

# Can the second law be compatible with time reversal invariant dynamics?

Leah Henderson

April 9, 2014

## Abstract

It is commonly thought that there is some tension between the second law of thermodynamics and the time reversal invariance of the microdynamics. Recently, however, Jos Uffink has argued that the origin of time reversal non-invariance in thermodynamics is not in the second law. Uffink argues that the relationship between the second law and time reversal invariance depends on the formulation of the second law. He claims that a recent version of the second law due to Lieb and Yngvason allows irreversible processes, yet is time reversal invariant.

In this paper, I attempt to spell out the traditional argument for incompatibility between the second law and time reversal invariant dynamics, making the assumptions on which it depends explicit. I argue that this argument does not vary with different versions of the second law and can be formulated for Lieb and Yngvason's version as for other versions. Uffink's argument regarding time reversal invariance in Lieb and Yngvason is based on a certain symmetry of some of their axioms. However, these axioms do not constitute the full expression of the second law in their system.

## 1 Introduction

It is a basic observation that many processes spontaneously occur in one direction, but not in reverse. For example, consider a gas confined to one side of a partitioned box. When the partition is removed, it spontaneously expands to fill the new available space. Yet the reverse does not occur. A gas in a box does not spontaneously contract to occupy a smaller volume. Similarly, ice melts in water, but ice does not spontaneously form out of water. Liquids mix together, but they do not spontaneously unmix themselves. In cases like these, it is possible to arrange for the reverse process to occur, by doing suitable manipulations on the system, but it does not happen on its own. The second law of thermodynamics is generally thought to codify the observation that macroscopic processes spontaneously proceed in one direction and to explain this in terms of permissible entropy changes.

Yet, the fundamental processes described by our theories of the microdynamics appear to be reversible in time, so where does the thermodynamic irreversibility expressed by the second law come from in a world where the microdynamics are time reversal invariant? This is a longstanding and notorious puzzle for the second law.

In recent work, Jos Uffink has argued that the origin of time reversal non-invariance in thermodynamics is not in the the second law [Uffink(2001)]. He argues that there are many subtle variations in what different versions of the the second law actually express in terms of time reversal

non-invariance and irreversibility. In some versions, he suggests, the second law expresses the irreversibility of natural processes. ‘But’, he says, ‘in many other formulations irreversibility or even time reversal non-invariance plays no role’ ([Uffink(2001)], p. 305). In particular, Uffink claims that the recent version of the second law due to Lieb and Yngvason [Lieb and Yngvason(1999)] is actually time reversal invariant. He therefore suggests that ‘it may be more fruitful to abandon the idea that time reversal non-invariance or irreversibility is essential to the second law’ (p. 94). One supposed advantage of doing so will be to defuse the tension between the second law and the dynamical theory:

What, then, do we gain with this proposal? The main advantage is, to my mind, that the second law would no longer represent an obstacle to the reconciliation of different theories of physics. More specifically, attempt to reduce thermodynamics to, or at least to harmonise it with, a mechanistic world picture would get a new lease of life. ([Uffink(2001)], p. 387).

This latter quotation suggests that Uffink’s claim about the lack of essential time reversal non-invariance in the second law is supposed to bear in some way upon the traditional puzzle of the relationship between second law and the time reversal invariant dynamics. My purpose in this paper is to examine the traditional puzzle on the one hand and Uffink’s claims on the other, and to explore the relationship between the two. My paper will therefore fall into two parts.

In the first, I will try to capture the traditional puzzle in a simple three-premise argument. A key concept in this argument is that of thermodynamic reversibility. Uffink claims that the concept of reversibility varies between versions of the second law. However, I will argue that there is a basic sense of reversibility which is common to the different versions of the second law. And thus, the traditional puzzle takes essentially the same form for different versions of the law.

In the second part of the paper, I will also examine Uffink’s arguments, particularly his claim that Lieb and Yngvason’s version of the second law is time reversal invariant. We will see that this claim is based on a certain symmetry of Lieb and Yngvason’s axioms. However I will argue that the significance of this symmetry is not that we have a time reversal invariant version of the second law. Rather it is that the subset of Lieb and Yngvason’s axioms under consideration do not constitute their full expression of the second law. There are other axioms in their system which play a crucial role. I will argue that the basic puzzle about the relationship between the second law and time reversal invariant dynamics, sketched in the first part of the paper, remains essentially untouched by Uffink’s arguments.

## 2 The traditional puzzle: time reversal invariant dynamics and the second law

In this section I will offer a formulation of the traditional puzzle of the tension between the second law and time reversal invariant dynamics. First, let us define the notion of time reversal invariance. This notion is usually applied to the fundamental microdynamical theories in physics. We take a ‘process’ to be a time-indexed sequence of instantaneous micro-states of a system  $P : \{s_i, \dots, s_f\}$ . In classical physics, such states would be points in phase space. The theory is said to be time reversal invariant if, for any process  $P$  allowed by the theory, the time reversed process  $P^T$  is also allowed. Here the time reversal of a process  $P : \{s_i, \dots, s_f\}$ , is  $P^T : \{Rs_f, \dots, Rs_i\}$ , the process found by first

applying a time reversal operator  $R$  to each instantaneous state in the sequence, and then reversing the order of the states.

As noted in the introduction, some thermodynamic processes, such as the free expansion of a gas, are never seen running backwards. However, we have the sense that time reversing the microdynamics behind a macroscopic process would produce a macroscopic reversal of that process. If so, we have a puzzle as to why we do not observe reversal of some processes, given that there seems to be no obstacle to reversal from the fundamental dynamical point of view.

Uffink raises the question of how the second law enters into this basic puzzle. According to the usual way of thinking, it is the second law which encodes our observations of irreversibility, and thus it is the second law which is in tension with time reversal invariance. But what the second law is generally thought to prohibit is reversal of certain processes in a quite specific, thermodynamic sense. This will produce tension with time reversal invariance if one assumes that time reversal produces thermodynamic reversal. Thus, the tension between time reversal invariant dynamics and the second law can be captured by an argument with the following premises:

1. The fundamental dynamics is time reversal invariant.
2. Suppose process  $P$  defined at the micro-level realises a thermodynamic process  $M$ . Then the time reversal  $P^T$  realises the thermodynamic reverse process  $M^*$ .
3. There exist macroscopic processes  $M$  such that the second law says that  $M^*$  is not possible.

