In Search of the Holy Grail: How to Reduce the Second Law of Thermodynamics

Robertson, Katie Robertson

In Search of the Holy Grail: How to Reduce the Second Law of Thermodynamics

Katie Robertson

Forthcoming in BJPS

21/4/2020

Abstract

The search for the statistical mechanical underpinning of thermodynamic irreversibility has so far focussed on the spontaneous approach to equilibrium. But this is the search for the underpinning of what Brown and Uffink (2001) have dubbed the 'minus first law' of thermodynamics. In contrast, the second law tells us that certain interventions on equilibrium states render the initial state 'irrecoverable'. In this paper, I discuss the unusual nature of processes in thermodynamics, and the type of irreversibility that the second law embodies. I then search for the microscopic underpinning or statistical mechanical 'reductive basis' of the second law of thermodynamics by taking a functionalist strategy. First, I outline the functional role of the thermodynamic entropy: for a thermally isolated system, the thermodynamic entropy is constant in quasi-static processes, but increasing in non-quasi-static processes. I then search for the statistical mechanical quantity that plays this role — rather than the role of the traditional 'holy grail' as described by Callender (1999). I argue that in statistical mechanics, the Gibbs entropy plays this role.

1 Introduction

'The second law is one of the all-time great laws of science, for it illuminates why anything — anything from the cooling of hot matter to the formulation of a thought — happens at all', or so claims Atkins (2007, preface). Yet controversy clouds the second law of thermodynamics

(TDSL). Whilst Snow (1959) famously said that an acquaintance with the second law is the scientific equivalent of having read a work of Shakespeare, Uffink (2001) discusses important nuances about the content of the TDSL. Further philosophical questions abound: Can the second law provide the naturalistic basis for the arrow of time, as Reichenbach (1956) thought? Can Maxwell's demon get its claws into it? But the most controversial question is how to find the microphysical underpinning of the second law. Addressing this latter question is my project in this paper.

This project is one of inter-theoretic reduction: the goal is to capture the behaviour described by one theory —thermodynamics (TD) — in terms of another theory —- statistical mechanics (SM). However, in what follows, my argument will not hang on the fine-grained details of any single account of reduction.[1] Nonetheless, I will, in section 2, emphasise how functionalism is useful for securing reductions. Under this functionalist lens, the goal becomes to find a SM realiser of the TD role. Much of the controversy resides in defining the correct role in a given case, and defining the TDSL role will form the heart of my argument. Indeed, if the idiom of functionalism is off-putting to the reader, the key argument can still be understood if you skip section 2 and read 'realiser' as 'reductive basis'.[2]

In section 3 I discuss the unusual nature of processes in TD. In section 3.4, following Uffink (2013), I describe three different types of irreversibility. Then, in section 4, I introduce the TDSL, and show how it implicitly defines the TD entropy, and codifies its behaviour.

In section 5, I articulate the role of TD entropy: for thermally isolated systems, S_{TD} is constant in quasi-static processes but increasing in non-quasi-static processes. In section 5.2, I emphasise how this role differs from the 'holy grail' — a non-decreasing function to call entropy, as outlined by Callender (1999). In section 6, I then search for the realiser of this role in quantum statistical mechanics, and I argue that the Gibbs entropy plays this role. But the

[1]Indeed, contending with the controversies of limits in reduction cf. Batterman (1995), or the nature of bridge laws cf. Sklar (1993) would leave no time for the main event.

[2]One caveat to this: one of my two replies to the objection to the Gibbs entropy relies on functionalism (section 7.1). Those allergic to functionalism can rely on the second reply in section 7.3.

Gibbs entropy has been criticised for its 'ensemble' nature, in section 7 I defend the Gibbs entropy, before concluding in section 8.

2 Functionalism: a tool for reduction

Functionalism — the view that 'to be X is just to play the X-role' — has risen to prominence in the philosophy of physics (e.g. Albert (2013), Wallace (2012, Ch. 2), Knox (2013)). For example, Knox (2013) advocates functionalism about spacetime; to be spacetime is just to play the spacetime role, that is: to pick out the inertial trajectories. The motivation for advocating functionalism in this, and other cases, is to understand inter-theory relations. As such, spacetime functionalism is used to compare spacetimes across different physical theories. If a theory of quantum gravity is 'non-spatiotemporal', prima facie it is difficult to see how general relativity (GR) can be reduced to this theory (or, in the physicists' idiom, how GR can be *recovered* from this theory). But the functionalist reminds us that spacetime need not be fundamental in a theory of quantum gravity (Lam 2018). Instead, we need only capture the relevant behaviour described by general relativity. That is, we only need to find something in a theory of quantum gravity that behaves, i.e. plays the role, of spacetime.

Functionalism makes behaviour centre stage, and it is 'the behaviour characteristic of the system [that] is the focus of reduction' (Rueger 2006, p. 343) (see also Rosaler (2019, p. 273)). Provided that real behaviour can be modelled by both theories, other differences may not matter. In this way, functionalism helps emphasise that there might be differences between the theory to be reduced T_t and the reducer T_b. Consequently, functionalism is useful as a strategy for overcoming scepticism about certain instances of reduction.

To a certain extent, the same point can be made in the Nagel-Schaffner account, in which only an approximation or 'close cousin', T_t^*, of the original theory T_t, must be deduced from T_b (Butterfield 2011a;b). However, understanding approximations is a notoriously thorny issue: in which ways are T_t and T_t^* allowed to differ? One might think that the difference should be minimised — but in certain cases this could lead the reductive project astray. For instance, in the thermodynamic limit, the probabilistic fluctuations of SM disappear, and the categorical nature of quantities regained, as is familiar from TD. Thus, in the limit the SM

description is closer to the original TD description. But bringing in the thermodynamic limit obscures the reduction (if indeed limits are allowed in Nagelian bridge laws, cf. Butterfield (2011a;b)), since no actual system is infinite. Moreover, in what follows, the thermodynamic limit is not required to consider the reduction of the TDSL (notably, unlike the case of phase transitions, cf. Batterman (2001), Ardourel (2018), Palacios (2019)).

Functionalism has the upper hand here as it specifies the differences that can be tolerated: the realiser can differ in ways that do not affect its playing the functional role. 'Being locked' is a functional property: it can be realised by various mechanisms — D-locks, padlocks, combination locks etc. These various mechanisms differ in many ways, such as their colour or whether a key is required, but these differences do not prevent them from playing the functional role.

Returning to the arena of thermal physics, here is an example of how functionalism can help overcome skepticism about reduction. Sklar raises the following concern: the 'temperature equals mean molecular kinetic energy' bridge law *identifies* a fundamentally non-statistical quantity with a fundamentally *statistical quantity*. How is this supposed to work?' (Sklar 1993, p.161) as quoted by Batterman (2010).

Sklar's worry is that mean kinetic energy and temperature have different features: the former is statistical and latter not, and thus this blocks the reduction. But if the non-statistical nature of temperature is not part of its functional role, then the same behaviour can be captured by a statistical property: provided they have the same relevant behaviour, and so mean kinetic energy plays the functional role of temperature in an ideal gas.

Of course, this then raises the question: should the non-statistical nature be a part of the functional role of TD temperature? Here I submit that the purely philosophical doctrine of functionalism is silent: only detailed engagement with the physical theory at hand will answer the question. Thus, substantive work in advocating functionalism in philosophy of physics is spelling out the functional roles (and this is what I do for TDSL in the first half of this paper). But, in particular case studies, cashing out which differences matter and which don't will be very controversial. To return to Sklar's example, being a statistical rather than non-statistical property could be a difference that does not matter, if, for example, the functional role of

temperature is to be a quantity that is numerically identical for two systems in mutual equilibrium. But in the case of spacetime, there is controversy over the correct functional role, cf. Knox (2019), Baker (2018), Read and Menon (2019).

In this way, functionalism is a useful strategy for considering, but not a solution to, vexed questions of reduction. Functionalism frames the debate, but doesn't singlehandedly resolve it.[3]

Whilst there are no in-principle restrictions about to which theories the functionalist strategy can be applied[4], thermodynamics lends itself especially naturally to a functionalist perspective, as suggested by Sklar (1999): 'In thermodynamics the concept of entropy is defined solely by its function in the theory. We have no direct phenomenal sense of entropy, nor are there devices that serve as direct entropy measurers' (Sklar 1999, p. 195), and so 'something akin to functionalist accounts of mental concepts is appealing' (Sklar 1999, p. 191). This is because many of its core arguments and notions, such as the Carnot cycle, are very abstract. Thermodynamic systems are described by only a few parameters and the microscopic details are purposefully not considered. As such, functional commonality amid diversity in the microstructure is a theme in thermodynamics — which is conducive to taking

[3]Consequently, I do not take functionalism to be a necessary component to reduction, contra Kim (1998; 1999). Naturally, there is a substantive project to connect functionalism in philosophy of physics to philosophy of mind, but these issues are not central to my project here. *Causal* roles are central to philosophy of mind, but seem inappropriate in physics (cf. Russell (1913), Norton (2009), Frisch (2014)). But independently of this, Kim's account faces problems: for example, see Rueger (2006) for an argument that the two quantities in question X_t and X_b will generally have different causal profiles.

[4]Lewis' approach to theoretical terms shows that all concepts in science can be considered to be functional concepts, (Lewis 1970). This formal point has an informal counterpart: 'Functionalism is the idea enshrined in the old proverb: handsome is as handsome does. Matter matters only because of what matter can do. Functionalism in this broadest sense is so ubiquitous in science that it is tantamount to a reigning presumption of all science' (Dennett 2001, p.233).

functionalist approach, since it echoes the slogan in philosophy of mind that 'functional commonality trumps physical diversity' (Levin 2018).

To sum up: If the higher-level concepts are functional role concepts, then the realiser just has to play the same role, i.e. have the same *behaviour*. Consequently, certain differences between the quantities of T_t and T_b that one might worry block reduction — might not matter.

Next I consider some important features of the theory to be reduced: thermodynamics.

3 The nature of thermodynamic 'processes'

Often a physical theory has two components: the kinematics and the dynamics. The kinematics specify the state-space: the possible states of the system. The kinematic component of thermodynamics is clear: in section 3.1, I describe the equilibrium state-space. I then consider the 'dynamics' in 'thermodynamics'. Usually, the evolution of a physical system is determined by the theory's equations of motion and its evolution can be represented by a curve through state-space parametrised by time. But this familiar situation is alien to thermodynamics. Thermodynamics is not a dynamical theory. Indeed, one might think that 'thermostatics' would be a more appropriate name. There are no equations of motion and no explicit time parameter. Furthermore, it is hard to see how a curve in an *equilibrium* state-space could represent any dynamical process, let alone which direction this process would occur. In section 3.2, I consider how any processes are possible at all, and then, in section 3.3, I discuss curves in equilibrium state-space. I discuss the sense in which they are reversible, and in section 3.4 I outline three types of time-asymmetry in thermal physics.

3.1 Equilibrium state-space

The state-space of thermodynamics, Ξ, is the space of equilibrium states, parametrised by two or more macrovariables. For a gas, the points of Ξ can be labelled by pressure and volume (p, V); for a film, they are labelled by surface tension and area; for a magnet, magnetic field and magnetization; and for a dielectric, electric field and polarization (e.g. Tong (2012, §4)).

