

## 6

# There Is No Thermodynamic Arrow

---

*Précis.* The structure of equilibrium thermodynamics, in harmony with statistical mechanics, does not contain a time asymmetry, except when it is trivially supplemented with one.

---

Our bodies, the planets and stars, and all traces of our existence will eventually dissipate into the cosmos. According to orthodoxy, this march towards equilibrium is *the* arrow of time.<sup>1</sup> I have tried to defend and clarify orthodoxy in earlier chapters of this book. But, in this case, I think orthodoxy is indefensible: in this chapter, I will argue that there is no non-trivial sense in which the approach to equilibrium in thermodynamics gives rise to a time asymmetry.

The basic argument that I will give, summarised in the next section, is that classical thermodynamics is not so different from its 'partner' theory, statistical mechanics, in how it bears on the arrow of time. It is widely agreed that in statistical mechanics, the approach to equilibrium is symmetric in time, in that the laws predict evolution towards equilibrium with roughly equal likelihood to the future and to the past. Or, in the language developed in this book: the time translations of statistical mechanics admit a representation of time reversal symmetry. The occurrence of statistical mechanical equilibrium in our future rather than in our past is not a fact about temporal symmetry but about contingent initial conditions, contingent boundary

<sup>1</sup> This is perhaps partly for historical reasons: Eddington (1928, pp.64–9) coined the term 'arrow of time' to describe irreversible thermodynamic processes; see also the remarks at the beginning of Chapter 5.

conditions, and our special limited perspective as observers.<sup>2</sup> I will argue that classical equilibrium thermodynamics is no different.

In contrast, a well-known folklore has it that classical equilibrium thermodynamics, if it were true, would require a robust arrow of time. This folklore was famously captured by Reichenbach:<sup>3</sup>

If the universe as a whole possesses at every moment a specific entropy, this value is subject to the general law of [thermodynamic] entropy increase . . . . Although this principle leads to the unwelcome consequence that someday our universe will be completely run down and offer no further possibilities of existence to such unequalized systems as living organisms, it at least supplies us with a direction of time: positive time is the direction towards higher entropy. (Reichenbach 1956, p.54)

Reichenbach himself dismissed classical thermodynamics in favour of the more fundamental and accurate tools of statistical mechanics. But, there are good reasons to return to classical thermodynamics and what it says about the arrow of time. In the first place, thermodynamics is a powerful tool in a wide variety of sciences, from biochemistry to stellar astrophysics. It has become an important technique for probing regimes in which the microphysics is not yet known, such as in black hole thermodynamics. And, a key requirement for understanding the reduction of thermodynamics to statistical mechanics is to make sense of the supposed difference between their temporal symmetries.

Uffink (2001, p.307) calls this problem of reconciling temporal symmetries “the most profound problem in the foundations of thermal and statistical physics”.<sup>4</sup> Ludwig Boltzmann famously spent most of his career trying to solve the problem. However, not everyone agrees about its significance. Callender (1999, 2001) concludes that we should give up some aspects of the reduction project: just use statistical mechanics to study what is fundamental, and deny that thermodynamics is “universally true and somehow independent of the statistics of the micro-constituents of thermal bodies” (Callender 2001, p.551). Others have recently argued that the difference in temporal symmetries remains a problem that philosophical accounts of reduction must face.<sup>5</sup>

I will argue that this is not a problem in the first place, because there is no non-trivial thermodynamic arrow. Some of my arguments will follow and

<sup>2</sup> An argument for this is set out in Section 5.3; see also Price (1996, Chapter 2).

<sup>3</sup> It is echoed by Hawking (1994, p.348), Sklar (1993, §2.I.3) and Zeh (2007, p.5), among others.

<sup>4</sup> The history of the problem is given by Brush (1976b); well-known philosophical discussions can be found in Batterman (2002, Chapter 5), Reichenbach (1956, Chapter III), and Sklar (1993, Chapter 9).

<sup>5</sup> Compare Dizadji-Bahmani, Frigg, and Hartmann (2010); North (2011); Robertson (2021); Valente (2021); Werndl and Frigg (2015).

(I hope) strengthen the work of Uffink (2001), who forcefully demonstrates that the thermodynamic second law stands little chance of rigorously establishing a thermodynamic arrow. My conclusion is more general: I can see no sense in which the structure of thermodynamics contains any basis at all for an arrow of time, except in the trivial sense of inserting it by hand.

Section 6.1 summarises the basic idea using the simple example of gas mixtures and addresses a proposal of Brown and Uffink (2001) to locate the thermodynamic arrow in the approach to equilibrium. Section 6.2 then motivates and summarises the mathematical structure of thermodynamics – at least, before a second law is introduced – including how it describes time and change. The remainder of the chapter will argue the temporal symmetry of thermodynamics remains even when the second law is introduced, in any of three classic formulations: that of Clausius (Section 6.3), of Kelvin and Planck (Section 6.4), and of Gibbs (Section 6.5).

## 6.1 Undirected Thermodynamics

*Doesn't the second law say that entropy increases to the future?* Reichenbach (1956, p.49) calls the second law of thermodynamics “a principle that makes possible the mathematical expression of a direction controlling the course of physical occurrences”.<sup>6</sup> The idea is to argue from the second law of thermodynamics that entropy is increasing and then define the direction of time as the direction of that increase.

However, virtually all formulations of thermodynamics agree that entropy increase is not an axiom of thermodynamics: it is at most derived as a consequence of the axioms, or in some cases taken to be the ‘essence’ of a thermodynamic system, all things considered. Of course, one can also skip these derivations entirely and just declare that entropy increases to the future. I fully support this practice when it serves the theoretician. But, when a time asymmetry is inserted in this way, as a bald addition to an otherwise time symmetric theory, I will refer to it as a *trivial* time asymmetry.<sup>7</sup> The thesis of this chapter is that, if thermodynamics has a time asymmetry, then it is trivial. Nothing about the formal structure of thermodynamics requires it, just as in a time symmetric theory like statistical mechanics or Newtonian gravitation. As Uffink (2007, p.938) aptly remarks: “[T]he Second Law has often been understood as demanding continuous

<sup>6</sup> Similar remarks can be found in Davies (1977, §2.1) and Sklar (1993, §2.I.3), among many other places.

<sup>7</sup> In Section 5.1, I called this kind of ‘arrow’ a *Heuristic misfire*: an informal extra-theoretic judgement, with no justification in terms of any well-supported physics.

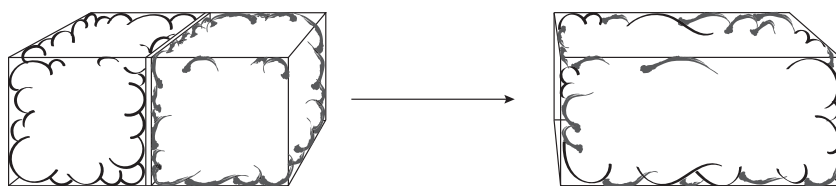


Figure 6.1 Two separated gases and their mixture.

monotonic increase of entropy in the course of time . . . . There is, however, no basis for this demand in orthodox thermodynamics”.

How can thermodynamics be symmetric in time? The basic idea can be illustrated using the simple example of a box that is isolated from any outside influences and which contains two gases in chambers separated by an impenetrable barrier. In a similar system with the barrier removed, where the system is assumed to be ‘settled down’ into equilibrium, the result is a mixture whose properties can be determined by the properties of the original gases, as shown in [Figure 6.1](#). For example, the mole number of the mixture can be determined on the basis of the component mole numbers; and the entropy of the mixture can be determined on the basis of the component entropies. All predictions in equilibrium thermodynamics have more or less this character: one set of equilibrium systems determines another.<sup>8</sup>

What I would like to point out is that these constraints are fundamentally directionless in time: in this example, as long as the component gases are viewed as ‘parts’ of the composite, they constrain it in exactly the same way, whether that composite is to the future or to the past. Nothing about the structure of thermodynamics determines that the arrow in [Figure 6.1](#) can only point in one temporal direction. If there is any time asymmetry in thermodynamics, it arise purely from the additional postulate that special initial conditions, or the special perspective of an observer, are associated with one direction in time. But this is just to assume at the outset what we were originally trying to prove.

The situation is strikingly similar to the temporal symmetry that one finds in statistical mechanics.<sup>9</sup> Typical counting arguments in statistical mechanics establish not only that a system will evolve with overwhelming probability to a higher-entropy equilibrium state but also that the system will have evolved from such a state in the past. Like the exploration of a

<sup>8</sup> This interpretation of thermodynamics has been developed in detail by Callen (1985); see Wills (2022) for philosophical remarks about it.