The argument is then the following. For any thermodynamic process  $M$ , take a dynamical process  $P$  which realises it. By (1), the time reversed process  $P^T$  is possible. By (2), the thermodynamic reverse process  $M^*$  is possible. But (3) says that for some  $M$ ,  $M^*$  is not possible according to the second law. Thus, for such processes, we have an outright contradiction. I will call this Argument I.

At this point the argument is quite schematic, and it will be necessary to be more precise about some of the concepts involved, particularly the concept of a thermodynamic process. Clearly a central concept in this argument is thermodynamic reversal. Uffink suggests that this is a concept that varies with different versions of the second law. If this is true, then we might anticipate that Argument I could also vary according to the version. In the following section, I will outline the traditional approach to the second law which I will call the ‘engineering approach’ In section 4 I will examine the form that Argument I takes in the engineering approach. In section 5, I give a brief description of the axiomatic tradition to which Lieb and Yngvason’s formulation belongs, and in section 6 consider Argument I in this approach. I will argue that the basic notion of thermodynamic reversibility does not fundamentally differ between these versions of the law. Furthermore, I will show that the traditional puzzle can be expressed by Argument I, for the Lieb and Yngvason version, just as for the traditional engineering version of the second law.

### 3 The engineering approach to the second law

The engineering approach takes as its starting point a principle codifying the limitations which have been found on engineering transformations of energy. The classic formulation is the Kelvin principle.

It is impossible to construct an engine that, operating in a cycle, will produce no effect

other than to extract heat from a heat reservoir and convert it to work.<sup>1</sup>

From this principle, a series of arguments lead to expression of the second law in the following form:

*2nd Law –Engineering version:*

1) There is a function of state, the entropy  $S$ , and the entropy change between two states  $X_i$  and  $X_f$  is defined as:

$$\Delta S = S(X_f) - S(X_i) = \int_i^f \frac{dQ_{rev}}{T} \quad (1)$$

where  $dQ_{rev}$  are the increments of heat supplied to the system in a reversible process between  $X_i$  and  $X_f$ , and  $T$  is the temperature of the reservoir supplying that increment of heat.

2) For *any* spontaneous process from  $X_i$  to  $X_f$

$$\int_i^f \frac{dQ}{T} \leq \Delta S \quad (2)$$

where  $dQ_i$  are increments of heat supplied to the system from a reservoir at temperature  $T$ .

The two parts of this law follow from two basic consequences of the Kelvin principle. The first consequence is that all reversible engines operating in a cycle between the same heat reservoirs have the same efficiency. In the proof of this consequence, an engine is taken to be reversible if it can run backwards with the directions of heat and work reversed – that is, if heat flows into the engine, then in the reverse, the same amount of heat flows out, and if work is done by the system, in the reverse, work is done on the system (and vice versa). This first consequence of the Kelvin principle means that the efficiency depends only on the temperature of the heat reservoirs between which the engine operates, and this allows a scale of thermodynamic temperature to be established. Then it can be shown that

$$\oint \frac{dQ_{rev}}{T} = 0 \quad (3)$$

From this, the first part of the second law follows. Here, the notion of ‘reversible’ as applied to a process is essentially just a generalisation of the notion applied to an engine operating in a cycle. To specify an arbitrary thermodynamic process  $M$  we need to know not just the initial and final equilibrium macrostates, but also the energy resources exchanged between the system and its surroundings: the heat flows into and out of the system, and the work which is done on or by the system. We can define the thermodynamic reverse of the process in the following way:

---

<sup>1</sup>Subject to certain assumptions, the formulation of the second law as Clausius’ principle can be shown to be equivalent. Clausius’ principle states

It is impossible to construct a device that, operating in a cycle, produces no effect other than the transfer of heat from a colder to a hotter body.

**Definition 1** The ‘thermodynamic reverse’ of process  $M$ , denoted by  $M^*$

a) restores the initial thermodynamic state of the original process from the final state, and  
b) reverses the resources required in terms of heat and work. That is, if in the original process, heat flowed into the system from the environment, in the reverse, it flows out of the system. Similarly, if work was done on the system by the environment, in the reverse process, the same amount of work was done by the system on the environment.

**Definition 2** A thermodynamic process is ‘reversible’ if its thermodynamic reverse is possible.

In order to be reversible, a process needs to be carried out under certain ideal conditions – in particular, the system must remain infinitesimally close to equilibrium at all times<sup>2</sup>.

The second basic consequence of the Kelvin principle is that the most efficient engine is one which is reversible. This gives rise to the Clausius inequality expressed in the second part of the second law.

The second law can also be stated in terms of adiabatic processes in which there is no heat exchange with the environment of the system ( $dQ = 0$ ). It then states that:

*2nd Law – Engineering version (stated in terms of adiabatic processes):*

For all adiabatic processes,  $\Delta S \geq 0$ . For processes that can occur adiabatically and reversibly, the entropy of the final state is equal to that of the initial state,  $\Delta S = 0$ . For processes which cannot be reversed adiabatically, the entropy of the final state is greater than that of the initial state,  $\Delta S > 0$ .

That is, there are surfaces in the space of states on which we can only pass from one state to another via an adiabatic reversible process. On these reversible adiabatic surfaces, the entropy is constant. An irreversible adiabatic process that moves off the surface can only proceed in the direction of increasing entropy.

## 4 Argument I for the engineering version of the second law

Returning to Argument I, premise (2) says that time reversing the microscopic process that realises a macroscopic process produces thermodynamic reversal of that macroscopic process, where thermodynamic reversal is defined as in Definition 1. This premise can be seen as supported by the following two assumptions:

i) The time reversal operator does not affect the macrostate. That is, the time reversal operator acts as the identity:  $\forall X, RX = X$ .

ii) Reversing the order of the states produces a process in which the directions of heat and work are reversed. In the case where the process is adiabatic, reversing the order of states produces another adiabatic process.