Thermodynamic equilibrium states are states in which the macrovariables no longer vary in time: the system (as described by thermodynamics) will sit there indefinitely. Naturally, the

absolute nature of thermodynamic equilibrium is an idealisation.[5] Nevertheless, the key point is that we get away with treating a system *as if* it were in thermodynamic, i.e. absolute, equilibrium (at least: for the cases where TD is empirically successful.)

Equilibrium is at the heart of thermodynamics, and it is a presupposition of the theory that systems will end up in equilibrium. Because this requirement that systems do in fact reach equilibrium is prior to the other laws, Brown and Uffink (2001) call it the 'minus first law' (but they also suggest that it is so central that the name 'the minus infinite law' is also appropriate (Brown and Uffink 2001, p. 529)).

> *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, p. 528).

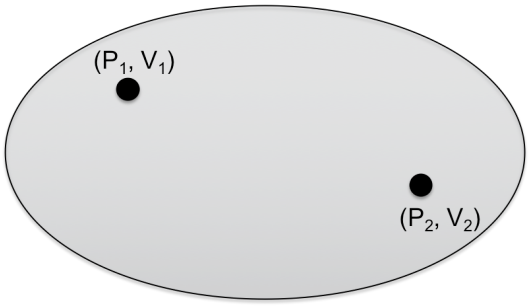

Figure 1: The equilibrium state-space Ξ appropriate for an ideal gas. The co-ordinates (P_1, V_1) label point x_1 and (P_2, V_2) label point x_2.

3.2 Interventions

By the very definition of an equilibrium state, once a system reaches such a state (and so is represented by a point x_1: (p_1, V_1) such as in Figure 1), it will remain there indefinitely — it

[5]Features of the underlying theories suggest that a system won't stay in equilibrium *forever*. For example, Poincaré recurrence suggests that systems will eventually return to earlier states. Furthermore, to take an example from (Wallace 2015, ft. 1): at room temperature, hydrogen and oxygen appear to be in equilibrium with one another, but if you strike a match, we see the system change dramatically: that equilibrium, also, wasn't forever.

cannot spontaneously move to another state, x_2: (p_2, V_2). Thus, for any change or process to occur, there must be an intervention on the system: e.g. inserting a partition, squeezing with a piston, placing the system in thermal contact with a heat bath or slowly varying a magnetic field (cf. Wallace (2014, p. 699)).

These are interventions on the system by external systems (that need not be agents in any thick sense). These interventions alter external parameters such as volume, or magnetisation — variables that would otherwise be unchanging for a system in thermal equilibrium.[6]

But if such interventions knock the system out of equilibrium, then its state is no longer represented in TD state-space, Ξ. However, the minus first law says that once the external parameter is no longer changing, the system will return to a — perhaps, new — equilibrium state.

To illustrate this, consider the following example: the Joule free expansion of a gas. The system is initially in equilibrium state x_1, represented by the point (p_1, V_1) in Figure 1. The partition is removed and the gas rapidly expands in an uncontrolled manner. After some short time, the gas settles down to a new equilibrium state, x_2, with a larger volume. Only the initial and final states of this process are represented in Ξ: thermodynamics is silent on what happens away from equilibrium. Therefore, Figure 1, but not Figure 2, represents the Joule expansion.

3.3 Curves: quasi-static processes

Considering a curve through the equilibrium state-space Ξ raises interpretational issues. Figure 2 shows an undirected, continuous curve from point x_1 to point x_2. How can such a set of points represent any process? Any intervention will knock the system out of equilibrium —

[6]Wallace (2014) uses the terminology 'control theory', and similar themes run throughout the foundational literature. Lavis (2018) discusses a similar control theory view, but in terms of adiabatic accessibility. Myrvold (2011) discusses Maxwell's means-relative view of thermodynamics, whereby certain quantities are relative to an agent's means. In the context of quantum theory, 'resource' theory views of thermodynamics are popular (Horodecki and Oppenheim 2013). I believe that these foundational views bring out what is already implicit in traditional presentations of thermodynamics: interventions are required.

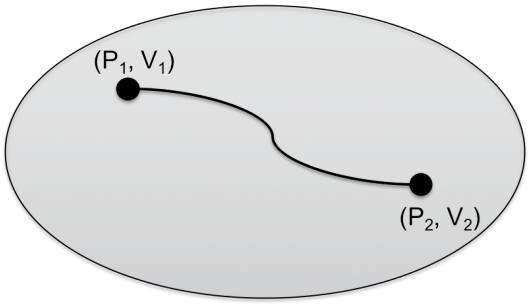

Figure 2: A curve through the above equilibrium state-space Ξ.

indeed, this is required for anything to happen. And we can't just ignore this problem. Although many processes in TD are like the Joule free expansion (i.e. will not be represented by such curves), much of thermodynamics involves examining curves through Ξ.[7]

But how should we consider such curves? This question has been at the heart of a recent controversy. In what follows, I outline the common thread to the three main recent papers on this controversy: Norton (2016), Lavis (2018), Valente (2017), who openly admit that there is not a vast difference between their resolutions.[8]

First, all agree no actual system will trace out the curve spontaneously. Hence, Tatiana Ehrenfest-Afanassjewa called these curves 'quasi-processes' to emphasise that they are unphysical, mathematical constructs (Ehrenfest-Afanassjewa 1925; 1956). But the orthodoxy is that we can make a series of very small interventions to external parameters, and the system will then arrive at a new, neighbouring equilibrium state without ever straying 'too far' from equilibrium. The orthodoxy is that intervening 'gently' or 'slowly enough' will ensure this closeness to equilibrium.[9]

[7]In particular, a common strategy is to integrate the small changes in parameters such as p, V along such curves to find new thermodynamic quantities, especially ones which are path-independent. This allows us to talk of the changes in the values of these quantities even in processes such as the Joule expansion — which involves the non-equilibrium goings-on of which TD is silent.

[8]For example: 'Granted, the two proposals do not seem to differ too much from each other ' (Valente 2017, p. 1777), and 'the work of this paper has similarities with that of Norton (2016)' (Lavis 2018, p. 137).

[9]Of course, there is an undesirable vagueness in the claim that the system is not 'too far'

Why is it assumed that performing the interventions slowly enough will help the system stay close to equilibrium? Equilibrium requires that the macroparameters are not changing in time. By perturbing the system *slowly* — e.g. inserting the piston slowly — the macroparameters won't change very quickly, and so the system will not be too far from equilibrium. But how should we evaluate 'fast'? Fast compared to what? There is no global, nor a priori answer, but to give a rough idea: in the case of the 'slow insertion' of the piston to intervene on the volume, the time taken to make a small change dV should be long compared to the timescale over which the molecules bounce between the piston and the wall. In this case, the process the system undergoes is a good approximation to that represented by the curve. The smaller the intervention the better this approximation that the curve represents the process occurring. But no actual process is perfectly represented by the curve. In the limit of smaller and slower interventions, nothing happens —- there is no 'process'. Rather the curve delimits or is the 'common frontier' (Lavis 2018, p. 139) of the set of sequences of processes, which approximate the quasi-static process. Thus, the term 'quasi-static' properly denotes a set of processes, whose sequence heads in the direction of the common frontier — the curve in Ξ — but never meets it.[10]

The bare curve can become a directed curve — the curve can be traversed in either direction, but different interventions are required for each direction. To travel in one direction pistons must be inserted, and in the other direction they must be removed. There is one further condition that must be mentioned: in order that a process can proceed in either direction, and so retrace its steps, there must be no friction (e.g. in the piston). Thus, standardly friction is excluded, cf. Tong (2012, p. 113), Uffink (2001, p. 365), Blundell and Blundell (2009, p.

from equilibrium. How far is too far? There is no satisfying answer to this question. As Valente (2017) notes, it is hard to make this precise: we can't appeal to a topology over non-equilibrium states to establish that they are close enough to equilibrium, since they are not described by TD. Instead —as is common with approximations— whether the system is 'close enough' is an empirical matter. Indeed, Afanassjewa-Ehrenfest's view is that we need an 'empirically grounded concept of "close enough to equilibrium" ' (Valente 2017, p. 1777).

[10]Norton (2016) and Lavis (2018) emphasise that Duhem (1902, p. 78) has a similar approach.

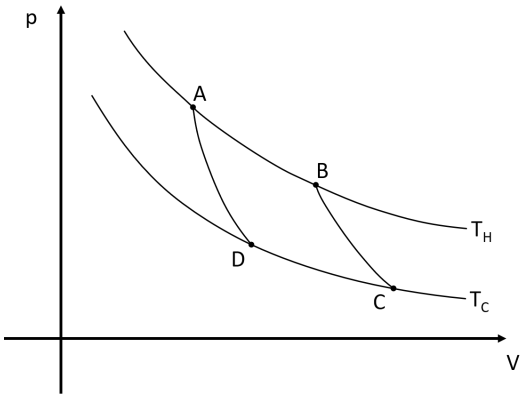

Figure 3: Quasi-static reversible processes represented in the p-V plane of equilibrium states.

120).[11]

Because the curve can be transversed in either direction (for example, in the Carnot cycle in Figure 3 the arrows can be drawn in either direction), there is a sense in which it is 'reversible'. But, next I note that there are different concepts of reversibility.

3.4 Concepts of Irreversibility

Before examining the TDSL, it is important to unravel the different concepts of reversibility. Uffink (2013) outlines three concepts of 'reversibility' in thermal physics:

1. **Time-reversal invariance (TRI)**: there exists a map $\mathcal{T}$ — frequently assumed to be the map $t \mapsto -t$ — that maps possible histories of the system to possible histories.

2. **Quasi-static reversible processes**: we previously saw that curves in TD state-space represent quasi-static processes, and are reversible in the sense that the arrows can be drawn in either direction on the curves in Figure 3: corresponding to expansions and compressions. But travelling in one direction is not straightforwardly the 'time reverse' in the TRI $t \to -t$ sense: one is not performing the same interventions in a different order, but rather performing different interventions (e.g. inserting rather than removing a piston). Furthermore, as previously discussed, this 'quasi-static reversibility' is a property of a sequence of processes, rather than of a single process.

[11] In the First law, writing that $đW = pdV$ requires that there is 'no friction or hysteresis' (Uffink 2001, p. 365).

The name given to this type of reversibility by Clausius and Planck is 'umkehrbar' (Uffink 2001, p. 343), a word for reversibility with connotations of 'unwinding'. Part of the idealisation of quasi-static (or umkehrbar) processes in thermodynamics is that there is no friction, or dissipation, so that the curve can be transversed in either direction. But, outside the context of thermal physics, a quasistatic process needn't be time-reversible at all. For example, discharge of a condenser through high resistance can be forced to happen very slowly, but nonetheless it is clearly an irreversible process (Uffink 2013, p. 277).

However, for the rest of this paper, we will stick to the usage in thermal physics; 'quasi-static processes' will denote the reversible processes represented by curves in Ξ discussed in the previous section, which exclude friction.

