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

***Harvey R Brown*[*]*and Jos Uffink*[^]**

(...)

2. What the Second Law does not do

Consider claims of the following kind: that the "fundamental fact of irreversibility is summarised in the Second Law of Thermodynamics"[2], or that we might think of the Second Law "as nature's way of driving systems towards equilibrium"[3]. Such claims are common enough. But are they correct?

Imagine a cylinder, within which a gas is contained by a frictionless piston, and which can have occasional diathermal contact with a single heat reservoir. At a given

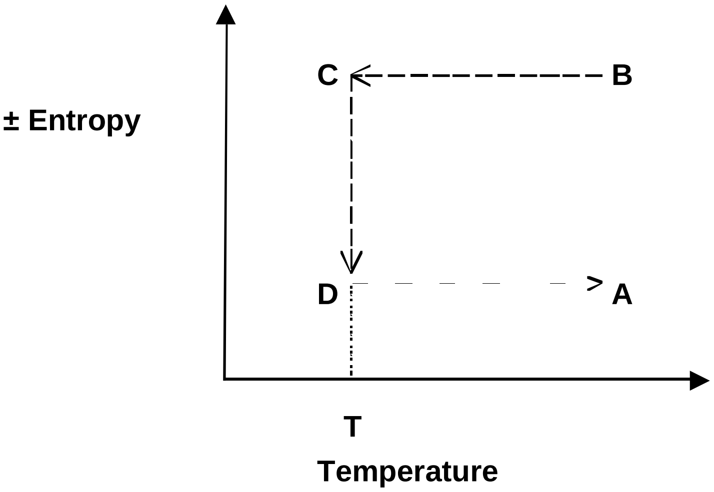
---

[2] Sklar (1993), p. 21.

[3] Davies (1999).

time t_0, the gas in found in the equilibrium state *A* and the piston is then made to undergo sudden motion, quickly increasing the volume of the cylinder by a definite amount. In the interval between t_0 and t_1, the gas undergoes adiabatic expansion until at it attains a new

state *B* of equilibrium. 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.



It is then arranged that the system undergoes a quasi-static, adiabatic process whereby the equilibrium state *C* is reached at time t_2, with the gas now at the temperature *T* of the heat reservoir. Between times t_2 and t_3, the cylinder is in contact with the reservoir, and the system undergoes a quasi-static isothermal change until the entropy associated with the new equilibrium state *D* at t_3 has retained the value it had for the initial state *A*. The cycle is completed when the system returns adiabatically and quasi-statically to state *A* at time t_4.

What does the Second Law of thermodynamics imply about the process of free expansion of the gas between t_0 and t_1? To some extent, the answer depends on the chosen formulation of the law, as we shall see later. Consider, for example, Kelvin's version of the Second Law[4]: no cyclic process, such as $A \to B \to C \to D \to A$ or $A \to D \to C \to B \to A$, can have the sole effect of extracting heat from a reservoir and producing a corresponding amount of work. It is commonly argued that it follows from this principle that the transition $A \to B$, if it occurs, is ‘irreversible’, i.e. the converse transition $B \to A$ 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 $B \to C \to D \to A$ and $A \to D \to C \to B$, allowing the above transitions to be part of a cycle, that the principle becomes applicable to them. But these assumptions are themselves not part of Kelvin’s principle.[5]

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

3. The “Minus First Law”

---

[4] This is the version of the Second Law comonly attributed to Kelvin; his original formulation is discussed in Uffink (2001), section 5.

[5] It might also be worth noting that Kelvin’s principle is not needed in this case to infer that the transition $A \to B$ involves an entropy increase. This fact can easily be seen to follow from the positivity of gas pressure, the invariance of internal energy in the transition and the First Law. However, if one considers a cycle of the kind given in the above entropy-temperature diagram, but without specifying the nature of the thermodynamic system, then Kelvin’s principle can be used to infer entropy increase in the transition $A \to B$, subject to the availability of the quasi-static processes that close the cycle; see for example Dugdale (1996) pp. 60-62.

The fact that in thermodynamics the tendency of systems to approach equilibrium is logically prior to the Second Law may not be universally appreciated, but it has not escaped the notice of a number of commentators. The existence of this tendency has sometimes been referred to as the "zeroth law of thermodynamics"[6], in an unfortunate competition with R. H. Fowler's famous usage concerning the quite distinct principle of the transitivity of inter-body thermal equilibrium. The latter forms part of the basis of the notion of temperature[7], but the tendency towards equilibrium is a more basic principle. This point has been stressed by Joseph Kestin, for example[8]:

> 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.

Let us articulate this phenomenological fact in the form of the following *Equilibrium Principle*:

> *When an isolated system finds itself in an arbitrary initial state within a finite fixed volume, it 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, the independent thermodynamic coordinates of the system are spatially homogeneous and remain thereafter constant in time, unless

---

[6] See Uhlenbeck and Ford (1963), p. 5, and Lebowitz (1994), p. 135.

[7] See, for example, Zemansky (1957), p. 6

[8] Kestin (1979), Vol I, p. 72; see also Uhlenbeck and Ford (1963), footnote 14.

the external conditions are changed. The claim that such states exist is not trivial—it rules out the possibility of spontaneous fluctuation phenomena[9].

(B) *The uniqueness of the equilibrium state*. The claim is that for any initial state of an isolated system bounded by a prescribed static surface, 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.)

The point we wish to stress here is that the time-asymmetry of thermodynamics arises, at the most basic level, through claim (A)[10]. 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.

---

[9] Readers who are squeamish about an existence claim being granted the status of a law (or part thereof) are invited to consider the case of Newton's first law of motion, which has as its content the claim that at least one ("inertial") reference frame exists, with respect to which the motion of *all* force-free bodies is uniform and rectilinear.

[10] One might debate the question whether an independent time-asymmetric element is involved in claim (B) and (C). The reason to think so is that the approach to equilibrium is typically a many-to-one transition: many different initial states will 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). However, although this asymmetry is arguably there, it is so to say 'non-malicious'. Or in other words, it does not necessarily lead to an 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 equilbrium state becomes one-to-one because the equilibrium state contains many more microstates than a non-equilibrium state.

The equilibrium state in thermodynamics is itself a time-asymmetric notion[11]— in contrast to a notion of equilibrium in statistical mechanics, as we shall see in the next section. Returning to the last section, it is seen that in the special case of the gas system the Equilibrium Principle not only pre-empts the Kelvin Principle in allowing just *one* of the processes $A \to B,\ B \to A$ to happen, it determines *which* one does.[12]

If lawlike status is to be conferred on the Equilibrium Principle, the existing appelation "Zeroth Law" clearly will not do; as we have mentioned it clashes with the now widely endorsed use of this nomenclature by Fowler for something that is logically distinct and less fundamental. Were the Equilibrium Principle the *most* fundamental tenet imaginable in thermodynamics, the term "Minus Infinite Law" might be appropriate. But is it? The Principle is, to repeat, a collection of claims, of which (A) is clearly the most basic. But (A) presupposes the ability to isolate the system of interest from the rest of the universe, and at least one author has argued that *this* is the most fundamental of all principles[13]. 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 adopt the term "Minus First Law"[14].

---

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

[12] 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 releasing the piston, expands to a greater volume, rather than contract to a smaller one, is not determined by the above Equilibrium Principle.

[13] Kestin (1979), vol II, p. 1.

[14] The two options in nomenclature—'law –1' and 'law $-\infty$'—were suggested in Uffink (2001), footnote 93. An interesting question is whether the Minus First Law, or