It follows from premises (1) and (2) of Argument I that all macroscopic processes can be thermodynamically reversed. This is what is denied by premise (3), which says that there exist macroscopic processes which cannot be thermodynamically reversed according to the second law. For the engineering version of the second law, there are processes allowed by the law for which the second law says the thermodynamic reverse is not possible. The engineering version says that for all adiabatic

---

<sup>2</sup>This condition for reversibility is sometimes called ‘quasi-static’, though in the literature, there are slightly different ways of defining quasi-static. Sometimes the term is used simply to mean ‘very slow’ eg. [Kestin(1966)]. For discussion of the historical development of this condition, the reader is referred to Uffink’s paper [Uffink(2001)].

processes,  $\Delta S \geq 0$ . Thus, there are processes – namely adiabatic ones with  $\Delta S > 0$  – which are permitted by the law but which the law says cannot be reversed (because the reversal would be adiabatic processes with  $\Delta S < 0$ ). Notice that premise (3) effectively adds the claim that some of these irreversible processes exist (in nature), as well as being allowed by the theory. However this existence claim is uncontroversial. If it were not so, our world would be such that all processes are thermodynamically reversible. Traditional versions of the second law are based on the idea, that comes from the Kelvin principle, that the reversible engine provides a limit on efficiency. It is logically compatible with that limit that all engines are reversible, just as it is logically compatible with everyone obeying a 100km/hr speed limit that everyone drives at exactly 100km/hr. However, this is not the world in which we live.

Thus, Argument I for the engineering version expresses the puzzle over the tension between time reversal invariance and the second law. To evade the tension, one would have to find grounds to reject one of the three premises. My concern here is not with how to resolve the puzzle, but rather with whether it takes a different form for different versions of the second law. In the next section, I turn to an alternative tradition to the engineering approach, which I will call the ‘axiomatic tradition’.

## 5 The axiomatic tradition

Some scientists in the history of thermodynamics have been unhappy with taking Kelvin’s principle as a starting point. Rather they have tried to lay down axioms which directly characterise the processes which can occur adiabatically. Historically, this tradition takes its lead from the work of Caratheodory [Caratheodory(1976)]. He took the following principle as a postulate:

In every open neighbourhood  $U_X$  of an arbitrarily chosen state  $X$  there are states  $Y$  such that for some open neighbourhood  $U_Y$  of  $Y$ : all states  $Z$  within  $U_Y$  cannot be reached adiabatically from  $X$ .

From this postulate, he deduced the existence of reversible adiabatic surfaces, and of a function of state  $S$  which is constant on each surface, for which  $dQ = TdS$  in a reversible process. Caratheodory then added an appeal to ‘experience’ in order to tell us which side of a surface can be reached by irreversible processes – namely the side of increasing entropy ([Caratheodory(1976)], p. 249).

In work following Caratheodory, the Caratheodory principle has been replaced with different sets of axioms about adiabatic processes. Some of this work has continued in Caratheodory’s topological vein (eg. [Boyling(1972)]). However, another approach is to be found in the work of Robin Giles, and subsequent authors, including Lieb and Yngvason [Gil(1964), Roberts and Luce(1968), Duistermaat(1968), Lieb and Yngvason(1999)]. In this approach, the problem of constructing the entropy becomes an instance of the problem of extensive measurement, that is of constructing a real-valued scale for a set of objects  $A$  on the basis of axioms satisfied by a triple  $\langle A, \prec, + \rangle$ , where  $\prec$  is a qualitative comparison relation and  $+$  is an addition operation. The aim then is to prove representation theorems which show how a structure  $\langle A, \prec, + \rangle$  satisfying certain axioms can be represented by a numerical structure which serves as a scale.

For a simple example of such a representation theorem, consider a domain of objects  $A = \{a, b, c, \dots\}$ , for which we can make qualitative comparisons such as ‘ $b$  is longer than  $a$ ’ which we denote  $a \prec b$ . Suppose we may also concatenate objects  $a$  and  $b$  to give object  $c$ , which we denote  $a + b = c$ . Under certain assumptions, these abilities allow us to construct a quantitative scale

for length, based on making qualitative comparisons of  $n$  copies of  $a$  with  $m$  copies of  $b$ . Other quantities than length, such as mass, duration, energy, etc. can also be constructed on similar foundations, as can the utility function in decision theory which represents qualitative preferences.

In the work of Lieb and Yngvason, the objects to compare are thermodynamic states,  $X, Y, Z...$  in a state space  $\Gamma$ . ‘Addition’ of two states  $X_1 \in \Gamma_1$  and  $X_2 \in \Gamma_2$  is given by the Cartesian product  $(X_1, X_2) \in \Gamma_1 \times \Gamma_2$ . Physically this means that the two systems are placed side by side and regarded as subsystems of one composite system. The relation  $\prec$  represents the possibility of an adiabatic process between states, which can be called ‘adiabatic accessibility’.

It is characteristic of work in the axiomatic tradition that explicit reference to heat as a primitive in the theory is avoided. Thus, Lieb and Yngvason, like Caratheodory and Giles, do not define adiabatic processes as those that involve no heat flow. They take a process to be adiabatic if it can be accomplished using interactions with auxiliaries in the environment of the system in which the only change is the possible lowering or raising of a weight. This condition would coincide with the usual definition of adiabatic if it excludes any changes in the environment of the system which would have happened had heat flowed between the environment and the system, and this is clearly the intention ([Lieb and Yngvason(1999)], pp. 23-24).<sup>3</sup> Thus, Lieb and Yngvason define  $\prec$  in the following way:

‘A state  $Y$  is adiabatically accessible from a state  $X$ ,  $X \prec Y$ , if it is possible to change the state from  $X$  to  $Y$  by means of an interaction with some device and a weight, in such a way that the device returns to its initial state at the end of the process whereas the weight may have changed its position in a gravitational field’ ([Lieb and Yngvason(1999)], p. 17)

We write  $X \sim^A Y$  if  $X \prec Y$  and  $Y \prec X$ , and  $X \prec\prec Y$  if  $X \prec Y$ , but not  $Y \prec X$ .