3. **Recoverability**: the process in question can be 'fully undone'. The system can be returned to its initial state K_i with no effect in the environment E. But the system need not retrace its steps — it can take a different path to its destination.[12] So process P is recoverable, if: writing $\langle K_i, E_i\rangle \xrightarrow{P} \langle K_f, E_f\rangle$ there is a process $P*$ such that $\langle K_f, E_f\rangle \xrightarrow{P*} \langle K_i, E_i\rangle$.

Having distinguished the different types of reversibility in thermal physics, we now turn to what is often claimed to be the source of irreversibility: the second law.

4 The Second Law Introduced

There are many statements of the TDSL (see Uffink (2001) for the relationships between them). One classic formulation of the TDSL is the Kelvin statement: 'it is impossible to perform a cyclic process with no other result than that heat is absorbed from a reservoir, and work is performed' (Kelvin (1882) as cited in (Uffink, 2001, p.328)). In this section, I show how the TDSL and the reversible quasi-static processes discussed in the previous section are

[12]Luczak (2018) adds the further condition that the process $P*$ must be one that *we* can implement — this is part of the Maxwellian view, which I cannot discuss further here but see Myrvold (2011).

used to define the thermodynamic entropy S_{TD}, and codify its behaviour.

The starting point is a formulation of the TDSL, known as the Carnot theorem (Blundell and Blundell 2009, p. 130). Carnot's theorem states that the Carnot cycle in Figure 3 (which operates between two reservoirs, one at a temperature T_h and the other at lower temperature T_c) is the most efficient, i.e. the best we can do, and so the efficiency $\eta = \frac{Q_h}{Q_c}$ is the same for all reversible engines, where Q_h is the heat absorbed in the isothermal expansion $A - B$ and Q_c is the heat emitted in the isothermal compression $C - D$. (See Clausius (1879, p. 80) for an argument that the efficiency is independent of the substance considered).

In a Carnot cycle:

$$\Sigma_{i=1}^{2} \frac{Q_i}{T_i} = 0. \tag{4.1}$$

This can be generalised to an arbitrary quasi-static reversible cycle in the equilibrium state-space Ξ, as shown in Figure 4.

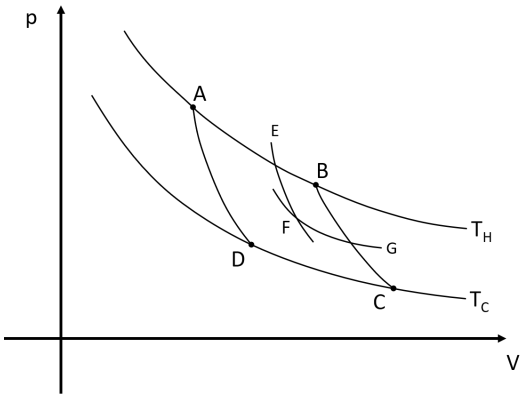

Figure 4: This diagram shows the original Carnot cycle, as well as another smaller Carnot cycle, EBGFE. By cutting more corners, i.e. by having many infinitesimal adiabats and isotherms, any quasi-static reversible cycle in the plane can be considered.

In this case,

$$\oint \frac{đQ}{T} = 0. \tag{4.2}$$

Thus, if there are two (or more) reversible paths (i.e. quasi-static curves) between equilibrium state A and equilibrium state B the change in $\int_A^B \frac{đQ}{T}$ is independent of the path taken. (See Figure 5).

This (along with a reference state 0) allows us to the define a new function of state which

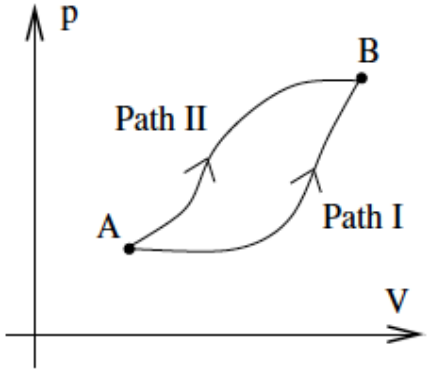

Figure 5: Two possible paths between two states in Ξ. Figure from Tong (2012).

only depends on the state variables p, V : the thermodynamic entropy S_{TD}.

$$\int_0^B \frac{đQ}{T} = S_{TD}(B) \tag{4.3}$$

Because entropy S_{TD} is a function of state, it is path-independent: it doesn't matter how we reached state B — quasi-statically or not, or whether the system was isolated or not — either way the entropy of state B is $S(B)$.[13]

Clausius' inequality generalises away from the quasi-static reversible cycle above to any cycle:

$$\oint \frac{đQ}{T} \leqslant 0 \tag{4.4}$$

$$\oint \frac{đQ}{T} = \int_1 \frac{đQ}{T} - \int_2 \frac{đQ}{T} \leqslant 0. \tag{4.5}$$

If path 1 is not quasi-static and path 2 is a quasi-static path from state A to B, and path 1 is adiabatic (so $đQ = 0$), then we learn that the thermodynamic entropy of a *thermally isolated* system cannot decrease:

$$S_{TD}(B) - S_{TD}(A) \geqslant 0. \tag{4.6}$$

[13]The Third law is needed to set the convention that at absolute zero, the entropy is zero, but since the third law is controversial (cf. Wald (1997), Masanes and Oppenheim (2017)), I leave it aside here.

From the quasi-static reversible curves in equilibrium state space, we've defined a new state function S_{TD}, and shown that for thermally isolated systems, if a process P is a quasi-static reversible process then $\Delta S_{TD} = 0$, and if P is non-quasi-static, then $\Delta S_{TD} > 0$.

4.1 What type of irreversibility does the TDSL describe?

The second law describes the irrecoverability of certain initial states — if the process P (in section 3.4's definition) is a non-quasi-static process. As such the Clausius relation provides a bridge from umkehrbar or quasi-static reversible processes to the definition of irrecoverability.

Generally, since irrecoverability is a modal notion, it requires we consult all possible processes to establish whether there exists a process P^* that takes $\langle K_f, E_f \rangle \xrightarrow{P*} \langle K_i, E_i \rangle$. Thus, determining whether P is recoverable is an epistemic challenge. But in the case of TD, the challenge is lessened, since the signature of irrecoverability is that there will be an increase in S_{TD} associated to the thermally isolated system.

Both concepts, recoverability and quasi-staticity, are central to the TDSL, but it is irrecoverability that captures the imagination. Irrecoverability seems like a widespread phenomenon: our inability to recapture lost youth, smashed wine glasses or split milk exemplifies the irrecoverability of these processes. But there's an open question whether we can assign a thermodynamic entropy S_{TD} to these processes, since it is unclear that there's an equilibrium state space description available, or there are the relevant quasi-static processes available. And in defining S_{TD}, we relied on quasi-static processes. Uffink emphasises their importance: 'if such processes did not exist then the entropy difference between these two states would not be defined' (Uffink 2006, p. 938)[14], adding that 'this warning that the increase of entropy is thus conditional on the existence of quasi-static transitions has been pointed out already by Kirchhoff (1894, p. 69)', as cited in (Uffink 2006, p. 938).[15] In a

[14]Of course, as discussed earlier, quasi-static processes only exist in the sense of being the limit of set of actual processes we can implement. But this is all that is needed to calculate various quantities, namely: $\frac{đQ}{T}$ — we don't need to be able to implement a perfect quasi-static process — as Norton (2016) emphasises, this is impossible!

[15]This casts doubt over Atkin's bold claim that the second law is responsible for a vast range of processes, including the 'formation of a thought', as quoted at the opening of this

nutshell, it is unclear thermodynamics applies to these everyday examples of irrecoverability in anything other than a metaphorical sense.

Of course, since S_{TD} is a state function (a path independent quantity), the entropy change during a non-quasi-static process is well-defined. But quasi-static processes are required to calculate ΔS_{TD}, and define it in the first place. Furthermore, adiabatic quasi-static processes provide the lower bound on the entropy change, $\Delta S_{TD} = 0$. This centrality of quasi-static processes will be important when we turn to the reduction of the TDSL, which is the topic of the next section.

5 Turning to statistical mechanics

Having outlined the TDSL and processes in TD we turn to the reductive project: what is the reductive basis of the TDSL in SM? Finding the distinction between heat and work in SM is at best complicated (cf. Maroney (2007), Prunkl (2018)) and at worst 'unnatural' (Knox 2016, p. 56) or 'anthropocentric' (Myrvold 2011).[16] Consequently, the Kelvin formulation does not have an obvious correlate in SM. Indeed, cyclic processes and Carnot engines do not have a starring role in SM, unlike TD. Transferring heat between bodies of different temperatures is not the main concern of SM either. Instead, non-equilibrium SM is concerned with qualitatively describing the approach to equilibrium. And equilibrium SM calculates various macroscopic quantities from the canonical probability distribution (and the partition function Z plays a starring role). As such, the focus of SM differs slightly from that of TD.

Thus, a natural way to connect these two subject matters in order to find the reductive basis of the TDSL within SM is this: the TDSL has the implication that S_{TD} cannot decrease (for a thermally isolated system). Hence, finding the SM realiser of S_{TD} is key to finding the

paper, since it is far from clear that the requisite quasi-static processes are available in the 'formation of a thought'.

[16]Maxwell claimed the distinction between heat and work is one of disordered and ordered motion, which 'is not a property of material things in themselves, but only in relation to the mind which perceives them' (Maxwell 1878, p. 221); (Niven 1965, p. 646) as quoted in Myrvold (2011).

reductive basis of the TDSL in SM. Indeed, Callender (1999) calls this the search for the 'the holy grail': find a SM function to call 'entropy' and establish that it is non-decreasing.

But I think that role of S_{TD} as defined by this holy grail does not capture the right features of the TDSL: I now argue that 'being a non-decreasing SM function' is not the right functional role for S_{TD}. Defining the functional role as 'non-decreasing' is in some respects too weak, and in other respects too strong. In the next section I emphasis how the reductive project at hand — the reduction of the TDSL — differs from the reduction of the minus first law. Then, in section 5.2, I criticise the old grail, and in doing so emphasise the essential features of the TDSL that lead to the correct functional role: the new grail.

5.1 The reductive project here

There is a feature about the TDSL that is important to emphasise for the reductive project at hand: the distinction between the second law and the minus first law. The spontaneous approach *to* equilibrium (from non-equilibrium) is distinct from the second law, which describes the thermodynamic entropy differences *between* equilibrium states. It is a presupposition of TD that systems *do* in fact reach a unique state of equilibrium, as discussed in section 3.1: this is the minus first law. Once the system reaches equilibrium then, by definition, it will not change — unless there is an intervention on an external parameter.

To emphasis the contrast: the second law tells us that certain interventions render the initial state irrecoverable, where as the minus first law tells us that systems spontaneously reach a state of equilibrium.

Finding the microphysical 'underpinning' for these two laws are distinct projects (cf. Luczak (2018)). The H-theorem and coarse-graining approaches in SM are concerned with quantitatively describing the approach to equilibrium. These foundational projects are concerned with establishing the circumstances under which a given system will approach equilibrium, rather than the quasi-static interventions on equilibrium states. That is, they are concerned with the underpinning of the minus first law, rather than the second law.

