Time and irreversibility in axiomatic thermodynamics

Robert Marsland III[a)]
Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139-4307

Harvey R. Brown[b)]
Faculty of Philosophy, University of Oxford, Radcliffe Humanities, Woodstock Road, Oxford OX2 6GG, United Kingdom

Giovanni Valente[c)]
Department of Philosophy, University of Pittsburgh, 1001 Cathedral of Learning, Pittsburgh, Pennsylvania 15260

Thermodynamics is the paradigm example in physics of a time-asymmetric theory, but the origin of the asymmetry lies deeper than the second law. A primordial arrow can be defined by the way of the equilibration principle ("minus first law"). By appealing to this arrow, the nature of the well-known ambiguity in Carathéodory's 1909 version of the second law becomes clear. Following Carathéodory's seminal work, formulations of thermodynamics have gained ground that highlight the role of the binary relation of adiabatic accessibility between equilibrium states, the most prominent recent example being the important 1999 axiomatization due to Lieb and Yngvason. This formulation can be shown to contain an ambiguity strictly analogous to that in Carathéodory's treatment.

I. THE ARROW OF TIME

In physical theories generally, time plays a multi-faceted role. There is the notion of temporal duration between events occurring at the same place (temporal metric, related to the ticking of an ideal inertial clock[1]), the comparison of occurrences of events at different places (distant simultaneity, registered by synchronized clocks), and the directionality, or arrow, of time. Thermodynamics is unusual, within the panoply of physical theories, in the double sense that a metric of time is not prominent, and that it is the only theory, apart from that of the weak interactions, that incorporates an arrow of time at a fundamental level. Let us consider these two aspects in turn.

It is sometimes said that thermodynamics has no clocks, in the sense that none of its fundamental laws contains derivatives with respect to time. For example, entropy is claimed never to decrease in adiabatic processes, but the theory gives no information about how quickly changes in entropy, if any, occur. It might be thought that a temporal metric and a privileged notion of simultaneity both lurk in the background, because thermodynamics always appeals to the mechanical notion of work. Whether this appeal to work introduces through the back door all the temporal structure of Newtonian time is far from clear. However, that may be, a noteworthy feature of Carathéodory's 1909 formulation of thermodynamics is the fact that time derivatives do appear explicitly in his Paper, as we shall see below.

As for the intrinsic arrow of time in thermodynamics, a reasonable question to ask is: what feature of the theory defines it? Consider the view expressed by Hawking:

> Entropy increases with time, because we define the direction of time to be that in which entropy increases.[2]

There may be much to be said for this view in the context of statistical mechanics, but in classical equilibrium thermodynamics natural doubts arise. Within the traditional, textbook approach to the theory, the mere introduction of a concept like a Carnot cycle presupposes a temporal ordering as applied to the equilibrium states within the cycle. The temporal direction of a Carnot cycle is taken for granted well before questions concerning the efficiency of such cycles in relation to other kinds of heat engine are raised, and hence before the second law is introduced. What does it mean to say that one state in the cycle is earlier than another? The claim that a certain process unfolds in such and such a way in time only makes sense in physics if some independent arrow of time is acting as a reference. What is it in this context?

That little attention is given to this question is not entirely surprising. Students learning, for example, Newtonian mechanics of systems of point particles are told that the state of the particles at a given time is given by the combination of the linear momenta of the particles and their positions at that time. That the ith particle has velocity $\mathbf{v}_i$ rather than $-\mathbf{v}_i$ relative to some inertial reference frame must again be referring to some background arrow of time.[3] Now given that the equations are time-reversal invariant, one might think that the choice of direction of time is mere convention. But it does not look like this to anyone trying to apply the theory to real systems in the world. A background arrow is being presupposed; though rarely made explicit, it plausibly is related to the thermodynamic arrow.

Rather than speculating as to what the 19th century fathers of thermodynamics would have meant by "before" and "after," if these terms were anything other than primitive,[4] it is tempting in physics generally to fall back on the psychological arrow of time, according to which observers remember the past and not the future. But within thermodynamics itself this position is unattractive. It seems plausible that formation of memories in the brain would be impossible without thermodynamic irreversibility, even if there is debate about the details. Maroney attempted to show in 2010 that the logical operations involved in computation do not *per se* determine an arrow of time.[5] But in a 2014 rejoinder, Smith

This is the accepted version of Marsland R., Brown H.R., Valente G., Time and irreversibility in axiomatic thermodynamics (American Journal of Physics, 2015, 83:7, 628-634)
The final publication is available at: http://dx.doi.org/10.1119%2F1.4914528

claimed that in the brain computational processes and in particular the formation of long-term memories in fact requires the existence of certain spontaneous diffusion/equilibration processes.[6] From the point of view of statistical mechanics, these processes correspond to local entropy increase. But from the point of view of thermodynamics, they are arguably tied up with a principle lying deeper in the theory than the second law.

