Pergamon

Stud. Hist. Phil. Mod. Phys., Vol. 32, No. 4, pp. 525–538, 2001

Printed in Great Britain
1355-2198/01/$ - see front matter

The Origins of Time-Asymmetry in Thermodynamics: The Minus First Law

Harvey R. Brown and Jos Uffink†*

This paper investigates what the source of time asymmetry is in thermodynamics, and comments on the question whether a time-symmetric formulation of the Second Law is possible. © 2001 Elsevier Science Ltd. All rights reserved.

Keywords: Thermodynamics; Entropy; Second law; Statistical mechanics.

1. Introduction

Many authors have addressed the topic of time-reversal invariance in thermodynamics and statistical mechanics. These discussions often assume that the origin of time-reversal non-invariance (or time asymmetry) of thermodynamics lies in its Second Law. However, in a recent paper one of us (JU) opposed this view. It was argued that modern, careful formulations of the Second Law do not presuppose or rely on asymmetric assumptions. This does *not* mean that thermodynamics is time-symmetric. Rather, the paper stated:

> It is often said that [the approach to equilibrium] [...] is accompanied by an increase in entropy, and [is] a consequence of the second law. But this idea

*Sub-Faculty of Philosophy, University of Oxford, 10 Merton Street, Oxford OX1 4JJ, U.K. (*e-mail*: harvey.brown@philosophy.ox.ac.uk).
†Institute for History and Foundations of Science, Utrecht University, P.O. Box 80.000, 3508 TA Utrecht, The Netherlands (*e-mail:* j.b.m.uffink@phys.uu.nl).

PII: S1355-2198(01)00021-1

> actually lacks a theoretical foundation. [...] This aspect of time-reversal non-invariance is woven much deeper in the theory (Uffink, 2001, p. 388).

Our main objective here is to spell this claim out in greater detail. Section 2 discusses a simple, concrete application of this law, demonstrating that it does not *legislate* the approach to equilibrium so much as *characterise* it. In Section 3, we state more precisely what the approach-to-equilibrium principle—which we shall call the 'Minus First Law'—is. We argue that its time-asymmetric component lies in the postulated notion of equilibrium itself. In Section 4, we contrast this with the corresponding 'problem of Boltzmann' in classical statistical mechanics, and raise what we call the 'second problem of Boltzmann', and its treatment by Boltzmann and Schrödinger.

Section 5 takes up the issue of whether entropy increase is a matter of definition, as Hawking has recently claimed. We contrast this to the formal approach to thermodynamics initiated by Carathéodory. Section 6 examines the claim that the counterpart of the Second Law in this approach, and its recent elaboration by Lieb and Yngvason, is time-symmetric.

2. What the Second Law Does Not Do

Consider the following claims: the 'fundamental fact of irreversibility is summarised in the Second Law of Thermodynamics' (Sklar, 1993, p. 21), and that we might think of the Second Law 'as nature's way of driving systems towards equilibrium' (Davies, 1999). Such claims are common enough. But are they correct?

Imagine a gas in a cylinder, fitted with a frictionless piston, which can have occasional diathermal contact with a single heat reservoir. At some initial time when the gas is in equilibrium state A, the piston is suddenly moved, quickly increasing the volume of the cylinder. During this stage, the gas undergoes adiabatic expansion until it attains a new equilibrium state B. The states between A and B are not equilibrium states, so no path can be drawn between them in the accompanying entropy-temperature diagram below (Fig. 1).

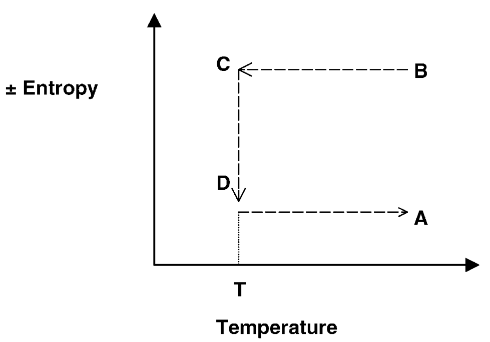

Fig. 1. *Entropy-temperature diagram for a gas confined to a cylinder.*

It is then arranged that the gas undergoes a quasi-static, adiabatic process whereby an equilibrium state C is reached, with the gas now at the temperature T of the heat reservoir. During the next stage, the system is in contact with the reservoir, and undergoes quasi-static isothermal change until it reaches a new equilibrium state D whose entropy has the same value as the entropy of the initial state A. The cycle is completed when the system returns adiabatically and quasi-statically to state A.