Of course, since equilibrium states are central to TD and the minus first law is baked deep into the nature of quasi-static processes, the two projects are connected (as we will see later).

But the crucial point to emphasise here is that even if we resolve the controversy around the underpinning of the minus first law, there is still a *further* project to find the underpinning of the second law. This project is rarely discussed but this is what is required to have a reduction of the TDSL.[17] And this isn't just nitpicking over the names of laws: the types of irreversibility captured by the minus first and second law differ. The minus first law is concerned with the more familiar type of irreversibility —non-TRI— exemplified in the spontaneous approach to equilibrium. But as we saw in section 3, the nature of processes in TD differs from this familiar spontaneous evolution: interventions and quasi-static processes are key.

5.2 Out with the old, in with the new: the search for the holy grail

I first describe why the old grail is too strong, and then discuss why it is too weak — before outlining the new grail.

The old grail claims that the SM realiser must be 'non-decreasing', but this is too strong: S_{TD} can decrease when the system is not thermally isolated from its environment. The state of the environment was key to the definition of irrecoverability, and the environment is a key feature of the TDSL. For example, it is important to emphasise the 'sole effect' part of the Clausius statement: otherwise, fridges would be a clear counterexample to the TDSL. Fridges transport heat from a colder to hotter body — at a cost. Such transport is only prohibited as the *sole effect*. Likewise, in section 4, we showed that the entropy S_{TD} of the system is only non-decreasing during adiabatic (i.e. thermally isolated) processes. Indeed, during an isothermal compression from C to D, the entropy of the system decreases. This is especially obvious when we view the Carnot cycle in the T-S plane, as shown in Figure 6. Of course, during an isothermal compression heat flows to the heat bath, i.e. the environment, and so during this process the *net* entropy change ΔS_{TD} is zero.

Here it is clear how central the system and environment split is in thermodynamics. We can

[17] One notable is Gibbs' 1902 textbook where he discusses the SM analogues of TD processes such as the Carnot cycle, (Gibbs 1902, Ch. XIII). But to Gibbs' eyes these are mere analogues rather than reductions, a point endorsed by Batterman (2010).

naturally take the original system and heat bath together as 'the system'. If *this* system is thermally isolated from all other systems, then the Carnot cycle is an adiabatic quasi-static process and the entropy change is zero — as expected. In this way, we might think that adiabatic processes are more foundational than isothermal processes. Henceforth, we will mainly consider thermally isolated systems (and so adiabatic processes), and return to heat in section 6.4.

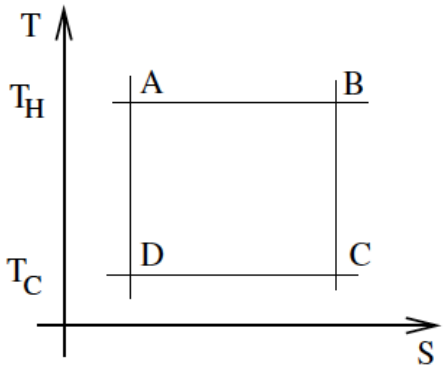

Figure 6: The Carnot Cycle represented in the T-S plane.

There is a second reason to think that the old grail — that the SM realiser must be non-decreasing — is too strong. From the outset I have emphasised that equilibrium is central to TD. Indeed, S_{TD} is only defined at equilibrium: it is silent about what happens away from equilibrium. And in this way the old grail too strong: if the SM entropy decreases away from equilibrium, this needn't be problematic. A decreasing SM entropy only conflicts with the TDSL if it decreases between isolated equilibrium states.

An importance consequence of the second law is the irrecoverability which it legislates; the signature of irrecoverability is often taken to be the increase of thermodynamic entropy (in a thermally isolated system). This highlights a way in which the old grail is too weak — 'non-decreasing' does not suffice to capture the right role because the realiser of the S_{TD} must *increase* in the right situations too. As such, the traditional holy grail — a non-decreasing function — does not suffice: a realiser of S_{TD} must also *increase* in the right situations — during non-quasi-static adiabatic processes.

Thus, the old grail does not capture the right role: 'non-decreasing SM entropy' is not the

right desiderata for microphysical realiser of the TDSL. But through criticising it we have considered the key features of the TDSL: the importance of quasi-static processes, the environment and the distinction between the minus first and second law. Thus, we are now in a position to state the correct functional role: the new grail.

> *The new grail:* find a SM realiser which, for thermally isolated systems, is increasing in non-quasi-static processes, but non-increasing in quasi-static processes, such as those represented by curves in Ξ.

Next, in section 6, I show how the realiser can be found in statistical mechanics (SM). I show how quasi-static processes can be modelled in SM, and then show how the Gibbs entropy plays the right 'new grail' role.

But, whilst part of daily workhorse of SM, the Gibbs entropy is unpopular in the foundational literature. The main complaint is that the Gibbs entropy is 'an ensemble property', rather than a property of the individual system. (This often frequently motivates a Boltzmannian approach to SM, instead of a Gibbsian one, cf. Callender (1999; 2001)). In section 7, I quell some of these worries about the Gibbsian approach, and defend the Gibbs entropy, S_G. However, this defence will not involve any criticism of the Boltzmannian entropy S_B — I leave it as a challenge to the neo-Boltzmannian to show that S_B can play the S_{TD} role as well as S_G does.[18]

6 Searching for the new grail in SM

The realiser of S_{TD} needs to behave differently in quasi-static and non-quasi-static processes. In this section I discuss how the distinction between slow, quasi-static processes and rapid, non-quasi-static processes can be made in SM.

[18] As such, I am leaving open the possibility that there is more than one realiser of the S_{TD} role. That is, the S_{TD} may be multiply realised - a view endorsed by Sklar (1999) and Wilson (1985). Whether multiple realisability is worrisome depends on issues in the metaphysics of properties, so I leave it to one side here.

SM is an umbrella term for classical SM (CSM) and quantum SM (QSM). Whilst the story I tell in this section runs in parallel for QSM and CSM, I focus on the QSM framework for two reasons: firstly, since quantum mechanics is considered to be the correct theory (to which classical mechanics is an approximation), QSM should be the priority (and happily, the key principle required for this section is less contentious in the quantum case than the classical case). Secondly, my focus on QSM over CSM in this section foreshadows my later argument (in section 7) that certain problems can be resolved (or dissolved) by considering the quantum rather than the classical.

In section 6.1, I first consider how interventions on external parameters influence the state of the system. Furthermore, I will demonstrate how, for thermally isolated systems, the Gibbs entropy S_G is constant in quasi-static processes (section 6.2), but increases in non-quasi-static processes (section 6.3) — and thus S_G can play the right role. In section 6.4 I will connect my claims about S_G back to heat.

6.1 Interventions in QSM

In QSM, like CSM, thermal equilibrium is represented by the canonically distributed state:

$$\rho_{can} = \Sigma_i w_i \left|E_i\right\rangle \left\langle E_i\right| \tag{6.1}$$

where

$$w_i = \frac{e^{-\beta E_i}}{Z}, \tag{6.2}$$

where Z is the partition function. Whilst in CSM, ρ_{can} is a probability density distribution over the phase space Γ, in QSM ρ_{can} is a density matrix.[19] ρ_{can} is a statistical mixture of energy eigenstates, where the probability of being a given energy $|E_j\rangle$ depends exponentially on the eigenvalue E_j of that state, and the temperature $\beta = k_B T$, where k_B is Boltzmann's constant.

Maroney (2007) gives an elegant justification for why ρ_{can} represents thermal equilibrium

[19] In QM, the density matrix $\rho = |\Psi\rangle \langle\Psi|$ is a more general object than the wavefunction Ψ, since it represents all that the wavefunction can and more – it can also represent statistical mixtures, cf. Sakurai and Commins (1995), Landau and Lifshitz (1964).

states familiar from thermodynamics[20], but here it will suffice to note two features:

1. The unitary evolution of a density matrix is given by the Liouville-von Neumann equation:

$$i\hbar\frac{\partial\rho}{\partial t} = [H,\rho] \tag{6.3}$$

Since ρ_{can} commutes with the time-independent H, it is unchanging in time:

$$\frac{d\rho_{can}}{dt} = 0. \tag{6.4}$$

2. The canonical ensemble (at a given total energy and temperature) maximises the Gibbs entropy:

$$S_G = -k_B Tr\rho ln\rho \tag{6.5}$$

which is the quantum analogue of the classical Gibbs entropy:

$$S_G = -k_B \int dqdp\rho(q,p)ln\rho(q,p). \tag{6.6}$$

At t_0, let us start with the system in the canonical ensemble, ρ_{can}, where the Hamiltonian, $H(t_0)$ is time-independent. When there is an intervention on an external parameter V in the period $t_0 < t < t_1$, the Hamiltonian will be time-dependent. At t_1, the parameter V has a new value V_1, and the Hamiltonian is once again time-independent.

For example, let us consider changing the volume of the box. The external parameter, V, determines the potential energy:

$$U_{box}(x_i,y_i,z_i)) = \begin{cases} 0 \text{ if } 0 < x_i < x(t), 0 < y_i < L_y, 0 < z_i < L_z \\ +\infty \text{ otherwise} \end{cases}$$

Changing an external parameter, like the volume of the box, changes $H(V(t))$. At the

[20]Here I am clearly working with Gibbsian SM. In the Boltzmannian picture, equilibrium is represented by the largest macrostate in phase space (or as Werndl and Frigg (2015a;b) suggest the state that the system spends the most time in).

beginning t_0, $H(V_0)$, and end of the process t_1, $H(V_1)$, the Hamiltonian is time-independent. When $t_0 < t < t_1$, the Hamiltonian is changing.

The energy eigenstates $|E_i\rangle$ in equation 6.1 are eigenstates of the initial Hamiltonian $H(V_0)$, and so are unchanging in time. In the period, $t_0 < t < t_1$, each eigenstate $|E_i\rangle$ evolves to a new state $|\Psi(t)\rangle$, which is written in this general form to emphasise that $|\Psi(t)\rangle$ might not be an eigenstate of the new Hamiltonian, and furthermore, is changing in time.

In the next two sections, I consider how the state of the system changes during the intervention in $t_0 < t < t_1$, and what the state at t_1 will be. In particular, we need to show that:

- If the change to the external parameter is quasi-static (i.e. $t_1 - t_0 \to \infty$) then S_G is constant: I do this in section 6.2.
- But if the intervention is non-quasi-static then S_G increases: I do this in section 6.3.

6.2 Quasi-static changes

In thermodynamics, a quasi-static process requires that the systems is very close to equilibrium at every stage. In QSM, this translates as the requirement that system is approximately canonically distributed, whilst an external parameter is altered very slowly.

One heuristic for thinking about this: each pure state (which is initially an energy eigenstate of $H(t_0)$) in the statistical mixture $\rho_{can}(t_0)$ evolves under the time-dependent Schrödinger equation, carrying its original weighting w_i with it.

