Does thermodynamics have a reversibility problem?

Bryan W. Roberts
b.w.roberts@lse.ac.uk

July 7, 2023

Abstract

No.

Contents

1. Introduction

Nagel reported feeling "acute intellectual discomfort" and "mystification" whenever two scientific theories that make different predictions turn out to be reducible to one another.[1] A particularly uncomfortable case is that of thermodynamics, the theory of heat and work, which is widely considered to reduce to statistical mechanics, the statistical analysis of classical and quantum mechanical systems in motion—and yet, these theories purportedly make incompatible predictions about the arrow of time. The laws of thermodynamics for mixing gases are said to distinguish past from future, while the mechanical laws for molecules making up those gases are said to be symmetric in time. Or, as Reichenbach (1956, p.29) has put it, "mechanical processes are reversible, whereas thermodynamical processes, apart from certain exceptions, are irreversible." This is known as the

[1]Cf. Nagel (1949, p.104) and Nagel (1961, p.340).

reversibility problem. It has been referred to as "the most profound problem in the foundations of thermal and statistical physics" (Uffink 2001, p.307).

Most philosophers at least tacitly assume that these two theories reach incompatible verdicts about the arrow of time. As Valente (2021) has pointed out, this presents a problem for reduction:

> "The Hamiltonian equations of motion in SM [statistical mechanics] are time-reversal invariant, whereas TD [thermodynamics] admits irreversible processes. Such issues undermine a reductive explanation of thermodynamical phenomena, and hence they must be addressed in order to determine if any version of statistical thermodynamics can be taken seriously" (Valente 2021, pp.178-179).

If that is right, then it leaves only two avenues of response. The first is Nagel's approach, to accept that only a rough analogue of thermodynamics actually reduces to statistical mechanics.[2] The second is Batterman's approach, to reject that thermodynamics reduces to statistical mechanics in any standard sense.[3] Here I will point out and defend a third avenue of response, that the assumption of Valente and most commentators is mistaken: there is no substantial difference between the arrow of time in thermodynamics and in statistical mechanics, and so there is no need for discomfort or mystification. Either both distinguish an arrow of time, or neither do.

My argument has two main steps. The first step shows that approaches to the thermodynamic arrow based on the second law imply the failure of energy conservation, which is typical when the description of a system is incomplete or 'open'. The second step shows that, since the failure of energy conservation is also associated with time asymmetry in statistical mechanics, there is no disharmony between the predictions of thermodynamics and statistical mechanics about the arrow of time. I then consider another approach to the thermodynamic arrow arising from equilibriation, and show that here too the asymmetry arises in a way that is harmonious between thermodynamics and statistical mechanics, from initial conditions associated with a special state in the past.

[2]Its seeds began in Nagel (1935), and well-known accounts are Nagel (1949, 1961), with influential adjustments due to Schaffner (1967). This aspect of Nagel's view is equally accepted by New Wave reductionists like Churchland (1979, 1985), Hooker (1981) and Bickle (1996), and by modern Nagelians like Callender (2001) and Dizadji-Bahmani et al. (2010).

[3]See Batterman (2002, especially Chapter 5) and Batterman (2021, especially Chapter 2).

I will argue for two main conclusions. First, the arrow of time is no barrier to the reduction of thermodynamics to statistical mechanics, although interesting foundational problems associated with its reduction may still remain. And second, thermodynamic asymmetry should really be detached from the laws of thermodynamics entirely, and instead associated either with the failure of energy conservation or with special initial conditions. I conclude with a deflated view of the thermodynamic arrow, as an artefact of descriptions of reality that are incomplete, or else as a contingent rather than a nomic fact.

2. A primer on reduction

'Reduction' describes the derivation of one scientific theory from another. It is naturally expected when two theories have overlapping domains of application, and has been discussed in a large philosophical literature going back at least to Nagel (1935). In the the philosophy of thermodynamics, there are two main party lines regarding the status of reduction.

The *unionist* party takes the reductive union of two theories to involve fudging, one hopes gently. For Nagel (1949) and Reichenbach (1956), and indeed for most scientists, this means embellishing statistical mechanics with temporally asymmetric bridge laws and auxilliary assumptions, such as Boltzmann's *Stoßzahlansatz* and special initial conditions of the kind pointed out by the founders of statistical mechanics (cf. Brush 1976, §14.5).[4] More generally, one replaces thermodynamics with a roughly analogous theory more amenable to derivation from a statistical mechanical theory. It is only in this sense that thermodynamics be derived from statistical mechanics, and even this is not uncontroversial. Some advocates like Albert (2000) have defended reduction using the 'past hypothesis' approach, recently rebranded I think for modesty purposes as 'The Mentaculus' (Loewer 2020).[5] Another unionist approach due to Robertson (2019, 2022) restricts attention to a pared-down theory that she calls the "functional form" of thermodynamics, and which she argues includes a concept

[4]The difficulties applying this fudge are chronicled by Earman and Norton (1998, 1999) and Uffink (2001). Bridge laws have been characterised in terms of everything from the infamous 'proxy function' of Quine (1964), to the bald analogies of Schaffner (1967), to the 'inter-theory deductions' of the New Wave approach; see (Endicott 1998) for an overview and critique of the latter, and Palacios (2019) for its application to the reduction of thermodynamic phase transitions.