What does the Second Law of thermodynamics imply about the expansion process between A and B? To some extent, the answer depends on the formulation of the law, as we shall see later. Consider, for example, Kelvin's version of the Second Law: no cyclic process can have the sole effect of extracting heat from a reservoir and producing a corresponding amount of work. It is commonly argued that this principle implies that the transition $A \rightarrow B$ is 'irreversible', i.e. the converse transition $B \rightarrow A$ is impossible. However, the logic of this argument is always conditional on further assumptions, often left implicit.

Clearly, Kelvin's principle only refers to *cycles* and does not assert the irreversibility of any non-cyclic process. It is only when one assumes the availability of both quasi-static processes $B \rightarrow C \rightarrow D \rightarrow A$ and $A \rightarrow D \rightarrow C \rightarrow B$, allowing the above transitions to be part of a cycle, that the principle becomes applicable to them. But these assumptions themselves are not part of Kelvin's principle.[1]

Hence, the implication of Kelvin's principle for the expansion process can be summarised as a conditional statement: *if* the gas spontaneously expands to a new equilibrium state, and *if* certain other processes are available, *then* the converse transition is impossible. But that this expansion occurs spontaneously is not part of the content of the Law. Therefore we resist the supposition that the Second Law drives systems towards equilibrium, and that the Second Law represents the most fundamental point of entry of time asymmetry into thermodynamics.

3. The 'Minus First Law'

The fact that in thermodynamics the tendency of systems to approach equilibrium is logically prior to the Second Law has been noticed by a number of commentators. Indeed, the existence of this tendency has sometimes been called the 'zeroth law of thermodynamics' (Uhlenbeck and Ford, 1963, p. 5; Lebowitz, 1994, p. 135), in an unfortunate competition with Fowler's famous

[1] Note that Kelvin's principle is not needed in this case to infer that the transition $A \rightarrow B$ involves an entropy increase. This fact already follows from the positivity of gas pressure and the First Law. However, if one considers a cycle of the kind given above for an unspecified thermodynamic system, then Kelvin's principle can be used to infer entropy increase in the transition $A \rightarrow B$, subject to the availability of the quasi-static processes that close the cycle; see Dugdale (1996, pp. 60–62).

nomenclature, whereby 'zeroth law' is used for the quite distinct principle of the transitivity of inter-body thermal equilibrium. The tendency towards equilibrium however is a more basic principle than that of transitivity (Uhlenbeck and Ford, 1963, footnote 14). As Kestin writes:

> The concept of temperature and our ability to perform reproducible temperature measurements rely on the fact that systems, however complex, which are made to interact across diathermal walls within a rigid adiabatic enclosure always reach a state of thermal equilibrium (Kestin, 1979, Vol. I, p. 72).

Let us articulate this phenomenological fact as the following *Equilibrium Principle*:

> *An isolated system in an arbitrary initial state within a finite fixed volume will spontaneously attain a unique state of equilibrium.*

The Equilibrium Principle can be broken into three distinct claims:

(A) *The existence of equilibrium states* for isolated systems. The defining property of such states is that once they are attained, they remain thereafter constant in time, unless the external conditions are changed. The claim that such states exist is not trivial—it rules out the possibility of spontaneous fluctuation phenomena.

(B) *The uniqueness of the equilibrium state*; i.e. for any initial state of an isolated system in a fixed given volume, there is exactly one state of equilibrium.

(C) *The spontaneous approach to equilibrium from non-equilibrium*. A non-equilibrium state will typically come about as the result of a removal of internal constraints, such as the rapid displacement of adiabatic walls separating two bodies. (No indication of the speed of approach to the new equilibrium state is given of course: thermodynamics provides no equations of motion.)

We emphasise that the time asymmetry of thermodynamics arises, at the most basic level, through claim (A).[2] The spontaneous motion towards equilibrium is time-asymmetric because of what equilibrium states are: once attained no spontaneous departure from them is possible without intervention from the environment. The equilibrium state in thermodynamics is itself a time-asymmetric notion[3]— in contrast to statistical mechanics, as we shall see in the next section. Returning to the example of the previous section, it is seen that the Equilibrium Principle not only pre-empts the Kelvin Principle in

[2] One might argue that an independent time-asymmetric element is involved in claims (B) and (C). Indeed, the approach to equilibrium is typically a many-to-one transition: many different initial states evolve to the same final state. The reversal of this transition would then be a one-to-many relation, which is excluded by claim (B). On the other hand, this asymmetry is so to speak 'non-malicious'. In other words, it does not imply incompatibility with an underlying time-symmetric theory. The reason is that in statistical mechanics many different microstates make up one thermodynamic macrostate. So at the microscopic level the transition from a non-equilibrium to an equilibrium state can still be one-to-one, since the equilibrium state contains many more microstates than a non-equilibrium state.