The key issue is why think that $\rho(t)$ will still be canonical under this evolution? For $\rho(t)$ to be canonical at any given time, it needs to be:

1. a statistical mixture of eigenstates of $H(t)$, whilst H changes in the period $t_0 < t < t_1$.
2. whose probability depends on the new energy eigenstate, E_i.

1. is ensured by a theorem, known variously as Ehrenfest's principle, or the quantum adiabatic

theorem[21] (cf. Griffiths and Schroeter (2018, Ch. 10), Messiah (1962, Ch. 17).[22]

> **Ehrenfest's principle**: If the energy eigenstates of $H(t)$ are non degenerate for times $t > t_1$, if $|E_i(t_1)\rangle$ is an energy eigenstate of $H(t_1)$, if $|E_i(t)\rangle$ is the state evolved from $|E_i(t_1)\rangle$ according to the Schrödinger equation, and if the external parameter changes very slowly, then $|E_i(t)\rangle$, for each time $t > t_1$, is very nearly an energy eigenstate of $H(t)$ at the corresponding time. In the mathematical limit of a finite change in the external parameter occurring over an infinite time interval, 'is very nearly' becomes 'is' (Baierlein 1971, p. 380).

Why should we think that the conditions of Ehrenfest's principle hold? Infinite time limits are contentious (cf. Palacios (2018)), and of course, only 'approximately' hold in real life situations. But, just like in the thermodynamic situation, an intervention is smooth and 'slow enough' if $t_1 - t_0$ is larger than the characteristic timescale of the particular system in question

[21]For our purposes, neither name is ideal. 'Adiabatic' here means 'very slow' rather than its usual TD meaning, and 'Ehrenfest's principle' may be confused with Ehrenfest's theorem, which relates the expectation value of position and momentum, and is related to the quantum-classical correspondence principle.

[22]There is an analog of the quantum adiabatic theorem in classical mechanics. In CM, a slow change to an external parameter (such as the length of a pendulum), cf. (Arnold 2010, p. 298), is called an adiabatic change (beware the different meaning of 'adiabatic' than in TD!). A property of a system that stays approximately constant when changes occur sufficiently slowly is called an adiabatic invariant. Rugh (2001) shows that under an ergodic hypothesis the entropy is an adiabatic invariant. Whilst these ideas date back to Hertz (1910), they are far less established than Ehrenfest's quantum adiabatic theorem. Furthermore, they depend on two contentious issues: (i) the ergodic hypothesis (which is hard to show holds of many realistic systems, cf. Earman and Rédei (1996)) and (ii) the nature of CSM probability: if one takes a Jaynesian approach, such dynamical considerations about changes in the Hamiltonian need to be connected to our knowledge (I will return to this latter issue about probability in CSM in section 7).

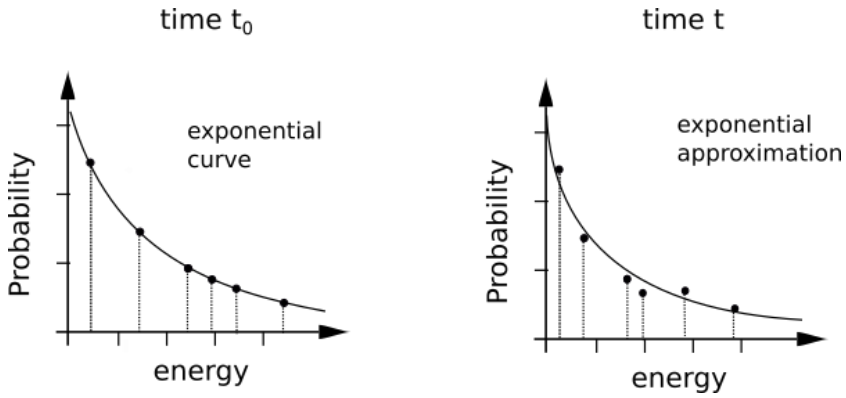

Figure 7: The graph on the left shows the canonical distribution at t_0, and the graph on the right shows the system approximately in the canonical distribution at the later time, diagram drawn following (Baierlein 1971, p. 380), depicting only six of the many states.

(see Messiah (1999) for more details).[23]

If Ehrenfest's principle applies, earlier eigenstates $|E_i(t_0)\rangle$ will be taken to new energy eigenstates, $|E'_i(t_1)\rangle$. Furthermore, if there is no degeneracy, then there will be no 'crossings' of the lines in Figure 7, and so the distribution is monotonically decreasing.[24] Thus, a radically different distribution (such as a Gaussian distribution) is not possible, since for such a radical difference, the eigenstates would need to cross (i.e. the originally highest probability lowest energy eigenstate must be shifted to the peak of the Gaussian distribution).

But even if Ehrenfest's principle ensures that the distribution will remain monotonically decreasing, there remains the question: is it canonically distributed – that is, is there an exponential dependence of the probability w_i on the energy eigenvalue, E_i?

Yes, provided that the energy eigenvalues of $H(t_0)$ and $H(t)$ are related in a particular way.

[23]In the case of a gas, the characteristic timescale is related to the mean free path: how far, on average, a given molecule travels before colliding. Baierlein gives the following suggestion for getting a handle on the timescale of 'fast': 'let us suppose that the piston is pulled out extremely rapidly, specifically, much faster than the speed of sound in the originally quiescent gas' (Baierlein 1971, p. 408).

[24]The assumption that the energy eigenstates are non-degenerate is contentious, especially for large systems. But the common justification is that any small perturbation will lift the degeneracy, cf. Tong (2012).

That is, if equation 6.7 holds for all i:

$$E_i(V(t)) = f(t).E_i(V(t_0)) \tag{6.7}$$

At t_0,

$$e^{-\frac{E_i(V(t_0))}{kT(t_0)}}, \tag{6.8}$$

which we can re-write in terms of equation 6.7

$$e^{-\frac{E_i(V(t))}{kT(t_0).f(t))}}. \tag{6.9}$$

Thus, if the temperature at t is a scaling of the earlier temperature: $T(t) = f(t).T(t_0)$, then we have a new canonical distribution:

$$e^{-\frac{E_i(V(t))}{kT(t)}}. \tag{6.10}$$

Thus, if the change to the external parameter is slow (i.e. quasi-static) and equation 6.7 holds, then the system will remain (close to) the canonical ensemble, with a varying temperature. Equation 6.7 has been shown to hold for a realistic gas (Katz 1967, p. 84-90), and the hope is that this result will generalise (Baierlein 1971, p. 380).[25]

Thus, we can model quasi-static processes in the QSM. But what of the Gibbs entropy $S_G(\rho)$? How does S_G change during such a process? Here the answer is immediate:

$$\Delta S_G = 0, \tag{6.11}$$

since the evolution is unitary (see Baierlein (1971, p. 379) for an extended discussion). This unchanging nature of S_G is wholly unsurprising, since the traditional problem with the Gibbs entropy is working out how it can *increase* — which is part of the project of the next section.

[25]This assumption is widespread, see (Wallace 2014, p. 714). Furthermore, Baierlein argues that assuming that equation 6.7 holds is reasonable: the new temperature $T(t)$ is determined from the proven constancy of S_G (Baierlein 1971, p. 384).

6.3 Rapid changes

If the change of the external parameter from V_0 to V_1 is rapid, then Ehrenfest's principle does not apply. In particular, re-writing the state of the system in terms of the later energy eigenbasis of $H(V_1)$ (which we denote $|E'_i(t_1)\rangle$) we see that ρ is not diagonal in this basis:

$$\rho(t) = \Sigma_{ij}\omega_{ij}\,|E'_i(t)\rangle\,\langle E'_j(t)| \tag{6.12}$$

Consequently, if $t_1 - t_0 \approx 0$, the system will not be in a statistical mixture of energy eigenstates of $H(V_1)$, and so will not in the canonical distributed state that represents thermal equilibrium. Thus, during the rapid change to the external parameter, the system is not even approximately canonically distributed. But what happens next, i.e. when $t >> t_1$?

In thermodynamics, we just *assume* that the system will settle down to a new equilibrium (the minus first law). In SM, there is a similar pragmatic move, which I consider first following Baierlein (1971), before seeking to justify it.

The pragmatic move: is just to *adopt* a new canonical distribution with energy eigenstates appropriate for the new volume, V_1. In other words, we coarse-grain:

$$\rho(t) = \Sigma_{ij}\omega_{ij}\,|E_i(t)\rangle\,\langle E_j(t)| \rightarrow \Sigma_i\omega_{ii}\,|E_i(t)\rangle\,\langle E_i(t)|\,, \tag{6.13}$$

where we assume that the off-diagonal terms $w_{ij}, i \neq j$ are small so

$$\Sigma_{ij}\omega_{ij}\,|E_i(t)\rangle\,\langle E_j(t)| \approx \Sigma_i\omega_{ii}\,|E_i(t)\rangle\,\langle E_i(t)|\,, \tag{6.14}$$

where t is a long time after the external parameter has stopped changing. Since we have coarse-grained, we expect

$$S_G[\rho_{can}(t_1)] - S_G[\rho_{can}(t_0)] > 0. \tag{6.15}$$

Within TD, the assumption that, after a while (i.e. when $t >> t_1$), systems settle down to a new equilibrium state has no justification, beyond the claim that this is indeed how many

systems in fact behave. But since SM goes beyond TD, we might hope it can do better.

The justification: Rather than just assuming that systems settle to a new equilibrium, the business of non-equilibrium SM is to quantitatively describe the approach to equilibrium. For example, Boltzmann's equation tells you how quickly a gas will settle down to the Maxwell-Boltzmann distribution.

Yet non-equilibrium statistical mechanics is riddled with controversy — and there are many different schools of thought. Should we justify the approach to equilibrium using Boltzmann's combinatoric reasoning (Albert 2000), the H-theorem (Brown *et al.* 2009), coarse-graining (Jancel 2013, Prigogine 1980, Prigogine and Stengers 1984)'s non-unitary dynamics, or some other framework? Here we enter a quagmire. But there's one saving grace for our concerns here: according to all approaches, the SM entropy increases in the approach to equilibrium from non-equilibrium.

But in what follows, I justify the pragmatic move above in terms of my preferred approach to SM: the ZZW coarse-graining framework (cf. Zwanzig (1960), Zeh (2007), Wallace (2011)), which applies to both QM and CM (see Wallace (2016)).

In equation 6.13, the full density matrix ρ evolved from $\rho_{can}(t_0)$ is replaced by a coarse-graining $\rho_{can}(t_1)$ corresponding to a new equilibrium. Many worry this amounts to replacing the true distribution with a distorted distribution (Grünbaum 1973, Redhead 1996, Denbigh and Denbigh 1985). But coarse-graining is not a form of distortion, but rather irrelevant details are thrown away – and so this is a case of abstraction. (See Robertson (2019) for more details, and Myrvold (2014) for a similar line).

More importantly, coarse-graining is used in the ZZW framework to construct the empirically successful irreversible equations that, inter alia, describe the approach to equilibrium. By banning coarse-graining, we would lose these empirically successful equations. Of course, finding an appropriate coarse-graining is hard, and depends on the details of the system at hand and particular initial conditions. But where successful, we can show that the details shown away are *truly irrelevant* for the future evolution of the system (over timescales less than the recurrence time, see Wallace (2011) for more details on when the discarded details are truly irrelevant).

