

REDUCING THERMODYNAMICS TO STATISTICAL
MECHANICS: THE CASE OF ENTROPY*

It has become increasingly clear in recent years that the foundations of *statistical mechanics* (SM) are almost as puzzling as—and in some senses are more puzzling than—the foundations of quantum mechanics. Lawrence Sklar's¹ important review of the field presents many different approaches to the foundational problems in SM. Here, the uninitiated encounter a bewildering variety of schemes aimed at explaining why SM works. What strikes one about these projects is that perhaps all of them crucially rely on surprisingly dubious assumptions. Many projects invoke indispensable “rerandomization” posits, some with Hamiltonian dynamics “turning on” and “turning off.” Others work only in the limit of infinite particle number and/or time. Still others rely on special dynamical properties that real systems are very unlikely to possess. The arguments supporting these assumptions are typically murky. Perhaps worse, they leave unclear their relevance to the project of providing a mechanical explanation of why classical or quantum-mechanical systems behave thermodynamically. Philosophically minded observers may feel at a loss when they cannot trace the reasoning from a foundational project to an explanation of why (say) particular gases tend to fill their available volumes. I argue here that we are often correct to feel this way. Most of the approaches to the foundations of SM have severed their link with the original foundational project, the project of demonstrating how real mechanical systems can behave thermodynamically.

My argument focuses on two different frameworks in SM: the Gibbsian and the Boltzmannian. I concentrate on their different definitions of entropy, since they underscore the more general conceptual differences between the two approaches. I argue against the prevailing view that the Holy Grail of modern foundational research

* Early versions of this paper were presented at Cambridge University's Σ Club, Oxford University's Colloquium in Honour of Robert Weingard, and the University of Hull's Conference on the Mathematical and Conceptual Foundations of Physics. I am grateful to the audiences at all three talks for useful discussion. I also want to thank Guido Bacciagaluppi, Bob Batterman, Joel Lebowitz, Michael Redhead, T.M. Ridderbos, Janneke van Lith and especially Shelly Goldstein and Jos Uffink. Finally, I thank the LSE's Centre for Philosophy of Natural and Social Sciences.

¹ *Physics and Chance: Philosophical Issues in the Foundations of Statistical Mechanics* (New York: Cambridge, 1993).

should be a Gibbsian entropy that strictly mirrors the behavior of the thermodynamic entropy. The search for the Grail has not only been in vain; it has also been misguided. Attempts to find a Gibbs entropy that avoids the so-called “paradox” of the conservation of the fine-grained entropy have proved unsuccessful. Worse, the very goal of these Gibbsian approaches is radically at odds with the potential reduction of thermodynamics to SM. Inasmuch as one seeks a reduction (in some sense) of thermodynamics to SM, one loses her way by pursuing the Grail. The Boltzmann entropy, neglected for much of this century, therefore deserves serious reconsideration in light of the failure of mainstream Gibbsian SM to tackle the above foundational project. The subsequent discussion more or less follows the order of this sequence of claims.

Before beginning, let me happily concede that for the practice of science, Gibbsian SM is usually to be preferred. Since the values of all the entropy functions I discuss agree at equilibrium, my arguments are necessarily philosophical in nature. My topic is the much-ignored *interpretation* of SM. How we interpret SM affects nearly all of the foundational problems plaguing the subject: for example, understanding reversibility, reduction, the role of probability. No less than the notorious problems in quantum mechanics, these problems cry out for philosophical investigation. Indeed, the demand for foundational investigation is especially pressing in SM; for it has long been recognized that, as E. T. Jaynes² expresses it, it “is the enormous *conceptual* difficulty of this field which has retarded progress for so long” (*ibid.*, p. 392). Yet philosophers and theoreticians have rarely tackled the subject. As Jos Uffink remarks, this neglect is no better illustrated than by the fact that Paul and Tatiana Ehrenfest’s classic 1911 survey article still represents, in many respects, the state of the art.³

I. REDUCTION AND IRREVERSIBILITY

Consider the canonical example of a gas confined by a partition to a section of a container. Remove the partition. In a short span of time, the gas will relax to equilibrium; that is, it will spread evenly throughout the box, possessing a uniform pressure and temperature. What accounts for this regularity? According to thermodynamics, the gas’s approach to equilibrium occurs mainly due to the

² “Gibbs versus Boltzmann Entropies,” *American Journal of Physics*, xxxiii (1965): 391-98.

³ Uffink, “Nought but Molecules in Motion,” *Studies in History and Philosophy of Modern Physics*, xxvii (1996): 373. The Ehrenfests’s classic article is translated and reprinted as *The Conceptual Foundations of the Statistical Approach in Mechanics* (New York: Dover, 1990).

Second Law. The Second Law states that an extensive state function, the total differential $\delta S = \delta Q/T$, defined only for equilibrium states, is such that $\Delta S \geq \int \delta Q/T$. Loosely put, for realistic systems, this implies that in the spontaneous evolution of a thermally closed system, the entropy can never decrease and that it attains its maximum value for states at equilibrium. We can therefore think of the Second Law as driving the gas to its new, higher entropy equilibrium state. Using this concept of entropy, thermodynamics is able to capture an extraordinary amount of diverse phenomena under one simple law. Remarkably, whether they are gases filling their available volumes, two iron bars in contact coming to the same temperature, or milk mixing in your coffee, they all have an (indirectly) observable property in common: their entropy increases.

Coffee, milk, and iron bars are not elementary individuals in our ontology, of course. Exactly what are is a matter of great controversy. But so long as one believes that quasi-periodic, *time reversal invariant* (TRI) laws govern elementary matter, we can frame our problem in terms of classical mechanics without fear that quantum mechanics will side-step our problem.⁴ (The Schrödinger equation is TRI and quasi periodic, so interpretations of quantum mechanics that do not modify or interrupt the Schrödinger evolution will face a problem isomorphic to the one discussed here. Other interpretations may face similar problems, but one must deal with these on a case-by-case basis.) In classical mechanics, the ontology is particulate, and the coordinates and momenta of all of the particles in the system determine the state of a system. Temporal evolution of states is given by Hamilton's equations of motion, $dq_i/dt = \partial H/\partial p_i$ and $dp_i/dt = -\partial H/\partial q_i$. I shall represent the microstates of N -particle systems by a point $X \in \Gamma$ and a $2Nf$ -dimensional phase space spanned by Nf momenta and Nf configuration axes, where f is the number of degrees of freedom of each particle. As the system evolves through time, this representative point will trace out a trajectory through Γ .

We thus have two descriptions of our gas: one mechanical and the other thermodynamical. We would like to know how they relate to one another. This kind of problem is quite familiar to philosophers.

⁴ Roughly put, TRI means (for deterministic systems) that, if we take any nominally possible sequence of states of a system, then the reverse sequence of "temporally reflected" states is dynamically necessary. That the equations are quasi periodic, or recurrent, means the following: for systems confined to a finite volume and a finite energy and for almost all (with respect to the standard measure) microstates X corresponding to these systems, if a system starts in some state X , then at some time in its subsequent evolution it will return either to X or to an arbitrarily small neighborhood Δ_X around X .

Philosophers of biology, psychology, and all of the special sciences are busy trying to demonstrate how the properties and concepts used by their science are (not) reducible in some sense to the properties and concepts of “lower level” sciences. The problem is one of intertheoretic reduction.