[5]But compare the critiques of Earman (2006), Wallace (2017) and Gryb (2021).

of entropy that matches the Gibbs entropy of a statistical mechanical system.[6] Dizadji-Bahmani et al. (2010, p.409) neatly summarise the unionist situation, of having to settle for a "watered down version" of thermodynamics to be deduced from statistical mechanics.

In contrast, the *home-rule* party rejects the reduction of thermodynamics to statistical mechanics, instead treating the two as autonomous. For example, Feyerabend (1963) argued that difference between thermodynamics and statistical mechanics amount to a discontinuity in the rational development of science in the sense of Kuhn (1962). Batterman (2002, 2021) and Morrison (2012) use thermodynamics and statistical mechanics to advocate for the radical autonomy or 'emergence' of theories at different scales.[7] Similar thinking led Mach (1896, pp.333-5) to advise that statistical mechanics should not be taken too seriously because it cannot capture all of thermodynamics. Callender (1999, 2001) suggests the opposite, that thermodynamics should not be taken too seriously despite its veneration by physicists, because among other things the reversibility problem prevents its reduction to the more fundamental statistical mechanics.

My account might be called the *European* view, where two theories retain some autonomy, but still happen to align in virtue of their own internal principles. I will argue that the arrow of time in thermodynamics and statistical mechanics has this character, in that each is structured in such a way that their temporal symmetries align, dissolving the reversibility problem.

To be clear, I do not wish to dismiss the remarkable challenge of explaining the powerful temporal asymmetry that pervades human experience. Without some care, a reversibility problem may still remain: by reversing a perfectly reasonable description of a disippating gas governed with time-reversal-invariant mechanical laws, one finds a totally unreasonable description of a spontaneously concentrating gas. What I will argue is rather that, insofar as there is a tension in this fact, it is not a tension for the reduction of thermodynamics and statistical mechanics.

[6] Her approach follows the Lewisian account of reductive functionalism, recently redeveloped for the general problem of reduction by Butterfield and Gomes (2022).

[7] Deflationary responses: Butterfield (2011), Norton (2012), Palacios (2019), and Wu (2021).

3. Thermodynamic asymmetry is non-conservative

Most commentators identify time asymmetry in thermodynamics with the second law.[8] Perhaps the most influential of these commentators was Planck:

> "If those [thermodynamic] processes are not irreversible, the entire edifice of the second law will crumble. None of the numerous relations deduced from it, however many may have been verified by experience, could then be considered as universally proved, and theoretical work would have to start from the beginning. ... It is this foundation on the physical fact of irreversibility which forms the strength of the second law." (Planck 1897, §114)

In this section I will show how, due to a little-known theorem of Affanasjewa and Jauch, the time asymmetry of the second law implies that energy is not conserved, in a sense usually associated with systems open to external influence.

Let me begin by dispelling a myth that may surprise non-specialists: it is not a law of thermodynamics that every isolated system has a property called 'entropy' that is universally non-decreasing to the future. Such an 'entropy principle' is a combination of over-simplified propaganda and wishful thinking. The propaganda goes right back to Clausius, one of the founders of thermodynamics, who claimed it is a law of nature that, "[t]he entropy of the universe tends to a maximum" (Clausius 1867, p.44). This claim was not argued for, not accepted even by his followers, and in a later republication of the same work Clausius seems to have asked that this passage be deleted (see Uffink 2001, pp.338-40). Of course, one might use this 'general entropy principle' as a metaphor for the observed dissipation of everyday systems as they approach equilibrium, which I will return to in Section 5. There are also indications that a generalised entropy principle may be true of black holes in semiclassical gravitation, where it is known as the Generalised Second Law.[9] But, it is not a universal fact about thermodynamics.

Of course, there is still a conceptually coherent entropy principle. To state it, let me first recall the origin of entropy in a model of thermodynamics. The state space of a thermodynamic system is a $(2n + 1)$-dimensional manifold M

[8] See Reichenbach (1956, §7), Sklar (1993), Callender (1999, 2001), Myrvold (2011, 2020a,b), and Valente (2021, §4.1), among others. I discuss dissenters like Brown and Uffink (2001) in Section 5.

[9] See Wald (1994, §7.2) for a classic introduction.

of states, together with an n-dimensional surface N on which the first law of thermodynamics holds, which is to say that there is a quantity dU representing a small amount of energy that satisfies,

$$dU = \xi + P_1 dX_1 + P_2 dX_2 + \cdots + P_{n-1} dX_{n-1}. \tag{1}$$

Here, U and each X_i is a smooth real-valued function, and $P_i = \partial U/\partial X_i$ for each $i = 1, \ldots, n-1$. The pairs of functions (P_i, X_i) represent accessible degrees of freedom that are associated with work, like pressure and volume, or electric charge and electrochemical potential.

