptions would be needed in order to reach the desired conclusion. As such, a fully satisfactory account of the truth of the second law would need to look more like Reichenbach's attempt, where several (possibly) independent and tendentious assumptions would be required.

Often, the problem of explaining the direction of time is taken to go hand in hand with a solution to the reversibility objection. The reversibility argument poses two separate, but related, problems. On the one hand, it demands a physical explanation of irreversible processes, asking why the entropy of isolated subsystems of the universe was most likely lower than it presently is, and thus perhaps providing some insight into the nature of temporal asymmetry. On the other hand, the reversibility argument poses a sceptical worry, apparently implying that the past was radically different from the way we remember it to have been, and furthermore undermining the claim that any records we presently have of the past did not come into being in the way that we naturally assume they did. These two dimensions of the reversibility objection are often run together, as if the solution to the epistemic worry can only be solved if the physical problem is.

However, if one denies that there is a genuine physical explanation that must be had on the basis of purely statistical mechanical considerations, then the problem becomes inverted. In characterizing statistical mechanics as a theory of inference rather than as a physical theory, the epistemic worry takes precedence over the physical one. The fundamental question then becomes: given the macroscopic states of affairs that presently obtain in the universe, is there good reason, in spite of the reversibility argument, to think that the thermodynamic entropy of isolated subsystems was once lower and that our records of the past are, by and large, veridical?

To be sure, the answer to this question will be contingent on the character of the universe's present macrocondition. In the final chapter, I argued that the actual present macrocondition of the universe is such that it vitiates the scepticism posed by the reversibility argument: the apparent correlations between the present macrocondition of the universe and the *representational* contents of our records are such that the records we possess are more than likely veridical. In the absence of these correlations, the reversibility argument would hold: if my memories of the past (assuming I had any) were radically at odds with my present surroundings, it would be reasonable to think that the presently observable macrocondition of the universe popped into existence, as it were, as a spontaneous fluctuation.

Notwithstanding the Jaynesian approach's ability to deal with the reversibility arguments (in part by denying the import of the problem), the central objection against this approach has been that it is conceptually unfit in a number of ways to serve as a reducing basis for the physical phenomena described by the laws of thermodynamics. In chapters 2 and 3, several of these charges were examined. Three such claims were:

1. Jaynesian statistical mechanics cannot describe non-equilibrium dynamical processes.
2. Jaynesian approaches cannot provide an adequate account of the reduction of the central theoretical terms of thermodynamics such as entropy.
3. Non-physical, epistemic probabilities are in principle unfit to explain the approximate truth of the laws of thermodynamics.

I have argued that each of these worries is misplaced. The first objection seems to imply that the fundamental dynamical microlaws and a probability distribution over microstates is insufficient to describe the non-equilibrium processes according to a macroscopic partition of macroscopic observables (which is all the Jaynesian approach posits). Although a complete description of non-equilibrium processes is complex and an ongoing object of study, from a foundational perspective nothing more should be needed.[117]

With respect to the question of reduction, Chapter 2 examined different standard notions of entropy, it being the most problematic theoretical term appearing in the formalism of thermodynamics to reduce to its statistical mechanical basis. Here I reviewed well-known problems associated with the Gibbs approach to entropy, and provided novel arguments against the Boltzmann conception of entropy. By elimination, I claimed that the Jaynesian understanding of entropy is the only one that satisfies several key *desiderata* for a reducing concept of entropy, where the increase of the thermodynamic entropy of isolated systems is to be accounted for by our practical inability to follow the dynamical evolution of the probability distribution.

As for the claim that epistemic probabilities are unsuitable for statistical mechanics, it was noted that there does not exist a clear and unproblematic interpretation of probability appropriate for statistical mechanics. Although this in itself might appear to be a *tu quoque* argument, I believe that much of the thrust of these criticisms of epistemic probabilities relies on placing upon the probabilities an explanatory burden that the Jaynesian denies, namely that one must provide a *physical* explanation of why it is that the laws of thermodynamics hold. Insofar as approaches based on the MEP method reject this presupposition, the objection falls flat.

As noted in the preface of this thesis, there exists no widespread agreement as to the appropriate formal structure to use in statistical mechanics, nor does there exist a consensus as to what problems need to be addressed in the foundations of statistical

[117] This point is in agreement with Albert (2000)

mechanics or even what would constitute an acceptable solution to them. Here I have attempted to defend the claim that statistical mechanics is best construed as a theory of inference against a broad range of criticisms, and to offer positive arguments demonstrating how the Jaynesian approach can be of use in solving or dispelling the reversibility objection, one of the most difficult and stubborn problems in statistical mechanics, while providing a neat and compelling foundational understanding of statistical mechanics and its relation to thermodynamics.

References

Adami, C. and N. J. Cerf (2000). "Physical Complexity of Symbolic Sequences." Physica D **137**: 62-69.

Albert, D. (2000). Time and Chance. Cambridge, MA, Harvard University Press.

Armstrong, D. (1983). What is a Law of Nature? Cambridge, CUP.

Batterman, R. (2002). The Devil in the Details: Asymptotic Reasoning in Explanation, Reduction, and Emergence. Oxford, OUP.