<sup>9</sup> See Section 5.3.

house with a thousand blue rooms and one red room – the next room you will visit is probably blue! One scenario is much more likely than the other, no matter which direction in time one looks in. The predictive mechanism of thermodynamics is similar in spirit: a collection of equilibrium systems determines another equilibrium system (their composite) that maximises entropy, no matter how it is directed in time. In this sense, there is complete harmony between the temporal symmetries of equilibrium thermodynamics and statistical mechanics.

*But, doesn't thermodynamics still say that equilibrium happens to the future?* A number of careful authors have noticed that the approach to equilibrium itself is not a consequence of the second law, but an independent assumption.<sup>10</sup> For example, in the mixing process for a pair of gases, one must first assume that the two unmixed gases can be associated with a third gas that is an equilibrium mixture of the two; only then can one apply reasoning typically associated with the second law in order to argue that the mixture maximises entropy.

This led Uffink (2001, fn. 93) and Brown and Uffink (2001) to introduce a new postulate into the axioms of thermodynamics, which they call “the Minus First Law”:

An isolated system in an arbitrary initial state within a finite fixed volume will spontaneously attain a unique state of equilibrium. (Brown and Uffink 2001)

The Minus First Law is unquestionably an important part of modern physics. In a sense, it is the starting point for a variety of approaches to thermodynamics developed in the twentieth century, often called ‘non-equilibrium thermodynamics’, which explicitly associate the approach to equilibrium with a continuous time parameter. This approach is used in the study of a variety of physical processes like the relaxation time, mass-flux, and biological cellular processes.<sup>11</sup> But, is the equilibration process it describes really asymmetric in time?

<sup>10</sup> See Uhlenbeck and Ford (1963, §I.3), and especially Brown and Uffink (2001), Marsland III, Brown, and Valente (2015), and Uffink (2001, 2007).

<sup>11</sup> The reciprocity relations of Onsager (1931a,b) provided a framework for studying thermal fluctuations away from equilibrium using techniques pointed out by Einstein (1910) and led to what is known as ‘Linear Irreversible Thermodynamics’ and ‘Extended Irreversible Thermodynamics’ (García-Colín and Uribe 1991). An alternative, axiomatic approach sometimes called ‘Continuum Thermodynamics’ or ‘Rational Thermodynamics’ was developed by Truesdell (1984), although it has been remarked that, “[i]f there is something ‘rational’ about Rational Thermodynamics it is certainly well hidden, and interest in it has not stood the test of time”. (Lavenda 2010, p.ix). See Haslach Jr. (2011) for a recent overview of non-equilibrium techniques, which are tangential to my purposes here.

Brown and Uffink conclude that it is. Their argument is that when a system approaching equilibrium is time reversed, the result is a system deviating from equilibrium; thus, insofar as this is prohibited by thermodynamics with a Minus First Law, the theory is not time reversal invariant:<sup>12</sup>

[T]ime 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 ... would correspond to a spontaneous adiabatic contraction. But this behaviour is inconsistent with ... the Minus First Law, which as we have seen rules out spontaneous deviations from equilibrium. (Brown and Uffink 2001, p.536)

Time asymmetry here depends on the assumption that a spontaneous decrease of entropy is “inconsistent” with the Minus First Law’s statement that equilibrium must be attained. But, inconsistency does not necessarily follow.

The Minus First Law, and any other axiom requiring that a physical system ‘attain’ equilibrium, is not intrinsically directed in time. The theoretician begins with one description: in equilibrium thermodynamics, it is a collection of equilibrium systems, like the separated gases in [Figure 6.1](#); in non-equilibrium thermodynamics, it is a system out of equilibrium. The Minus First Law only guarantees that this description determines a unique state of equilibrium. This is perfectly consistent with that equilibrium state appearing in both time directions: the initial description is associated with an equilibrium state, which may occur in either the past or the future. The box of separated gases determines an equilibrium state associated with their mixture, whether it occurs in the future or in the past, as in [Figure 6.2](#). Time asymmetry plays no essential role in the requirement that such an equilibrium state exists.

There is no mystery as to why we typically associate one of these scenarios with the future: in practice, one commonly inserts temporally directed language to this description, by referring to an ‘intervention’ or ‘manipulation’

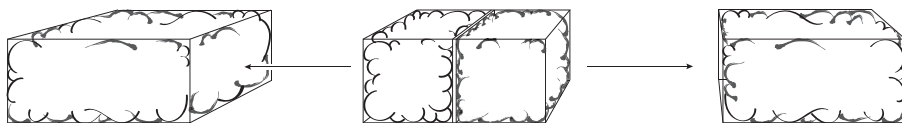


Figure 6.2 An equilibrium state is determined no matter which temporal direction is chosen.

<sup>12</sup> This argument has enjoyed some support: compare Marsland III, Brown, and Valente (2015), Myrvold (2020b, Forthcoming), and Valente (2021).

that removes the barrier between the gases in a temporally-directed way. When such language is included, one may be inclined to call the mixture the ‘effect’ or ‘future’ of the intervention. There is nothing wrong with that. However, it would be misleading to call such an insertion a ‘thermodynamic arrow’, since there is nothing about the structure of equilibrium thermodynamics that encodes it. Inserting the language of intervention reflects a contingent human perspective, of the kind discussed extensively in [Chapter 5](#), and not an asymmetry of time itself.

A position very similar to mine has been developed by Myrvold (2020c). Viewing thermodynamics as a theory that relates heat reservoirs to interventions on “manipulable variables”, Myrvold develops all the essential structure of thermodynamics, including the zeroth, first, and second laws, in a way that makes no mention of the approach to equilibrium. For my purposes, I prefer to avoid the language of ‘manipulation’, which might suggest a temporally directed structure that does not exist in the structure of equilibrium states. However, Myrvold is right to identify Brown and Uffink’s Minus First Law as lying outside the “province” of thermodynamics, properly conceived. Of course, time asymmetric statements may still be formulated using thermodynamic language.<sup>13</sup> But, these invariably arise from the assumption that the approach to equilibrium is associated with a temporal direction, which is not part of the structure of equilibrium thermodynamics. His diagnosis of the widespread confusion about this is similar to that of Brown and Uffink (2001):

There is a tendency to conflate the second law of thermodynamics with the tendency of systems to relax to a state of thermal equilibrium, and this has encouraged the idea that the study of equilibration does fall within the scope of thermodynamics. These are not the same thing, however. (Myrvold 2020c, p.1243)

That said, one might still choose to add a time asymmetric axiom to thermodynamics: an axiom which states that systems are associated with not just an equilibrium state but one that occurs in one time direction and not the other. This assumption is implicit in Brown and Uffink’s discussion of the Minus First Law. However, this amendment to the ‘temporally symmetric Minus First Law’ that I have characterised above is no longer a deep structural requirement of thermodynamic theory but the insertion of a time asymmetry of a kind that I call ‘trivial’. This is not enough to

<sup>13</sup> Myrvold (2020c, §9) identifies his statistical version of the Clausius inequality as an example (his Proposition 2) but points out that all the thermodynamic assumptions that go into it are temporally symmetric. Thus, temporal asymmetry enters when a temporal direction is associated with the approach to thermal equilibrium, which lies outside the scope of the theory.

distinguish the temporal symmetries of thermodynamics from other time symmetric theories.

In a nutshell, this is the sense in which I will argue that there is no thermodynamic arrow of time. To make this argument precise, and to square it with the many thermodynamic arrows that have been proposed, a more careful treatment of equilibrium thermodynamics will be needed. In particular, I will aim to show that on a rigorous understanding of the structure of equilibrium thermodynamics, including the first and second laws, there is no thermodynamic arrow of time. So, we begin in the next section with a brief review of the structure of thermodynamics.

## 6.2 Equilibrium Thermodynamics

The Soviet mathematical physicist Vladimir Arnol'd once began an article by declaring, "Every mathematician knows it is impossible to understand an elementary course in thermodynamics". This quote was often passed around by other mathematically-inclined philosophers and physicists when I was a graduate student, as evidence that thermodynamics is a chimera beyond our reach.<sup>14</sup> But, when I finally read the article myself, I found that Arnol'd goes on in the second sentence to say: "The reason is that the thermodynamics is based – as Gibbs has explicitly proclaimed – on a rather complicated mathematical theory, on the contact geometry" (Arnol'd 1990, p.163). In other words, it *is* possible to understand an *advanced* course in thermodynamics!

So, in the remainder of this chapter, I will try to discuss the structure of equilibrium thermodynamics in a mathematically and conceptually rigorous way, at the cost of sometimes introducing a little more differential geometry than is typical.<sup>15</sup> I begin with a brief overview of the structure of thermodynamic state space; I will then discuss two senses in which time and change enter into thermodynamic thinking.