The reason we associate the first law with *thermal* phenomena is that it contains a term ξ representing heat. Mathematically, heat ξ is a one-form.[10] Sometimes, though not always, this structure may admit a pair of smooth functions S and T (where $T = \partial U/\partial S$) such that,

$$\xi = TdS. \tag{2}$$

When they exist, T is called *temperature* and S is called *entropy*. Otherwise, their existence might be restricted to some subset of state space like a particular region or curve. But, their meaning is always defined by their ability to describe heat, as in Equation (2).

To derive an entropy principle, we now consider a finite, piecewise-smooth curve γ through thermodynamic state space, with initial and final endpoints representing the initial and final states of some process.[11] A curve γ will be called *reversible* whenever entropy on it is well-defined, in that heat satisfies $\xi = TdS$ for some S along γ; otherwise the curve is called *irreversible*.[12] finally posit the existence of a smooth function T that we wish to think of as 'temperature', and such that the following two postulates are satisfied:

(i) *(adiabatic process)* the heat ξ is zero along the curve γ.

(ii) *(no perpetual motion)* there is a reversible curve λ restoring the system to

[10] I adopt some standard differential geometry terminology; see e.g. Marsden and Ratiu (2010).

[11] There is some subtlety in how to interpret a curve through state space when all the states are in equilibrium. Textbooks generally follow Carathéodory (1909) in viewing such a curve as representing very slow or 'quasistatic' change; see Norton (2014, 2016) for an alternative interpretation. For present purposes, only the initial and final endpoints are need to represent the initial and final stage of a physical process.

[12] The term 'Reversibility' is unfortunately used in a variety of ways; note that my definition here is purely mathematical and follows the geometric tradition of contact geometry approaches.

its initial state, and without any positive accumulation of heat per unit temperature, $\int_c \xi/T \leq 0$, where c is the closed loop formed γ followed by λ, as shown in Figure 1.

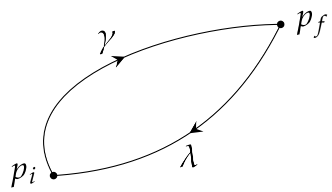

Figure 1: The closed curve c.

It is an elementary exercise[13] to show that these two assumptions imply the existence of a function S, which 'behaves like entropy' on the reversible curve λ in the sense that Equation (2) holds, and such that $S(p_i) - S(p_f) = \int_c \xi/T$, where p_i and p_f are the initial and final endpoints of the process γ. Combining this latter result with our postulate (ii) that $\int_c \xi/T \leq 0$, we find that $S(p_i) \leq S(p_f)$. This fact, first observed by Clausius, is called the *Clausius entropy principle*. In summary, it says:

> *there is a quantity S whose value at the end of the process is not less than it was at the beginning, and which 'behaves like' entropy during a reversible process restoring that initial state.*

The assumptions of this argument can be challenged, although I will not do so here.[14] Instead, I would like to point out a simpler problem with using this entropy principle as the basis for a time asymmetry: the inequality is not strict. It is still possible that the final and initial entropies are equal, $S_i = S_f$, which would eliminate the time asymmetry completely.

The strict inequality $S_i < S_f$ needed for an arrow of time requires a third assumption:

(iii) *(irreversibility)* the process γ is irreversible.

To see why, recall again that an 'irreversible' curve γ is one for which no function S behaves like entropy in the sense that $dS = \xi/T$ along the curve. Now, it is a

[13]*Solution:* Since λ is reversible, $\xi = TdS$ on λ for some S. A closed curve c now runs from p_i to p_f along γ, and then back from p_f to p_i along λ. Therefore, since γ is an adiabat (so $\int_\gamma \xi/T = \int_\gamma \xi = 0$), we have $S(p_i) - S(p_f) = \int_\lambda dS = \int_\lambda \xi/T = \int_\lambda \xi/T + \int_\gamma \xi/T = \int_c \xi/T$.

[14]Challenges can be found in Roberts (2022, pp.157-159) and Uffink (2001, pp.338-342).

purely mathematical fact that no such function exists if and only if the one-form ξ/T is not conserved around a closed loop.[15] So, condition (iii) is equivalent to the claim that $\int_c \xi/T$ is non-zero, and hence that $\int_c \xi/T \leq 0$ is actually a strict inequality $\int_c \xi/T < 0$, which through the argument above is equivalent to a strict entropy inequality,

$$S(p_i) < S(p_f). \tag{3}$$

Thus, it is exactly when there is an irreversible curve that one can use a strict entropy inequality to derive a thermodynamic arrow of time.

What I would like to point out is that this third assumption *also* implies that mechanical energy is not conserved. In rough physical terms, this arises from the fact that ξ/T is strictly lost after the completion of a process containing an irreversible component that cycles back to where it started, such as the functioning of a realistic engine. So, if the cyclic process conserves work, as every isolated mechanical system does, then the energy associated with that heat must have been strictly lost.