It has occasionally been noted in the literature that a fundamental principle that underlies all thermodynamic reasoning (including the zeroth law concerning the transitivity of equilibrium) is this:

> An isolated system in an arbitrary initial state within a finite fixed volume will spontaneously attain a unique state of equilibrium.

This *equilibration principle* is the entry point in thermodynamics of time asymmetry: an isolated system evolves from non-equilibrium into equilibrium, but not the reverse. Already in 1897 Planck had emphasised the independence of this principle from the second law,[7] and in subsequent litera-ture, it has been variously called the "zeroth law" (a particu-larly unfortunate title, given that it standardly refers to the transitivity of equilibrium between systems), the "minus first law,"[8] and the "law of approach to equilibrium."[9] The suggestion we wish to make is that *all* references, implicit and explicit, to the temporal ordering of events in thermodynamics can be understood in relation to the arrow of time defined by this process of spontaneous equilibration.

Such an approach is by no means compulsory; in principle an appeal to, say, the cosmological arrow of time (defined by the expansion of the universe) can serve the same purpose. In particular, our attempt in what follows to clarify certain temporal issues arising in modern axiomatic formulations of thermodynamics that do not rely on such notions as Carnot cycles does not strictly depend on the choice of the background arrow, as long as the role of the arrow is not overlooked. However, the suggestion we are making to use the equilibration principle in this context seems to us an elegant solution to the problem raised above in relation to the temporal direction of Carnot cycles (and, as we shall see, of adiabatic accessibility): one does not have to appeal to an arrow of time outside of thermodynamics itself.

II. AXIOMATIC THERMODYNAMICS

For the purpose of elucidating the source and consequences of the arrow of time in thermodynamics, the standard formulations of the theory given in most undergraduate classes and textbooks are inadequate: the rigorous analysis of the heat engine concept involving the Carnot cycle proves to be very complicated,[10] obscuring these issues still further. Moreover, such a cycle requires, in the case of a two-dimensional state space for a simple system (see below), that adiabats and isotherms in the space of equilibrium states are curves intersecting only at single points. An example of a situation in which this is not the case is the region of the triple point of water, where the adiabats for a range of entropy values coincide partly with the 273.16 K isotherm.[11]

The first attempt to put equilibrium thermodynamics on a rigorous conceptual and mathematical footing without appeal at the fundamental level to cyclic heat engines, and in particular Carnot cycles, was found in the 1909 work of Constantin Carathéodory.[12] A number of subsequent careful formulations of thermodynamics owe much to this work; the most prominent recent example is that due to Elliott Lieb and Jacob Yngvason, published in a lengthy Paper in 1999.[11] These authors follow Carathéodory in basing their approach on the notion of adiabatic accessibility but do without the machinery of differential forms that Carathéodory had used in his reasoning. A penetrating analysis of the Lieb-Yngvason formulation was published by Jos Uffink in 2001, principally with a view to determining which axioms pro-posed by these authors were time symmetric and which not.[13] In the present Paper, we are concerned with a related but distinct issue. It is well known that Carathéodory's for-mulation contained an ambiguity, or incompleteness, which Carathéodory himself highlighted, and which is connected with the fact that his postulates lead to a version of the sec-ond law that is weaker than the traditional version due to Kelvin and Planck. These postulates permit the existence of two possible worlds: one in which entropy is non-decreasing for adiabatic processes, and another in which it is non-increasing. We argue that by referring to the arrow of time defined by the equilibration principle, it becomes clear that these worlds are indeed empirically distinct. The ambiguity in question arises in many Carathéodory-inspired approaches to thermodynamics; some, whether of the formal[14] or infor-mal variety,[15] add an extra empirical postulate to remove the ambiguity. We argue that this is what Lieb and Yngvason do in their approach, though not with complete transparency.

III. CARATHÉODORY

In his seminal 1909 reformulation of thermodynamics, Carathéodory realized that heat need not be introduced as a primitive notion, and that the theory could be extended to systems with an arbitrary number of degrees of freedom using generalized coordinates analogous to those employed in mechanics. In doing so, he provided the first satisfactory enunciations of what are now called the zeroth and first laws of thermodynamics.[16] In particular, by defining an adiabatic enclosure in terms of its capacity to isolate the thermodynamic variables of the system of interest from external disturbances, Carathéodory was the first to base the first law, and thus the existence of internal energy, on Joule's experi-ments (under the assumption that Joule's calorimeter was adiabatically isolated). Heat is then defined as the change in internal energy that is not accounted for by the work being done on or by the system, the existence of heat being a consequence of the first law and the conservation of energy.