[3] This has been noted by Price (1996), p. 24.

allowing just *one* of the processes $A \to B$, $B \to A$ to happen, it determines *which* one does.[4]

If lawlike status is conferred on the Equilibrium Principle, the name 'Zeroth Law' will not do; it clashes with the widely endorsed nomenclature proposed by Fowler for something that is logically distinct and less fundamental. Were the Equilibrium Principle the *most* fundamental tenet in thermodynamics, the term 'Minus Infinite Law' might be appropriate. But is it? The Principle presupposes the ability to isolate systems from the rest of the universe, and at least one author (Kestin, 1979, Vol. II, p. 1) has argued that *this* is the most fundamental principle. On the other hand, the term 'Minus First Law' might (falsely) suggest that no further fundamental assumptions are needed 'between' it and the Zeroth Law. Be that as it may, in this paper we will call our Equilibrium Principle the 'Minus First Law'.[5]

4. Statistical Mechanics and the Problems of Boltzmann

In Boltzmann's approach to statistical mechanics, thermodynamical states correspond to so-called 'macrostates', i.e. certain subsets in a partition of the phase space Γ of the mechanical system with a fixed total energy E and particle number N. The equilibrium state is identified with the subset with the largest volume under these constraints.

Uhlenbeck and Ford characterised as 'the problem of Boltzmann' in statistical mechanics the following challenge:

> [...] in its simplest form, one must 'explain' in which sense an isolated (that is conservative) mechanical system consisting of a very large number of molecules approaches thermal equilibrium, in which all 'macroscopic' variables have reached steady values (Uhlenbeck and Ford, 1963, p. 5).

The hope of many authors is that this problem can be solved by 'combining dynamics with phase space volume considerations' (Lebowitz, 1994, p. 138).

Such latter *combinatorial* considerations involve the recognition that the volume $|\Gamma_M|$ of the region Γ_M associated with the equilibrium macrostate M_{eq} $\left(|\Gamma_M| = \int_{\Gamma_M} \prod_{i=1}^{N} \mathrm{d}r_i \mathrm{d}v_i\right)$ is vastly larger than that associated with any other macrostate M when N is large. The *dynamical* considerations, on the other hand, are arguably necessary to justify the claim that almost any phase point in an arbitrary non-equilibrium macrostate is overwhelmingly likely to

[4] Some care must be taken in interpreting this last claim. The Equilibrium Principle discriminates between (i) a process in which a system, initially in equilibrium state A, is perturbed by some external intervention, and then evolves to a final equilibrium state B, and (ii) a process in which a system initially in equilibrium state B spontaneously evolves into a non-equilibrium state and is then, by external intervention, brought to equilibrium state A. This latter process is ruled out. The principle remains, however, neutral in deciding between process (i) and (iii) a process in which a system initially in state B is perturbed and then spontaneously evolves to state A. So the fact that a gas, after the piston is released, expands to a greater volume, rather than contracts to a smaller one, is not determined by the above Equilibrium Principle.

[5] The two names —'law –1' and 'law $-\infty$'—were suggested in Uffink (2001, footnote 93).

wander into the large equilibrium region of phase space. There is still debate as to what precise form these dynamical considerations must take (in particular, whether some type of ergodic behaviour is required[6]), and even as to whether *any* dynamical input apart from the Hamiltonian flow on phase space is necessary in solving the problem of Boltzmann. Moreover, there is debate over whether these two ingredients are sufficient. Indeed, as the words 'overwhelmingly likely' indicate, some independent probabilistic assumption seems to be needed as well. (Some authors argue that a notion of 'typicality' might take the place of 'probability'.) But for our purposes, these debates can be put aside.

The main point to note here is that whatever the solution to the problem is, it need not involve any time-asymmetric elements. Indeed, the notion of the equilibrium macrostate as defined above is time-symmetric,[7] unlike its counterpart in thermodynamics. The equilibrium state cannot, in general, be permanent: fluctuations out of equilibrium will occur spontaneously sooner or later for almost every initial microstate. Thus, Claim (A) above is not valid here. (So part of the problem of Boltzmann is specifying 'in what approximate sense' claim (A) can hold.)

The fact that Boltzmann saw $\log|\Gamma_M|$ as a generalisation of the thermodynamic notion of entropy—applicable to non-equilibrium as well as equilibrium states—is not strictly relevant to the present considerations, as we are interested here in the mechanical counterpart of the Minus First Law, not the Second. But it is worth noting that in Boltzmann's approach, the mechanical counterparts of these two laws are not logically independent. The spontaneous approach to equilibrium automatically involves an increase in Boltzmann entropy—a point we shall return to below.