Lieb and Yngvason then impose axioms on  $(\cup_\alpha \Gamma_\alpha, \prec, +)$ . The basic six axioms are the following:

- A1. Reflexivity.  $X \sim^A X$
- A2. Transitivity.  $X \prec Y$  and  $Y \prec Z$  implies  $X \prec Z$ .
- A3. Consistency.  $X \prec X'$  and  $Y \prec Y'$  implies  $(X, Y) \prec (X', Y')$ .
- A4. Scaling invariance. If  $X \prec Y$ , then  $tX \prec tY$  for all  $t > 0$ .
- A5. Splitting and recombination. For  $0 < t < 1$ ,  $X \sim^A (tX, (1-t)X)$ .
- A6. Stability. If  $(X, \varepsilon Z_0) \prec (Y, \varepsilon Z_1)$  holds for a sequence of  $\varepsilon$ 's tending to zero and some states  $Z_0, Z_1$ , then  $X \prec Y$ .

Lieb and Yngvason also make use of the ‘comparison hypothesis’ (CH), which holds for a state space if any two states in the space are comparable, meaning either  $X \prec Y$  or  $Y \prec X$ . When two states are comparable, an adiabatic transition between the states is possible in one direction or the other (or possibly both).

From these six axioms and the CH, Lieb and Yngvason derive what they call the ‘Entropy Principle’:

- There is a real-valued function on all states of all systems (including compound systems), called **entropy** and denoted by  $S$  such that
- a) Monotonicity: If  $X \prec\prec Y$ , then  $S(X) < S(Y)$ , and if  $X \sim^A Y$ , then  $S(X) = S(Y)$ .

---

<sup>3</sup>However, see [Uffink(2001)], footnote 90, p. 382 for a discussion of this point.

b) Additivity and extensivity: If  $X$  and  $Y$  are states of some (possibly different) systems and if  $(X, Y)$  denotes the corresponding state in the composition of the two systems, then the entropy is additive for these states, i.e.  $S((X, Y)) = S(X) + S(Y)$ .  $S$  is also extensive, i.e. for each  $t > 0$  and each state  $X$  and its scaled copy  $tX$ ,  $S(tX) = tS(X)$ .

For convenience, I will refer to A1-A6 and the CH collectively as the ‘EP-axioms’, since these are used to derive the Entropy Principle (EP), even though strictly speaking CH is not an axiom for Lieb and Yngvason. Lieb and Yngvason regard the comparison hypothesis as unsatisfactory as an axiom in its own right, so they devote a great deal of energy to deriving it from a series of further axioms.

Let us now compare the Entropy Principle to the engineering version of the second law stated at the end of section 3. Both versions say that if a process from  $X$  to  $Y$  can occur adiabatically, then the entropy difference  $\Delta S = S(Y) - S(X)$  is non-negative. If processes from  $X$  to  $Y$  and  $Y$  to  $X$  are both possible adiabatically, then  $\Delta S = 0$ . The key difference between the two is that Lieb and Yngvason’s Entropy Principle only states that there is some function (with properties of additivity and extensivity) which serves as a measure of the adiabatic accessibility. It doesn’t yet connect the entropy function to heat or temperature via the relation  $dQ = TdS$ . In fact, neither heat nor temperature have at this stage in Lieb and Yngvason’s treatment been defined. It also doesn’t yet give any information about which processes are achievable adiabatically. Lieb and Yngvason introduce further axioms to make these connections. I will return to this in section 9.

## 6 Argument I for Lieb and Yngvason’s version of the second law

If the same two assumptions are made to support of premise (2) of Argument I as for the engineering version, the time reversal of an adiabatic process from  $X$  to  $Y$ , with  $X \prec Y$ , would produce an adiabatic process from  $Y$  to  $X$ , and hence  $Y \prec X$ . From premises (1) and (2) then, it would follow that all adiabatic processes can be thermodynamically reversed. This is denied by premise (3) which says that there exist processes that cannot be thermodynamically reversed, according to the second law. Suppose we take a process  $M$  that is specified not just by its initial and final states  $X$  and  $Y$  but also by being adiabatic. By definition of the adiabatic accessibility relation  $\prec$ , it follows that  $X \prec Y$ . There are then two possible cases allowed by the EP-axioms: either i)  $X \prec\prec Y$  or ii)  $X \sim^A Y$ . By the entropy principle, in case i)  $S(X) < S(Y)$  and in case ii)  $S(X) = S(Y)$ . The thermodynamic reversal of  $M$  is an adiabatic process from  $Y$  to  $X$ . By the definition of  $\prec$ , if the thermodynamic reversal is possible we have  $Y \prec X$ . Again by the entropy principle, this means  $S(Y) \leq S(X)$ . In case i) this produces a contradiction. Therefore there are allowable processes which the EP-principle says are not thermodynamically reversible. Premise (3) for this version of the law states that there *exist* macroscopic processes for which the EP-principle says the thermodynamic reverse is not possible. Again, this existence claim goes beyond the law itself, but in this it is no different from the engineering version of the second law. Thus, we find that Argument I is based on essentially the same considerations for Lieb and Yngvason’s version of the second law as for the engineering version.

## 7 Uffink's argument

We now turn to Uffink's argument that some versions of the second law, most notably Lieb and Yngvason's version, are time reversal invariant. Uffink sets up the issue differently than how it has been set up in Argument I. Argument I expresses the clash between what two different theories (or levels of theory) say is possible, given a premise about how time reversal affects thermodynamic processes. It follows the usual approach of taking the concept of time reversal invariance to be a property of the fundamental dynamical theory, and applying it to processes consisting of sequences of microstates. Reversibility, on the other hand, is understood in the context of thermodynamics as a property of macroscopic processes. Uffink departs from this usual picture by applying the notion of time reversal invariance directly to macroscopic processes. Both time reversal invariance and reversibility then are properties of the thermodynamic theory. Uffink emphasises that they are distinct concepts, and thus opens up the possibility that a version of thermodynamics could be time reversal invariant and yet allow thermodynamically irreversible processes. He then argues that Lieb and Yngvason's version of the second law is such a version.