One might expect the present reduction would be comparatively easy to complete successfully. After all, few take the thermodynamic properties of gases to demand a new ontology. The thermodynamic properties of gases are, broadly speaking, material properties, and we simply need to determine the mechanical properties to which they correspond. The trouble we immediately face, however, is that the properties of fundamental physics are much different from the properties of thermodynamics. The classical mechanical equations of motion are TRI and quasi periodic, yet the Second Law of thermodynamics is not. Focusing on reversibility, Hans Reichenbach⁵ puts the problem the following way:

The elementary processes of statistical thermodynamics, the motions and collisions of molecules, are supposed to be controlled by the laws of classical mechanics and are therefore reversible. The macroprocesses are irreversible, as we know. How can this irreversibility of macroprocesses be reconciled with the reversibility of microprocesses? It is this paradox which the physicist has to solve when he wishes to account for the direction of thermodynamical processes (*ibid.*, p. 109).

Strictly speaking, this is perhaps not the best way to frame the problem. The fallacy of composition threatens. In philosophy of mind, for example, few think that component physical systems of the brain ought to possess the same properties as the brain (for example, intentionality). Similarly, we should not expect aggregates of particles to have the same properties as individual particles. There is no paradox here. In principle, values of momenta and positions exist that will reproduce a thermodynamic system’s motion, at least if one is willing to posit conspiratorial initial conditions.

The worry is really about whether thermodynamics can emerge in some natural way from mechanics. Thermodynamic systems—like chairs, tables, and similar systems picked out by our common object language—are nothing more than complicated arrangements of physical properties. Very few would disagree with this. Thermodynamics does not threaten physicalism. In this very weak sense, thermodynamics is *already* “ontologically reduced” to mechanics. Unlike chairs, however, the thermal features of objects play a role in an ex-

⁵ *The Direction of Time* (Berkeley: California UP, 1956).

traordinarily powerful science: thermodynamics. We have no hope of defining 'chair' in the language of physics—we do not believe it has a simple nondisjunctive mechanical counterpart. The concept is just too coarse. But thermodynamics defines its concepts comparatively precisely, and it is closer to the level of fundamental physics than is our common object language. It is such a powerful tool in organizing a vast range of phenomena, so striking and explanatorily powerful, that we suspect we can find a simple, nondisjunctive counterpart for it at the (statistical) mechanical level. We suspect the thermodynamic entropy is either identical (a posteriori) to this mechanical correlate or at the very least supervenient upon it. Indeed, we *hope* SM has *already* found this correlate. In this stronger, more Nagelian sense of reduction, where the reduced theory is a "definitional extension" of the reducing theory, or at least an approximation to it, the jury is still out concerning the reduction of thermodynamics to SM.

As Reichenbach recognizes, the problem of macroscopic irreversibility is a warning shot for those seeking such a reduction. For as we know from a century's discussion of Boltzmann's notorious H-Theorem, the TRI and quasi periodicity of the dynamics place severe restrictions on any mechanical definition of entropy. Rather than rehash what should be a familiar debate, let me briefly state what I take to be their main point. Consider the following assumptions:

- (A) Entropy is a function S of the dynamical variables $X(t)$ of an individual system.
- (B) $S(X(t)) = S(X^*(t))$, where '*' indicates a temporal reflection.
- (C) The system is closed (Γ is bounded).

If (A), (B), and (C) hold, then the TRI of Hamilton's equations implies S cannot increase monotonically for all initial conditions; and if (A) and (C) hold, the quasi periodicity of the solutions to these equations implies S cannot increase monotonically for all time. In short, if S is a function of the dynamical variables of an individual system, then S cannot exhibit monotonic behavior. This captures the essence of Joseph Loschmidt's, Ernest Zermelo's, and Edward Culverwell's objections to the H-Theorem (see Sklar, pp. 39ff.). As we shall see, it is symptomatic of the confusion in the field that the lessons drawn from this implication are so varied.

Perhaps the most common response in the physics literature is to take this implication as effectively a reductio of assumption (A). (Assumption (B) is never challenged; in section III, (C) will be challenged.) The thermodynamic entropy is understood as displaying monotonically increasing behavior; therefore, many reason, the SM

analogue of the entropy cannot be a function of the dynamical variables of an individual system. We should instead conceive of entropy in terms of some function of a collection of systems. Clearly, the fact that $S(X(t))$ cannot monotonically increase is compatible with some *other function* defined on an *ensemble* of systems monotonically increasing. The move to ensembles is sometimes advertised as a “way around” the reversibility and recurrence properties of the classical dynamics.⁶ Indeed, Jerome Rothstein explicitly treats it as such in a paper aptly entitled “Loschmidt’s and Zermelo’s Paradoxes Do Not Exist.”⁷ Noting that Loschmidt and Zermelo’s paradoxes apply to individual systems, Rothstein argues that *since* entropy is a function rightly applied to ensembles, the paradoxes do not threaten proofs of entropy increase. With similar reasoning, many circles have come to see denying (A) as the “answer” to the reversibility and recurrence paradoxes. The reader of Sklar’s survey will recall, for instance, the denial of (A) as a refrain repeated with the introduction of nearly every approach.

Of course, this foundational problem is not the primary reason for denying (A). The primary reason is that Gibbs’s approach to SM denies (A), and the Gibbsian approach dominates the field (which is the reason why Rothstein feels he can use as a premise the claim that entropy is a function on ensembles). Before criticizing the move of avoiding the paradoxes by denying (A), we need to compare the Gibbsian and Boltzmannian approaches to SM.

II. GIBBS AND BOLTZMANN ENTROPIES

In his famous work of 1902, J. W. Gibbs⁸ worries about the paradox of the specific heats, a problem plaguing the field at the time. He famously reacts to it by developing a very abstract approach to the subject, one not so dependent upon the actual constitution of classical systems. When he arrives at his various definitions of entropy, he

⁶ More than denying A is necessary to avoid recurrence in quantum mechanics. If a quantum system has a finite number of degrees of freedom, then recurrence among individuals is matched by recurrence *of the ensemble as a whole*. This is the lesson of a theorem by S. Ono, *Mem. Fac. Kyushu University*, xi (1949): 125; and (independently) Ian Percival, “Almost Periodicity and the Quantal H Theorem,” *Journal Of Mathematical Physics*, ii (1961): 235-39. The theorem shows that the expectation value $\langle F \rangle_t$ of any dynamical variable F in an arbitrary Gibbs ensemble is an “almost periodic” function of t . This implies that the probability of a particular value of a macroscopic state must come arbitrarily close to its initial value at arbitrarily large times. To avoid this, one usually goes to the so-called “bulk limit,” in which $N \rightarrow \infty$ while keeping N/V finite. Doing this, however, only serves to aggravate the problems already saddling the denial of (A); see section III.

⁷ *Foundations of Physics*, iv (1974): 83-89.

⁸ *Elementary Principles in Statistical Mechanics* (New Haven: Yale, 1902; reprint, New York: Dover, 1962).

tentatively offers them as “analogies” to the thermodynamic entropy rather than as mechanical reductions of it.

On Gibbs’s approach, one considers a fictitious infinite ensemble consisting of all the microstates compatible with a given macrostate. The macroscopic parameters thus pick out a distribution of possible representative points, which we treat as a fluid in Γ . Upon this ensemble we specify a normalized density function $\rho(p, q, t)$. The mean values of a phase function with respect to ρ describe the system’s thermodynamic properties, except for the entropy and temperature. For the entropy, Gibbs chose the expression:

$$(1) S_{EG}(\rho(X)) = -k \int \rho(X) [\log \rho(X)] d\Gamma$$

where the integral is over Γ and k is Boltzmann’s constant. This is the so-called “fine-grained” entropy. For isolated systems, we use the microcanonical probability distribution in (1) and this will match (up to an additive constant) the value for the thermodynamic entropy. (1) will also allow us to define a temperature. Remarkably, using these definitions Gibbs can recover the familiar thermodynamic relations for systems in equilibrium.