The above problem of Boltzmann refers to the behaviour of an isolated macro-system subsequent to the preparation of a non-equilibrium state; i.e. it refers to *prediction*. The reason that an awkward time-asymmetric element enters the picture is that if the above (time-symmetric) considerations are applied to the problem of *retrodiction*, an unacceptable conclusion seems inevitable: at any time *prior* to the instant at which the system is found out of equilibrium, the system was, with high probability, closer to equilibrium. This conclusion is, to say the least, hard to reconcile with the information stored in our memories, and no one seems to believe it. This difficulty—how to prevent unacceptable retrodictions once one has acceptable predictions—we shall call *the second problem of Boltzmann*.[8] It was nicely expressed by Schrödinger:

[6] Sometimes even individual commentators waver on this issue; compare Lebowitz (1994, p. 135) and Lebowitz (1993, p. 3). A critical analysis of Lebowitz's treatment of the problem of Boltzmann is found in Ridderbos (2000).

[7] However, not all approaches to statistical mechanics share this feature. In a Gibbsian approach, one can define notions of equilibrium which, like the thermodynamical notion, are time-asymmetric (Van Lith, 1999).

[8] For recent discussions of the second problem, see Price (1996, Ch. 2), and Schulman (1997, Sec. 2.6).

> The following is known and is universally agreed upon: the overwhelming majority of all those micro-states that would impress our crude senses as the same observable (=macro-) state do lead to identical, moreover to the actually observed *consequences*. What ails us only, that we can equally well scan the *antecedents*. And they are—again for an overwhelming majority—entirely wrong, inasmuch as the antecedents are the mirror image in time of the aforesaid consequences; it would thus appear that the system has reached its momentary state by an 'anticipation' of its actual future history in reversed order (Schrödinger, 1950, pp. 189–190).

It is fairly well known that Boltzmann addressed this second problem in a number of brief passages (Boltzmann, 1897a, 1897b, 1898). In his (1897a) he offered a choice between two options. The first is to assume as a contingent fact that at present the entire universe is in a very low entropy state. The second option is that the observable world is the result of a localised, spontaneous fluctuation from a permanent equilibrium state of a vastly larger and older universe.

Critics have noticed that even granting this picture of a 'great enough' universe, by far the most likely fluctuation consistent with our observations would still put the lowest entropy state at the present moment, and not into the ancient past, thus making our present memories strictly illusory (cf. Price, 1996, p. 35; Lebowitz, 1994, p. 142). Boltzmann's second suggestion does not really provide a clean solution to the retrodiction problem. Moreover, it has of course been overtaken by developments in modern cosmology. In all fairness, however, it should be remarked that Boltzmann never took his own suggestions very seriously.[9]

We conclude this section with two final remarks. The first is, again, that the time-asymmetric element, if any, that enters Boltzmann's picture only does so in addressing the second, and not the first problem of Boltzmann. (It is essentially this consideration that has led Price (1996, p. 39) to complain that 'the problem of explaining why entropy increases has been vastly overrated'.) This assertion is the mechanical counterpart of our claim in the previous section that no time asymmetry was necessarily being introduced in claim (C) of the Minus First Law. Of course, what prevents anything like the second (retrodiction) problem even arising in thermodynamics is the time-asymmetric nature of equilibrium states in this theory: no retrodictive approach to equilibrium is possible because it would imply the existence of fluctuations away from equilibrium, in conflict with claim (A).

Secondly, the common response to the second problem of Boltzmann is to attribute a very special, low-entropy state to the universe immediately after the Big Bang. If this is the correct response, it would seem that the origins of time asymmetry are different in kind in thermodynamics and statistical mechanics.

[9] Thus, Boltzmann (1897b) offers the idea for the benefit of those 'who cannot resist the temptation to phantasise about the universe', and his (1898) claims as the only virtue for his speculation that they might stimulate 'the mobility of thoughts'.

In the former, the asymmetry is built into the theory by way of the notion of equilibrium. In the latter, it would appear that the asymmetry comes from global initial or final conditions. The price to be paid, as is widely appreciated, is that these conditions are staggeringly ungeneric, or 'improbable'. It is conceivable, however, that a situation closer to that of thermodynamics may re-emerge in canonical quantum gravity.[10]

5. Is Entropy Increase True by Definition?

All traditional formulations of the Second Law presuppose the distinction between past and future (or 'earlier' and 'later', or 'initial' and 'final'). To which pre-thermodynamic arrow(s) of time were the founding fathers of thermodynamics implicitly referring? It is not clear whether this was a question they asked themselves, or whether, if pushed, they would not have fallen back on psychological time. Moreover it is not clear whether such a pre-thermodynamic arrow need be provided by fundamental laws. Any appropriate contingent facts would do for this purpose, and indeed, as Earman teasingly noted, one might simply define an arrow by the order in which most people eat main course and dessert.[11]