However, what is of particular relevance for our purposes is that Carathéodory did to the equilibrium state space something akin to what his ex-teacher Hermann Minkowski had done to space-time a year earlier. Assuming that the space Γ of equilibrium states is an N-dimensional differentiable manifold equipped with the usual Euclidean topology, Carathéodory introduced the relation of adiabatic accessibility between pairs of points, a notion clearly analogous to that of the causal connectibility relation in Minkowski space-time.[17] An adiabatic process is one taking place within an adiabatic enclosure. It is time-directed; the arrow of time can be that defined by the equilibration principle, though Carathéodory himself was silent on the matter. He famously postulated that in any neighborhood of any point p in Γ, there exists at least one point p' that is not adiabatically ac-cessible from p.[18] We shall refer to this axiom as the *inac-cessibility principle*.

Carathéodory's main result concerns "simple" systems, whose states can be described by a single thermal coordinate along with an arbitrary number of "deformation" coordinates, sometimes called work or configuration coordinates, which depend on the external shape of the system and on any applied fields. This rules out systems comprised of a collection of subsystems adiabatically isolated from each other. Carathéodory also assumed that simple systems show no internal friction or hysteresis in sufficiently slow (quasi-static) processes, in the definition of which he referred to derivatives with respect to time.[19] As a result of these and other assumptions, he showed that quasi-static processes involving simple systems can be represented by continuous curves in the state space, where the external work associated with the process can be determined solely by the forces required to maintain equilibrium at all times. (Carathéodory made a point of *proving* that quasi-static adiabatic processes of a simple system are reversible.) By appealing to a result in the theory of Pfaffian forms, he was further able to show that given the inaccessibility principle, the differential form for heat for quasi-static processes has an integrating factor. In other words, there exist functions T and S on the state space such that the heat form can be expressed as TdS, where dS is an exact differential. Further considerations show that T and S are related to the absolute temperature (which depends on empirical temperature as defined by way of the zeroth law[20]) and entropy of the system.[21]

IV. THE AMBIGUITY

In Sec. 9 of his 1909 Paper, devoted to irreversible processes, Carathéodory introduced a terse argument related to simple systems that has often been repeated and/or elabo-rated in the literature.[22] The conclusion of the argument is that, given the inaccessibility principle and certain continuity assumptions,[23] then for any two points p and p' not con-nected by a reversible quasi-static path, when p' is adiabati-cally accessible from p, always either $S(p') > S(p)$ or $S(p') < S(p)$. (Quasi-static adiabatic processes involve no change in entropy.) Regarding this ambiguity, Carathéodory emphasized both that it persists even when the entropy is defined so as to make the absolute temperature positive, and that it can only be resolved by appeal to experiment:

> Experience (which needs to be ascertained in relation to a single experiment only) then teaches that *entropy can never decrease*.[24]

It is important for our purposes to note first that the prior existence of an entropy function is not in fact intrinsic to the argument or rather that a related ambiguity can be derived in a more general way. The single thermal coordinate for the simple system in question could be chosen instead to be internal energy (whose existence is a consequence of the first postulate in Carathéodory's Paper). In this case, the inaccessibility principle and the same continuity assumptions can be shown to result in the existence of a foliation of Γ (subject to a qualification to be clarified below), such that on each hypersurface of the foliation any continuous curve represents a reversible, quasi-static adiabatic process involving a continuous change in the deformation coordinates. In the case of an *arbitrary* adiabatic process from p to a distinct state p', the final state p' will generally *not* lie on the same hypersur-face as p, but it can be shown from Carathéodory's postu-lates that all possible final states p' must lie on the same *side* of this hypersurface. In particular, when p and p' share the same deformation coordinates, p' will either always have greater internal energy than p, or always have less internal energy, independently of the choice of the initial state p. Let us call this the *energy ambiguity* for adiabatic processes.