Although Gibbs himself originally introduced ensembles as a kind of mathematical trick useful for calculating observed values of macroscopic parameters, influential physicists such as D. Ter Haar and R. C. Tolman⁹ dissented from this view. They claimed that the adoption of ensembles is not merely pragmatically desirable but is warranted by the inevitable imprecision of our measurements. Today, it is fair to say that this is the orthodox opinion. Their use in SM is so ingrained that, as D. Lavis¹⁰ expresses it, one needs “a considerable act of will...to avoid all references to ensembles” (*ibid.*, p. 260). Oliver Penrose,¹¹ for instance, is able to begin a major review of the foundations of SM with the claim that the fundamental problem facing the subject is justifying the choice of the correct ensemble.

How much, if at all, the popularity of Gibbs’s methods owes to the denial of (A) as a response to the recurrence and reversibility paradoxes is a historical question I am not competent to judge. Surely, the calculational ease of the Gibbs approach is the primary reason

⁹ Ter Haar, “Foundations of Statistical Mechanics,” *Reviews of Modern Physics*, xxvii (1955): 289; and Tolman, *The Principles of Statistical Mechanics* (New York: Oxford, 1938; reprint, New York: Dover, 1979).

¹⁰ “The Role of Statistical Mechanics in Classical Physics,” *British Journal for the Philosophy of Science*, xxviii (1977): 255-79.

¹¹ “Foundations of Statistical Mechanics,” *Reports on the Progress of Physics*, xlii (1979): 1937-2006.

why it dominates other approaches. But Gibbs's worries about the paradox of the specific heats surely struck a chord with many like-minded scientists. Scientists often cited this paradox as a partial justification of instrumentalism and forms of positivism.¹² It would be interesting to know whether these philosophical ideas played any role in the subsequent popularity of Gibbs's theory.

Let us now turn to Ludwig Boltzmann's¹³ approach, in particular, his later approach after the H-theorem. In contrast to the Gibbs entropy, we define the Boltzmann entropy S_B for the actual microstate of an individual system. Consider some microstate X . X corresponds to a macrostate $M(X)$, which, in turn, is compatible with many different microstates. We wish to determine the relative volume in Γ corresponding to all the microstates giving rise to M . To accomplish this, we must partition Γ into compartments, such that all of the microstates X in a compartment are macroscopically indistinguishable, that is, they share the same thermodynamic features. We do this however we can, but for the ideal gas we use six-dimensional μ -space. Here, we can write the total energy of the system as a sum of one-particle energy functions on μ . The μ -space description is important only for discovering which microstates are macroscopically indistinguishable. Once we determine this, we are then able to associate with each M a certain volume of Γ , $|\Gamma_M|$, which is the integral of the time invariant Liouville volume element over Γ_M . Up to additive constants, S_B is then given by:

$$(2) S_B(X) = k \log |\Gamma_{M(X)}|$$

where $|\Gamma_M|$ is the volume in Γ associated with M . S_B can be thought of as providing a relative measure of the amount of Γ corresponding to the M determined by X .

The volume corresponding to a particular macrostate depends, of course, on the macrostate in question. By far the compartment with

¹² See Henk de Regt, "Philosophy and the Kinetic Theory of Gases," *The British Journal for the Philosophy of Science*, XLVII (1996): 31-62, and references therein.

¹³ See, for instance, *Vorlesungen über Gastheorie*, Volumes 1 and 2 (Leipzig: Barth, 1896-98)—translated by S. G. Brush as *Lectures on Gas Theory* (New York: Dover, 1964); *Theoretical Physics and Philosophical Problems: Selected Writings*, B. McGuinness, ed. (Dordrecht: Reidel, 1974). For recent work on Boltzmann's theory, see J.L. Lebowitz, "Boltzmann's Entropy and Time's Arrow," *Physics Today* (September 1993): 32-38, and "Macroscopic Laws, Microscopic Dynamics, Time's Arrow and Boltzmann's Entropy," *Physica A*, CXCIV (1993): 1-27; J. Bricmont, "Science of Chaos or Chaos in Science?" in *The Flight from Science and Reason: Annals of the New York Academy of Science*, DCLXXV (1996): 131-75; S. Ito, M. Mizutani, and T. Niwa, "On the Time Evolution of the Boltzmann Entropy," *Journal of Mathematics of Kyoto University*, XXVI, 1 (1986): 1-11.

the greatest volume in Γ is the one each of whose points corresponds to a Maxwellian distribution (in μ -space). When N is large enough for the system to be macroscopic ($\approx 10^{23}$) it occupies almost all of Γ . If we think of a gas in a box, it is easy to see why this is so. There are simply many more ways for a gas to be uniformly distributed in a box, with uniform temperature and pressure, than for a gas to occupy a small corner of the box, with nonuniform temperature and pressure. Almost all of the available Γ , therefore, corresponds to a compartment representing a macrostate in thermal equilibrium. If we adopt the "natural" Euclidean probability metric over Γ , we may say that thermal equilibrium is the most probable state of an individual system. At equilibrium, for macroscopic N at a given energy and volume, S_B is approximately equal to the thermodynamic entropy. In section v, I shall return to our discussion of S_B in order to clarify how it works and how it deals with various difficulties.¹⁴

What is the relationship between the Boltzmann and Gibbs entropies? As the Ehrenfests and others have pointed out, they take the same values at equilibrium. If the generalized microcanonical ensemble associated with M is defined as $|\Gamma_M|^{-1}$ if $X \in \Gamma_M$ and 0 otherwise, then the two are identical at equilibrium (Lebowitz). This is not surprising, of course, for both approaches essentially seek to calculate the region of motion in Γ available to a system.

Nevertheless, *conceptually* the two entropies are *very* different. Most significantly, S_B is a function of the microstate of an individual system, in contrast to S_{FG} , which is a function of a probability distribution on a fictional ensemble. This is the cause of many other deep differences between the Gibbsian and Boltzmannian approaches. For instance, the two approaches conceive of equilibrium quite differently and, as a consequence, they also conceive of the SM counterparts of the thermodynamic laws differently. For Gibbs, equilibrium is characterized by a particular probability distribution over Γ . At equilib-

¹⁴ The μ -space procedure is not in general useful for nonideal gases. Unfortunately, due to the neglect of the Boltzmann approach, no work (to my knowledge) has been done on providing a substitute for nonideal gases. Nevertheless, since work began on this paper two different approaches have been suggested to me, one by Uffink and Janneke van Lith and another by Sheldon Goldstein. To object to the Boltzmann approach for this deficiency is therefore premature. It would be to mistake a lack of *practical sophistication* (due to its neglect) for an *in principle* difficulty. The difficulty is clearly not a problem of principle with the definition of S_B . There is a fact of the matter about which subsets of Γ are macroscopically indistinguishable. Furthermore, one will focus on macroscopic functions that are invariant under permutation of particles. Each subset of Γ will have a volume proportional to the number of permutations compatible with the macrostate, and when N is large one or more of these subsets will become vastly larger than the rest.

rium S_{FG} will not fluctuate. By averaging over an ensemble, it has already built into its conception of equilibrium all of the improbable members of the ensemble whose S_B spontaneously decreases. Thus, in Gibbs approach, it is correct to say one has fluctuational phenomena *at equilibrium*, so the entropy at equilibrium remains constant as in thermodynamics. On Boltzmann's approach, by contrast, one fluctuates *away from* equilibrium, that is, away from the most probable value. This is because we characterize equilibrium by a particular proper subset in Γ , the one whose points all correspond to a nearly Maxwell-Boltzmann velocity distribution. We know that S_B must fluctuate away from equilibrium simply because it is a function of X . The quasi periodicity of the dynamics for X implies that (for almost every microstate) S_B cannot *always* remain in equilibrium.