Uffink starts with the following definitions ([Uffink(2001)], p. 314):

**Definition 3** *A macroscopic process is a sequence of equilibrium macro-states:  $\{X_i, \dots, X_f\}$ .*

By analogy with the microscopic case, the time reversal of the macroscopic process is defined as:

**Definition 4** *The time reverse of a macroscopic process  $M : \{X_i, \dots, X_f\}$  is  $M^T : \{RX_f, \dots, RX_i\}$  where  $R$  is the time reversal operator. If  $R$  acts as the identity,  $M^T$  is  $\{X_f, \dots, X_i\}$ .<sup>4</sup>*

**Definition 5** *(A version of) the theory of thermodynamics is time reversal invariant if for any macroscopic process  $M$  which is allowed by the theory, the time reversed process  $M^T$  is also allowed.*

Uffink stresses the difference between the notion of time reversal invariance and the notion of thermodynamic reversibility. Uffink suggests that there have been many shifts in the meaning of thermodynamic 'reversibility' throughout history. Nonetheless, he claims that there is a basic sense of thermodynamic irreversibility which is very prevalent. He describes this as a notion of 'irrecoverability' – that is, of not being able to fully undo the transition from one state to another. Uffink traces this idea back to Planck's conception of an 'irreversibel' process<sup>5</sup>.

When Planck introduces the notion of irreversibel, he says:

A process which can in no way be completely reversed is termed *irreversibel*, all other processes *reversibel*. That a process may be irreversible, it is not sufficient that it cannot be directly reversed. This is the case with many mechanical processes which are not irreversible. The full requirement is, that it be impossible, even with the assistance of all agents in nature, to restore everywhere the exact initial state when the process has once taken place. ([Planck(1945)], p. 84).

---

<sup>4</sup>Uffink claims that 'In a theory like classical thermodynamics, in which the state does not contain velocity-like parameters one may simply take  $R$  to be the identity transformation' ([Uffink(2001)], p. 314). Thus, the time reversed process  $M^T$  would indeed be  $\{X_f, \dots, X_i\}$ .

<sup>5</sup>I follow Uffink in using the German term to denote Planck's concept.

Uffink argues that there are several important differences between the thermodynamic notion of irreversibility or irrecoverability, as found in Planck, and the notion of time reversal invariance ([Uffink(2001)], pp. 316-7). In particular, he claims that the thermodynamic notion of irrecoverability is based on a different meaning of ‘possible’ than just what is allowed by the theory. This sense of possibility is ‘availability in the actual world’, and this is supposed to be a type of possibility that ‘transcends theoretical boundaries’ ([Uffink(2001)], p. 312). Uffink reads the clause ‘even with the assistance of all agents in nature’ in Planck’s definition as evidence that Planck thought that the recovery must be ‘available in our actual world, not merely in some model of the theory’. This distinction appears to be the basis for his general claim that time reversal invariance does not imply thermodynamic reversibility ([Uffink(2001)], p. 317). Thus, it would be possible, in Uffink’s picture, for a theory to be time reversal invariant, yet allow irreversible processes.

Uffink claims that some versions of the second law are time reversal invariant, whereas others are not. In particular, he claims that unlike the engineering version of the second law, Lieb and Yngvason’s ‘version of the second law can be obtained without invoking time reversal non-invariance at all’ ([Uffink(2001)], p. 387). I will now outline his argument for this conclusion.

Uffink first claims that the following is a natural way to understand the effect of the time reversal operator  $R$  ([Uffink(2001)], p. 371)<sup>6</sup>:

**Definition 6** *The time reversal operator acts in such a way that  $Rs = s$  and  $R(\prec) = \succ$ .*

He then offers two ways of analysing the question of time reversal invariance for Lieb and Yngvason. The arguments are brief, so I will quote them in full.

The first [way of analysing] is to consider all structures  $\langle \cup_{\alpha} \Gamma_{(\alpha)}, (, \prec) \rangle$  as candidate models, and look upon the axioms as singling out a class of possible worlds. In that case it is easy to show, using the implementation of time reversal [just given], that the six general axioms, as well as the comparability hypothesis, are completely time-symmetric!

[footnote: This conclusion cannot be extended to the complete set of axioms proposed by Lieb and Yngvason. For example, axiom S1 [Irreversibility. For each  $X \in \Gamma$  there is a point  $Y \in \Gamma$  such that  $X \prec \prec Y$ .] is already time-asymmetric. However, the time-asymmetry introduced by this axiom is only temporary. In the course of their presentation, axiom S1 is subsumed by a stronger axiom (called ‘Transversality’) which restores time-symmetry. (Transversality entails that for all  $s$  there is also a state  $r$  such that  $r \prec s$  and  $s \not\prec r$ .) Yet, there are two other axioms (called A7 and T1) which address mixing and equilibrium processes. These axioms are explicitly time-asymmetric (I thank Jakob Yngvason for pointing this out to me.) Note, however, that as far as the entropy principle is concerned, these axioms are needed only in the derivation of the (time-symmetric) compatibility hypothesis.]

Another way of analyzing the question is to start from the interpretation proposed by the authors [Lieb and Yngvason] for the relation  $\prec$  and note that it invoked the term “possible”. One may regard this as a modal relationship, to be understood in terms of a “possible worlds” semantics. On this reading, the statement “ $s \prec t$ ” does not express a manifest property of one single world, but rather commits one to the existence of possible worlds in which the state  $s$  can be transformed into  $t$  without leaving changes in auxiliary systems except the displacement of a weight. It does not, however, commit us to the existence of a possible world in which  $t$  is transformed into  $s$  under the same conditions. Thus, the class of possible world allowed by such a statement is time-asymmetric.

[footnote: In this view, the role of the axioms would then be to characterise a kind of second-order possibility, namely, to determine which relations between possible worlds are possible (allowed by the theory).]

---

<sup>6</sup>See also [Brown and Uffink(2001)] ‘a natural option is to call a law or set of laws time-symmetric iff *they continue to hold under reversal of the adiabatic accessibility relation on the space  $\Gamma$  of equilibrium states.*’ (p. 535).

Therefore, the answer to the question whether this approach is time-symmetric or not depends on whether one analyzes the question from the point of view of the formalism or its interpretation. Nevertheless, the fact that it is not necessary to introduce time-asymmetry into the formalism to obtain the second law, is very remarkable. ([Uffink(2001)], p. 383).

On the first analysis, a ‘model’ of the theory roughly consists of all the states and adiabatic accessibility relations between the states. As a simple example, suppose we have a model with just two states  $X$  and  $Y$  and the adiabatic accessibility relation  $X \prec Y$ . Then we also have a model of the EP-axioms with states  $X$  and  $Y$  and adiabatic accessibility relation  $Y \prec X$ . Uffink takes the fact that the set of models of the EP-axioms has the property of being invariant under the given definition of time reversal to show that the EP-axioms are time reversal invariant. However, in the particular model which represents the actual world, it may still be that  $X \prec Y$ . Thus, in Uffink’s account, thermodynamic irreversibility may coexist with time reversal invariance of the theory.