Let me formulate this as a corollary of a little-known theorem due to Ehrenfest-Afanassjewa (1925) and Jauch (1972). Given the first law, the theorem shows the following:[16]

> *If work is conserved on all closed adiabatic processes (energy conservation), then there are functions S and T in a neighbourhood of each point such that heat can be written $\xi = TdS$.*

An adiabatic curve is one along which heat remains zero. So, since total energy is just the sum of heat and work, the only energy available on an adiabat is work, meaning that energy conservation is equivalent to work conservation. Energy conservation is a plausible assumption to about an isolated local system, and was used by Jauch (1972) and Roberts (2022, §6.2.1) to motivate the existence of a global entropy function. However, for the argument I would like to make, the contrapositive form of the theorem is what matters. That argument is the

[15] More precisely, a one-form ω is exact if and only if it is conservative (see e.g. Lee 2013, p.292, Theorem 11.42).

[16] More precisely: Let $(U, X_1, \ldots, X_{n-1})$ be a complete set of coordinate functions of a manifold N of dimension n, and let ξ satisfy the first law $dU = \xi + \sum_{i=1}^{n-1} P_i dX_i$ for some functions $(P_1, \ldots, P_{n-1})$. Suppose that work $W := \sum_{i=1}^{n-1} P_i dX_i$ is 'conserved on adiabats', in that for every closed, piecewise-smooth curve c with tangent vector field $\bar{c}$ satisfying $\xi(\bar{c}) = 0$, we have $\int_c W = 0$. Then in a neighbourhood of every point there exist smooth functions $S : N \to \mathbb{R}$ and $T : N \to \mathbb{R}$ such that $\xi = TdS$. For the proof see Jauch (1972), and for commentary see Roberts (2022, p.149).

following.

We have seen that a thermodynamic arrow requires the existence of an irreversible process, which in turn requires neighbourhoods in which heat cannot be expressed as $\xi = TdS$ for any smooth functions T and S. Such systems violate energy conservation, by an equivalent statement of the Afanassjewa-Jauch theorem:

> *If there are no functions S and T in any neighbourhood of a point such that heat can be written $\xi = TdS$, then work is not conserved on all closed adiabatic processes (violation of energy conservation).*

The combined result is that a thermodynamic arrow of time only arises in models where energy is not conserved, such as those that are open to external influence. This is truly remarkable, given how much heat has been lost on the thermodynamic arrow. For, as I will recall in the next section, time asymmetry in the absence of energy conservation is in complete harmony with the laws of mechanics, and so is no barrier to reduction.

4. Non-conservative mechanics is asymmetric

It is only when we are speaking loosely about mechanics that we say that it is generally symmetric in time. Both classical and quantum mechanics are used to model systems that are temporally asymmetric, such as a damped harmonic oscillator, a fluid with internal friction, or an electric current with resistance. These models are all 'open' in that they provide incomplete descriptions of the systems they refer to: they ignore degrees of freedom into which energy can be transferred, resulting in a formal violation of energy conservation. An arrow of time is thus a typical feature of statistical mechanics, but only so long as energy conservation is violated, just as I have argued is the case for thermodynamics.

For example, a classical damped harmonic oscillator is a system in which an oscillating mass slows to a stop, but in which the temporally reversed motion of a spontaneously oscillating mass is prohibited. The oscillating mass m has position $x(t)$ and velocity $v = dx/dt$ that satisfy the following equation for some positive constants k and c:

$$m\frac{dx^2}{dt^2} = -kx - cv. \tag{4}$$

It is easy to confirm that if we are given a solution $(x(t), v(t))$ to Equation (4), the time-reversed trajectory $(x(-t), -v(-t))$ is not generally a solution. The energy E of changes at a rate of,[17]

$$\frac{dE}{dt} = -cv^2, \tag{5}$$

which is non-zero whenever the system has non-zero initial velocity, violating conservation. Of course, for most applications one would not refer to this as a 'fundamental' failure of conservation, but rather one associated with the missing degrees of freedom into which energy dissipates. This kind of arrow of time has been said to fall prey to a "missing information misfire" (Roberts 2022, §5.1).

Thus, time asymmetry arises in mechanics in exactly the sort of situation that it does for thermodynamics, when energy conservation is violated. One can go further and ask whether this is a general fact about the arrow of time in mechanics, the way that we have shown it is in thermodynamics. There are various ways to make this idea precise. One of them is to consider classical mechanics as characterised by Newton's second law, $F = md^2x/dt^2$, and to suppose the force is characterised as a vector field that depends only on position and 'falls off quickly' as one moves away from some compact region in space. The latter implies[18] that F can be expressed as the sum of a divergence and a curl component,

$$F = \nabla\varphi + \nabla \times A, \tag{6}$$

for some scalar field φ and some vector field A. Such a system is *time reversal invariant* if and only if the force F is preserved under the time reversal transformation, since this is what guarantees that time reversal preserves solutions to Newton's equation $F = mdx^2/dt^2$. Moreover, the system is *conservative* whenever work is conserved on all closed loops. Our aim is to show that such a system is time reversal invariant if and only if it is conservative.