### 6.2.1 Heat, Entropy, and Temperature

In a simple mechanical system, like a mass sitting on top of a hard block of volume  $V$ , energy  $U$  is completely determined by the force of pressure:  $dU = -PdV$ . This situation changes strikingly when the block is replaced

<sup>14</sup> Cf. Uffink (2001, p.310).

<sup>15</sup> A 'light' version of this approach can be found in the textbook of Callen (1985). Myrvold (2020c) builds a conceptual foundation for thermodynamics that is very close to what I present here. A more detailed development of geometric thermodynamics can be found in Arnol'd (1990), Bravetti (2018), Hermann (1973), or Mrugała (1978); see J. Wills (2021, "Representing mixtures", Unpublished manuscript) and (2022) for philosophical remarks.



with a gas: when we apply pressure to the gas, it gets hotter to our touch! Since this heating effect can be harnessed to produce energy, local energy conservation would require that there be further ‘hidden’ degrees of freedom into which energy can flow. And of course nowadays we know that there are, such as the microscopic motions of the molecules in the gas.<sup>16</sup> But, never mind the nature of those hidden degrees of freedom: let us embrace our ignorance and refer to those hidden contributions to energy as ‘heat’.

Classical equilibrium thermodynamics is the study of the relationship between contributions to energy like pressure and volume, which are directly observable and are called *work*, and those that are not directly observable in this way, which are called *heat*. Burke (1985) calls the latter ‘unexamined’ degrees of freedom; I will refer to them as ‘hidden’. The state space used to describe these degrees of freedom and their structure is what characterises a model of thermodynamics. And, it is this state space that I will argue has no non-trivial time asymmetry.

A *state* of a thermodynamic system is a point in a smooth (for our purposes  $C^\infty$ ) real manifold  $N$  of dimension  $n$ . Each state represents a physical description of the system when it is ‘settled down’ into stable equilibrium. The *energy* of the system is an exact one-form<sup>17</sup>  $dU$ , where we view  $U$  as a smooth function of the manifold. The various contributions of work to energy, such as pressure and volume, chemical potential and mole number, and so on, all have the form  $P_i dX_i$  for some smooth functions  $X_i$  and  $P_i$ . (So, our previous  $V$  and  $P$  are an example: say,  $P_1 dX_1 := PdV$ .) We assume at this stage that the functions representing degrees of freedom that are observable or ‘not hidden’,  $(U, X_1, \dots, X_{n-1})$ , form a coordinate system for  $N$ , with each  $P_i : N \rightarrow \mathbb{R}$  a smooth function. Since heat is by definition a contribution to energy, the discussion above suggests that it can be represented by a one-form too; a standard convention is to write it with a ‘slash’, as  $dQ$ , to emphasise that it might not be exact. We thus arrive at an expression of our conclusion above, that energy can be written as a sum of heat  $dQ$  and work  $\sum_{i=1} P_i dX_i$ , namely,

$$dU = dQ + \sum_{i=1} P_i dX_i. \quad (6.1)$$

Eq. (6.1) is called the *first law of thermodynamics*.<sup>18</sup>

<sup>16</sup> Feynman’s ‘ping pong ball’ analogy explains this in simple terms: the compression of a piston into the gas, like a moving ping pong paddle, increases the kinetic energy of the particles that collide with it (Feynman 1963a, §1–2).

<sup>17</sup> A *one-form* on a manifold  $N$  is a smooth field of covectors, which at each point maps each vector at that point to a real number. A one-form  $\omega$  is *exact* if and only if there exists a smooth function  $f : N \rightarrow \mathbb{R}$  such that  $\omega = df$ , where  $d$  is the exterior derivative.

<sup>18</sup> If  $dQ$  is any one-form on a manifold for which coordinate variables are given by  $(U, X_1, \dots, X_{n-1})$ , it can always be expressed as in Eq. (6.1). So, the argumentation above can be viewed as establishing

For our purposes, an important formal question is whether the heat one-form  $dQ$  can be expressed in the same way as the other contributions to energy  $P_i dX_i$ , in the sense that there exist smooth functions  $S : N \rightarrow \mathbb{R}$  and  $T : N \rightarrow \mathbb{R}$  such that  $dQ = TdS$ . The function  $T$  is then formally known as an *integrating factor*. The physical significance of this property is that, whatever degrees of freedom give rise to the ‘hidden’ contributions to energy in the form of heat, they can all be represented by single parameter  $S$  called thermodynamic *entropy* and a parameter  $T$  called thermodynamic *temperature*. These functions might not exist in general: Myrvold (2020c) presents thermodynamics in a way that does not assume this. However, many applications of thermodynamics apply only to situations in which they do exist.

Carathéodory (1909) gave an influential argument that a one-form representing physical heat should have an integrating factor, given the existence of “adiabatically inaccessible points” in every neighbourhood of each point. His assumption, known as *Carathéodory’s principle*, is that in every neighbourhood of every point  $p \in N$ , there exists a point that is not accessible by any piecewise-smooth curve  $\gamma$  satisfying the adiabatic (‘no heat exchange’) condition  $dQ(\dot{\gamma}) = 0$  for every tangent vector  $\dot{\gamma}$ . Intuitively, this encodes a sense in which processes involving heat exchange are ‘always available’ in thermodynamics; and, from this it was shown that there are indeed smooth functions  $T$  and  $S$  satisfying  $dQ = TdS$  to characterise temperature and entropy, respectively.<sup>19</sup> This relation of adiabatic accessibility is unexpectedly powerful: in a remarkable series of papers, Lieb and Yngvason (1998, 1999, 2000, 2013) showed that nearly all of thermodynamics can be recovered from axioms characterising adiabatic accessibility, including Carathéodory’s principle itself.

Since adiabatic accessibility can be an asymmetric relation, there has been some discussion as to whether it might give rise to the time asymmetry of thermodynamics (Uffink 2001, §9). In short, it does not. One of the central results of Lieb and Yngvason (1999) is a representation theorem, according to which an ‘entropy’ function  $S$  exists that increases with the adiabatic

that, whatever the nature of heat is,  $N$  is indeed the correct manifold on which to formulate it, as a one-form obeying Eq. (6.1).

<sup>19</sup> The proof of Carathéodory (1909) that  $dQ = TdS$  was significantly simplified by Bernstein (1960), who pointed out that Carathéodory only established his conclusion in some local neighbourhood of each point. Conditions for a global integrating factor were obtained by Bernstein (1960) and by Boyling (1968, 1972).

accessibility relation, in that  $S(p) \geq S(q)$  if and only if state  $p$  is adiabatically accessible from state  $q$ . But, as Uffink (2001) has emphasised, the axioms used to establish this are completely symmetric in time.<sup>20</sup> Marsland III, Brown, and Valente (2015) have pointed out that as a consequence of this, the representation theorem equally establishes the existence of a decreasing entropy function  $-S$  as well.

More importantly for our purposes, there is a different and more perspicuous argument for the existence of an integrating factor, which uses only energy conservation. No asymmetric orderings are needed for entropy and temperature to be defined. This clear-headed approach is due to Jauch (1972) and based on an argument of Tatiana Ehrenfest–Afanassjewa (1925). It can be formulated as follows.

**Proposition 6.1** (Afanassjewa–Jauch) *Let  $(U, X_1, \dots, X_{n-1})$  be a complete set of smooth coordinate functions of a manifold  $N$  of dimension  $n$ , and let  $\delta Q$  be a one-form on  $N$ , which implies that  $dU = \delta Q + \sum_{i=1}^{n-1} P_i dX_i$  for some smooth functions  $(P_1, \dots, P_n)$  of  $N$ . Suppose that  $W := \sum_{i=1}^{n-1} P_i dX_i$  is ‘conserved on adiabats’, in that for every closed, piecewise-smooth curve  $\gamma$  with tangent vector field  $\bar{\gamma}$  satisfying  $\delta Q(\bar{\gamma}) = 0$ , we have,*

$$\int_{\gamma} W = 0. \quad (6.2)$$

*Then there exist smooth functions  $S : N \rightarrow \mathbb{R}$  and  $T : N \rightarrow \mathbb{R}$  such that  $\delta Q = T dS$  and  $T = \partial U / \partial S$ .*

The interpretation of the theorem is as follows. Recalling that ‘work’ degrees of freedom are not enough to capture all the contributions to energy in a thermodynamic system when heat is present, the Afanassjewa–Jauch theorem assumes conversely that, roughly speaking, there are no further contributions to energy besides work and heat. More precisely: when heat exchange is absent, the total work on a closed curve vanishes, as in the classic expression of energy conservation.<sup>21</sup> From this assumption, which I have presented above as the very definition of heat in the formulation of the first law, the theorem ensures that entropy and temperature exist as the

<sup>20</sup> Lieb and Yngvason (1999, §2.6) formulate an additional axiom to guarantee entropy is a concave function, which gives rise to an entropy maximisation principle in the sense of Gibbs (Section 6.5). They also formulate a ‘thermal contact’ axiom for the description of equilibration (Lieb and Yngvason 1999, §4.1). I conjecture that, insofar as either of these additional axioms are time asymmetric, the time symmetry can be relaxed along the lines suggested in Section 6.1.