Let us now move from thermodynamics to statistical mechanics. Boltzmann's speculations contain a startling suggestion which has attracted considerable attention. Consider a localised fluctuation, on the scale of a galaxy or so. From the time symmetry of statistical mechanics it follows that it is equally likely that the local state will develop by evolving towards equilibrium as by moving away from it. However, a living being who happens to find itself within an environment that shows a local entropy gradient might simply define a time direction as going from lower to higher entropy.[12]

Perhaps the most elaborate development of this position was proposed by Schrödinger (1950). Schrödinger noted that it might happen that a large system in a non-equilibrium state separates into several subsystems, which, at least for some period of time, remain isolated from each other. These 'branch' systems can then be assigned their own individual phase spaces, and hence their own individual Boltzmann entropies. Now although for each branch system an evolution away from equilibrium is still as likely as an evolution towards equilibrium, Schrödinger claimed that in all probability the collective behaviour is such that these entropies change (if at all) monotonically *in the*

[10] This would be the case if the speculation of Barbour (1994; 2000, Ch. 22), concerning the role of time capsules in the solution of the Wheeler–DeWitt equation, is borne out.

[11] Earman (1974). For discussions of non-thermodynamic arrows of time, see Zeh (1992), Price (1996), and Schulman (1997).

[12] See the excerpt from (Boltzmann, 1897a) in Barbour (1999, p. 342); see also the discussion in Price (1996, pp. 32–37).

same way, i.e. decreasing or increasing together.[13] The time-reversal symmetry of statistical mechanics implies that both behaviours are equally probable.

Schrödinger proceeded to propose what he called *the entropy law* for any pair of such branch systems:

$$(S_{1B} - S_{1A})(S_{2B} - S_{2A}) \geqslant 0, \tag{1}$$

where S_{1B} is the entropy of system 1 at time t_B, etc., and $t_B > t_A$. This time-symmetric law allows us to define 'phenomenological time' differently from the time t that appears in statistical mechanics: it is, as Schrödinger stressed, neither t nor $-t$ but rather 'either t or $-t$'! His point was that from the perspective of phenomenological time, entropy always increases, whether or not such increase occurs with increasing t. In other words, phenomenologically the entropic behaviour of the world is the same whether one is dealing with the ascent phase of Boltzmann's cosmic fluctuation or the descent phase—as Boltzmann himself appreciated.

The idea that the direction of time might be reduced to an entropy gradient has recently been advocated by Hawking, who bases it on the assertion that our brains act like computers, which supposedly incur an entropic cost in the process of using memory.[14] It follows from this assertion that the states of the world we remember are those with lower entropy than present and future states, since (to repeat Schrödinger's point), all subsystems of the universe partake in the same entropic flow. For Hawking, this 'reasonable' notion that 'the psychological arrow of time' coincides with the thermodynamic arrow of entropy increase leads to a striking conclusion:

> So the second law of thermodynamics is really a tautology. Entropy increases with time, because we define the direction of time to be that in which entropy increases (Hawking, 1994, p. 348).

In our opinion, some care must be exercised in interpreting this claim. It is probably right to interpret Hawking's point as similar to Boltzmann's speculation regarding what Schrödinger called 'phenomenological time': as long as entropy is changing, in a world sufficiently rich to contain observers it will appear to such observers that entropy is increasing.[15] But note that Boltzmann's claim is made in the context of statistical mechanics. Even in that

[13] This is the so-called 'branching hypothesis'. For some acute comments about the latter, see Price (1996, pp. 44–46). It is interesting to note that although this idea is usually attributed to Reichenbach (1956), Schrödinger (1950) predates this source. Indeed, it appears, together with the term 'abgezweigte Teilsysteme' already in Schrödinger (1933).

[14] Hawking (1994, p. 348) states that the process of recording information in memory in computers must involve an increase of entropy, but the claim need not take this precise form. Others argue that rather memory *erasure* involves an entropic cost. (For a recent discussion of this question, see Bub (2001).) Be that as it may, the relevant claim is that in practice the use of memory in the widest sense of the term is entropically costly. Doubts about this claim are expressed in Earman (1974, p. 34).

[15] This is distinct from the Leibnizian claim defended by Reichenbach (1956) and Gold (1956), that the model of the universe created by taking the total history of our universe, say, and inverting it

context, the conclusion depends on the validity of contingent facts (e.g. the branching hypothesis): so the Second Law is not strictly tautological (or perhaps better: analytic).