Nonetheless, to reiterate, regardless of whether you endorse the ZZW framework, all schools of non-equilibrium SM agree that the SM entropy *increases* in the approach to equilibrium. Thus, we have achieved our goal: S_G *increases* in rapid, non-quasi-static adiabatic processes, but is *constant* in quasi-static processes.

Thus, I conclude (in agreement with Maroney (2007)): S_G is the realiser of S_{TD} since it plays the right role — and so the TDSL is reduced to SM.

The conceptual hard work is done, but now we can enjoy a corollary of this approach: we can connect the discussion back to heat, and so come full circle.

6.4 Heat and the Gibbs entropy

In thermodynamics, the relationship between heat Q and entropy S_{TD} is:

$$dS_{TD} = \frac{đQ}{T_{TD}} \tag{6.16}$$

(Throughout this section, I will use the subscript TD to make clear that these quantities are defined in thermodynamics.) The first law of thermodynamics states that $dE_{TD} = đQ + đW$, and so

$$dS_{TD} = \frac{1}{T_{TD}}(dE_{TD} + p_{TD}dV) \tag{6.17}$$

In Gibbsian QSM, we find this relationship between heat and entropy as follows.

$$S_G(\rho_{can}) - k_B\Sigma_i p_i ln p_i = -k_B(\beta\langle E\rangle + lnZ) \tag{6.18}$$

If the external parameter V is changed slowly enough that the system remains in the canonical distribution, then the differential form is:

$$dS_G = k(d\beta\langle E\rangle + \beta d\langle E\rangle + \frac{\partial lnZ}{\partial\beta}\partial\beta + \frac{\partial lnZ}{\partial V}dV) \tag{6.19}$$

$$= \frac{1}{T}(d\langle E\rangle + \langle p\rangle dV) \tag{6.20}$$

Since equation 6.20 and equation 6.17 represent the same functional interdependencies, S_G

bears the right relation to 'heat'. Of course, much more could be said about heat and work in QSM: here, I direct the reader to Prunkl (2018), Maroney (2007).

There is one obvious difference between equation 6.20 and 6.17: in QSM, we are dealing with expectation values. In the next section I consider the vexed issue of probability and the associated objections to S_G. But here note that discussing expectation values is not a detraction to this account. The variance from the mean can be calculated, and this gives us useful information about fluctuations (Wallace 2015). Here SM goes beyond TD, and so is the successor theory to TD.

A successor theory often limits the domain — or scope — of the older theory, and this is the case with TD. Since Maxwell (1891), all hands admit that the TDSL can be violated.[26] But nonetheless the TDSL seems to capture something true about our world; greater-than-Carnot efficiency engines are hardly a dime a dozen.[27] Thus, the key issue is to establish under what circumstances the TDSL can be violated, and then restrict the scope of TDSL to exclude those circumstances. Here the orthodoxy is that the TDSL must be weaken to a probabilistic statement, at the very least.[28] Fluctuation phenomena imply that heat can spontaneously flow from colder to hotter bodies (with no other effect), but *on average* there will be no net such flow. Thus, a weakening the TDSL to a probabilistic version, as reflected in the use of expectation values in SM, is appropriate.

Thus, I conclude that the Gibbs entropy can play the right role, since it increases in non-quasi-static processes but is constant in quasi-static processes. Furthermore, S_G is connected to heat in the right way, and the presence of expectation values is a feature, not a

[26]The idea that the TDSL is not a strict law was suggested by Maxwell: 'Hence the TDSL is continually being violated, and that to a considerable extent, in any sufficiently small group of molecules belonging to a real body ' Maxwell (1891) as quoted by Cercignani (1998).

[27]And if there were even an glimmer of hope that a greater-than-Carnot engine is possible, it would be a hive of research, since it would help us solve the energy crisis.

[28]There are more severe possible restrictions in its scope. For example, the Maxwellian view discussed by Myrvold (2011), restricts the TDSL to suitably 'large' systems. But see Linden *et al.* (2010) for a discussion of the smallest possible thermal systems. Here I leave aside the interesting questions about the size and type of systems that TD applies to.

bug.

7 Defending the Gibbs entropy

In philosophical circles, the Gibbs entropy far more unpopular than its cousin, the Boltzmann entropy.[29] The main objection to the Gibbs entropy is that it is a property of an *ensemble*, rather than an individual system. S_G is a function of the canonical distribution, commonly known as the canonical ensemble. In CSM, the canonical ensemble is a probability measure over the $6N$ phase space of possible states, which is understood to represent how many members of the imaginary infinite ensemble have that state. This breeds puzzlement. Why should an infinite ensemble—and, furthermore, one that is *imaginary*—be helpful? And how on earth is it connected to the individual system whose thermodynamic entropy can be measured in the laboratory?

But the 'infinite imaginary ensemble' can be demystified. It is just a vivid way to give the probabilities in SM a frequency interpretation. (For a canonical example of this frequentist understanding of SM probabilities, see Gibbs (1902, p. 5)). The probability of a given state is just the number of (imaginary) systems in that state. But there are many other positions in the philosophy of probability aside from frequentism. Thus, the canonical ensemble is just a probability distribution which needn't be given this imaginary ensemble interpretation.

As such, the ensemble worry is not strictly about ensembles, but rather about probability in SM. In particular, the concern is the S_G is not a property of the possessed microstate of the system but a property of a probability distribution over possible microstates. (However, to fit with the rest of the literature, I will continue to call this objection 'the ensemble worry', but in what follows 'ensemble property' is used interchangably with 'property of a probability distribution'.)

Why worry that S_G is a property of a probability distribution rather than a microstate? In

[29] I call them cousins, since they are related to one another. In particular, each can be derived from the other, despite conceptual differences (see Frigg and Werndl (2011) for more details on this, and Wallace (2018) for an argument that Boltzmannian SM is a special case of, rather than an alternative to, Gibbsian SM).

section 7.1, I defuse a common but ill-motivated answer: that there is a mismatch with S_{TD}. In section 7.2, I then give a better reason to be concerned: if S_G depends not only on the microstate, then in CSM, it depends on something else. I discuss how this can affect the status of S_G — in particular, S_G may consequently appear anthropocentric. But, as I will argue in section 7.3, the situation is radically different in the quantum setting: the 'ensemble vs individual' property problem does not even arise, and there is no reason to think that S_G is anthropocentric.

7.1 A bad objection: mismatches

Why should the 'ensemble nature' of the Gibbs entropy worry us? As Callender (2001) emphasises, S_{TD} is a feature of the individual system, and so S_G does not match S_{TD}. In contrast, the Boltzmann entropy $S_B = k_B ln\Omega$ *is* a property of the individual system. Thus, Boltzmannians (cf. Callender (1999; 2001), Goldstein and Lebowitz (2004), Frigg (2010)) claim that S_B is superior, since it is a function of the microstate of the system.[30]

As such, there is a mismatch between S_G and S_{TD}. Yet this mismatch is a bad reason to worry about the ensemble nature of S_G. Mismatches are not problematic solely in virtue of revealing differences between the higher and lower-level quantities. As discussed in section 2, the higher-level quantities X_t need not always exactly match the lower-level quantities X_b.[31] S_G is not bad merely in virtue of not matching S_{TD} exactly. According to functionalism, differences between quantities are not instantly a problem that blocks reduction. Provided S_G plays the role of S_{TD}, then other differences are tolerated. Such as, if 'being a property of the

[30]Note however that S_B is a *modal* property: it depends on the number of microstates within the macrostate partition, and as such it measures the number of microstates the system *could* have been in, but actually isn't, whilst still having the same macroproperties.

[31]Indeed, given the two concepts are embedded in distinct theories, some differences are to be expected. Two theories will inevitably employ different concepts. They are different theories, after all. Furthermore, in order to secure a reduction, the lower-level theories' quantities must only capture the relevant, or crucial, features of the higher-level theories quantities.

individual system' or 'being non-probabilistic' is not part of the essential role of S_{TD}, then the ensemble nature of the Gibbs entropy is not worrying.

As discussed in section 2, the realiser can differ in ways that do not affect its playing the functional role. Being an ensemble property doesn't seem to prevent S_G playing the S_{TD} role (for isolated systems, increasing in non-quasi-static processes, but remaining constant in quasi-static processes).

Of course, those who levy the ensemble objection against S_G can just reply that the functional role of S_{TD} is as I've defined in terms of quasi-static processes *plus* the requirement that it is a property of the individual system. However, I see no reason to alter the functional role in this way. The role I've defended required careful consideration of the nature of processes in TD, the minus first law and the types of irreversibility. Thus, the onus is on 'ensemble objector' to say why 'being a property of an individual system' is an integral part of TD in particular, rather than a general suspicion of mismatches and probabilities (which after all, form a large part of the scientific enterprise, even if they are philosophically contentious).

7.2 A better objection: the nature of probability in CSM

Indeed, it is the philosophical issues with probabilities that provide a better reason to be worried that the Gibbs entropy is a function of ρ, a probability distribution over possible microstates. In CSM, since S_G is not just a property of the microstate of the system, it depends on something extra outside of the system too. What this 'something extra' is depends on your interpretation of the CSM probabilities. In the case of the ensemble interpretation, S_G depends not only on the state of the individual system but also on the other members of the ensemble. Thus, S_G seems like a mysterious quantity. Of course, earlier I claimed that ρ needn't be given a frequentist interpretation in terms of an imagined ensemble. Shorn of this ensemble gloss, we might prefer a different view of probability — but none of the available options render S_G a full-blooded anthropocentrism-free quantity.

Jaynes, for instance, thought that ρ represented our ignorance of the system's exact microstate. Here, the probability distribution ρ depends not only on the state of the system but on our epistemic situation. If the probability distribution depends on our ignorance, then if we

were to learn the exact microstate of the system, we would assign probability 1 to this state — and, since $ln1 = 0$, the Gibbs entropy would vanish! Thus, it would seem the Gibbs entropy is to do with what is going on inside our heads — rather than a bona fide feature of reality independent of us.[32] On this interpretation, the Gibbs entropy is thoroughly anthropocentric; a mirage stemming our ignorance.

But whilst Jaynes' view is popular, it is important to flag that ρ needn't be given a subjective interpretation following Jaynes (1957), since CSM probabilities can be considered to be 'almost objective' following Myrvold (2012), whose work is in the spirit of the 'objectified credences' tradition, cf. Poincaré (1896). Here the dynamics play a crucial role by washing out differences in our initial credences, such that there is intersubjective agreement about the right probability distribution to assign to the system. Thus, unlike the Jaynesian view, on this interpretation of CSM probability, S_G is not just 'in our heads'. Yet moving from the actual microstate of the system (given by the underlying dynamical theory CM) to a probability distribution assigned by CSM, requires an additional ingredient, credence. Since this is a hybrid view that mixes epistemic and ontic considerations, a vestigial tail of anthropocentrism remains.