The proof is simple: the curl $\nabla \times A$ of a vector field generally reverses sign under time reversal, whereas the divergence $\nabla\varphi$ does not.[19] So, given Equation (6), time reversal invariance is equivalent to the statement that $F = \nabla\varphi$. Moreover, that statement is known to hold if and only if the system is conservative (Arnol'd 1989, p.29 §6B). Therefore, *the system is time reversal invariant if and only if it is*

[17]Energy is $E = mv^2/2 + kx^2/2$, which implies $dE/dt = m\frac{dx^2}{dt^2}v + kxv = (-kx - cv)v + kxv = -cv^2$.

[18]This is a consenquence of the Helmholtz-Hodge theorem (Arfken 1985, §1.15).

[19]This can be seen through arguments like those of Malament (2004). Roughly, an axial vector field like $\nabla \times A$ makes use of a temporal orientation for its definition, which must therefore be reversed when time is reversed. Without this link between axial vector fields and time orientation, the link between time reversal invariance and energy conservation can fail (cf. Roberts 2013).

conservative.

This argument applies to most common systems of classical mechanics, though not all mechanical systems.[20] However, my aim is only to capture a piece of common physics lore in somewhat precise terms, that an arrow of time is neither problematic nor unexpected when a physical system is non-conservative, in mechanics or otherwise. Given this, reversibility problem for reduction as it is usually posed is dissolved.

5. The Minus-First Law

Not everyone identifies the thermodynamic arrow with the second law.[21] Brown and Uffink (2001) have argued for an entirely different origin for thermodynamic time asymmetry, in the approach to equilibrium. Suppose that the states of a thermodynamic system can be partitioned into two types, called *non-equilibrium* and *equilibrium*, and that some sense of dynamical evolution in time is available to describe such states.[22] The proposal is that a law of nature governing equilibrium must be added to the laws of thermodynamics, called the *Minus-First Law* to indicate its priority before the others, and which states:

> *"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, p.528)

Brown and Uffink break this principle into several parts, and clarify that the part which they consider to be temporally asymmetric is just the following, which they call *Claim (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" (Brown and Uffink 2001, p.528).

[20]Indeed, its generality does not extend to electroweak interactions, which are both conservative and time asymmetric. A 'new' reversibility problem might thus arise in exactly the opposite sense as it is usually presented: a weakly interacting system is temporally asymmetric in statistical mechanics, but when it is in thermal equilibrium it would seemingly still be temporally symmetric in thermodynamics! I do not have space to discuss this curious issue here.

[21]For heterodox arguments, see Uffink (2001) and Roberts (2022, Chapter 6).

[22]These assumptions are left as somewhat vague additions to the standard mathematics of thermodynamics: as Wallace (2014, p.699) has pointed out, thermodynamics "is not in the business of telling us how those equilibrium states evolve if left to themselves" in spite of the awkward appearance of "dynamics" in the name.

Unlike the aspects of the Minus-First Law postulating uniqueness and the approach to equilibrium, Brown and Uffink consider Claim (A) to be essentially time asymmetric, and thus conclude that in thermodynamics, the "time-asymmetric component lies in the postulated notion of equilibrium itself" (Brown and Uffink 2001, p.526). This is taken to be in stark contrast with Boltzmann's definition of equilibrium in statistical mechanics—namely, as the macrostate of largest volume—which they consider to be "time-symmetric, unlike its counterpart in thermodynamics" (Brown and Uffink 2001, p.530).

This is an entirely different origin for thermodynamic time asymmetry as compared to the standard one discussed above. Brown and Uffink explicitly formulate the Minus-First Law for systems that are isolated, and thus for which energy conservation should be expected. So, my response to to the approach that makes use of the second law seemingly does not apply here. Moreover, Brown and Uffink claim that this principle is inconsistent with the kind of spontaneous deviations from equilibrium associated with statistical mechanics:

> "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 claim (A) of the Minus First Law, which as we have seen rules out spontaneous deviations from equilibrium." (Brown and Uffink 2001, p.536)

If they are right, then it seems impossible to strictly deduce this temporally asymmetric aspect of thermodynamics from statistical mechanics, and the reversibility problem reappears.

However, there is a missing component to this story. Despite what Brown and Uffink suggest, the Minus-First law is not guaranteed to be temporally asymmetric, even with the clarification of Claim (A). The situation is comparable to that of the entropy principle, which requires subtle adjustment to guarantee a strict inequality. In the case of the Minus-First Law: equilibrium states are only defined to be fixed in one temporal direction—"thereafter constant in time"—which still allows temporal symmetry to occur if those states are fixed in the other temporal direction as well. Strictly speaking, the Minus-First Law as Brown and Uffink have stated it is compatible with equilibrium states remaining fixed for all times, and non-equilibrium states approaching equilibrium in both temporal directions.