Thermodynamics states that once an isolated system achieves equilibrium, it stays in equilibrium forever (until disturbed by external influences). Gibbsian SM agrees. Gibbsians suppose that the probability distribution characterizing equilibrium is time independent. Boltzmannian SM disagrees, however, as it abandons the idea that equilibrium is stationary in time. The Boltzmann approach balances this affront to thermodynamics by retaining the idea that equilibrium and entropy are properties of individual systems. The Gibbs approach pays for its strict agreement with the thermodynamic laws by relinquishing the idea that entropy and equilibrium are properties of individual systems. We are confronted with a difficult choice. As Sklar puts it, the choice is of "sticking with a concept most like the thermodynamic concept and exchanging the law of thermodynamics for a statistical regularity [Boltzmann] or of moving to an ensemble concept and keeping the unexceptionless law [Gibbs]" (*op. cit.*, p. 365). In the next two sections, I argue that there are some reasons for preferring the Boltzmannian single-system approach to the Gibbsian one when thinking about the reduction of thermodynamics to SM.

III. GETTING THE GIBBS ENTROPY MOVING

No sooner does one begin thinking about Gibbs fine-grained entropy than one notices the following consequence of Liouville's equation. Liouville's equation, $d\rho/dt=0$, a consequence of the Hamiltonian equations of motion, states that the swarm of representative phase points representing an ensemble move like an incompressible fluid. The Hamiltonian flow preserves the measure of any set of points if and only if their corresponding density function satisfies Liouville's equation. Liouville's equation implies, however, that the fine-grained entropy is invariant in time:

$$(3) \quad dS_{FG}(\rho)/dt = -k \int d\rho/dt [1 + \log \rho] d\Gamma = 0$$

S_{FG} does not change, yet the thermodynamic entropy clearly does. This is the so-called “paradox” of the conservation of the fine-grained entropy.¹⁵

Although this “paradox” is standardly taken to present a huge difficulty for the theory, it is not immediately clear why this is so. The thermodynamic entropy is defined only at equilibrium. The entropy gradient holds only between equilibrium states. Explaining this fact does not require that we discover a function that is well defined and increasing during the interval between two equilibrium states. Finding a function defined only at equilibrium but whose value at a later equilibrium state is higher than at an earlier equilibrium state is enough. The Gibbs approach has no problem doing this. Since the macroscopic parameters change between two such states, we define a new Gibbsian ensemble for this state and this will match the thermodynamic entropy. The knee-jerk reaction to the conservation of the fine-grained entropy therefore needs re-examination. While, in the words of J. Bricmont (*op. cit.*), it may be “interesting or useful” to extend the concept of entropy to nonequilibrium situations, it is not necessary in order to account for thermodynamic irreversibility.

If there is a problem with S_{FG} , it is instead that it is “*not fair* to choose a new ensemble description of the system at a later time” (Sklar, p. 54; my emphasis). The ensemble at the later time should be, it seems, the Hamiltonian time development of the earlier ensemble; otherwise, the statement that Hamilton’s equations govern the system is a fiction. Unfortunately, as Gibbs found, no interesting function of the earlier ensemble approaches the value of that function for the later ensemble. The time development of the earlier ensemble will not lead it to approach the later one in any way. This fact is a more serious worry about the fine-grained entropy, not that it does not monotonically increase in the intervals between equilibrium states. The problem, therefore, really has much more to do with whether the fine-grained entropy can provide a plausible reduction basis for the thermodynamic entropy than it does with the simple fact that S_{FG} does not move.

In any case, my present target is not the fine-grained entropy itself, but rather the multitude of attempts to get S_{FG} moving between equilibrium states. Beginning when the Ehrenfests first complained about Gibbs’s treatment of irreversibility, the last eighty-five years have witnessed the near universal acceptance of the idea that to explain “irreversibility,” we must find some SM entropy function that behaves monotonically in time. These approaches either retain S_{FG} ,

¹⁵ A. Wehrl, “General Properties of Entropy,” *Reviews of Modern Physics*, 1 (1978): 221-58, here p. 227.

and therefore modify the dynamics to escape Liouville's theorem, or they define a new Gibbs-like entropy and try to show that it behaves monotonically with time. The goal, which has become the Holy Grail of modern research in the foundations of SM, is to find an entropy function that strictly mirrors the thermodynamic entropy. In other words, the goal is to find some function that always increases until reaching a maximum value at which it remains forever. Despite producing many interesting results, I believe these approaches are unsuccessful, unnecessary, and wrong-headed. Showing that this is the case is the task of the remainder of this section and the next.

Before getting into the arguments, let us first ask *why* one would demand having a *strict* statistical mechanical version of the Second Law, one exhibiting non-TRI behavior. Why believe nomic irreversibility is necessary for the task at hand? One often finds the claim that thermodynamic entropy increase is strictly irreversible, and therefore, that we must base any mechanical counterpart of entropy on something not TRI. This belief is based on a deep confusion about what it takes for one theory to correspond to another. All we need to do is explain how something *apparently irreversible* emerges from something reversible, and we *need not* accomplish this through non-TRI laws. We merely need to reproduce the *image* of thermodynamics in SM at the appropriate scale, not the actual mathematical properties of the thermodynamic laws. When trying to recover one higher-level theory from a lower-level theory, we merely need to recover the phenomena supporting the theory of interest, not a literal statement of the reduced theory. TRI laws plus temporally asymmetric initial conditions can of course mimic the phenomena governed by thermodynamics, so non-TRI laws are not necessary.¹⁶

Despite the great effort put into getting S_{FC} moving, no successful approach has yet emerged, nor does any seem likely to emerge. Un-

¹⁶ Some—like N. Krylov, *Works on the Foundations of Statistical Physics* (Princeton: University Press, 1979)—object that the reliance on initial conditions in the recovery of the thermodynamic laws would threaten the lawfulness of thermodynamics. If we found the laws of thermodynamics to be dependent upon special mechanical initial conditions, they would not be universally valid and would therefore lose their explanatory force (for example, their ability to support counterfactuals). This strikes me as an overreaction. If the laws of thermodynamics can only be recovered mechanically by adopting special initial conditions, then we shall merely have discovered that thermodynamics is a special science. Thermodynamics would be like biology in its need of special mechanical initial conditions. But that does not imply that thermodynamics loses its modal force any more than the same situation implies biological generalizations have no modal force. (Or if it does imply that, then the problem is already a huge one. In either case, the point is that there is no *special* problem here for thermodynamics. See my “What Is ‘The Problem of the Direction of Time?’” *Philosophy of Science* (Supplement), LXIII, 2 (1997): 223-34.)

derstanding irreversibility this particular Gibbsian way puts one in a hole out from which one cannot plausibly dig. Obviously, I cannot do justice to the rich variety of such attempts in one essay, never mind one section of an essay. Fortunately, most of the major objections to the alternatives are already well documented. For this reason, I shall quickly remind the reader of some of these difficulties, refer her to the relevant readings, and then make a more general objection. I think it is primarily important that the reader get a flavor of the *difficulty and cost* of extracting oneself from the "problem" of the conservation of S_{FG} .