Naturally, interpreting probabilities in SM is a large project, especially justifying Gibbs phase averaging (Malament and Zabell 1980). Not only is the project large, it is also pressing given the indispensability of probabilities in SM (cf. Wallace (2015; 2018)). However, lack of space is not the only reason why I won't dwell further on the issues with probability in CSM here: the main reason is that understanding probability in SM in completely transformed in the (foundationally more important) quantum context (Wallace 2016). Crucially for our discussion, the ensemble objection does not even get off the ground in the QSM context.

7.3 Quantum of solace

In CSM, there is a gap between the possessed microstate of the system and ρ, which opens the door to claims that S_G is mysterious, or anthropocentric. But in QSM there is no such gap

[32] In this way, a Jaynesian view of SM probability seems incompatible with standard scientific realism that requires our scientific descriptions be mind-independent.

between the 'microstate' of the system and ρ — and thus, no room for ignorance, credence or anthropocentrism to sneak in. There is no analogous 'gap' in QSM because the underlying microdynamics, QM, is already probabilistic. Furthermore, as I will now argue, there is no distinction between the 'microstate' of the system and a probability distribution over these microstates: both are density matrices. And consequently, the distinction between a property of a probability distribution (an 'ensemble property') or a property of the individual system does not arise in the first place.

The density matrix is arguably the best mathematical object to represent the state of the individual system (rather than the wavefunction ψ), since $\hat{\rho}$ is a more general object than ψ. Quantum systems rapidly become entangled with their environment — which means that the individual system cannot be described by a wavefunction, but instead must be a (reduced) density matrix (by tracing over the environment). Since the density matrix formalism is more general, and sometimes *required*, the density matrix should be taken to be more fundamental. (See Wallace and Timpson (2010), Wallace (2011), Chen (2018) and Maroney for more on this point). Thus the individual state of the system in QM is not represented by a ray in Hilbert space (the quantum equivalent of a point in phase space), but a density matrix.

A probability distribution over these 'fundamental microstates' of QM, density matrices, just gives...another density matrix! Furthermore, we should not be misled: ρ is not straightforwardly a probability distribution over states, one of which the system is 'really in', because ρ is degenerate: the same ρ can represent distinct probability distributions over different (even incompatible) pure states, see Hughes (1989). (See Popescu *et al.* (2006) for a discussion of how the density matrix ρ_{can} representing the canonically distributed state can be derived — free from any claims about ignorance). Thus, there is no difference in the mathematical object that represents the state of the individual system, and a probability distribution over it. Thus, in QSM, the dichotomy between 'being a property of a probability distribution' and 'being a property of the individual system' never arises.

Whilst this removes the dichotomy upon which the 'ensemble worry' about S_G rests, insofar as this topic stemmed from the mystery mongering about probability in SM, there is bad news. Understanding the nature of probability in QSM any further involves tangling with

the quantum measurement problem, since the status of probability in QM is interpretation-dependent. And so in this way, we are out of the frying pan but into the fire.

One consolation: taking QSM rather than CSM as the conceptual starting point not only defuses the ensemble worry, it also removes one of Sklar (1999)'s concerns about the reduction of TD. In the classical case, probabilities are a new conceptual ingredient that have to be added to the microdynamics in order to construct CSM and so find the regularities of TD. Sklar is concerned that probability is a new, autonomous posit[33], and so may spell trouble for reduction because too much has been bundled into the bridge laws.[34] Regardless of whether we share Sklar's worry that we may be helping ourselves to 'too much', note that the problem does not arise in QSM. Probability is already inherent in the 'microdynamics', and so is not a new ingredient (cf. (Wallace 2016, p. 6)).

To sum up: Gibbsian SM can be shorn of the ensemble metaphor, which just indicates a frequentist interpretation of the probability in SM. But the popular Jaynesian alternative makes probabilities a reflection of our ignorance, and consequently endangers taking S_G to be subjective. Even objective interpretations of CSM probabilities create a distance between being a possessed property of the system, and a property of a probability distribution. This problem does not arise in QSM, since a density matrix such as ρ_{can} can be considered the fundamental 'microstate' description of the system, and thus the 'ensemble vs individual system' objection does not get off the ground in QSM.

[33] 'It is, in fact, the status of these probabilistic assumptions, central to the theory and *possibly not importable into it from other physical theories*, that is the most problematic element when one asks whether we ought to claim a reductive relationship between thermodynamics and statistical mechanics' (Sklar 1999, p. 190) emphasis added.

[34] Sklar is concerned that if we are too liberal with what is allowed in a bridge law, the reduction is trivialised. However see Uffink (1996) for a robust reply to this issue on bridge laws. Briefly, in practice no matter how many conceptual resources we help ourselves to performing a reduction in a particular case study is tricky — and far from trivial!

8 Conclusion

Because of the different concepts — most importantly heat and work — in thermodynamics, finding a statistical mechanical correlate to the classic formulations of the second law is not straightforward. The traditional approach is to try to find a non-decreasing entropy function: what Callender dubs the search for the holy grail. But I argued that this holy grail is too weak in some respects, and in other respects too strong: it does not capture the functional role of S_{TD}. Instead we need the new grail: an SM entropy function that, for thermally isolated systems, is constant during quasi-static processes and increasing in non-quasi-static processes.

To find the new grail, I took a Gibbsian approach. By using Ehrenfest's principle, for thermally isolated systems, we found that the Gibbs entropy is constant during a quasi-static process, but increases during a non-quasi-static process. Thus, I argued that the Gibbs entropy plays the requisite role.

I then defended the Gibbs entropy against the objection that it is an 'ensemble property' rather than a property of an individual system. The functionalist strategy allows the theory being reduced to differ (to an extent) from its realiser (reductive base): thus, this mismatch between S_{TD} and S_G is not a problem solely in virtue of being a mismatch. Furthermore, when we consider the more fundamental theory QSM, rather than CSM, we see that the dichotomy between being a property of an ensemble or an individual system never arises, thus removing this main objection to S_G.

Acknowledgements

My thanks to Carina Prunkl, Al Wilson, Erik Curiel and two anonymous referees for helpful comments on an earlier draft. This work forms part of the project A Framework for Metaphysical Explanation in Physics (FraMEPhys), which received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 757295).

Katie Robertson

Department of Philosophy

University of Birmingham

UK

krobertson0706@gmail.com

References

Albert, D. Z. [2000]: *Time and chance*, Cambridge, Mass: Harvard University Press.

Ardourel, V. [2018]: 'The infinite limit as an eliminable approximation for phase transitions', *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, **2**, pp. 71–84.

Arnold, V. I. [2010]: *Mathematical methods of classical mechanics*, vol. 60 of *Graduate Texts in Mathematics* New York: Springer Science & Business Media, 2nd edition edition.

Atkins, P. [2007]: *Four laws that drive the universe*, Oxford: Oxford University Press.

Baierlein, R. [1971]: *Atoms and information theory: An introduction to statistical mechanics*, San Francisco: W. H. Freeman.

Baker, D. [2018]: 'On Spacetime Functionalism', Http://philsci-archive.pitt.edu/14301/.

Batterman, R. [2001]: *The devil in the details: Asymptotic reasoning in explanation, reduction, and emergence*, Oxford: Oxford University Press.

Batterman, R. [2010]: 'Reduction and renormalization', in A. Hütteman (*ed.*), *Time, Chance, and Reduction: Philosophical Aspects of Statistical Mechanics*, Cambridge: Cambridge University Press, pp. 159–179.

Batterman, R. W. [1995]: 'Theories between theories: Asymptotic limiting intertheoretic relations', *Synthese*, **103**(2), pp. 171–201.

Blundell, S. J. and Blundell, K. M. [2009]: *Concepts in thermal physics*, Oxford: Oxford University Press.

Brown, H. R., Myrvold, W. and Uffink, J. [2009]: 'Boltzmann's H-theorem, its discontents, and the birth of statistical mechanics', *Studies In History and Philosophy of Science Part B: Studies In History and Philosophy of Modern Physics*, **40**(2), pp. 174–191.

Brown, H. R. and Uffink, J. [2001]: 'The origins of time-asymmetry in thermodynamics: The minus first law', *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, **32**(4), pp. 525–538.

Butterfield, J. [2011a]: 'Emergence, reduction and supervenience: a varied landscape', *Foundations of Physics*, **41**(6), pp. 920–959.

Butterfield, J. [2011b]: 'Less is different: emergence and reduction reconciled', *Foundations of Physics*, **41**(6), pp. 1065–1135.

Callender, C. [1999]: 'Reducing thermodynamics to statistical mechanics: the case of entropy', *The Journal of Philosophy*, **97**(7), pp. 348–373.

Callender, C. [2001]: 'Taking thermodynamics too seriously', *Studies In History and Philosophy of Science Part B: Studies In History and Philosophy of Modern Physics*, **32**(4), pp. 539–553.

Cercignani, C. [1998]: *Ludwig Boltzmann: the man who trusted atoms*, Oxford: Oxford University Press.

Chen, E. K. [2018]: 'Quantum mechanics in a time-asymmetric universe: On the nature of the initial quantum state', https://doi.org/10.1093/bjps/axy068.

Clausius, R. [1879]: *The Mechanical Theory of Heat*, Cambridge: Cambridge University Press.

Denbigh, K. and Denbigh, J. [1985]: *Entropy in relation to incomplete knowledge*, Cambridge: Cambridge University Press.

Dennett, D. [2001]: 'Are we explaining consciousness yet?', *Cognition*, **79**, pp. 221–237.

Duhem, P. M. M. [1902]: *Termodynamique et chimie: leçons élémentaires*, Paris: A. Hermann & Fils.

Earman, J. and Rédei, M. [1996]: 'Why ergodic theory does not explain the success of equilibrium statistical mechanics.', *The British Journal for the Philosophy of Science*, **47**, pp. 63–78.

Ehrenfest-Afanassjewa, T. [1925]: 'Zur Axiomatisierung des zweiten Hauptsatzes der Thermodynamik', *Zeitschrift für Physik*, **33**(34), pp. 933–945.

Ehrenfest-Afanassjewa, T. [1956]: *Die Grundlagen der Thermodynamik.*, Leiden: E. J.Brill.

Frigg, R. [2010]: 'A field guide to recent work on the foundations of statistical mechanics', in D. Rickles (*ed.*), *The Ashgate companion to contemporary Philosophy of Physics*, Aldershot: Ashgate.

Frigg, R. and Werndl, C. [2011]: 'Entropy- a guide for the perplexed', in S. H. C. Beisbart (*ed.*), *Probabilities in physics*, Oxford: Oxford University Press.

Frisch, M. [2014]: *Causal reasoning in physics*, Cambridge: Cambridge University Press.

Gibbs, J. W. [1902]: *Elementary principles in statistical mechanics*, New York: Dover (1960).

Goldstein, S. and Lebowitz, J. L. [2004]: 'On the (Boltzmann) Entropy of non-equilibrium systems', *Physica D*, pp. 53–66.

Griffiths, D. J. and Schroeter, D. F. [2018]: *Introduction to quantum mechanics*, Cambridge: Cambridge University Press.

Grünbaum, A. [1973]: 'Is the coarse-grained entropy of classical statistical mechanics an anthropomorphism?', in *Philosophical problems of space and time*, Netherlands: Springer, pp. 646–665.