On the second reading, we can only say that an adiabatic process  $\{X, Y\}$  is allowed by the theory, relative to a given model. Then if  $X \prec Y$  in that model, then an adiabatic process  $\{X, Y\}$  is allowed, but an adiabatic process  $\{Y, X\}$  is not.

Uffink does not discuss the relative merits of these two analyses, but in drawing his overall conclusion, he clearly thinks the first is very significant. In a subsequent paper, only the first way of analysing the issue is discussed, and the conclusion is again that ‘time reversal invariance is not needed to obtain a rigorous statement of the second law’ ([Brown and Uffink(2001)], p. 537).

## 8 Critique of Uffink’s argument

We have seen that Uffink defines a notion of time reversal invariance which can be applied directly to versions of the second law. However, Definition 3 underspecifies a thermodynamic process. To specify a process  $M$  we need to know not just the initial and final macro-states, but also the energy resources exchanged between the system and its surroundings: the heat flows into and out of the system, and the work which is done on or by the system. If Definitions 4 and 5 are strictly applied, there is no interesting question about whether different versions of the second law are time reversal invariant. All versions are time reversal invariant in this sense. This is because, according to these definitions, the question of whether a version of the theory is time reversal invariant amounts to the question of whether for all allowed processes  $\{X_1, \dots, X_f\}$ , the reverse sequence of states  $\{X_f, \dots, X_1\}$  is allowed by the theory. But as long as the states in question are equilibrium states, there is nothing in equilibrium thermodynamics to preclude a transition from one equilibrium state to another. We can get from any state to any other, provided appropriate resources are supplied from the environment of the system. For example, the free expansion of a gas to twice its volume is thermodynamically irreversible when it occurs without work or heat flow. It is not possible to compress the gas back to the original state without work or heat flow. But it is quite possible to go from the final to the initial state of the free expansion by doing work isothermally on the system and allowing heat to flow out into a reservoir.

In practice, Uffink’s analysis of Lieb and Yngvason’s version of the second law, does not rely only on his official definition 5, but uses definition 6, which does tell us about the effect of time reversal on the adiabatic accessibility relation  $\prec$ . This definition effectively grants that time reversal produces thermodynamic reversal of adiabatic processes, as claimed by premise (2) of Argument I. This is because the definition grants the two assumptions which would support the premise, since

it assumes that the time reversal operator acts as the identity on the macrostate, and the adiabatic accessibility relation is reversed. Uffink claims that Lieb and Yngvason's entropy principle is time reversal invariant. But he does not deny that the entropy principle allows irreversible processes. That is, he does not deny premise (3) of Argument I either. In summary then, Uffink's analysis does not challenge any of the three premises of Argument I and so it leaves the traditional puzzle essentially untouched.

I have claimed that, given the definitions used, a version of the second law is time reversal invariant iff all processes that it allows can be thermodynamically reversed. Yet Uffink claims that Lieb and Yngvason's entropy principle is time reversal invariant, without denying that it allows irreversible processes. How can this be? One possible explanation could be that Uffink is not using irreversible in the sense of Definition 1, but rather in the sense of his notion of irrecoverable, as used by Planck. In particular, he thinks that irrecoverability may be based on a different notion of possibility than simply what is allowed by the laws of the theory.

Definitions of thermodynamic reversibility in Planck's sense are quite common in textbooks on thermodynamics.<sup>7</sup> And I will now suggest that this Planckian notion is not a different notion than reversible as in Definition 1. It is rather just a different way of expressing the same basic idea. Definition 1 talks about work and heat exchanges with the environment. However Planck expresses these exchanges in terms of their effects on auxiliary systems. Thus, work is done by the system when a weight is raised in an auxiliary system, and work is done on the system when a weight is lowered. Heat exchanges with auxiliary systems produce 'internal changes' in those systems. For reversibility, Planck requires that it be possible to restore the initial state of the system and all the auxiliaries with which it has interacted. Thus, if a weight was raised, it must be lowered again – this amounts to reversing the direction of work. And auxiliary systems must be put back into the same internal state – this amounts to reversing the direction of heat flow. Planck was pivotal in inaugurating a tradition in thermodynamics which avoided direct appeal to the concept of heat. But the basic notion of reversibility he is working with is otherwise continuous with the notion of reversibility in Definition 1, which is used in Carnot's basic proofs.

Uffink reads the clause 'even with the assistance of all agents in nature' as evidence that the sense of possibility involved in the notion of recoverability is more than just what is allowed by the theory. However, an alternative reading would be that Planck is pointing out, what we have seen before, which is that simply reversing the order of states of the system is insufficient for thermodynamic reversal – the energy exchanges with the auxiliary systems must also be undone in some way or another but without producing further changes in other auxiliaries.

Setting aside any particular reading of Planck, the broader point is that thermodynamic reversibility need not be tied to an unusual reading of possibility. The modal claims in the theory of thermodynamics are supposed to capture the modal facts in our world just as much as the modal claims of any other theory. The tension expressed in Argument I between time reversal invariance of the microdynamics and thermodynamic irreversibility arises from a conflict between what the microdynamical theory says is possible and what the thermodynamic theory says is possible. The microdynamics suggests that it is possible to reverse certain microscopically defined processes which would result in macroscopic processes which the thermodynamic theory says is impossible. However, 'possible' has the same meaning in both cases. Overall, then, appealing to different senses

---

<sup>7</sup>For example, 'a reversible process is one that is performed in such a way that, at the conclusion of the process, both the system and the local surroundings may be restored to their initial states, *without producing any changes in the rest of the universe*' ([Zemansky(1968)], p. 192). Other examples are to be found in [Finn(1993)], p. 16, [Bowley and Sanchez(1996)], p. 16.

of possibility or reversibility does not help us to evade Argument I.