Coarse-graining. The usual response to the conservation of S_{FG} is to devise new notions of entropy and equilibrium, in particular, the coarse-grained entropy and a notion of equilibrium suitable for it. The motivation for these new notions is solely as a means of escaping the above "paradox," though it is usually defended with appeals to the imprecision with which we observe systems.

In the coarse-grained approach, one divides Γ into many small finite cells of volume Ω_i . One takes the average of ρ over these cells, obtaining the coarse-grained probability $\bar{\rho}$, and attributes the value of $\bar{\rho}$ to all $X \in \Omega_i$. By substituting $\bar{\rho}$ for ρ in (1), one defines the coarse-grained entropy $S_{CG}(\bar{\rho}(\rho, \Omega_i, t))$. This entropy need not be constant in time. Due to Liouville's theorem, the volume ρ remains constant; however, the density of ρ in each cell Ω_i will change. Equilibrium is now defined as the state in which ρ has fibrillated uniformly throughout the available Γ . One then hopes to demonstrate that S_{CG} will monotonically approach equilibrium.

One such attempt is Tolman's generalized H-theorem. Tolman shows that, if one assumes that $\bar{\rho}_0 = \rho_0$, where $\bar{\rho}_n$ means $\bar{\rho}$ at time $t=n$, then it will follow that $S_{CG}(\bar{\rho}_0) \leq S_{CG}(\bar{\rho}_1)$, and that $S_{CG}(\bar{\rho}_0) \leq S_{CG}(\bar{\rho}_2)$, and so on. The proof is notoriously weak, however. First, it is not a consequence that $S_{CG}(\bar{\rho}_1) \leq S_{CG}(\bar{\rho}_2)$. Nothing follows about the relative values of two S_{CG} s if both are later than the initial condition. Second, the assumption that $\bar{\rho}_0 = \rho_0$ is a very strong and contentious one. Third, and perhaps most worrying, the irreversible behavior of S_{CG} arises almost solely due to the coarse-graining. Although coarse-graining will undoubtedly be part of our explanation of entropy increase, due to the large-scale differences between micro and macro, surely it is not the whole story. Thermodynamic behavior does not depend for its existence on the precision with which we measure systems. Even if we knew the positions and momenta of all the particles in the system, gases would still diffuse through their available volumes.

Others have tried to resurrect this style of justification by using arguments that assume a Markov condition. If the transition probabilities for \bar{p} do not depend on the actual past state of the system, then it is possible to show that this kind of rerandomization will spread \bar{p} uniformly throughout Γ . The question is whether this sort of rerandomization is compatible with the underlying deterministic dynamics. To answer this, many physicists and mathematicians turn to the idea that the underlying dynamics is “mixing.”

Mixing. In an attempt to make plausible the idea that a Markovian postulate can describe correctly the ensemble evolution, many researchers appeal to certain features of the dynamics. The property of mixing is particularly popular. Loosely put, an ensemble is mixing if, as $\bar{p}(t)$ spreads out in Γ , its ‘threads’ are uniformly distributed in the limit when $t \rightarrow \infty$. Advocates of this approach hope that real systems are mixing. If they are, this is supposed to warrant the use of a Markov postulate in the coarse-grained approach.

Despite many beautiful results, the relevance of mixing to the approach to equilibrium is doubtful. Apart from the problems of being valid only in the infinite time limit, the mixing justification also suffers from the well-known “measure zero” problem. Perhaps worse, mixing is also neither necessary nor sufficient for thermodynamic behavior. Many systems exhibiting no “irreversible” behavior have the property of mixing (for example, a particle in a rigid one-dimensional box).¹⁷ And due to the KAM theorem, we know many real thermodynamic systems are not mixing systems (Sklar, pp. 169ff.).

Master equations. Master equations are a kind of generalization of Boltzmann’s kinetic equation. They are linear Markovian rate equations for the probability distribution of a system. The idea here is to show that the probability distribution for a system approaches the equilibrium distribution under certain conditions. These demonstrations are successful only when one uses the infinite time limit, continually rerandomizes the system, and posits special random initial ensembles. The justification for any of these conditions and the relevance of the result if successful is obscure.¹⁸

Interventionism. No realistic thermodynamically closed system is mechanically closed. Treating them as such is clearly an idealization. A real gas in a box is subject to all sorts of ignored, uncontrollable perturbations: for example, the gravitational pull of the moons of Jupiter, quantum fluctuations of the walls of the container, cosmic rays, and so

¹⁷ S.K. Ma, *Statistical Mechanics* (Singapore: World Scientific, 1985), p. 446.

¹⁸ See Sklar for critical review.

on. Interventionists take these perturbations as essential to why entropy increases. They hold that these uncontrollable perturbations warrant abandoning Hamiltonian dynamics and hence Liouville's theorem. This rejection of the classical dynamics allows them to write an equation for the fine-grained entropy which does not have the effect of leaving it an invariant of the motion. Modeling the "random" external influences with a stochastic equation of motion, perhaps as P. Bergmann and J.L. Lebowitz¹⁹ do, they are able to circumvent the conservation of the fine-grained entropy. They therefore have no need to "brave the ambiguities of the coarse-grained entropy formulae," as Penrose puts it (*op. cit.*, p. 1938). (In terms of our earlier problem, interventionists deny assumption (C), but most deny (A), too.)

As an answer to the foundational questions in SM, however, I think interventionism is a nonstarter. If classical or quantum mechanics is the fundamental theory of the world, then it governs the environment of closed systems, too. If so, then Hamilton's and Liouville's equations govern the environment, too, and thus the total fine-grained entropy of system + environment does not increase. One might then refer to the fact that the system + environment is itself subject to uncontrollable external perturbations from its environment, and that with respect to these perturbations, the entropy of the combined system increases, and so on a la Reichenbach. But clearly this is a mug's game. Unless one is willing to bite the bullet and claim that the TRI quasi-periodic laws of mechanics only govern the universe as a whole, and not its subsystems, then this method of pushing the explanation back further and further will always be in vain. But to do this is to claim that the laws of nature do not govern the very items that are evidence for believing the laws of nature, that is, the subsystems of the world. The fundamental problem with interventionism is that it wants to have it both ways. It does not want to offer new fundamental laws (for example, as Ilya Prigogine or David Albert²⁰ do) but it does not wish to be constrained by the current TRI deterministic laws either.

¹⁹ "New Approach to Nonequilibrium Processes," *Physical Review*, xcix (1955): 578-87. See T.M. Ridderbos and M. Redhead, "The Spin Echo Experiments and the Second Law of Thermodynamics," *Foundations of Physics*, xxviii (1998): 1237-70, and references therein for defenses of interventionism.

²⁰ Albert points out that according to the GRW interpretation of quantum mechanics, the indeterministic and non-TRI laws of nature may be helpful to an understanding of irreversibility; see "On the Foundations of Quantum Mechanics and the Approach to Thermodynamic Equilibrium," *British Journal for the Philosophy of Science*, xlv (1994): 669-77. Prigogine, by contrast, wants to change even classical mechanics; see "A Unified Foundation of Dynamics and Thermodynamics," *Chémica Scripta*, iv (1973): 5-32.

In addition, the approach founders on justifying the stochastic terms used in modeling the environment and on justifying the necessary time asymmetric effect of the environment. If the environment does not operate in a random way, interventionism will not work. We would have no guarantee that systems will tend toward equilibrium. (If we do not want such a guarantee, or we have other reasons to think systems will tend toward equilibrium, then interventionism is unnecessary.) But we cannot justify these random influences (as they often are) by appealing to our inability to control or predict the effects of the environment. Uncontrollable and unpredictable influences, we know well, are not the same as genuinely random influences. If deterministic laws govern the environment, they cannot be genuinely stochastic. A similar problem arises concerning time asymmetry. If the approach is to avoid implying that entropy increases toward the past, time asymmetric laws must govern the environment if the approach is not to beg the question. Again, unless we adopt new, in this case non-TRI, laws of nature for the environment, this assumption will not be warranted. Finally, we can ask whether it is really plausible *or necessary* to suppose that (say) the reason why gases expand through their available volumes is in part due to the locations and sizes of the moons of Jupiter.