<sup>21</sup> See Roberts (2013b) for an interpretation of this statement and its relation to time reversal invariance.

formal constituents of that heat. It turns out that Carathéodory's principle of adiabatic inaccessibility follows as a corollary of this theorem.<sup>22</sup>

### 6.2.2 Thermodynamics on Contact Manifolds

This completes the conceptual underpinnings of heat in thermodynamic systems. We are now ready to set out the formal structure of thermodynamics, as it might appear in an 'advanced' course. When energy is conserved, the description of a thermodynamic system requires a manifold  $M$  of dimension  $m = 2n + 1$ : in local coordinates there is a coordinate function  $U$  characterising energy and a set of  $2n$  independent coordinate functions  $P_i$  and  $X_i$ , where the special variable  $P_0 = T$  is the temperature and  $X_0 = S$  is the entropy. Thermodynamics is about situations in which those variables satisfy the first law,  $dU = \sum_{i=0}^n P_i dX_i$ , with each  $P_i = \partial U / \partial X_i$  characterising the 'rate of change of energy' with respect to the degree of freedom  $X_i$ . For example, pressure  $P = \partial U / \partial V$  is the rate of change of energy with respect to volume, while temperature  $T = \partial U / \partial S$  is the rate of change of energy with respect to entropy.

We will characterise these statements in terms of a particular one-form  $\theta$ , called the *Gibbs one-form*,

$$\theta := dU - \sum_{i=0}^n P_i dX_i. \quad (6.3)$$

At each point of the manifold, there is a special subspace of vectors  $v$  of dimension  $n$  such that  $\theta(v) = 0$ , and it is on this subspace that the first law of thermodynamics holds. So, it would be appropriate to formulate thermodynamics on an  $n$ -dimensional submanifold  $N \subset M$  whose vectors  $v$  have this property,  $\theta(v) = 0$ . The contact manifold approach to thermodynamics is built around this idea.

We begin with some background definitions, in order to identify the essential properties of a surface on which the Gibbs one-form vanishes. When each point of  $M$  is smoothly assigned a codimension-1 hypersurface, meaning a subspace of dimension  $2n$  of the 'total'  $(2n + 1)$ -dimensional tangent space of that point, the result is called a *field of hypersurfaces*. It can always be locally characterised by a one-form  $\eta$  such that  $\eta(v) = 0$  for each vector  $v$  in a hypersurface. A *contact structure* is a field of hypersurfaces

<sup>22</sup> The converse of Carathéodory's principle is that if a heat form  $dQ$  has an integrating factor  $dQ = TdS$ , then there are adiabatically inaccessible points in every neighbourhood of every point. This well-known statement has a comparatively simple proof (cf. Boyling 1968).

satisfying a non-degeneracy condition called<sup>23</sup> ‘maximum non-integrability’:  $(d\eta)^n \wedge \eta \neq 0$ . The pair  $(M, \eta)$ , where  $M$  is a smooth manifold and  $\eta$  is a contact structure, is called a *contact manifold*.

The physical significance of maximal non-integrability is that it selects a surface  $N$  that is ‘as large as it can be’, while still satisfying the first law on a manifold of dimension  $2n+1$ ; and, it turns out that  $N$  must have dimension  $n$ , capturing an essential property of the first law. If a submanifold  $N \subset M$  is in ‘contact’ with a field of hypersurfaces, in that each of its tangent vectors  $v$  is tangent to a hypersurface  $\eta(v) = 0$ , then it is called an *integral submanifold* of the field. An integral submanifold of maximal dimension is called a *Legendre submanifold*. And, on a  $(2n+1)$ -dimensional contact manifold, a Legendre submanifold always has dimension  $n$  (cf. Arnol’d 1989, Appendix 4). So, this is just what we need to define an  $n$ -dimensional submanifold on which the first law holds.

In particular, the Gibbs one-form  $\theta$  of Eq. (6.3) is an example of a contact structure; conversely, every contact structure  $\eta$  can be expressed in local coordinates in the form of the Gibbs one-form (Arnol’d 1990, p.167). So, the essential structure of the first law really is captured by a contact manifold of dimension  $2n+1$ , together with a Legendre submanifold of dimension  $n$ , which characterises a maximal surface on which the first law holds,  $\theta = 0$ . The contact manifold  $(M, \theta)$  is often referred to as a *thermodynamic phase space* and the choice of a Legendre submanifold as a ‘representation’, such as the energy or entropy representation. This allows for a precise definition of a model of thermodynamics, originally proposed by Hermann (1973, p.264):

**Definition 6.1** An ‘equilibrium thermodynamic system’ is a triple  $(M, \theta, N)$ , where  $M$  is a  $(2n+1)$ -dimensional manifold,  $\theta$  is a contact structure, and  $N$  is an  $n$ -dimensional Legendre submanifold.

Writing the contact structure in local coordinates as the Gibbs one-form, and expressing the energy variable  $U$  as a function of the  $n$  coordinate variables on a Legendre submanifold, we now get what Gibbs (1873) called the ‘fundamental relation’ of a thermodynamic system:

$$U = f(X_1, \dots, X_n). \quad (6.4)$$

Taking derivatives, we get  $dU = (\partial f/\partial X_1)dX_1 + \dots + (\partial f/\partial X_n)dX_n$ , and defining  $P_i := \partial U/\partial X_i$  for each  $i = 1, \dots, n$ , we recover the first law.

<sup>23</sup> This condition gets its name from the concept of an ‘integrable’ field of hypersurfaces, characterised by the condition that  $d\eta \wedge \eta = 0$ . A manifold can only satisfy maximal non-integrability if it is of odd dimension; for an introduction, see Arnol’d (1989, Appendix 4).

The  $n$  equations for  $P_i$  are known as ‘equations of state’. For example, a simple gas has thermodynamic phase space given by the five-dimensional manifold  $M$ , with the Gibbs one-form  $\theta = -dU + TdS - PdV$ . The two-dimensional Legendre submanifold is characterised by the coordinate functions  $(V, S)$ , and the equations of state are determined by calculating  $P = \partial U/\partial V$  and  $T = \partial U/\partial S$ . The statement that  $V$  and  $S$  are ‘extensive’ or scale with the energy of the system is captured by a further assumption<sup>24</sup> that  $U$  is a first-degree homogeneous function,  $f(\lambda V, \lambda S) = \lambda f(V, S) = \lambda U$  for all  $\lambda > 0$ . The remaining variables  $P$  and  $T$  are called ‘intensive’.

Virtually all aspects of equilibrium thermodynamics can be treated using this geometric foundation as a starting point. Gibbs (1873) himself pointed out that even the description of phase transitions can be captured by the study of the boundaries of convex surfaces, of the kind discussed in [Section 6.5](#).<sup>25</sup> More recent work has approached this through the study of an additional metric and its curvature singularities (cf. Quevedo 2007; Quevedo et al. 2011). But, this is more or less where the consensus about the foundations of thermodynamics ends. We have so far said nothing about dissipation or the second law, or their bearing on the arrow of time. We will review these matters in the remaining sections. But first, let me develop one more piece of background material: the nature of time and time reversal in thermodynamics.

### 6.2.3 *Time and Time Reversal in Thermodynamics*

It is well-known that the description of time is an awkward part of equilibrium thermodynamics: in spite of the name, the theory does not describe time and change in the way that most dynamical theories do:

In contrast to mechanics, thermodynamics does not possess equations of motion. This, in turn, is due to the fact that thermodynamical processes only take place after an external intervention on the system (such as: removing a partition, establishing thermal contact with a heat bath, pushing a piston, etc.). (Uffink 2001, p.315)

This latter fact led Wallace (2014) to argue that thermodynamics is “mis-named” and better viewed as part of the general study of interventions using control theory. In a similar spirit, Myrvold (2011) formulates it as a theory relative to the “means” available to the experimentalist, referring to thermodynamics as a “resource theory” (Myrvold 2020b, Forthcoming).