## 9 The significance of the symmetry of the EP-axioms

Uffink's argument that Lieb and Yngvason's formulation of the second law is time reversal invariant is based on the fact that the models of the EP-axioms are invariant under reversal of the adiabatic accessibility relation  $\prec$ . In other words, if we take a model of the EP-axioms and reverse all the  $\prec$ , the result is another model of the EP-axioms. The significance of this symmetry is not that the second law has been formulated in a time reversal invariant way, it is rather that the entropy principle (or EP-axioms) is not the full second law. The entropy principle in Lieb and Yngvason is just a representation theorem that says that there is some real-valued (and additive) function which does not decrease in a process from  $X_i$  to  $X_f$  iff such a transition is possible adiabatically. It doesn't say anything about which adiabatic processes are possible. This is similar to how the axioms that give the conditions for the existence of a utility function representing qualitative preferences do not specify the content of the preferences. For example, there are models of the preference axioms in which the agent prefers apples to oranges, and others where the agent prefers oranges to apples.

In Lieb and Yngvason's system, the information about which processes can be achieved adiabatically comes from further axioms that go beyond the axioms used to derive the entropy principle. The key further axiom in Lieb and Yngvason's system is A7. Lieb and Yngvason introduce coordinates for the states which give the state space convex structure. That is, if  $X$  and  $Y$  are any two points in  $\Gamma$  and if  $0 \leq t \leq 1$ , then the point  $tX + (1-t)Y$  is a well-defined point in  $\Gamma$ . This is applicable in particular to 'simple' systems, whose state space can be identified with some open convex subset of  $\mathbb{R}^{n+1}$  with a distinguished coordinate denoted by  $U$  called the 'energy', and additional coordinates  $V \in \mathbb{R}^n$ , called 'work coordinates'. Then axiom A7 says:

Convex combination. Assume  $X$  and  $Y$  are states in the same state space,  $\Gamma$ , that has a convex structure. If  $t \in [0, 1]$  then  $(tX, (1-t)Y) \prec tX + (1-t)Y$ .

Lieb and Yngvason also assume three axioms S1-S3 for simple systems, and five axioms T1-T5 which deal with thermal equilibrium. A large part of their project is to avoid merely postulating the comparison hypothesis CH, but to derive it from these other axioms. As we saw, Uffink did mention that in a footnote that some of these other axioms, notably A7 and T1 do not possess the symmetry under reversal of  $\prec$  that the EP-axioms possess. However, he dismisses them as only used in derivation of CH. The role of A7 is much more significant than that. Even if it were not used in deriving CH, A7 still plays a role. Setting aside the derivation of CH, Lieb and Yngvason show ([Lieb and Yngvason(1999)], pp. 34-35) that it follows from the EP-axioms and A7 that the entropy function is concave on the state space  $\Gamma$ :

$$S(tX + (1-t)Y) \geq tS(X) + (1-t)S(Y). \quad (4)$$

This concavity in turn allows us to establish the direction in which various processes spontaneously proceed.

For example, consider the free expansion of a gas with internal energy  $2U$  and  $2N$  molecules from a volume of  $V$  to  $2V$ . By concavity of the entropy

$$2S(U, V, N) \geq S(U + \Delta U, V + \Delta V, N + \Delta N) + S(U - \Delta U, V - \Delta V, N - \Delta N) \quad (5)$$

Let  $\Delta U = U$ ,  $\Delta V = 0$ ,  $\Delta N = N$ , and using additivity of the entropy on the left hand side, we obtain

$$S(2U, 2V, 2N) \geq S(2U, V, 2N) + S(0, V, 0) \quad (6)$$

Consider the system before the free expansion. It consists of two subsystems. The first subsystem is a box of volume  $V$  containing a gas with energy  $2U$  and  $2N$  molecules – this has entropy  $S(2U, V, 2N)$ . The second subsystem is a box of volume  $V$  with no molecules, and entropy  $S(0, V, 0)$ . By additivity the total entropy is  $S(2U, V, 2N) + S(0, V, 0)$ , which is the right hand side in the above inequality. We assume that the energy and number of molecules is conserved in the expansion. After the expansion, the system has energy  $2U$ ,  $2N$  molecules and volume  $2V$ . Its entropy thus corresponds to the left hand side of the above inequality. The inequality tells us that the entropy is not lower after the expansion. Thus, by the entropy principle, an adiabatic transition consisting of an expansion of the gas is possible, whereas the reverse process, contracting the gas adiabatically, is not.

Lieb and Yngvason also show how to establish the direction of heat flow between two subsystems placed in thermal contact ([Lieb and Yngvason(1999)], p. 72). First, the temperature is defined

$$\frac{1}{T} = \left( \frac{\partial S}{\partial U} \right)_{V,N} \quad (7)$$

Then two subsystems placed in thermal contact will reach a final equilibrium state in which the temperatures of the two subsystems are equal. Suppose we assume that an object at a higher temperature is ‘hotter’ than one at lower temperature. We can then show that the direction of heat flow is from ‘hot’ to ‘cold’. Take two subsystems with states  $(U_1, V_1)$  and  $(U_2, V_2)$ . There is no change in the volume or number of molecules, only an exchange of energy between subsystems, with conservation of the total energy. Thus far, heat has not been explicitly mentioned, but it can be taken to be the energy transferred when work is not permitted. Suppose the first subsystem is hotter than the second,  $T_1 > T_2$ . That is  $\left( \frac{\partial S_1}{\partial U_1} \right)_{V_1, N_1} > \left( \frac{\partial S_2}{\partial U_2} \right)_{V_2, N_2}$ . Now suppose, for reductio, that the final energy  $U'_1 > U_1$ , which means, by conservation of energy, that  $U'_2 < U_2$ . Because  $S$  is concave, the temperatures  $\left( \frac{\partial S_1}{\partial U_1} \right)_{V_1, N_1}$  and  $\left( \frac{\partial S_2}{\partial U_2} \right)_{V_2, N_2}$  are both decreasing functions of  $U_1$  and  $U_2$  respectively. Thus, the above inequalities of the energy imply that  $T'_1 > T_1$  and  $T_2 > T'_2$ . Thus, it is impossible for the two subsystems to equilibrate at the same temperature. Hence we conclude that  $U'_1 < U_1$  and thus that the hot body has lost energy. Therefore the direction of the energy flow has been from hot to cold.

The usual formula for Carnot efficiency as an upper bound for the efficiency of a heat engine in a cyclic process can also be demonstrated ([Lieb and Yngvason(1999)], pp. 72-73). In sum, then, the directionality of possible processes follows from the entropy principle and the concavity of the entropy, which depends in turn on axiom A7.