Radical approaches. Given the problems with all of these approaches, it is not surprising that some physicists turn to quite radical approaches to the problem of irreversibility. One is due to Jaynes. This is the view that entropy increase is due to the informational state of the observer and the nature of statistical inference. The other is due to the so-called “Brussels School” led by Prigogine and his theory of subdynamics. This approach alters the fundamental notion of a state of a system, replacing trajectories by probability distributions. While one can sympathize with both approaches and agree that they are both ingenious in different ways, in my opinion neither view can be sustained.²¹

Finally, though each approach has its own particular difficulties to overcome, they all (with the possible exception of the Brussels school) share a very general problem, what Sklar dubs the “parity of reasoning problem.” Either the reasoning behind the approaches also works for the past direction of time as well as the future, in which case the theory predicts entropy increase toward the past, or it smuggles in an unwarranted temporally asymmetric assumption,

²¹ For criticism of the second view, see Robert Batterman, “Randomness and Probability in Dynamical Theories: On the Proposals of the Prigogine School,” *Philosophy of Science*, LVIII (1991): 241-63. For criticism of both views, see Sklar.

analogous to Boltzmann's notorious *Stosszahlansatz*. This assumption typically either conflicts with the Hamiltonian dynamics or merely ends up pushing the question back a step. Falling victim to the first horn of this dilemma implies the theory of interest is wrong half the time; falling victim to the second half implies the theory is either inconsistent or incomplete. For this reason, along with the multitude of less general difficulties, I consider it unlikely that any of these approaches ever will yield a satisfying explanation of irreversibility.

IV. REDUCING THE THERMODYNAMIC ENTROPY

We have just seen how hard it is to find a Gibbsian entropy function that monotonically increases in time. But even if one manages to skirt the above objections and define an entropy with the proper behavior, that alone does not make a function suitable as a mechanical counterpart of the thermodynamic entropy; for, as Jaynes reminds us when discussing Boltzmann's H-function: "Merely to exhibit a mathematical quantity which tends to increase is not relevant to the second law unless one demonstrates that this quantity is related to the entropy as measured experimentally" (*op. cit.*, p. 392). With this sentiment in mind, I wish to make a sweeping objection against any approach that defines an entropy function that increases strictly monotonically with time.

Monotonically increasing ensemble entropies are generally regarded as the mechanical version of the thermodynamic entropy. But are they suitable for this role? Sklar notes well the difference between entropy as a feature of individual systems and entropy as a feature of probability distributions. He cautions us not to move too quickly in our reduction of the thermodynamic entropy to the ensemble entropy, remarking that this identification is fraught with difficulty. He claims, rightly I believe, that to assert an "identity" between features of ensembles and thermodynamic features would be to misrepresent the relationship between the two theories (*op. cit.*, p. 361). But he never adjudicates the debate between the single- and many- system approaches as they relate to reduction. One suspects Sklar thinks of the Gibbsian reconception of entropy as a property of ensembles rather than a property of individual systems as simply an instance of concept extension or revision. The idea of concept extension or revision is a familiar one to philosophers. When applying a concept to a new context, we may slightly reform its meaning. The reformed concept may not preserve the original term's intension or extension, yet its meaning may not have changed sufficiently to warrant calling it a new concept. What counts as sufficient? Clearly, there is no hard and fast rule telling us how to distinguish reformed

concepts from new concepts; rather, we make intuitive judgments about whether the reformed concept changes from the original only in “peripheral” features. Undoubtedly, we should expect such an extension in the present case. Whenever we explain the concepts of one theory in terms of those of another we inevitably use concepts with features different from the reduced theory’s concepts. When we move from thermodynamics to SM, for instance, we extend the concept of temperature to absolute negative values for spin systems. On a SM treatment of temperature, this *seems* to make sense (but see Uffink, *op. cit.*). Since “being positive” arguably is not essential to the thermodynamic temperature (like its relationship to pressure is), this appears to be a case of concept extension and not concept substitution. Yet is the switch from an entropy defined for individual systems to one defined for ensembles merely an instance of a surprising concept extension?

It may be. Perhaps we can understand features of ensembles as properties of single systems. Gibbs himself seems to have thought it possible to eliminate reference to ensembles, although it is not entirely clear what he had in mind.²² Nevertheless, denying that entropy is a property of individual systems *in order to avoid recurrence and define a monotonically increasing entropy function* does not so much elastically stretch the original concept as it does completely break from it. We can see this in a variety of more or less striking ways. The simple argument that follows strongly suggests that no monotonically increasing function of time should be identified as the reduction basis of the thermodynamic entropy.

Suppose one is able to define a function of an ensemble, $G(\rho)$, that increases monotonically with time until reaching a maximum value, at which it remains forever. G is defined for an ensemble, for we know from above that if it were not (that is, if assumption (A) were affirmed) there could be no such G . Finding a function like G , or a proof of such behavior, is the goal of many in the foundations of SM. If we could find such a G , we could identify it with the thermodynamic entropy, and the thermodynamic Second Law would remain a strict, unexceptional law.

Suppose we wish to describe a gas diffusing throughout its available volume into a new equilibrium state. We know that, though it does not matter to the definition of G , the actual gas has a microstate X . We also know that X , whatever it is, gives rise to the

²² See Jan von Plato, *Creating Modern Probability* (New York: Cambridge, 1994), p. 83, and references therein.

macrostate M we see before us. These are merely the assumptions we make when we say thermodynamics is in some sense reducible to mechanics. They are completely uncontroversial. Surely, the gas has a microstate, and surely whatever microstate it occupies corresponds to the macrostate we see. Now, if we make one more meager assumption, the irrelevance of G is assured. That assumption is that classical or quantum mechanics always govern the evolution of X . For Poincaré's recurrence theorem implies that for almost all microstates, at some subsequent time in their evolution they will return either to X or to an arbitrarily small neighborhood Δ_X around X . Since this neighborhood can be arbitrarily small, the difference between X and $X + \Delta_X$ will not matter to the macroscopic description of the state. X and $X + \Delta_X$ will correspond to the same M , *whatever* it is.

We can now exhibit the problem with the proposed identification. The behavior of our function G supposedly accounts for the Second Law, even though this explanation is compatible with the real microstate of the gas being one that subsequently leads to a macrostate that the system nearly *already occupied*. G will continue to increase, or remain constant, even if the direction of the thermodynamic entropy turns around. How do we know the thermodynamic entropy would turn around in such a circumstance? We know this because the thermodynamic entropy is a function of merely the endpoints of the evolution and of the macrostate of the system. Let S_1 be the thermodynamic entropy of macrostate M_1 , which corresponds to the gas confined to a small section of a container. Let S_2 be the thermodynamic entropy of macrostate M_2 , which corresponds to the subsequent state of the gas after it has spread evenly throughout its new volume. The thermodynamic entropy, whatever it is, is given by $\Delta S_A = S_2(M_2) - S_1(M_1)$. But when the system recurs to an earlier macrostate—in this case from M_2 to M_1 —its thermodynamic entropy, whatever it is, is given by $\Delta S_B = S_1(M_1) - S_2(M_2)$. The changes ΔS_A and ΔS_B cannot be of the same sign. The evolution of the thermodynamic entropy therefore will change direction during recurrence. This is obvious, especially when we consider a film of gas diffusing run backward, but it helps to go through the reasoning to expose the uncontroversial premises of the argument. One might object and point out that the thermodynamic entropy *does not* in fact decrease. Yet that misses the point: *if* the system in question is a mechanical system it *can* recur (and *will* recur, given enough time). It is physically possible, maybe even actual, that the value of the thermodynamic entropy not even approximately match the value of G .