<sup>24</sup> We will return to this assumption in Section 6.5.

<sup>25</sup> For an introduction, see Wightman (1979).

Other authors refer to large parts of thermodynamics as ‘thermostatistics’ (cf. Tribus 1961). Nevertheless, there are at least two ways that time evolution is routinely represented in thermodynamics, which I will refer to as the *quasistatic process approach* and the *constraint approach*.

My discussion of each approach will adopt the perspective developed in earlier chapters, which I have called the Representation View: instead of studying time as a coordinate variable, we study time translations and view time evolution in a physical theory as a representation of those time translations on a state space.<sup>26</sup> In this way, studying the properties of time translations on state space can be used to determine the symmetries of time itself. As in earlier chapters, we take a representation of time translations to be a homomorphism  $\varphi$  from a (possibly local) Lie group of time translations  $G$  to the automorphisms of a theory’s state space. The set of automorphisms of an equilibrium thermodynamic system  $(M, \theta, N)$ , which we denote by  $\text{Aut}(M, \theta, N)$ , consists of *contact transformations*, or smooth maps  $\phi : M \rightarrow M$  that preserve the contact structure,  $\phi^*\theta = \theta$ ; we assume also that they preserve the Legendre submanifold  $N$ .

### *The Quasistatic Process Approach*

The quasistatic process approach, advocated by Carathéodory (1909) and Ehrenfest–Afanassjewa (1925); Ehrenfest–Afanassjewa (1956), views a piecewise-smooth curve through thermodynamic phase space  $(M, \theta)$  as an approximate representation of ‘quasistatic’ change: a process for which the system can at all times be approximated as being in an equilibrium state. This occurs, for example, when the volume of a box of gas is very slowly increased. There is some philosophical work to do in clarifying exactly how this representation is supposed to work (see Lavis 2018; Norton 2014, 2016b).<sup>27</sup> However, that debate is tangential to my purposes, which will aim to determine whether there is any sense of irreversibility in this perspective on time translations.

Viewing time as associated with the Lie group  $G = (\mathbb{R}, +)$ , a representation of time translations of this kind is given by any map  $t \mapsto \varphi_t$  to the contact transformations<sup>28</sup> of  $(M, \theta, N)$  that is continuous, preserves  $N$ , and satisfies  $\varphi_{t+t'} = \varphi_t \circ \varphi_{t'}$ . This allows us to apply the account of time reversal developed in [Chapter 2](#), as an extension of the group of time translations to include an

<sup>26</sup> See Section 2.3.

<sup>27</sup> Norton (2016b) points out that this concern goes back at least to Duhem (1903, §58) and offers an interesting clarification; see Valente (2017) for an alternative view.

<sup>28</sup> Strictly speaking, it is only each smooth part of the flow  $\varphi_t$  that defines a (local) representation of time translations; the piecewise-smooth flow  $\varphi_t$  may then represent a finite sequence of representations of time translations corresponding to a sequence of successive processes.

element that reverses time translations,  $t \mapsto -t$ . When our representation can be extended to include time reversal, it transforms a curve representing a quasistatic process to one that ‘flows in the opposite direction’ with the reverse tangent vector field, and we say that the theory is time reversal invariant; otherwise, it is time reversal violating. This recovers one standard approach to time reversal in thermodynamics (cf. Uffink 2001, §3).

Most who argue that thermodynamics is time asymmetric adopt the quasistatic process approach to time and time reversal: this is the case for both Clausius (discussed in [Section 6.3](#)) and Planck (discussed in [Section 6.4](#)). In the next sections, I will argue that they do not succeed. But first, let me review a very different approach to change in thermodynamics.

### *The Constraint Approach*

The constraint approach is an alternative view of thermodynamic prediction, which finds its natural home in the work of Gibbs (1876, 1877), and more recently in the textbook by Callen (1985), who writes:

The single, all-encompassing problem of thermodynamics is the determination of the equilibrium state that eventually results after the removal of internal constraints in a closed, composite system. (Callen 1985, p.26)

This approach is fundamentally about the relationship between two sets of equilibrium systems. The first is a collection of thermodynamic systems, often interpreted as subject to some ‘constraint’; the second is a single equilibrium system that is taken to be determined by the first collection, usually by the removal of that constraint. We saw a paradigm example of this above: two gases separated by a barrier can be viewed as determining a mixture when the barrier is removed and once the gases have settled down into an equilibrium state again.

The constraint approach to equilibrium thermodynamics has the advantage of avoiding the awkward appearance of non-equilibrium states. Instead, we describe a process in two stages: one characterising a collection of separate equilibrium systems, and another characterising the single equilibrium system that they determine. We can make predictions by associating these two stages with two moments in time, as when we first observe two separate gases and then later observe a mixture. We can even introduce structure that captures the fact that the entropy of the unconstrained system maximises the entropy of the constrained ones (see [Section 6.5](#)). However, this introduction of a temporal direction is entirely independent of the constraint approach because nothing about the relationship it describes between systems requires them to occur in any particular time-order. On the contrary,



a collection of equilibrium systems determines another equilibrium system, regardless of whether the latter occurs at an earlier or later time.

Remarkably, the constraint approach provides a ‘timeless’ view of thermodynamics. We can use it to model changes in time if we wish and can even supplement it with a temporal direction. But, these are at best trivial additions of a thermodynamic arrow, in the sense of ad-hoc additions not required by the otherwise undirected structure of equilibrium states. In the remaining sections, I will argue that even the classic expressions of the second law have this character: when they can be rigorously formulated at all, they are fundamentally undirected in time.

### 6.3 The Clausius Inequality

Most arguments in support of a thermodynamic arrow begin with the second law and then argue that entropy is non-decreasing in just one temporal direction. Perhaps the oldest and most well-known argument of this kind is due to Clausius.<sup>29</sup> The Clausius argument for increasing entropy suffers from conceptual problems, as Uffink (2001) has pointed out. In this section I will give a precise reconstruction of it and argue that its use as evidence for a thermodynamic arrow is completely unfounded.

The Clausius approach adopts less structure than the complete thermodynamic state space introduced in Section 6.2.2. In particular, we begin with a manifold  $N$  of dimension  $n$  on which the first law holds but without assuming at the outset that there is a global entropy function  $S$  that can be viewed as a coordinate variable. Thus, as in Section 6.2.1, we write the first law as  $dU = \delta Q + \sum_{i=1}^{n-1} P_i dX_i$ , where  $(U, X_1, \dots, X_{n-1})$  are coordinate functions for  $N$ , each  $P_i$  is a smooth functions of  $N$ , and  $\delta Q$  is a one-form representing heat. A central part of this argument is now to assume that it is possibly the case that  $\delta Q \neq TdS$  for all smooth functions  $T : N \rightarrow \mathbb{R}$  and  $S : N \rightarrow \mathbb{R}$  – and so, by the Afanassjewa–Jauch theorem, energy is not necessarily conserved in the sense that Eq. (6.2) fails! But let us press forward for now. Then, the *Clausius inequality* says: given any ‘physically possible’ cycle described by a closed curve  $c$ , together with a function  $T$  representing temperature and a one-form  $\delta Q$  representing heat, we have that

$$\int_c \delta Q/T \leq 0. \quad (6.5)$$

<sup>29</sup> This argument was introduced by Clausius (1865); see Uffink (2001) for a detailed review and Henderson (2014) and Valente (2021, §4.1) for recent supportive remarks.

Clausius argued that this inequality follows from what is now known as the ‘Clausius version’ of the second law, that “[h]eat cannot of itself pass from a colder to a hotter body without some other change, connected herewith, occurring at the same time” (see Uffink 2001, p.333).

Two definitions are now needed: we say that a curve  $\gamma$  is *thermodynamically reversible* with respect to a one-form  $dQ$  if and only if  $dQ$  has an integrating factor when it is restricted to that curve, in that there are smooth functions  $T$  and  $S$  such that  $dQ = TdS$  at least on every tangent vector to  $\gamma$ , though not necessarily in general. Otherwise, we will say that  $\gamma$  is *thermodynamically irreversible*. If a closed curve  $c$  is thermodynamically reversible, then, since the integral of an exact one-form is path-independent, the Clausius inequality becomes identically zero. In short, strict inequality only holds on a closed curve that is thermodynamically irreversible with respect to  $dQ$ .