An alternative approach would be to include a variational principle such as principle that the entropy is maximum in equilibrium. This is essentially what Giles does, and he then draws out the directionality of processes in a similar fashion to Lieb and Yngvason ([Gil(1964)], Chapter 9).

A useful comparison may be drawn with an alternative ‘postulational’ approach to thermodynamics which in some sense reverses the order of presentation of the engineering approach [Cal(1960)]. Rather than, as in the engineering approach, starting with some generalisations of our experiences with macroscopic systems and drawing out the theoretical apparatus of the entropy, on the postulational approach, postulates are made about the existence and nature of the

unobservable entropy and the observational basis falls out as a consequence. The subject of thermodynamics is thus developed with an eye to its eventual subsumption within the statistical mechanical framework. On this approach, the existence and additivity of the entropy are postulated, and so is a variational principle. Lieb and Yngvason differ from the postulational approach in that they are not content to merely postulate the existence of an additive entropy function. They rather derive it from axioms. However, a further postulate is then made which plays a similar role to the variational principle in the postulational approach.

Thus, it is particularly clear in axiomatic versions that part of the explanatory apparatus of the second law can be constructed without specifying which direction allowed processes go. However, this can also be seen in more traditional versions. One can split the derivation of the engineering second law from the Kelvin principle into two stages, as described in section 3. First, one proves directly from the Kelvin principle that there exist reversible adiabatic surfaces and an entropy function which satisfies  $dQ_{rev} = TdS$ .<sup>8</sup> At this stage, nothing has been said about which direction irreversible adiabatic processes proceed in. The claim that the adiabatic processes that occur are those for which the entropy increases comes from the Clausius inequality, itself a direct consequence of the Kelvin principle. In the traditional approach, then, both the existence of reversible adiabatic surfaces and the direction of processes are derived ultimately from the Kelvin principle, whereas in axiomatic approaches, one set of axioms are used for the former and further axioms are added to determine the latter.

## 10 Conclusion

Uffink has suggested that versions of the second law vary in their implications for time reversal non-invariance and irreversibility. In particular, he claims that unlike traditional versions, the Lieb and Yngvason version of the second law is time reversal invariant. He questions then whether time reversal non-invariance is essential to the second law. It would seem that these claims would have implications for the long-standing puzzle of the relationship between the second law and time reversal invariance of the dynamics.

However, in this paper, I have argued that the traditional problem of tension between time reversal invariant dynamics and the second law can be stated as a simple three-premise argument, which is based on considerations which do not vary with different versions of the second law. In fact, the puzzle still exists for the Lieb and Yngvason version of the second law, which Uffink claims to be time reversal invariant.

We have also seen that Uffink's claim that Lieb and Yngvason's version is time reversal invariant is based on the symmetry of the axioms for their entropy principle under reversal of the adiabatic accessibility relation. The significance of this symmetry however is not that we have a time reversal invariant version of the second law. It is that the entropy principle makes no specification of the direction of natural processes. But the entropy principle is not the full expression of the second law in Lieb and Yngvason's system. Their system also includes other axioms which do specify the direction of processes.

In general, the content of the second law of equilibrium thermodynamics can be thought of as having two components:

i) an explanatory structure consisting of a state space which is foliated by reversible adiabatic surfaces and a function, the entropy, which is constant on those surfaces.

---

<sup>8</sup>A particularly direct method for this is shown in [Zemansky(1968)], Chap. 8.

ii) a statement about the direction in state space that natural processes take when they move off reversible adiabatic surfaces.

In traditional versions, both components flow, via a series of arguments, from the Kelvin principle. But in axiomatic formulations, the first component is derived from separate axioms than the axioms that determine the direction of the second component.

## 11 Acknowledgements

I thank Jos Uffink for detailed comments on multiple versions of this manuscript. I also thank Robert Batterman, John Earman, John Norton, Mark Wilson and the Fellows of the Center for Philosophy of Science reading group, Pittsburgh, 2014, for useful feedback on previous versions. I am grateful to Bryan Roberts for discussion.

## References

- [Cal(1960)] , 1960. Thermodynamics: an introduction to the physical theories of equilibrium thermostatics and irreversible thermodynamics. John Wiley and sons, New York.
- [Gil(1964)] , 1964. Mathematical foundations of thermodynamics. MacMillan Company, New York.
- [Bowley and Sanchez(1996)] Bowley, R., Sanchez, M., 1996. Introductory statistical mechanics. Oxford Science Publications.
- [Boyling(1972)] Boyling, J. B., 1972. An axiomatic approach to classical thermodynamics. Proceedings of the Royal Society of London 329, 35–70.
- [Brown and Uffink(2001)] Brown, H. R., Uffink, J., 2001. The origins of time-asymmetry in thermodynamics: The minus first law. Studies in History and Philosophy of Modern Physics 32, 525–538.
- [Caratheodory(1976)] Caratheodory, C., 1976. Investigation into the foundations of thermodynamics. In: The second law of thermodynamics. Dowden, Hutchinson and Ross Inc., Pennsylvania.
- [Duistermaat(1968)] Duistermaat, J. J., 1968. Energy and entropy as real morphisms for addition and order. Synthese 18, 327–393.
- [Finn(1993)] Finn, C. B. P., 1993. Thermal Physics. Stanley Thornes Publishers Ltd.
- [Kestin(1966)] Kestin, J., 1966. A course in thermodynamics. Blaisdell Publishing Co.
- [Lieb and Yngvason(1999)] Lieb, E. H., Yngvason, J., 1999. The physics and mathematics of the second law of thermodynamics. Physics Reports 310, 1–96.
- [Planck(1945)] Planck, M., 1945. Treatise on thermodynamics, 3rd Edition. Dover.
- [Roberts and Luce(1968)] Roberts, F. S., Luce, R. D., 1968. Axiomatic thermodynamics and extensive measurement. Synthese 18, 311–326.
- [Uffink(2001)] Uffink, J., 2001. Bluff your way in the second law of thermodynamics. Studies in History and Philosophy of Modern Physics 32, 305–394.

[Zemansky(1968)] Zemansky, M. W., 1968. Heat and thermodynamics, 5th Edition. McGraw-Hill Book Company.