If Hawking's reference to *thermodynamics* above is taken literally, his (Hawking's) claim is even more questionable. This can best be seen by considering the formal, axiomatic approach to thermodynamics of Carathéodory (1909, 1925), which is based on the assumption of an 'adiabatic diabatic accessibility' relation defined on a differentiable space Γ of equilibrium states. The version of the Second Law obtained in this approach is known as Carathéodory's Theorem, which states that for so-called 'simple systems', state functions exist on Γ which play the role of entropy and absolute temperature. In his 1909 work, Carathéodory discusses an arbitrary adiabatic process of a simple system (say: the free adiabatic expansion of a gas), and argues that it both is irreversible and involves a change of entropy—but that his principles are incapable of determining whether the entropy increases or decreases.

What is determinate in this approach is that entropy is defined consistently for independent simple systems, i.e. it will either increase or decrease for all such systems. For Carathéodory, it was explicitly a matter of *appealing to experience* to determine whether entropy increases or decreases. More important is that one thus obtains a version of Schrödinger's entropy law: all isolated simple systems undergoing adiabatic processes suffer a change of entropy of the same sign.

Yet there are some subtle points regarding this claim; and we should point out that the brief presentation of Carathéodory's 1909 discussion on irreversibility in Uffink (2001, Sec. 9) may be slightly misleading. For the sake of concreteness, that discussion was framed in terms of the example of the adiabatic expansion of a gas. Unfortunately, this may suggest that the argument *presupposes* that gases spontaneously expand, rather than contract, in adiabatic processes, and that all that Carathéodory's axioms leave undecided is whether entropy increases or decreases in a process of *expansion*. But these two possibilities, surely, cannot be regarded as each other's time reversal. (Perhaps an analogy might help. One can imagine a theory of gravitation which leaves it to experiment whether this interaction is attractive or repulsive. Obviously, these two possibilities are not each other's time reversal.) The possibility of entropy decrease in this context would correspond to a world vastly different from ours in which it is not even clear that sentience is possible.

However, this is not the correct construal of Carathéodory's argument. The statement that a gas expands during adiabatic processes does not follow from his axioms. In fact, this is exactly the empirical input which he uses to gauge the sign of entropy changes. The difference between Carathéodory's view and that of Boltzmann and Schrödinger is that Carathéodory relies on a presupposed

(*footnote continued*)
with respect to time, is just our universe under a different description. This latter position comes under sustained attack in Earman (1974, Sec. 4).

distinction between earlier and later, and uses the behaviour of a gas to fix the sign of changes in thermodynamic entropy; whereas the latter authors start from an explicitly defined sign for the Boltzmann entropy, and then use the behaviour of a gas (or any other irreversible process) to gauge the direction of phenomenological time.

So to sum up: in the context of Carathéodory's thermodynamics, it seems far from analytic that entropy increases.

6. Time-Symmetric Versions of the Second Law

Now, even though Carathéodory's approach to the Second Law leaves it to experience whether entropy increases or decreases in adiabatic processes of simple systems, this is not the same as saying that his Second Law is time-symmetric.

In order to see whether a law is time-symmetric or not, one of course needs to state a clear criterion for the notion. In an influential paper, Earman (1974, p. 23) formulated a condition for a theory to be time-reversal invariant. The idea is here that one specifies a transformation * on the states of the theory, implementing their reversal, and a theory is time-reversal invariant if the class of dynamically allowed models is invariant under the mapping: $s(t) \to s*(-t)$.[16]

Applying this criterion to thermodynamics is not completely straightforward. Thermodynamics has no equations of motion and it is hard to specify its dynamically allowed models. The only clear instance where reference to time is made explicitly in Carathéodory's approach is in the distinction between 'initial' and 'final' that appears in the adiabatic accessibility relation. Hence, a natural option is to call a law or set of laws time-symmetric iff *they continue to hold under reversal of the adiabatic accessibility relation on the space Γ of equilibrium states*.

Now, Carathéodory's version of the Second Law says, roughly speaking, that in the neighbourhood of every equilibrium state there exists another, which is adiabatically inaccessible to it. That is: there exists no state from which it is possible to reach all neighbouring states by some adiabatic process. Under the reversal of adiabatic accessibility, this would transform into the statement that there exist no state which can be reached by all its neighbouring states. Carathéodory's formalism, however, allows the latter situation. Thus Carathéodory's Second Law is not time-reversal invariant.

However, both cases mentioned here are somewhat pathological. The first situation would obtain if the entropy function has an absolute or local minimum (say a state of zero absolute temperature). The second case would

[16] Not all commentators are in sympathy with such a criterion of symmetry based on inter-world, rather than *intra*-world considerations, for instance Brown and Sypel (1995) and Budden (1997). But let us put such doubts aside.

obtain if its entropy function had a local maximum (i.e. a state in which the system was unable to absorb more heat). But neither possibility occurs in ordinary systems. Hence one can, at little risk, strengthen Carathéodory's principle by excluding *both* cases to obtain a time-symmetric formulation.