The Clausius argument that ‘entropy does not decrease’ can now be reconstructed as follows:

**Proposition 6.2** *Let  $T$  be a smooth function on a manifold  $N$ , and let  $dQ$  be a one-form. Let  $\gamma$  be a smooth curve from  $p_i$  to  $p_f$  that is adiabatic ( $dQ(\bar{\gamma}) = 0$  for all tangent vectors  $\bar{\gamma}$ ) but not necessarily reversible with respect to  $dQ$ . If the Clausius inequality is satisfied,  $\int_c dQ/T \leq 0$  on all closed curves, and if there exists a curve  $\gamma'$  from  $p_f$  to  $p_i$  along which  $d(dQ) = 0$ , then  $dQ = TdS$  on  $\gamma'$  for some smooth function  $S$ , and any such function must satisfy  $S(p_i) \leq S(p_f)$ .*

*Proof* The fact that  $d(dQ/T) = 0$  along  $\gamma'$  says that  $dQ/T$  is a closed one-form on the contractible manifold  $\gamma'$ , and so the Poincaré lemma implies that  $dQ/T$  is exact on  $\gamma'$ , in that there exists a smooth function  $S$  such that  $dQ/T = dS$  for all tangent vectors  $\bar{\gamma}'$ .

A closed curve  $c$  now runs from  $p_i$  to  $p_f$  along  $\gamma$  and then back along  $\gamma'$  from  $p_f$  to  $p_i$ , as in Figure 6.3. Therefore, applying our adiabatic assumption and the Clausius inequality,

$$\int_{\gamma'} dQ/T = \underbrace{\int_{\gamma} dQ/T}_{=0} + \int_{\gamma'} dQ/T = \int_c dQ/T \leq 0. \tag{6.6}$$

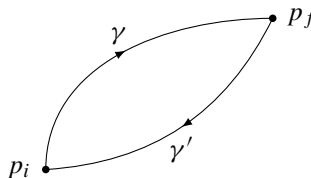


Figure 6.3 The closed curve  $c$ .

But, we also have by definition that

$$\int_{\gamma'} dQ/T = \int_{S(p_f)}^{S(p_i)} dS = S(p_i) - S(p_f). \quad (6.7)$$

Combining Eqs (6.6) and (6.7) thus establishes that  $S(p_i) \leq S(p_f)$ . ■

In summary, the precise sense in which Clausius argues that entropy is non-decreasing is: if thermodynamics is equipped with the structure of a second law in the form of the Clausius inequality, then any (possibly thermodynamically irreversible) adiabatic process  $\gamma$  from  $p_i$  to  $p_f$  can be supplemented with a smooth function  $S$ , which behaves like entropy with respect to a different, reversible process  $\gamma'$  and is such that  $S(p_i) \leq S(p_f)$ .

As Uffink (2001, p.339) has emphasised, this argument is not perfectly general because of its restriction to adiabatic processes; and, one might be concerned about whether a reversible path  $\gamma'$  exists on which entropy is defined or, if it does, whether it represents any physically realistic process. I think the situation for Clausius is much worse: the conceptual problems with this argument are so severe that I do not see any way to justify, on this basis, Clausius' conclusion that entropy tends to increase. Here are three problems with the argument.

#### *Problem 1*

The Clausius argument applies only to systems with a heat form  $dQ$  that does not admit global temperature and entropy functions for which  $dQ = TdS$ . By definition, this is required for the existence of an irreversible curve. The problem is that this also implies a violation of local energy conservation, since otherwise we would have  $dQ = TdS$  by the Afanassjewa–Jauch theorem. That might be taken as evidence that entropy can only increase in 'open' systems, for which local energy conservation can be expected to fail. But, then the often-repeated dictum of Clausius that "[t]he entropy of the universe tends to a maximum" would be false, insofar as the universe is by definition a closed system.<sup>30</sup> This means that it is not an argument that establishes a direction of time for the universe as a whole, but at best for local subsystems.

<sup>30</sup> Clausius writes, "Die Entropie der Welt strebt einem Maximum zu" (Clausius 1867, p.44). Followers of Clausius like Planck were aware of difficulties with this statement; see Uffink (2001, pp.338–9). Earman (2006) has argued that such statements about the entropy of the universe are not even false, but meaningless.

## Problem 2

A second, more serious problem is that the Clausius argument only establishes the existence of an increasing entropy function with respect to the curve  $\gamma'$ : why is the function  $S$  associated with this reversible curve chosen, and not some other? Even worse, the smooth function  $S$  constructed to represent the 'entropy' of an irreversible curve  $\gamma$  in Proposition 6.2 is questionable: it deserves the name at best on the *reversible* curve  $\gamma'$  for which  $dQ = TdS$  but not on the irreversible curve  $\gamma$  for which  $dQ$  does not generally take this form. When this  $dQ$  is not exact, it is not associated with any global entropy function at all. So, the choice to interpret this particular function  $S$  as the entropy of a 'possibly irreversible' adiabatic process is totally arbitrary, since a different choice of reversible curve  $\gamma$  would have led to a different entropy function. That this arbitrary function is standardly referred to as unqualified 'entropy' is a remarkable sleight of hand.

Let me make the arbitrariness stark: we could equally have chosen the function  $S' = -S$  to represent entropy along the (possibly) irreversible curve  $\gamma$ . This is a function for which entropy is non-*increasing* from  $p_i$  to  $p_f$ . The very same curve  $\gamma'$  running in the opposite direction from  $p_f$  to  $p_i$  undergoes a positive increase in entropy with respect to this new function  $S'$ . It is hard to see why this reversed path should be impossible from the perspective of the Clausius argument, since  $\gamma'$  represents a reversible process by construction. Of course, one could avoid the problem by postulating that  $S' = S = 0$ . But, then the argument would just result in entropy being constant and not that it 'tends to increase'.

## Problem 3

This last problem suggests that on general symmetry considerations, the Clausius inequality cannot establish a general increase of entropy, even for adiabatic processes. Suppose we say that every 'physically possible' cycle is described by a triple  $(dQ, T, c(t))$  that satisfies the Clausius inequality  $\int_{c(t)} dQ/T \leq 0$ , where I now write the explicit parametrisation of the curve  $t \mapsto c(t)$ . For each such system, simple properties of integrals<sup>31</sup> imply that  $(dQ, -T, c(-t))$ ,  $(-dQ, T, c(-t))$ , and  $(-dQ, -T, c(t))$  satisfy the Clausius inequality as well, where  $t \mapsto c(-t)$  denotes the parametrisation of the curve  $c(t)$  in the reverse direction. The former two of these triples describe a process unfolding along the same curve  $c$ , but in the reverse direction. So, if entropy is increasing along some portion of that curve in the

<sup>31</sup> Namely,  $\int_{c(t)} dQ/T = -\int_{c(-t)} dQ/T = \int_{c(-t)} dQ/(-T) = \int_{c(-t)} (-dQ)/T$ .

cycle ( $dQ, T, c(t)$ ), then entropy is decreasing in these other two cycles. This problem was pointed out by Ehrenfest–Afanassjewa (1925), who ruled it out by the further postulate of a positive temperature function  $T > 0$ . However, I do not see any way to rule out any of these systems without assuming by fiat what we were originally trying to prove: that entropy must increase to the future, not to the past.

I agree that the Clausius inequality does provide a substantial addition to the structure of thermodynamics. However, I can find no physically meaningful sense in which it establishes an increase in entropy, let alone a general arrow of time. Indeed, although Clausius suggested that his inequality implies that the entropy of the universe tends to increase in his earlier papers, it is not surprising that he appears to have deleted any mention of this some years later when those papers were collected into a book.<sup>32</sup>

#### 6.4 Planck's Argument

Another popular approach to arguing that entropy is non-decreasing is due to Planck (1897b). Planck produced a series of related arguments in the eleven successive editions of his influential book on thermodynamics, which have been carefully reconstructed and evaluated by Uffink (2001, §7, §10.1). Planck's arguments are even less rigorous than the Clausius argument I have presented above and suffer from similar conceptual problems. It is enough for my purposes to review one of them here. For, as Uffink (2001, p.376) points out, even the final version does not give a substantial improvement on the earlier ones.

Like Clausius, Planck adopts the continuous process approach to describing the passage of time. So, if he can show that a continuous curve from one state to another is such that no physical process exists that can restore the original state, then he will have established an arrow of time. Planck calls such a curve 'irreversible', which differs from the usage of Clausius, who adopts this word specifically to describe curves on which the heat form fails to admit an integrating factor,  $dQ \neq TdS$ .