A related ambiguity holds when the thermal coordinate is chosen to be temperature (empirical or absolute in Carathéodory's terms, but assumed to be positive). Indeed, the underlying ambiguity in Carathéodory's formulation of thermodynamics—prior to the performance of the "single experiment" referred to above and given the positivity of temperature—can also be stated as: Either heat always flows from a hot body to a cold body or the converse. When considering cyclic processes, the ambiguity can be expressed in two further ways:

(1) Either it is always impossible to create a cyclic process that converts heat entirely into work or it is always impossible to create a cyclic process that converts work entirely into heat;[25] and
(2) In relation to a Carnot cycle, any other type of cyclic process either always has lower efficiency or always has a greater efficiency.[26]

Statement 1 is clearly weaker than the traditional Kelvin-Planck form of the second law in thermodynamics; indeed Carathéodory's inaccessibility principle above is easily seen to be a consequence of the latter (the first possibility in 1), but the converse implication does not hold.[27]

The argument in Sec. 9 of Carathéodory's Paper presupposes that adiabatic accessibility is a transitive relation between states (so that if state q is adiabatically accessible from state p, and r is adiabatically accessible from q, then r is adiabatically accessible from p). It is obviously reflexive (any p is adiabatically accessible from itself), so it satisfies the conditions for being a *preorder*. The qualification men-tioned earlier in relation to Sec. 9 is that, as originally noted by Bernstein,[28] the argument is of a local, not global, nature in the state space; indeed entropy itself is a local notion in Carathéodory's approach.[29] Hence, the adiabatic accessibility for Carathéodory is locally, not globally, a preorder.

Returning to Carathéodory's point that experiment is needed to determine the sign of the entropy gradient, it should be clear that the notion only makes sense if a back-ground arrow of time is specified. Indeed, it is easily seen that the two possible Carathéodory worlds are not simply the temporal inverses of each other, because, to repeat, the adia-batic (in)accessibility relations that are postulated to hold themselves are defined with respect to a background arrow of time. Again, we suggest it can be that defined by the equilibration principle. Whether in adiabatic processes entropy is universally non-decreasing or non-increasing relative to the arrow defined by spontaneous equilibration is a clear-cut empirical matter and not a matter of convention.

V. LIEB AND YNGVASON

A. Introduction

In 1999, Lieb and Jakob Yngvason proposed[11] a new axiomatization of thermodynamics, which owed much to the work of Carathéodory and that of later writers such as Robin Giles. The central concept is again the binary relation on the state space associated with adiabatic accessibility,[30] now designated by $\prec$, and assumed to be *globally* a preorder. An

attempt is made by the authors to provide a treatment of entropy and its essential properties based on "maximum principles instead of equations among derivatives," so that real systems where some of these derivatives fail to be well-defined at certain points (such as the triple point of water mentioned above) pose no special problems for the theory. Another notable and unusual feature is the attempt to provide a proof of the *comparison hypothesis*, normally tacitly assumed to be an essential property of a well-behaved thermodynamic system, which states that for any states X and Y in the state space, then either $X \prec Y$ (Y is adiabatically accessible from X) or $Y \prec X$. The Lieb-Yngvason (L-Y) formulation of thermodynamics is a *tour de force* of physical and mathematical reasoning.

It should be noted that two further incentives behind the formulation are the desire to banish the notion of heat altogether from thermodynamics, and a shift of emphasis from *impossible* processes (as in traditional formulations) to *possible* ones.[31]

B. Entropy

Our main concern lies more with energy than entropy, but a word about the L-Y treatment of the latter is in order. This treatment, remarkably, provides what is effectively a representation Theorem for the preorder $\prec$ on the state space in terms of a numerical "entropy" function on the space. It will be recalled that adiabatic accessibility is a temporally ordered concept, and the question arises whether and how the monotonic temporal behavior of entropy in adiabatic processes is connected with natural constraints on this preorder, without appeal to an assumption as strong as Carathéodory's inaccessibility principle. Lieb and Yngvason introduce six plausible axioms governing the $\prec$ relation holding for single and compound systems, and assuming (without proof at this stage) the comparison hypothesis, show that there exists a real-valued function S on all states of all systems such that $X \prec Y$ if and only if $S(X) \leq S(Y)$. Furthermore, S has the properties of additivity and extensivity that one expects of the entropy function. "In a sense it is amazing," Lieb and Yngvason write (p. 14), "that much of the second law follows from certain abstract properties of the relation among states, independent of physical details (and hence of concepts such as Carnot cycles)."[11] It should not be overlooked, however, that the very definition of entropy in this construction requires the existence of at least one pair of states X_0 and X_1 such that $X_0 \prec\prec X_1$, i.e., $X_0 \prec X_1$ but not the converse. Like the traditional notion of entropy, the L-Y notion is not meaningful in a world without irreversibility of some sort.[32]

The "entropy principle" is striking, and its proof is ingenious. But it is important to note that the temporal monotonicity associated with this numerical representation of $\prec$ does not resolve the kind of ambiguity found in Carathéodory's system. The question, recall, is whether the physical entropy is non-increasing or non-decreasing relative to the arrow of time determined by the equilibration princi-ple. In the L-Y formulation, a formal definition of S is con-structed such that, given all the assumptions, S cannot decrease in adiabatic processes. But the representation Theorem of course holds just as well for the function $\tilde{S} \equiv -S$, in which case $X \prec Y$ if and only if $\tilde{S}(X) \geq \tilde{S}(Y)$. There is nothing in the theorem *per se* that distinguishes between S and $\tilde{S}$ in terms of physical import.