Breaking the symmetry requires a further assumption, that some equilibrium states are not fixed for all time. In addition to the Minus-First law, one must assert that some equilibrium states began in a non-equilibrium state at some point in the past. This is not a deviation from statistical mechanics, and on the contrary is exactly the situation that one finds in the 'past hypothesis' approach. To illustrate, let me briefly recall how it works there.

One of the central results of Boltzmannian statistical mechanics is that equilibrium states are stupendously more common than non-equilibrium ones.[23] The situation is analogous to a house with 10^{22} blue rooms, analogous to equilibrium, and just one that is red, analogous to non-equilibrium. When walking into an arbitrary room, one should expect that it will very likely be blue, and indeed that nearly all rooms that one visits in the future will be blue as well. But, this argument runs the same way in the reverse direction: when walking *out of* an arbitrary room, one should expect to have just left a blue room as well.

An early and prominent way that this temporal symmetry was broken was through the postulate that a non-equilibrium state exists in the past, such as in the first moments after the big bang. This is what Albert (2000) calls 'past hypothesis'. The past hypothesis helps to ensure that equilibrium is approached to the future and not the past, breaking the time asymmetry in the statistical argument above. It is not my aim to discuss the status of this controversial argument. What matters for my purposes is that the past hypothesis is exactly analogous to the assumption of a past non-equilibrium state in thermodynamics, which is implicitly responsible for the temporal asymmetry in Brown and Uffink's Minus-First Law.

The result is a remarkable harmony in the way that both thermodynamics and statistical mechanics treat the arrow of time, no matter how one describes the thermodynamic arrow. If it is identified with the Clausius entropy principle, then the thermodynamic arrow arises from a failure of energy conservation, in just the way that one expects an arrow of statistical mechanics to arise. If it is identified with the approach to equilibrium captured by the Minus-First Law, then the thermodynamic arrow requires postulating a special non-equilibrium state in the past, in just the way that the arrow of statistical mechanics arises out of the past hypothesis. In either case, the reversibility problem is dissolved.

[23] A commonly-cited figure for a litre of gas in a tank is that the states *not* in Boltzmannian equilibrium occupy a portion of the total number of states equal to just $1/10^{22}$ (see e.g. Penrose 2004, p.693).

6. Conclusion

Tatiana Afanassjewa, one of the founders of modern thermodynamics, took the main lesson of her great work to be the decoupling of the theory's foundations from the arrow of time:

> "If I dare to produce one more book among so many... then it is to show how one can free the derivation of the fundamental thermodynamic equations from the question regarding the direction of natural phenomena and from the dissipation of energy." (Ehrenfest-Afanassjewa 1956, p.142)

What I have argued for here is a special case of this lesson. Thermodynamics does not imbue its subjects with an arrow of time. As in many other theories, the temporal asymmetries of thermodynamics arise in special environments, such as those that give rise to the Clausius entropy principle, or those that render the Minus-First Law asymmetric. The reversibility problem only appears when one ignores the fact that those special environments are needed. When they are made explicit, the underlying mechanisms behind the thermodynamic arrow are not just clarified, but made harmonious with the arrow of statistical mechanics. Reversibility is thus no barrier to the reduction of thermodynamics to statistical mechanics: these theories are reversible and irreversible in almost exactly the same situations.

Given this, the arrow of time in thermal and statistical physics should perhaps be viewed in less lofty terms than it is usually presented. It does not have the deep, nomic status of a law of motion like Schrödinger's equation. The thermodynamic arrow is a rather more prosaic fact, which arises from the human condition of describing Nature incompletely in the case of the entropy principle, or using contingent facts about the special nature of conditions in our past. This should not displace its importance, or the sense of wonder that arises in seeking to understand it more deeply. But, let part of that wonder be this remarkable sense in which the reversibility problem is dissolved.

Acknowledgments. Thanks to Foad Dizadji-Bahmani for conversations that were both scintillating and adiabatic.

References

Albert, D. Z. (2000). *Time and chance*, Cambridge, MA: Harvard University Press.

Arfken, G. (1985). *Mathematical Methods for Physicists*, 3rd edn, San Diego: Academic Press, Inc.

Arnol'd, V. I. (1989). *Mathematical methods of classical mechanics*, 2nd edn, Springer-Verlag New York, Inc.

Batterman, R. W. (2002). *The Devil in the Details: Asymptotic Reasoning in Explanation, Reduction, and Emergence*, New York: Oxford University Press.

Batterman, R. W. (2021). *A Middle Way: A Non-Fundamental Approach to Many-Body Physics*, New York: Oxford University Press.

Bickle, J. (1996). New wave psychophysical reductionism and the methodological caveats, *Philosophy and Phenomenological Research* **56**(1): 57–78.

Brown, H. R. and Uffink, J. (2001). The Origins of Time-Asymmetry in Thermodynamics: The Minus First Law, *Studies in History and Philosophy of Modern Physics* **32**(4): 525–538. http://philsci-archive.pitt.edu/217/.