Planck's style of argument requires one to say which continuous curves are 'physically impossible', which is not found in the structure of thermodynamics presented in Section 6.2. For this, Planck, adopts an expression of

<sup>32</sup> This was pointed out by Kuhn (1987); see Uffink (2001, pp.339–340).

the second law of thermodynamics, which he takes to prohibit the existence of a *perpetuum mobile* of the second kind. This is known as:

*Kelvin's Principle*: No cyclic process exists whose effect on the environment "consists of nothing other than the production of work and the absorption of the equivalent heat."<sup>33</sup>

Here, one considers a description of two thermodynamic systems, one representing the 'target' of investigation and the other the 'environment'. In the environment, since the energy one-form is exact, a closed curve will satisfy  $\int_c dU = 0$ , and so the first law implies that  $-\int_c \delta Q = \int_c W$ , where  $\delta Q$  and  $W = \sum_i P_i dX_i$  represent heat and work done on the environment, respectively. Kelvin's principle of no work or heat loss can now be interpreted formally as the statement that  $-\int_c \delta Q = \int_c W \leq 0$ . Like in the Clausius argument, we must assume here that the heat one-form  $\delta Q$  is not exact, violating local energy conservation by the Afanassjewa–Jauch theorem, since otherwise the inequality would be identically zero.

Planck's style of argument is now to consider a process along which entropy increases and to present a *reductio ad absurdum* of the hypothesis that some physical process is described by a curve that leads back to the original low-entropy state. To be concrete, I will review just one of Planck's examples. Consider the mixing of two gases of the same temperature, one of which has higher pressure, and which are initially separated by a barrier, as in the example of Section 6.1. Here is a summary of the process Planck (1897b, §118–124) describes and which is illustrated in Figure 6.4.<sup>34</sup>

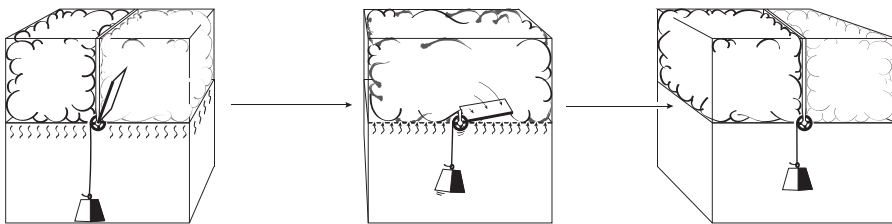


Figure 6.4 Planck's procedure for producing work violating Kelvin's principle from an entropy-decreasing process.

<sup>33</sup> Planck (1897b, p.81) writes: "eine Veränderung, die in nichts Anderem besteht als in Erzeugung von Arbeit und Absorption der äquivalenten Wärme". For a discussion of its origin in the work of Clausius and Kelvin, see Uffink (2001, §5).

<sup>34</sup> Uffink (2001, p.352) refers to this argument as Planck's "Lemma 6".

- *First stage:* Two unmixed gases are assumed to have the same temperature but one having higher pressure, and the 'environment' is assumed to be associated with some large amount of heat.
- *Second stage:* The gases are slowly (quasistatically) transformed into a mixture, while absorbing heat from the environment and losing work in equal amounts, with the temperature of the gases remaining the same throughout. The environment correspondingly loses heat and gains work in equal amounts, which we interpret as 'heat and work exchange' between the system and its environment.
- *Third stage:* The mixture is now completely isolated from the environment, and the *reductio* hypothesis is applied: we assume there is an entropy-decreasing process that restores the mixture to its original low-entropy state as two unmixed gases. The environment remains in its previous state of having lost heat and gained work.

The result is that the gases complete a cycle, and the environment loses heat and gains work in the sense that  $-\int_c dQ = \int_c W > 0$ , thus violating Kelvin's principle. The conclusion by *reductio* is that the entropy-decreasing process that restores the original state is impossible, and so time reversal invariance is violated.

As Uffink (2001, §7.5) points out, there are a number of problems with this approach: for example, Planck does not manage to give meaning to the concept of 'entropy' in arbitrary non-equilibrium systems; and, his formulation of the second law is not clearly justified for arbitrary thermodynamic processes.

However, the problems with the Clausius argument discussed in [Section 6.3](#) apply to Planck's argument as well. In the first place, the heat one-form  $dQ$  is not exact by assumption, and so Planck's systems do not satisfy local energy conservation. This means that his argument at best establishes a time asymmetry of open systems and not for closed systems or for the universe as a whole. In the second place, the fact that there is not generally a function  $S$  such that  $dQ = TdS$  means that the association of an entropy function with Planck's proposed irreversible descriptions is questionable at best. Although his discussion is less formal than the one of the previous section, this does not relieve him of the burden of giving meaning to concepts he uses.

Finally, there is a symmetry argument here too, which makes it hard to see how Planck's thinking could ever give rise to a time asymmetry. The key assumption, Kelvin's principle, says that whatever initial thermodynamic system we begin with, we must attain a new system in which there is no

net increase in work and heat loss in the environment. Formulated in this way, Kelvin's principle is similar to the Minus First Law's requirement that every system 'attain' a state of equilibrium. This sort of statement is not intrinsically directed in time: it can be consistently applied whether the system attained is in the past or in the future. At its core, the principle only expresses that a target and environment system determine a 'new' target and environment system, in which there is no net increase in work and heat loss in the new environment. But, if this principle can be applied at all, then it can be applied whether that 'new' pair of systems is in the future or in the past.

In other words, the structure of thermodynamics can be retained, even with the addition of Kelvin's principle, without introducing a temporal direction. If one particular temporal direction is chosen, then this appears to be due to a special human understanding of what it means to be an intervention, or to special initial and boundary conditions. But, this is what I have called a 'trivial' introduction of a time asymmetry, which arises from heuristic considerations from outside the theory.

### **6.5 Concavity, Convexity, and Gibbs**

Our final discussion of time asymmetry in thermodynamics involves the Gibbs formulation of the second law. Gibbs (1876, 1877) set out his great vision of thermodynamics in a monograph-length work, which spanned over 300 pages in two journal issues. The mathematician Robert Hermann had a positive view of this work and an impressively negative view of the work that followed:

After much reading on the subject, I would say that Gibbs understood, in about 1870, the mathematics of thermodynamics – even in its most 'modern' form – better than almost all of the authors who followed him. Unfortunately, physicists and chemists think that a reasonable statement of the mathematics which they are trying to apply is an 'axiomatization,' and therefore should be avoided as a bad thing which inhibits the creative mind. (Hermann 1973, p.261–2)

In particular, the Gibbs formulation of the second law provides an elegant and conceptually clear addition to the mathematics of thermodynamics, which retains the basic character of the second law as a 'principle of entropy maximisation'. However, the prospects for this approach to establish temporal asymmetry have been succinctly captured by Uffink (2001, p.361): "Obviously, there are no implications for the arrow of time in the second law as formulated by Gibbs". I agree. However, it is helpful to briefly review



the Gibbs approach, as a step towards clarifying how it is possible for thermodynamics to have no arrow of time.

A modern view of the Gibbs approach begins with an equilibrium thermodynamic system  $(M, \theta, N)$ . The Legendre submanifold  $N$  is described in terms of a fundamental relation, which I will write in coordinate form as  $U = f(S, X_1, \dots, X_{n-1})$ , where  $f$  is smooth and is a bijective function of entropy  $S$  when all the other variables are held fixed. I will continue to make this argument using coordinate variables, since this will allow me to make the point in relatively simple language. However, deep approaches also exist that are entirely geometrical.<sup>35</sup>

To motivate a second law, we will make just three physical assumptions about a system that has attained equilibrium. I will state these assumptions first and then comment on their interpretation. Perhaps surprisingly, these are all that we need to derive an entropy maximisation principle:

1. *Temperature is positive*  $T := \partial U / \partial S > 0$ .
2. *Energy is (positively) first-degree homogeneous*, in that when viewed as a function of the other variables  $U(S, X_2, \dots, X_n)$ , it satisfies  $U(\lambda S, \lambda X_1, \dots, \lambda X_{n-1}) = \lambda U(S, X_1, \dots, X_{n-1})$  for all  $\lambda > 0$ .
3. *Energy has a strongly stable ground state* with a global minimum and no other local minima or maxima; in other words,  $U$  is a *convex function*. A convex function of one variable is defined by the property that  $U(\alpha S_1 + (1 - \alpha)S_2) \leq \alpha U(S_1) + (1 - \alpha)U(S_2)$  for all  $\alpha \in (0, 1)$ , which implies that it has the expected shape illustrated in [Figure 6.5](#).

Let me now interpret these statements. The requirement that temperature  $T := \partial U / \partial S$  is positive implies that energy is a monotonically increasing function of entropy. As far as I can see, whether it is increasing or decreasing is conventional. However, the statement that temperature is monotonic is a substantial empirical postulate about the degrees of freedom associated with heat, which by definition are not available to intervention. Postulate 2 says that when energy changes with respect to those degrees of freedom, the entropy function  $S$  used to capture them never reaches a local extremum. This need not be true of all physical systems;<sup>36</sup> however, equilibrium thermodynamics can be viewed for our purposes as a theory about those systems for which it is.