Hertz, P. [1910)]: 'Über die mechanischen Grundlagen der Thermodynamik', *Annalen der Physik*, **33**, pp. 225–274, 537–552.

Horodecki, M. and Oppenheim, J. [2013]: '(Quantumness in the context of) Resource Theories.', *International Journal of Modern Physics B*, **27**.

Hughes, R. I. [1989]: *The structure and interpretation of quantum mechanics*, Cambridge, Mass: Harvard University Press.

Jancel, R. [2013]: *Foundations of Classical and Quantum Statistical Mechanics: International Series of Monographs in Natural Philosophy*, vol. 19 Amsterdam: Elsevier.

Jaynes, E. T. [1957]: 'Information theory and statistical mechanics', *Physical review*, **106**(4), pp. 620.

Katz, A. [1967]: *Principles of statistical mechanics: the information theory approach*, New York: W. H. Freeman.

Kelvin, W. T. B. [1882]: *Mathematical and Physical Papers: By Sir William Thomson*, Cambridge: At the University Press.

Kim, J. [1998]: *Mind in a physical world: essays on the mind-body problem and mental causation*, Cambridge, Mass: MIT press.

Kim, J. [1999]: 'Making sense of emergence', *Philosophical studies*, **95**(1-2), pp. 3–36.

Kirchhoff, G. [1894]: *Vorlesungen über die theorie der waerme*, Leipzig: Teubner.

Knox, E. [2013]: 'Effective spacetime geometry.', *Studies in History and Philosophy of Modern Physics*, **3**(44), pp. 346–356.

Knox, E. [2016]: 'Abstraction and its Limits: Finding Space For Novel Explanation', *Noûs*, **50**(1), pp. 41–60.

Knox, E. [2019]: 'Physical Relativity from a Functionalist Perspective.', *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, **67**, pp. 118–124.

Lam, C., V. Wüthrich [2018]: 'Spacetime is as spacetime does', *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, **64**, pp. 39–51 ArXiv:1803.04374v2.

Landau, L. D. and Lifshitz, E. M. [1964]: *Quantum mechanics: non-relativistic theory*, vol. 3 London: Pergamon.

Lavis, D. A. [2018]: 'The Problem of Equilibrium Processes in Thermodynamics', *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, **62**, pp. 136–144.

Levin, J. [2018]: 'Functionalism', in E. N. Zalta (*ed.*), *The Stanford Encyclopedia of Philosophy*, Metaphysics Research Lab, Stanford University, fall 2018 edition.

Lewis, D. K. [1970]: 'How to define theoretical terms', *The Journal of Philosophy*, **67**(13), pp. 427–446.

Linden, N., Popescu, S. and Skrzypczyk, P. [2010]: 'How Small Can Thermal Machines Be? The Smallest Possible Refrigerator', *Physical Review Letters*, **105**(130401).

Luczak, J. [2018]: 'How Many Aims Are We Aiming At?', *Analysis*, **78**(2), pp. 244–254.

Malament, D. B. and Zabell, S. L. [1980]: 'Why Gibbs phase averages work–the role of ergodic theory', *Philosophy of Science*, **47**(3), pp. 339–349.

Maroney, O. [2007]: 'The physical basis of the Gibbs-von Neumann entropy', *arXiv preprint quant-ph/0701127*.

Masanes, L. and Oppenheim, J. [2017]: 'A general derviation and quantification of the Third law of thermodynamics', *Nature communications*, **8**(1), pp. 14538.

Maxwell, J. C. [1878]: 'Diffusion', in *Encyclopedia Britannica (Nineth Ed.)*, vol. 7, Cambridge: Cambridge University Press, pp. 214–221.

Maxwell, J. C. [1891]: *Theory of heat*, New York: Longmans, Green.

Messiah, A. [1962]: *Quantum mechanics, Vol. II*, New York: Wiley.

Myrvold, W. [2014]: 'Probabilities in statistical mechanics', in C. Hitchcock and A. Hajek (*eds*), *Oxford Handbook of Probability and Philosophy*, Oxford: Oxford University Press.

Myrvold, W. C. [2011]: 'Statistical mechanics and thermodynamics: A Maxwellian view', *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, **42**(4), pp. 237–243.

Myrvold, W. C. [2012]: 'Deterministic laws and epistemic chances', in Y. Ben-Menahem and M. Hemmo (*eds*), *Probability in Physics*, New York: Springer, pp. 73–85.

Albert, D. Z. [2013]: 'Wave Function Realism', in Ney, A. and Albert, D. Z (*eds*), *The Wave Function: Essays in Metaphysics of Quantum Mechanics.*, New York: Oxford University Press, pp. 52–57.

Niven, W. (*ed.*) [1965]: *The Scientific Papers of James Clerk Maxwell*,, vol. 2 New York: Dover Publications, reprint of CUP edition of 1890 edition.

Norton, J. D. [2009]: 'Is there an independent principle of causality in physics?', *The British Journal for the Philosophy of Science*, **60**(3), pp. 475–486.

Norton, J. D. [2016]: 'The impossible process: thermodynamic reversibility', *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, **55**, pp. 43–61.

Palacios, P. [2018]: 'Had we but world enough, and time... but we don't!: Justifying the thermodynamic and infinite-time limits in statistical mechanics', *Foundations of Physics*, **48**, pp. 526–541.

Palacios, P. [2019]: 'Phase Transitions: A Challenge for Intertheoretic Reduction?', *Philosophy of Science*, **86**(4), pp. 612–640.

Poincaré, H. [1896]: *Calcul de Probabilities*, Paris: Gauthier-Villars.

Popescu, S., Short, A. and Winter, A. [2006]: 'The foundations of statistical mechanics from entanglement: Individual states vs. averages. eprint', *Nature Physics*, **2**, pp. 754–758.

Prigogine, I. [1980]: *From being to becoming: time and complexity in the physical sciences*, San Francisco: W. H. Freeman.

Prigogine, I. and Stengers, I. [1984]: *Order out of chaos: Man's new dialogue with nature*, London: Flamingo.

Prunkl, C. [2018]: 'The Road to Quantum Thermodynamics', in D. B. C. Timpson (*ed.*), *Quantum Foundations of Statistical Mechanics*, Oxford: Oxford University Press.

Read, J. and Menon, T. [2019]: 'The limitations of inertial frame spacetime functionalism', *Synthese*, **https://doi.org/10.1007/s11229-019-02299-2**.

Redhead, M. [1996]: *From physics to metaphysics*, Cambridge: Cambridge University Press.

Reichenbach, H. [1956]: *The direction of time*, Berkeley, California: University of California Press.

Robertson, K. [2019]: 'Asymmetry, abstraction and autonomy: justifying coarse-graining in statistical mechanics', https://doi.org/10.1093/bjps/axy020 Forthcoming in The British Journal for the Philosophy of Science.

Rosaler, J. [2019]: 'Reduction as an a posteriori relation', *The British Journal for the Philosophy of Science*, (1), pp. 269–299.

Rueger, A. [2006]: 'Functional reduction and emergence in the physical sciences', *Synthese*, **151**(3), pp. 335–346.

Rugh, H. H. [2001]: 'Microthermodynamic formalism', *Physical Review E*, **64**(5), pp. 055101.

Russell, B. [1913]: 'On the notion of cause', *Proceedings of the Aristotelian society*, **13**, pp. 1–26.

Sakurai, J. J. and Commins, E. D. [1995]: *Modern quantum mechanics, revised edition*, San Francisco, California.: Addison-Wesley.

Sklar, L. [1993]: *Physics and chance: Philosophical issues in the foundations of statistical mechanics*, Cambridge: Cambridge University Press.

Sklar, L. [1999]: 'The reduction (?) of thermodynamics to statistical mechanics.', *Philosophical Studies*, **95**(1), pp. 187–202.

Snow, C. [1959]: *The Two Cultures: And A Second Look.*, Cambridge: Cambridge University Press, 2nd 1964 edition.

Tong, D. [2012]: 'Statistical Physics', Cambridge Lecture Notes, available at http://www.damtp.cam.ac.uk/user/tong/statphys.html.

Uffink, J. [1996]: 'Nought but molecules in motion', *Studies in History and Philosophy of Modern Physics*, **27**(3), pp. 373–387.

Uffink, J. [2001]: 'Bluff your way in the second law of thermodynamics', *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, **32**(3), pp. 305–394.

Uffink, J. [2006]: 'Compendium of the Foundations of Classical Statistical Physics,', in J. Butterfield and J. Earman (*eds*), *Handbook for Philosophy of Physics*, Amsterdam: Elsevier, pp. 924–1074.

Uffink, J. [2013]: 'Three concepts of irreversibility and three versions of the second law', in M. S. F. Stadler (*ed.*), *Time and History. Zeit und Geschichte*, vol. 1, Frankfurt: Ontos Verlag, pp. 275–287.

Valente, G. [2017]: 'On the Paradox of Reversible Processes in Thermodynamics', *Synthese*, **196**, pp. 1761–1781.

Wald [1997]: 'The "Nernst Theorem" and black hole thermodynamics', *Phys. Rev. D*, **56**(10), pp. 6467–6474.

Wallace, D. [2011]: 'The logic of the past hypothesis', http://philsci-archive.pitt.edu/8894/.

Wallace, D. [2012]: *The emergent multiverse: Quantum theory according to the Everett interpretation*, Oxford: Oxford University Press.

Wallace, D. [2014]: 'Thermodynamics as Control Theory', *Entropy*, **16**(2), pp. 699–725.

Wallace, D. [2015]: 'The quantitative content of statistical mechanics.', *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, **52**, pp. 285–293.

Wallace, D. [2016]: 'Probability and irreversibility in modern statistical mechanics: Classical and quantum', in D. Bedingham, O. Maroney and C. Timpson (*eds*), *Quantum Foundations of Statistical Mechanics*, Oxford: Oxford University Press.

Wallace, D. [2018]: 'The Necessity of Gibbsian Statistical Mechanics', Http://philsci-archive.pitt.edu/15290/.

Wallace, D. and Timpson, C. [2010]: 'Quantum Mechanics on Spacetime I: Spacetime State Realism.', *The British Journal for the Philosophy of Science*, **61**(4), pp. 697–727.

Werndl, C. and Frigg, R. [2015a]: 'Reconceptualising equilibrium in Boltzmannian statistical mechanics and characterising its existence', *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, **49**, pp. 19–31.

Werndl, C. and Frigg, R. [2015b]: 'Rethinking boltzmannian equilibrium', *Philosophy of Science*, **82**(5), pp. 1224–1235.

Wilson, M. [1985]: 'WHAT IS THIS THING CALLED "PAIN"?—THE PHILOSOPHY OF SCIENCE BEHIND THE CONTEMPORARY DEBATE', *Pacific philosophical quarterly*, **66**(3-4), pp. 227–267.

Zeh, H. D. [2007]: *The physical basis of the direction of time*, Berlin: Springer, 5th edition.

Zwanzig, R. [1960]: 'Ensemble method in the theory of irreversibility', *The Journal of Chemical Physics*, **33**(5), pp. 1338–1341.