Brush, S. G. (1976). *The kind of motion we call heat: A history of the kinetic theory of gases in the 19th century*, Vol. 1: Physics and the Atomists, Amsterdam: Elsevier Science B.V.

Butterfield, J. (2011). Less is different: Emergence and reduction reconciled, *Foundations of physics* **41**(6): 1065–1135. http://philsci-archive.pitt.edu/8355/.

Butterfield, J. and Gomes, H. (2022). Functionalism as a species of reduction. Forthcoming in: *Current Debates in Philosophy of Science*, Edited by Cristian Soto, Springer, http://philsci-archive.pitt.edu/18043/.

Callender, C. (1999). Reducing thermodynamics to statistical mechanics: The case of entropy, *The Journal of Philosophy* **96**(7): 348–373.

Callender, C. (2001). Taking thermodynamics too seriously, *Studies in history and philosophy of modern physics* **32**(4): 539–553. http://philsci-archive.pitt.edu/289/.

Carathéodory, C. (1909). Untersuchungen über die Grundlagen der Thermodynamik, *Mathematische Annalen* **67**: 355–386.

Churchland, P. M. (1979). *Scientific realism and the plasticity of mind*, Cambridge: Cambridge University Press.

Churchland, P. M. (1985). Reduction, qualia, and the direct introspection of brain states, *The Journal of Philosophy* **82**(1): 8–28.

Clausius, R. (1867). *Abhandlungen Über Die Mechanische Wärmetheorie*, Braunschweig: F. Vieweg.

Dizadji-Bahmani, F., Frigg, R. and Hartmann, S. (2010). Who's afraid of nagelian reduction?, *Erkenntnis* **73**(3): 393–412. http://philsci-archive.pitt.edu/5323/.

Earman, J. (2006). The 'Past Hypothesis': Not Even False, *Studies in History and Philosophy of Modern Physics* **37**(3): 399–430.

Earman, J. and Norton, J. D. (1998). EXORCIST XIV: The wrath of Maxwell's demon, Part I: From Maxwell to Szilard, *Studies In History and Philosophy of Modern Physics* **29**(4): 435–471.

Earman, J. and Norton, J. D. (1999). EXORCIST XIV: The Wrath of Maxwell's Demon, Part II: From Szilard to Landauer and Beyond, *Studies In History and Philosophy of Modern Physics* **30**(1): 1–40.

Ehrenfest-Afanassjewa, T. (1925). Zur Axiomatisierung des zweiten Hauptsatzes der Thermodynamik, *Zeitschrift für Physik* **33**: 933–945. Translated into English by Marina Baldissera Pacchetti as, 'On the Axiomatization of the Second Law of Thermodynamics', in Uffink et al. (2021) *The Legacy of Tatiana Afanassjewa*, Springer Nature Switzerland AC, pp.170-183.

Ehrenfest-Afanassjewa, T. (1956). *Die Grundlagen der Thermodynamik*, Leiden: E.J. Brill. Translated into English by Marina Baldissera Pacchetti as, 'Foundations of Thermodynamics', in Uffink et al. (2021) *The Legacy of Tatiana Afanassjewa*, Springer Nature Switzerland AC, pp.141-170.

Endicott, R. P. (1998). Collapse of the new wave, *The Journal of Philosophy* **95**(2): 53–72.

Feyeraband, P. K. (1963). How to be a Good Empiricist: A Please for Tolerance in Matters Epistemological, *in* B. Baumrin (ed.), *Philosophy of Science: The Delaware Seminar*, New York: Interscience Press, pp. 3–39.

Gryb, S. (2021). New difficulties for the past hypothesis, *Philosophy of Science* **88**(3): 511–532. https://philsci-archive.pitt.edu/17253/.

Hooker, C. A. (1981). Towards a general theory of reduction, Parts I, II and III, *Dialogue: Canadian Philosophical Review/Revue canadienne de philosophie* **20**(1): 38–59.

Jauch, J. M. (1972). On a new foundation of equilibrium thermodynamics, *Foundations of Physics* **2**(4): 327–332.

Kuhn, T. S. (1962). *The structure of scientific revolutions*, University of Chicago Press.

Lee, J. M. (2013). *Introduction to Smooth Manifolds*, Graduate Texts in Mathematics, second edn, Springer Science+Business Media New York.

Loewer, B. (2020). The Mentaculus Vision, *in* V. Allori (ed.), *Statistical mechanics and scientific explanation: Determinism, indeterminism and laws of nature*, Singapore: World Scientific Publishing Co. Pte. Ltd., chapter 1, pp. 3–29.

Mach, E. (1896). *Die Principien Der Wärmelehre*, Leipzig:Verlag von Johan Ambrosius Barth. From the 1986 English translation, *Principles of the Theory of Heat*, Brian McGuinness (Ed.), Dordrecht, Holland: D. Reidel Publishing Company.