Bennett, C. (2003). "Notes on Landauer's Principle, Reversible Computation, and Maxwell's Demon." Studies in History and Philosophy of Modern Physics **34**: 501-510.

Boltzmann, L. (1898). Lectures on Gas Theory. Trans. S. Brush. Berkeley, University of California Press.

Bricmont, J. (2001). Bayes, Boltzmann and Bohm: Probabilities in Physics. Chance in Physics: Foundations and Perspectives, Lecture Notes in Physics. J. Bricmont, D. Dürr, M. C. Galavotti et al. Berlin, Springer-Verlag. **574:** 3-24.

Callender, C. (1999). "Reducing Thermodynamics to Statistical Mechanics - The Case of Entropy." Journal of Philosophy **96**(7): 348-373.

Clark, P. (2001). Statistical Mechanics and the Propensity Interpretation of Probability. Chance in Physics: Foundations and Perspectives, Lecture Notes in Physics. J. Bricmont, D. Dürr, M. C. Galavotti et al. Berlin, Springer-Verlag. **574:** 271-281.

Davies, P. (1974). The Physics of Time Asymmetry. Berkeley, University of California Press.

Earman, J. (1974). "An Attempt to Add a Little Direction to 'The Problem of the Direction of Time'." Philosophy of Science **41**: 15-47.

Earman, J. (forthcoming). "The Past Hypothesis: Not Even False." Studies in History and Philosophy of Modern Physics.

Earman, J. and J. Norton (1998). "Exorcist XIV: The Wrath of Maxwell's Demon, Part I." Studies in History and Philosophy of Modern Physics **29**: 435-471.

Earman, J. and J. Norton (1999). "Exorcist XIV: The Wrath of Maxwell's Demon, Part II." Studies in History and Philosophy of Modern Physics **30**: 1-40.

Earman, J. and M. Redei (1996). "Why Ergodic Theory Does Not Explain the Success of Equilibrium Statistical Mechanics." British Journal for Philosophy of Science **47**: 63-78.

Friedman, K. and A. Shimony (1971). "Jaynes's Maximum Entropy Prescription and Probability Theory." Journal of Statistical Physics **3**: 381-384.

Frisch, M. (forthcoming). Causation, Counterfactuals and the Past-Hypothesis. Russell's Republic: The Place of Causation in the Constitution of Reality. H. Price and R. Corry. Oxford, OUP.

Garrido, P. L., S. Goldstein, et al. (2004). "Boltzmann Entropy for Dense Fluids not in Local Equilibrium." Physical Review Letters **92**(5): 056021-056024.

Gibbs, J. (1902). Elementary Principles in Statistical Mechanics. New York, Dover.

Giere, R. (1976). "A Laplacean Formal Semantics for Single-Case Propensities." Journal of Philosophical Logic **5**: 321-353.

Goldstein, S. (2001). Boltzmann's Approach to Statistical Mechanics. Chance in Physics. J. Bricmont, D. Dürr, M. C. Galavotti et al. Berlin, Springer-Verlag. **574:** 39-54.

Goldstein, S. J. L. (2004). "On the (Boltzmann) entropy of non-equilibrium systems." Physica D **193**: 53-66.

Grad, H. (1961). "The Many Faces of Entropy." Communications on Pure and Applied

Mathematics **14**: 323-354.

Grünbaum, A. (1963). Philosophical Problems of Space and Time. Dordrecht, Reidel.

Guttmann, Y. M. (1999). The Concept of Probability in Statistical Physics. Cambridge, Cambridge University Press.

Hagar, A. (2005). "Discussion: The Foundations of Statistical Mechanics—Questions and Answers." Philosophy of Science **72**(3): 468-478.

Hajek, A. (1997). "'Mises Redux'—Redux: Fifteen Arguments Against Finite Frequentism." Erkenntnis **405**: 209-227.

Horwich, P. (1988). Asymmetries in Time. Cambridge, Mass., MIT Press.

Jaynes, E. T. (1965). Gibbs vs. Boltzmann Entropies. Papers on Probability, Statistics and Statistical Physics. R. D. Rosenkrantz. Dordrecht, Reidel.

Jaynes, E. T. (1967). Delaware Lecture. Papers on Probability, Statistics and Statisitcal Physics. R. D. Rosenkrantz. Dordrecht, Reidel.

Jaynes, E. T. (1973). The Well-Posed Problem. Papers on Probability, Statistics and Statisitcal Physics. R. D. Rosenkrantz. Dordrecht, Reidel.

Jaynes, E. T. (1978). Where Do We Stand on Maximium Entropy? Papers on Probability, Statistics and Statistical Physics. R. D. Rosenkrantz. Dordrecht, Reidel.

Jaynes, E. T., (1983). Papers on Probability, Statistics and Statistical Physics R. D. Rosenkrantz. Dordrecht, Reidel.

Jaynes, E. T. (1985). Macroscopic Prediction. Complex Systems - Operational Approaches. H. Haken. Berlin, Spinger-Verlag.

Jaynes, E. T. (1992). The Gibbs Paradox. Maximum-Entropy and Bayesian Methods. P. N. G. Erickson, and C. R. Smith. Dordrecht, Kluwer**:** 1-22.