A similar consideration applies to the recent axiomatisation of thermodynamics by Lieb and Yngvason.[17] This latter approach is related to that of Carathéodory, again being based on a relation of adiabatic accessibility on the state space Γ (though with a somewhat different interpretation of 'adiabatic'). The main result Lieb and Yngvason establish is the existence, for all 'simple' systems (again, for a somewhat different interpretation of 'simple') of an entropy function which increases under adiabatic processes. The principal advantages of their approach are that it does not presuppose the differentiability of Γ—and so is capable of handling phase transitions—and more importantly that it guarantees that entropy is defined *globally* on Γ.

Remarkably, the axioms needed to obtain their version of the Second Law (called 'the Entropy Principle') are time-reversal invariant in the sense indicated above. In the case of Lieb and Yngvason, what non-invariant axioms there are, are not directly related to the derivation of their Entropy Principle (Uffink, 2001, Sec. 11). Of course, the invariance of these axioms implies trivially that under the same conditions, there exists also a function that decreases under adiabatic processes.

As in the case of Carathéodory, the Lieb–Yngvason approach actually obtains more than just the result that an entropy function exists for simple systems, which increases (or decreases) for adiabatic processes. An important additional feature is that these entropy functions can be chosen *consistently* for all simple systems, i.e. as either increasing or decreasing for all such systems. Furthermore, these entropy functions are additive, which implies the highly non-trivial fact that the possibility of adiabatic processes in combined systems is determined just by the sum of their individual entropies. This in turn implies Schrödinger's entropy law: for any two non-interacting simple systems, entropy changes will always occur in the same sense.

Some final remarks are in order. Recall that in the above analysis, time reversal is implemented by the reversal of the adiabatic accessibility relation. Such a reversal should then correspond to a world in which processes occur which look like those occurring in a film of our world but played backwards. In particular, the reversal of the spontaneous adiabatic expansion of a gas discussed in Section 2 above would correspond to a spontaneous adiabatic contraction. But this behaviour is inconsistent with claim (A) of the Minus First Law, which as we have seen rules out spontaneous deviations from equilibrium. Indeed, if one starts from the notion of equilibrium of Section 3, which has time asymmetry built into it, one may well argue that the time reversal of an equilibrium state is not an equilibrium state. From this point of view, we would not be entitled to put $s = s^*$, and this analysis would fail.

[17] See Lieb and Yngvason (1998, 1999, 2000). A review is given in Uffink (2001, Sec. 10).

This leaves us with several options. We could ignore claim (A) and merely focus on the formal structure of the axioms, in which the elements of the state space Γ are given little or even no physical interpretation. The disadvantage of this option is that it is less clear how physically significant this symmetry is.

Alternatively, we could turn to some of the additional axioms in the Lieb–Yngvason approach (in particular, those called A6 and T4). These axioms, roughly speaking, postulate the existence of special kinds of adiabatic processes (namely, mixing and thermal equilibration) whose properties are non-invariant under the reversal of the adiabatic accessibility. In this option, the Entropy Principle is embedded in a more general time-asymmetric theory.

Or, third, we could rely on the intended meaning of equilibrium and, in particular, on the validity of claim (A). In that case, the time reversal of an equilibrium state is no longer an equilibrium state, and the earlier analysis is no longer applicable.

Regardless of which option one prefers, we emphasise that, in accordance with Earman's criterion, time asymmetry is not needed to obtain a rigorous statement of the Second Law; it is required only in the interpretation of the notion of equilibrium.

Acknowledgements—We are grateful to Yemima Ben-Menahem and Itamar Pitowsky for their kind invitation to and hospitality during the May 2000 International Workshop in Jerusalem. We also thank Julian Barbour for providing the 1897a Boltzmann reference, and Jeremy Butterfield for detailed suggestions towards improving presentation. HRB thanks Michael Lockwood for useful discussions, and gratefully acknowledges the support of the British Academy and the Leverhulme Trust.

References

Barbour J. (1994) 'The Emergence of Time and Its Arrow from Timelessness', in Halliwell *et al.* (1994), pp. 405–414.

Barbour, J. (1999) *The End of Time* (London: Weidenfeld & Nicolson).

Boltzmann, L. (1897a) 'Zu Hrn Zermelos Abhandlung "Über die mechanische Erklärung irreversibler Vorgänge"', *Wied. Ann.* **60**, 392–398.