Let us now reconsider the mechanical explanation of the gas relaxing to equilibrium provided by function G . Apparently, its explanation of why the system approaches equilibrium will “work” even if the diffused gas suddenly starts to concentrate in one small corner of the container! This is the upshot of avoiding the recurrence theorem. Any theory that explains entropy increase with a function such as G will be forced to say its SM entropy is increasing, or remaining constant, even while its thermodynamic entropy decreases. The only other option is to give up one of the meager assumptions that created this predicament. But since these are so minimal, to abandon one of these is to give up the game.

Clearly, even without going into all the gory details concerning the many uses of the term ‘reduction’, we cannot say that the thermodynamic entropy reduces to G in any sense.²³ The entropy to be reduced does not supervene upon the reducing entropy, yet supervenience is usually considered a minimal condition for any reduction. Lacking supervenience, it is hard to imagine how the other metaphysical relationships philosophers sometimes desire of reduction could hold. We cannot view G as *causing* the thermodynamic entropy’s behavior, since the two do not even covary. Nor can G be a posteriori *identified* with the thermodynamic entropy. There are nomologically possible worlds where the two entropies do not have the same behavior. Since any scientific identification holds across all physically (and perhaps metaphysically) possible worlds, the fact that there is a possible world where the two entropies differ implies they cannot be identical. So on any theory of reduction remotely Nagelian in spirit, where one looks for a correspondence (even approximate correspondence) between a concept at the reducing level and a concept at the reduced level, we do not have a reduction. This is the price of making G ’s evolution independent of the individual system’s evolution: G becomes irrelevant to the individual system. The drive to find a function that behaves like G , we see, runs directly contrary to the desire to have a mechanical reduction of the thermodynamic entropy.

V. BOLTZMANN REVISITED

SM need not and should not search for a mechanical counterpart for entropy that exhibits monotonic behavior with time. For when we find such an entropy, we immediately know that it is irrelevant to the behavior of real individual systems in the world. If this is right, how should we react to this conclusion?

²³ For some of the gory details, see M. Spector, *Concepts of Reduction in Physical Science* (Philadelphia: Temple, 1978) and references therein.

One reaction might be to take it (either happily or sadly, depending on one's other views in philosophy of science) as announcing that we cannot have a SM counterpart of the thermodynamic entropy. The thermodynamic entropy, then, would be viewed as a concept more like 'chair' than like 'carbon'. One might then either adopt a nonreductivist or an eliminativist stance toward the thermodynamic entropy. The options and moves available here are familiar from the philosophy of mind literature on folk-psychological terms.

I think it is a bit early, however, to give up the search for a SM counterpart of the thermodynamic entropy. The thermodynamic entropy is a real (indirectly) observable property of individual systems. That the major Gibbsian approaches to SM have not found it does not imply a mechanical counterpart does not exist. Rather, if we cannot find it in these approaches, then we ought to look elsewhere.

One entropy worthy of more investigation is the *unaltered* Gibbsian fine-grained entropy S_{FG} . We saw that the conservation of S_{FG} is not the problem many believe it to be. Although there remain questions about it, I think it premature to close the door on S_{FG} from the perspective of reduction.

Another entropy worthy of much more thought is the largely neglected Boltzmann entropy—in particular, the reconstructed, post-H-Theorem entropy recently championed by Lebowitz (*op. cit.*) and others. The philosophical literature on SM almost completely ignores this entropy. And despite a few articles articulating the theory, I believe it is fair to say that many prominent physicists remain confused about this approach. Certainly, the letters replying to the Lebowitz article make this abundantly clear.²⁴ The reason for the neglect lies in a complicated web of different influences. Boltzmann's later theory is often confused with his earlier, infamous H-theorem.²⁵ Moreover, his own writing is difficult to follow and his reasoning is often convoluted. When combined with several common philosophical mistakes regarding Zermelo and Loschmidt's objections to Boltzmann, and the generally confused state of the field, such misunderstanding is not terribly surprising. In any case, though I cannot guarantee that this approach is problem free or even well developed, I believe it represents a promising candidate for under-

²⁴ *Physics Today* (November 1994): 113-15.

²⁵ Jaynes shows that the H-function differs (nonnegligibly) from S_{FG} whenever the potential energy of a system is relevant (that is, when there are interactions). Consequently, since S_{FG} agrees with S_b at equilibrium, the H-function differs (nonnegligibly) from S_b too.

standing why mechanical systems tend to approach thermal equilibrium. Due to its neglect and the confusion surrounding it, I believe it is worth describing the position in greater detail than I did in section II. I would like to do what I can in this section to shed some light on potentially confusing points. I shall not claim that the theory yet offers a satisfactory reduction of the thermodynamic entropy. Although I believe it to be more promising than mainstream Gibbsian approaches, I shall show that it still faces some challenges.

What kind of behavior should we expect from S_B ?²⁶ Unlike S_{FG} , S_B is *not* an invariant of the motion, despite Liouville's theorem. To see this, consider the volume of macrostate M in Γ at $T = 0$, $|\Gamma_M|$. I shall represent the time development of the points in $|\Gamma_M|$ at $T=t$ as $|\Gamma_M^t|$. Liouville's theorem implies that $|\Gamma_M| = |\Gamma_M^t|$. But this does not imply that the volume of the *new* macrostate M^* in Γ at $T=t$, $|\Gamma_{M^*}^t|$, is equal to $|\Gamma_M^t|$. Indeed, because most of the states in $|\Gamma_{M^*}^t|$ will not usually have evolved from the relatively tiny region $|\Gamma_M|$, $|\Gamma_{M^*}^t|$ is not only *unequal* to $|\Gamma_M^t|$, but it is typically *not even close* to as large as $|\Gamma_{M^*}^t|$. Thus, S_B is free to vary with time.

Furthermore, it is overwhelmingly likely to increase. One can see this qualitatively as follows. Consider a gas in equilibrium confined to a small portion of a box. It will correspond to a certain volume $|\Gamma_M|$. Now, when we remove the partition holding the gas inside that small section, the values of the macroscopic variables will change, opening up a large new volume in Γ available to the system. Since the new equilibrium distribution occupies *almost all* of the newly available phase space, almost all of the microstates originating in the smaller volume will evolve into this region $|\Gamma_{M^*}|$ of the larger phase space. Once the representative phase point of the system evolves into this larger region, it stands virtually no chance of evolving backward. Since S_B is the relative measure of the amount of Γ corresponding to a macrostate, we can see that S_B will increase when the volume corresponding to M increases from $|\Gamma_M|$ to $|\Gamma_{M^*}|$. Notice also that this explanation, unlike many in SM, works only for large N , which is reassuring. J.C. Maxwell²⁶ expressed the essence of this explanation when he wrote: "The Second Law has the same degree of truth as the statement that if you throw a tumblerful of water into the sea you cannot get the same tumberful of water out again" (*ibid.*, p. 583).