Jeffreys, H. (1967). Theory of Probability. Oxford, OUP.

Kemeny, J. G. and P. Oppenheim (1956). "On Reduction." Philosophical Studies **5**: 6-19.

Khinchin, A. (1949). Mathematical Foundations of Statistical Mechanics. New York, Dover.

Lavis, D. A. (2004). "The Spin-Echo System Reconsidered." Foundations of Physics **34**: 669-688.

Lebowitz, J. (1994). Time's Arrow and Boltzmann's Entropy. Physical Origins of Time Asymmetry. J. J. Halliwell, J. Perez-Mercador and W. H. Zurek. Cambridge, Cambridge University Press**:** 131-146.

Lebowitz, J. (1995). Microscopic Reversibility and Macroscopic Behaviour: Physical Explanations and Mathematical Derivations. 25 Years of Non-Equilibrium Statistical Mechanics. J. J. Brey, J. Marro, J. M. Rubi and M. S. Miguel. Berlin, Springer-Verlag. **445:** 1-20.

Leff, H. S. and A. F. Rex, Eds. (1990). Maxwell's Demon: Entropy, Information, Computing. Princeton, Princeton University Press.

Lewis, D. K. (1994). "Humean Supervenience Debugged." MInd **103**: 473-490.

Loewer, B. (2001). "Determinism and Chance." Studies in History and Philosophy of Modern Physics **32**(4): 609–620.

Malament, D. B. and S. L. Zabell (1980). "Why Gibbs Phase Averages Work: The Role of Ergodic Theory." Philosophy of Science **47**: 339-349.

Merleau-Ponty, M. (1945). Phénoménologie de la Perception. Paris, Gallimard.

Morrison, M. (2000). Unifying Scientific Theories: Physical Concepts and Mathematical Structures. Cambridge, CUP.

Nagel, E. (1961). The Structure of Science. New York.

Popper, K. (1959). "The Propensity Interpretation of Probability." British Journal for the

Philosophy of Science **10**: 25-42.

Price, H. (1996). Time's Arrow and Archimedes' Point. Oxford, OUP.

Reichenbach, H. (1956). The Direction of Time. Berkeley, University of California Press.

Ridderbos, K. (2002). "The Coarse-Graining Approach to Statistical Mechanics: How Blissful is our Ignorance?" Studies in History and Philosophy of Modern Physics **33**: 65-77.

Ridderbos, T. M. and M. L. G. Redhead (1998). "The Spin Echo Experiments and the Second Law of Thermodynamics." Foundations of Physics **28**: 1237-1270.

Rosenberg, A. (1994). Instrumental Biology or the Disunity of Science. Chicago, University of Chicago Press.

Savitt, S. (1990). "Epistemological Time Asymmetry." PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association **1**: 317-324.

Seidenfeld, T. (1986). "Entropy and Uncertainty." Philosophy of Science **53**: 467-491.

Shannon, C. and W. Weaver (1949). The Mathematical Theory of Communication. Urbana, University of Illinois.

Shenker, O. (2000). Logic and Entropy, Phil-Sci Archive. **2006** http://philsci-archive.pitt.edu/archive/00000115/.

Sklar, L. (1967). "Types of Inter-Theoretic Reduction." British Journal for Philosophy of Science **18**(2): 109-124.

Sklar, L. (1993). Physics and Chance. Cambridge, Cambridge University Press.

Smart, J. C. C. (1967). Time. Encylopedia of Philosophy. P. Edwards. New York, Macmillan.

Sober, E. (1984). The Nature of Selection. Cambridge, Mass., MIT Press.

Uffink, J. (1995). "Can the Maximum Entropy Principle be explained as a Consistency

Requirement?" Studies in History and Philosophy of Modern Physics **26**: 223-261.

Uffink, J. (1996a). "The Constraint Rule of the Maximum Entropy Principle." Studies in History and Philosophy of Modern Physics **27**: 47-79.

Uffink, J. (1996b). "Nought but Molecules in Motion." Studies in History and Philosophy of Modern Physics **27**: 373-387.

van Fraassen, B. (1989). Laws and Symmetry. Oxford, Oxford University Press.

van Kampen, N. G. (1994). The Gibbs Paradox. Essays in Theoretical Physics in Honor of Dirk ter Haar. W. E. Parry. Oxford, Pergamon.

van Lith, J. (2001). Stir in Stillness:A Study in the Foundations of Equilibrium Statistical Mechanics, Utrecht University.

Vranas, P. B. M. (1998). "Epsilon-Ergodicity and the Success of Equilibrium Statistical Mechanics." Philosophy of Science **65**: 688-708.

Winsberg, E. (2004a). "Can Conditioning on the "Past Hypothesis" Militate Against the Reversibility Objections?" Philosophy of Science **71**: 489-504.

Winsberg, E. (2004b). "Laws and Statistical Mechanics." Philosophy of Science **71**: 707-718.

Yi, S. W. (2003). "Reduction of Thermodynamics: A Few Problems." Philosophy of Science **70**(5): 1028-1038.