Boltzmann, L. (1897b) 'Über einige meiner weniger bekannten Abhandlungen uber Gastheorie und deren Verhältnis zu derselben', *Verh. der 69. Vers. D. Naturf. und Ärzte* (Braunschweig), pp. 19–26.

Boltzmann, L. (1898) *Vorlesungen über Gastheorie,* Vol. II, (Leipzig: Barth).

Brown, H. R. and Sypel, R. (1995) 'On the Meaning of the Relativity Principle and Other Symmetries', *International Studies in the Philosophy of Science* **9**, 235–253.

Bub, J. (2001) 'Maxwell's Demon and the Thermodynamics of Computation', *Studies in History and Philosophy of Modern Physics* **32** (4), 569–579.

Budden, T. (1997) 'Galileo's Ship and Spacetime Symmetry', *British Journal for the Philosophy of Science* **48**, 483–516.

Carathéodory, C. (1909) 'Untersuchungen über die Grundlagen der Thermodynamik', *Mathematische Annalen* **67**, 355–386. English translation in Kestin (1976), pp. 229–256.

Carathéodory, C. (1925) 'Über die Bestimmung der Energie und die absoluten Temperatur mit Hilfe von reversiblen Prozessen', *Sitzungsberichte der Preussischen Akademie der Wissenschaften,* pp. 39–47.

Davies, P. (1999) 'Is the Flow of Time an Illusion?', *International Framtider* **9**, 4–8.

Dugdale, J. S. (1996) *Entropy and its Physical Meaning* (London: Taylor and Francis).

Earman, J. (1974) 'An Attempt to Add a Little Direction to "The Problem of the Direction of Time" ', *Philosophy of Science* **41**, 15–47.

Gold, T. (1956) 'Cosmic Processes and the Nature of Time', in R. Colodny (ed.), *Mind and Cosmos* (Pittsburgh: University of Pittsburgh Press), pp. 311–329.

Halliwell, J. J., Pérez-Mercador, J. and Zurek, W. H. (eds) (1994) *Physical Origins of Time Asymmetry* (Cambridge: Cambridge University Press).

Hawking, S. (1994) 'The No Boundary Condition And The Arrow Of Time', in Halliwell *et al*. (1994), pp. 346–357.

Kestin, J. (ed.) (1976) *The Second Law of Thermodynamics* (Stroudsberg, PA: Dowden, Hutchison and Ross).

Kestin, J. (1979) *A Course in Thermodynamics,* Vols. I, II, (New York: McGraw-Hill).

Lebowitz, J. (1993) 'Macroscopic Laws and Microscopic Dynamics', *Physica A* **194**, 1–27.

Lebowitz, J. (1994) 'Time's Arrow and Boltzmann's Entropy', in Halliwell *et al*. (1994), pp. 131–146.

Lieb, E. and Yngvason, J. (1998) 'A Guide to Entropy and the Second Law of Thermodynamics', *Notices of the American Mathematical Society* **45,** 571–581.

Lieb, E. and Yngvason, J. (1999) 'The Physics and Mathematics of the Second Law of Thermodynamics', *Physics Reports* **310**, 1–96; **314** (1999), 669; cond-mat/9708200.

Lieb, E. and Yngvason, J. (2000) 'A Fresh Look at Entropy and the Second Law of Thermodynamics', *Physics Today* (April issue); math-ph/0003028.

Lith, J. van (1999) 'Reconsidering the Concept of Equilibrium in Classical Statistical Mechanics', *Philosophy of Science* **66** (Proceedings), S107–S118.

Price, H. (1996) *Time's Arrow and Archimedes' Point* (New York: Oxford University Press).

Reichenbach, H. (1956) *The Direction of Time* (Los Angeles: University of California Press).

Ridderbos, T. (2000) 'The Arrow of Time in Statistical Mechanics and Quantum Cosmology', Ph.D. Thesis, Cambridge University.

Schrödinger, E. (1933) 'Über den zweiten Hauptsatz der Thermodynamik', *Sitzungsberichte der Preussischen Akademie der Wissenschaften*, 165.

Schrödinger, E. (1950) 'Irreversibility', *Proceedings of the Royal Irish Academy* **53**, 189–195.

Schulman, L. S. (1997) *Time's Arrows and Quantum Measurement* (Cambridge: Cambridge University Press).

Sklar, L. (1993) *Physics and Chance* (Cambridge: Cambridge University Press).

Uffink, J (2001) 'Bluff Your Way in the Second Law of Thermodynamics', *Studies in History and Philosophy of Modern Physics* **32** (3), 305–394.

Uhlenbeck, G. E. and Ford, G. W. (1963) *Lectures in Statistical Mechanics* (Providence: American Mathematical Society).

Zeh, H.-D. (1992) *The Physical Basis of the Direction of Time* (New York: Springer).