Malament, D. B. (2004). On the time reversal invariance of classical electromagnetic theory, *Studies in History and Philosophy of Modern Physics* **35**: 295–315. Preprint: http://philsci-archive.pitt.edu/1475/.

Marsden, J. and Ratiu, T. (2010). *Introduction to mechanics and symmetry: a basic exposition of classical mechanical systems*, 2nd edn, Springer-Verlag New York, Inc.

Morrison, M. (2012). Emergent physics and micro-ontology, *Philosophy of science* **79**(1): 141–166.

Myrvold, W. C. (2011). Statistical mechanics and thermodynamics: A Maxwellian view, *Studies in History and Philosophy of Modern Physics* **42**(4): 237–243. http://philsci-archive.pitt.edu/8638/.

Myrvold, W. C. (2020a). Explaining Thermodynamics: What remains to be done?, *in* V. Allori (ed.), *Statistical mechanics and scientific explanation: Determinism, indeterminism and laws of nature*, Singapore: World Scientific Co. Pte. Ltd., pp. 113–143. http://philsci-archive.pitt.edu/16217/.

Myrvold, W. C. (2020b). The science of $\theta\delta^{cs}$, *Foundations of Physics* **50**(10): 1219–1251. http://philsci-archive.pitt.edu/17609/.

Nagel, E. (1935). The logic of reduction in the sciences, *Erkenntnis* **5**: 46–52.

Nagel, E. (1949). The meaning of reduction in the natural sciences, *in* R. C. Stauffer (ed.), *Science and Civilization*, Madison: University of Wisconsin Press, pp. 99–135.

Nagel, E. (1961). The structure of science: Problems in the logic of scientific explanation.

Norton, J. D. (2012). Approximation and idealization: Why the difference matters, *Philosophy of Science* **79**(2): 207–232. https://philsci-archive.pitt.edu/8622/.

Norton, J. D. (2014). Infinite idealizations, *in* M. C. Galavotti, E. Nemeth and F. Stadler (eds), *European Philosophy of Science - Philosophy of Science in Europe and the Viennese Heritage*, Vol. 17, pp. 197–210. http://philsci-archive.pitt.edu/9028/.

Norton, J. D. (2016). The impossible process: Thermodynamic reversibility, *Studies in History and Philosophy of Modern Physics* **55**: 43–61. http://philsci-archive.pitt.edu/12341/.

Palacios, P. (2019). Phase transitions: A challenge for intertheoretic reduction?, *Philosophy of Science* **86**(4): 612–640. http://philsci-archive.pitt.edu/15998/.

Penrose, R. (2004). *The Road to Reality: A Complete Guide to the Laws of the Universe*, London: Jonathan Cape.

Planck, M. (1897). *Vorlesungen über thermodynamik*, Leipzig: Verlag von Veit & Comp. https://www.google.co.uk/books/edition/Vorlesungen_%C3%BCber_Thermodynamik/o07PAAAAMAAJ.

Quine, W. V. O. (1964). Ontological reduction and the world of numbers, *The Journal of Philosophy* **61**(7): 209–216.

Reichenbach, H. (1956). *The direction of time*, Berkeley: University of California Press. Edited by Maria Reichenbach.

Roberts, B. W. (2013). When we do (and do not) have a classical arrow of time, *Philosophy of Science* **80**(5): 1112–1124.

Roberts, B. W. (2022). *Reversing the Arrow of Time*, Cambridge: Cambridge University Press. Open Access: https://doi.org/10.1017/9781009122139.

Robertson, K. (2019). *Reductive Aspects of Thermal Physics*, PhD thesis, Pembroke College, University of Cambridge. https://doi.org/10.17863/CAM.36677.

Robertson, K. (2022). In Search of the Holy Grail: How to Reduce the Second Law of Thermodynamics, *The British Journal for the Philosophy of Science* **73**(4). https://philsci-archive.pitt.edu/17516/.

Schaffner, K. F. (1967). Approaches to reduction, *Philosophy of science* **34**(2): 137–147.

Sklar, L. (1993). *Physics and chance: Philosophical issues in the foundations of statistical mechanics*, 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. http://philsci-archive.pitt.edu/313/.

Valente, G. (2021). Taking up statistical thermodynamics: Equilibrium fluctuations and irreversibility, *Studies in History and Philosophy of Science Part A* **85**: 176–184.

Wald, R. M. (1994). *Quantum field theory in curved spacetime and black hole thermodynamics*, University of Chicago Press.

Wallace, D. (2014). Thermodynamics as control theory, *Entropy* **16**(2): 699–725. http://philsci-archive.pitt.edu/9904/.

Wallace, D. (2017). The nature of the past hypothesis, *in* K. Chamcham, J. Silk, J. D. Barrow and S. Saunders (eds), *The Philosophy of Cosmology*, Cambridge: Cambridge University Press, pp. 486–499.

Wu, J. (2021). Explaining universality: infinite limit systems in the renormalization group method, *Synthese* **199**: 14897–14930. https://philsci-archive.pitt.edu/.