<sup>35</sup> For geometric approaches to these same ideas, see especially the treatment of convexity by Hermann (1973, §6.6), as well as Bravetti (2018), Mrugała (2000), and Weinhold (1975).

<sup>36</sup> Onsager (1949) predicted both positive and negative statistical mechanical temperatures associated with a local entropy maximum in fluids with no stable steady flow, and Ramsey (1956) postulated them for classical thermodynamics, although they are not associated with any equilibrium states and so fall outside the scope of equilibrium thermodynamics.

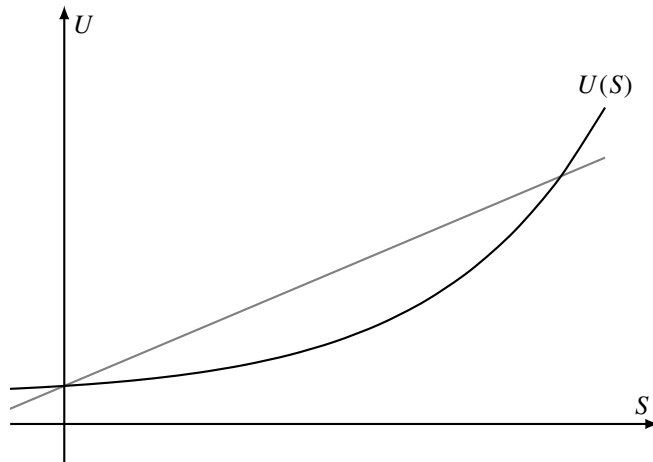


Figure 6.5 A convex function  $U(S)$ .

The statement that energy is a first degree homogeneous function is also substantial. But it is a natural assumption about thermodynamic systems that have attained equilibrium. Roughly speaking, it expresses the idea that such a system consists of ‘similar stuff’ in all of its parts, with all inhomogeneities smoothed out. First-degree homogeneity expresses this property in a clever way: we assume that when we scale a thermodynamic system by some factor, such as by doubling the volume and entropy of an ideal gas, the energy is scaled by the same factor.

Finally, we have assumed that energy has a stable ground state in a certain strong sense, expressed by the requirement of convexity. This assumes, as one would typically do, that energy has a global minimum. But, our assumption demands an even stronger sense of stability: that no amount of perturbation will send the system into a ‘local’ ground state characterised by a local minimum or maximum. Again, there is no requirement that every physical system will satisfy this assumption. But, it is a plausible restriction on the domain of application of thermodynamics; let us restrict our study to systems that are stable enough to satisfy this requirement.

These assumptions, remarkably, are enough to establish an entropy maximisation principle. To manage notation, let me write  $X = (U, X_1, \dots, X_{n-1})$  to denote a set of values of energy and the other non-entropy extensive variables. Then we have the following result.<sup>37</sup>

<sup>37</sup> Compare Wightman (1979, pp.xiv–xvi) and Wills (2022) for similar ideas.

**Proposition 6.3** Let  $U = f(S, X_1, \dots, X_{n-1})$  be a smooth function, bijective in  $S$ , such that,

1. (positive temperature)  $T = \partial U / \partial S > 0$ ;
2. (homogeneity)  $U(X) = f(S, X_1, \dots, X_{n-1})$  is first-degree homogeneous; and
3. (strong stability)  $U = f(S, X_1, \dots, X_{n-1})$  is convex.

Then  $S(X) = S(U, X_1, \dots, X_{n-1})$  is concave, and for any fixed value of  $X = (U, X_1, \dots, X_{n-1})$ ,

$$S(X) = \sup \{S(\bar{X}) + S(\tilde{X})\}, \quad (6.8)$$

where the supremum ranges over all points  $\bar{X}$  and  $\tilde{X}$  such that  $X = \bar{X} + \tilde{X}$ .

*Proof* The function  $U$  is a smooth, convex bijection of  $S$ , and so it is strictly monotonic in  $S$ . Since  $\partial U / \partial S > 0$ , this means that it is strictly increasing, which by an application of the definitions implies that its inverse  $S(U, X_1, \dots, X_{n-1})$  is concave. Since  $S$  is the inverse of a homogeneous function, it is also homogeneous.

Now, viewing  $S$  as a function of  $X := (U, X_1, \dots, X_{n-1})$ , an elementary property of first-degree homogeneous functions  $S(X)$  guarantees that  $S$  is concave if and only if

$$S(\bar{X} + \tilde{X}) \geq S(\bar{X}) + S(\tilde{X}), \quad (6.9)$$

for all values of  $\bar{X}$ ,  $\tilde{X}$  (Rockafellar 1970, Theorem 4.7). For any fixed value of  $X$ , this means that  $S(X)$  is an upper bound for the set  $\{S(\bar{X}) + S(\tilde{X}) \mid \bar{X} + \tilde{X} = X\}$ . Moreover, if  $u \leq S(X)$  is any other upper bound, then since  $S(X/2) + S(X/2)$  is an element of the set, homogeneity implies

$$u \geq S(X/2) + S(X/2) = S(X)/2 + S(X)/2 = S(X) \quad (6.10)$$

and hence that  $u = S(X)$  is a supremum. ■

Let me comment briefly on the interpretation of this theorem. For a given thermodynamic state with fixed values of energy and the 'work' degrees of freedom,  $X = (U, X_1, \dots, X_n)$ , we can consider any pair of states  $\bar{X}$  and  $\tilde{X}$  that sum to the original ( $\bar{X} + \tilde{X} = X$ ) to be 'possible contributions' to that state. We might now imagine that each of these various possible contributions could in fact occur and note that when we sum the entropies  $S(\bar{X}) + S(\tilde{X})$  of each pair of contributing states, we will get a variety of different values. But, the theorem says that when systems achieve equilibrium in the strong sense of our three postulates – positive temperature, homogeneity, and a strong sense of stability – then the only possible contributions that actually occur

are those that render the total entropy as large as it can be. The entropy of a given state is the supremum of the possible contributing entropies.

Entropy maximisation in this ‘Gibbs sense’ has a wide range of applications in equilibrium thermodynamics. It is sufficient, for example, to identify the entropy increase of a pair of mixing gases (cf. Wightman 1979). Wills (2022) has shown that it can be applied in the resolution of a well-known thermodynamic puzzle, the Gibbs paradox.

However, entropy maximisation in the Gibbs sense *has no implications about the direction of time*. The principle might be viewed as describing how the entropy of possible contributions or ‘parts’ of a system are related to the whole. But, this is a statement about what philosophers call ‘mereology’, or the nature of parthood relations, and not about the way things change over time.

As a result, when Gibbs entropy maximisation is applied to describe time developments, it remains temporally symmetric. To go back to our paradigm example from [Section 6.1](#): if a mixture in equilibrium is viewed as arising from a pair of separate gases, that mixture will achieve a maximum entropy state, whether it is viewed as occurring in the future or in the past. That is, Gibbs entropy maximisation is most naturally viewed from the ‘constraint’ perspective on prediction in thermodynamics introduced in [Section 6.2.3](#), that a collection of thermodynamic systems constrains another equilibrium system to achieve maximum entropy, whether that maximum occurs in the future or in the past.

## 6.6 Summary

In spite of the heterodox thesis of this chapter, most of what I have said is already suggested by the work of Uffink (2001) and Brown and Uffink (2001): a direction of time is not encoded in the formal structure of thermodynamics in its first or second laws. Insofar as a time asymmetry is found there, it is ‘trivial’, in the sense of an ad hoc addition that is not required by the theory. Their argument is only further supported by the presentation I have given of the formal structure of thermodynamics: of the work of Carathéodory, and of the second law as it appears in the work of Clausius, Planck, and Gibbs. When the latter arguments are made precise, any semblance of a thermodynamic arrow of time disappears.

To reach my conclusion, one only needs to take Brown and Uffink’s observations about the second law and apply them equally to the ‘approach to equilibrium’ encoded in their Minus First Law. Like the second law, the

association of each collection of thermodynamic systems with a unique equilibrium system does not require a temporal direction. All the formal foundations and predictive structure of thermodynamics can be constructed in its absence. Of course, one can still postulate temporal asymmetry in a trivial way, as in every temporally symmetric theory, by supplementing it with the explicit postulate that change occurs asymmetrically in time. But, if we do not indulge in trivialities, then there is no thermodynamic arrow of time.