We need one more assumption before our sketch of Boltzmann's theory is complete; for as the reader has probably noticed, I have not

²⁶ P.M. Harman, ed., *The Scientific Letters and Papers of J.C. Maxwell, Volume II: 1862-1873* (New York: Cambridge, 1995).

yet addressed Sklar's "parity of reasoning" problem. It is still true that for typical microstates S_B will increase both toward the future and toward the past from present nonequilibrium states. That is, given a nonequilibrium state, we predict that S_B will increase toward the future and that the system *evolved from* a system with higher S_B . The solution to this problem, as Boltzmann and many other luminaries saw, lay in appealing to temporally asymmetric boundary conditions. Richard Feynman²⁷ notes this clearly: "it is necessary to add to the physical laws the hypothesis that in the past the universe was more ordered, in the technical sense, than it is today" (*ibid.*, p. 116). By assuming that what we call the earlier states of the universe are of comparatively low entropy with respect to what we call the later states, we remove the unwelcome retrodiction. Earlier states do not have higher entropy than present states simply because we assume that sometime shortly after the Big Bang entropy was almost unimaginably low. Following Albert, let us call this cosmological hypothesis the *past hypothesis*:

Past hypothesis: the initial (or very early) macrostate of the universe is one of extraordinarily low entropy.

The past hypothesis solves the problem of the direction of time. Whether it adequately *explains* the direction of time is another question (see my *op. cit.*). It also tells us that we can live with recurrence. Given that the universe currently resides so far from equilibrium, it is overwhelmingly probable that entropy did and will increase for all relevant observation times.

Before turning to two worries about the Boltzmann approach, let me try to dispel the idea that there is anything unpleasantly subjective about S_B . The coarse graining used to define S_B does not imply that S_B is subjective, only that it is relational. The thermodynamic variables are used to characterize the macrostates of interest. Which microstates give rise to a given thermodynamic state is an objective matter. It does not rely on the precision with which we can measure a state or our knowledge of a state. If a Laplacian intelligence informed you of the exact microstate of some macrostate, that would not affect the value of S_B one jot. You would have more information about the actual microstate of the system than S_B could provide, but S_B is completely independent of how much you know. This is true despite S_B 's being relational. As Jaynes (*op. cit.*) points out, a crystal of salt may have two different entropies: one expressed as a function

²⁷ *The Character of Physical Law* (Cambridge: MIT, 1965).

of temperature, pressure and volume, $S(T, P)$, another expressed, say, as a function of temperature, the strain tensor and electric polarisation, $S(T, e, E)$. But this 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 already existing in thermodynamics. How many microstates correspond to a particular thermodynamic description is still an objective matter, even if a system admits of more than one such description.²⁸

S_B seems to have the credentials we want in a reduction basis for the thermodynamic entropy. It is an objective function of the state of individual systems. The actual microstate of a gas matters to the Boltzmannian explanation of thermodynamic behavior, in the sense that we can tell whether an individual system is in equilibrium or not. Further, the entropy behaves the right way. For these reasons, and some other mathematical features (for example, its additivity), S_B seems very promising.²⁹ Before we get too excited, however, let me finish with three questions about S_B which are not often discussed.

First, for the Boltzmannian approach to work the actual microstate needs to be a generic or *typical* one for M (in the language of Lebowitz). We know from the argument of section I that we cannot expect entropy increase from all microstates all of the time. We also know, however, that "most" microstates will tend toward equilibrium, since the size of the regions deviating from thermal equilibrium (according to Lebesgue measure) becomes negligible for large N . As $N \rightarrow \infty$, the proportion of these typical states in Γ goes to one. But since $N \neq \infty$ for real systems, the set of atypical points has positive measure. What justification do we have for ignoring microstates in this set? And are we not just begging the question if we assume the actual microstate is always in the set of typical points? Have we really accomplished anything? (This problem is similar in some respects to the so-called "measure-zero" objection to ergodic theory—only here the atypical set is not measure zero.)

It is important not to overreact to this problem. The worry is essentially asking for an independent justification of the imposition of our "natural" probability metric on Γ , which is something we do not have. But we can only try to solve one problem at a time, and any-

²⁸ Another potential source of subjectivity lay in the fact that to define S_B we must "discretize" Γ into small cells. So long as the number of cells grows slowly with respect to N , however, this will not affect the behavior of S_B .

²⁹ It is worth noting that we can extend this approach to quantum mechanics. The volume corresponding to a macrostate has a natural analogue in Hilbert space, namely, the dimension of the projector on the macrostate.

way, the problem of justifying the “natural” probability metric is a very large one common to all of the different approaches to SM.³⁰ Indeed, it is a problem wherever we use probabilities, even outside physics (for example, in games of chance). But, surely, it is *interesting and striking* to discover that according to a very *natural* measure, Lebesgue measure, *nearly all* of the initial conditions compatible with a macrostate are such that Hamiltonian evolution will take them to states closer to thermal equilibrium. And, surely, it is interesting and striking (for instance) that it can be rigorously shown that this dominant set of microstates will, in fact, have their time evolution approximated by generalizations such as the Boltzmann equation.³¹ To be sure, the typicality assumption is somewhat undesirable, but given the lesson of the reversibility and recurrence objections, I do not know how we can escape ultimately making such an assumption without changing the dynamical laws.

Furthermore, intuitively, the evolution of atypical microstates is very contrived, as these states systematically avoid evolving into the larger compartments of Γ . One hopes that some mild dynamical constraints that will prohibit this “contrived” behavior. Computer models and toy models such as the Kac ring model already suggest that these conditions will be weaker than ergodicity or mixing. Development of the Boltzmannian approach should include a statement of precisely what dynamical constraints one needs.

This point raises my second question: How should we understand the “accessible regions $|\Gamma_M|$ ” corresponding to X ? After all, if our system is not ergodic (metrically transitive), then Γ will decompose into subsets with boundaries that cannot be crossed by X . In what sense, then, is $|\Gamma_M|$ accessible to X , if $|\Gamma_M|$ overlaps one or more of these subsets? As Uffink expressed the problem to me, what dynamical relevance does the relative size of $|\Gamma_M|$ have? One partial answer might be that the compartments of Γ representing equilibrium are *so* large that, even if X is trapped in some subset of Γ , still within that subset “most” of the states are ones either representing equilibrium or states closer to equilibrium. Another hope might be that the dynamical constraints mentioned in the above paragraph will show that most of $|\Gamma_M|$ is, in fact, accessible to X .

Finally, and more philosophically, it is commonly thought that one theory reduces to another only if the reducing theory in some

³⁰ One potential solution to this problem is found in the GRW interpretation of quantum mechanics—see Albert.

³¹ See Oscar Lanford III, “The Hard-Sphere Gas in the Boltzmann-Grad Limit,” *Physica A*, *cvi*, 70 (1981): 70-76.

sense explains the potentially reduced theory. If we understand ‘explain’ as ‘causally explain’, for instance, can S_B play a role in a causal explanation of why systems tend toward thermal equilibrium? While playing such a role may not be necessary for the success of S_B , certainly it would be attractive. One reason to think it cannot fill this role is that S_B essentially measures the number of microstates the system is *not in* but could have been without us noticing. As this number does not say much about the nature of X itself, at first glance it hardly recommends itself as the sort of thing useful to a causal explanation. But as Peter Railton³² points out: “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” (*ibid.*, p. 251). To be sure, it would be wrong to think that the number of states X *is not* “drives” X toward equilibrium. But finding out about the (typical) “inevitability” of thermodynamic behavior does carry with it modal and explanatory force. S_B quantifies this modal force.

CRAIG CALLENDER

London School of Economics

³² “Probability, Explanation, and Information,” *Synthese*, XLVIII (1981): 233-56.