C. The ambiguity again

The L-Y framework takes on a different, more geometrical tone after the treatment of the entropy principle. The systematic treatment of irreversibility in simple systems requires additional axioms in order to derive an analog of Carathéodory's inaccessibility principle and notably the *global* foliation of the state space defined by adiabats. Recall now the energy ambiguity in Carathéodory's theory outlined in Sec. III above for simple systems. Precisely this issue is addressed in Sec. III C of the L-Y Paper (Ref. 11). The authors first adopt the view that within their framework of axioms, it is "conventional" whether in an adiabatic process between states with the same deformation (work) coordinates the internal energy never decreases, or never increases.[33] (Unsurprisingly, they adopt the former option.) This is a curious stance, difficult to reconcile with subsequent remarks that take into account the definition of adiabatic accessibility peculiar to Lieb and Yngvason (p. 44):

> From a physical point of view there is more at stake, however. In fact, our operational interpretation of adiabatic processes involves either the raising or lowering of a weight in a gravitational field and these two cases are physically distinct. Our convention, together with the usual convention for the sign of energy for mechanical systems and energy conservation, means that we are concerned with a world where adiabatic process at fixed work coordinate can never result in the raising of a weight, only in the lowering of a weight. The opposite possibility differs from the former in a mathematically trivial way, namely by an overall sign of the energy, but given the physical interpretation of the energy direction in terms of raising and lowering of weights, such a world would be different from the one we are used to.[11]

This seems to be an admission that, as Carathéodory claimed (admittedly in the context of entropy not energy), two distinct physical possibilities are at stake, so it is hard to see how the issue is merely one of convention, in the usual sense of the term. It is noteworthy that Lieb and Yngvason state as a *Theorem*, which they call *Planck's principle*, that:

> If two states, X and Y, of a simple system have the same work coordinates, then $X \prec Y$ if and only if the energy of Y is no less than the energy of X.[34]

The authors make a point of saying (p. 46) that this principle (or rather a consequence of it) is "clearly stronger than Carathéodory's principle, for it explicitly identifies states that are arbitrarily close to a given state, but not adiabatically accessible from it."[11] Note that Lieb and Yngvason give as the reason for calling the mentioned Theorem "Planck's principle" that "Planck emphasized the importance for thermodynamics of the fact that "rubbing" (i.e., increasing the energy at fixed work coordinate) is an irreversible process."

It is quite true that the Planck principle is stronger than anything derivable from Carathéodory's inaccessibility principle along with the continuity assumptions mentioned in Sec. III above. Indeed, within some Carathéodory-inspired formulations of thermodynamics, the empirical fact appealed to by Planck that frictional rubbing under fixed deformation coordinates leads to an increase of internal energy is chosen

as precisely the extra empirical ingredient needed to resolve the ambiguity in Carathéodory's original theory.[35] On p. 46 of their 1999 Paper, Lieb and Yngason claim that Planck's principle, and as a consequence the standard Kelvin-Planck version of the second law, follow from their first nine axioms.[11] (These include the convex combination axiom A7 to which we return below.) But this is not strictly the case. We now attempt to further clarify the nature of the extra *factual* (not conventional) ingredient needed over and above the first nine axioms required in order to recover the standard second law of thermodynamics within the L-Y scheme.

D. The ambiguity exposed

If the question is whether there is *any* component of the L-Y scheme that favors one side of the adiabats over the other, then the answer is actually yes. A key assumption in the L-Y treatment of simple systems and their irreversible behavior is their convex combination axiom A7. This asserts that for any two states $X_1 = (U_1, V_1)$ and $X_2 = (U_2, V_2)$ of a simple system (U being the internal energy), a fraction t of X_1 can be adiabatically combined with a fraction $(1 - t)$ of X_2 to form a new state $Y = [tU_1 + (1 - t)U_2,\ tV_1 + (1 - t)V_2]$.[36] This axiom leads immediately to the theorem that in the case of a single simple system the set of points in the state space adiabatically accessible from a given point X—the "forward sector" associated with X—must be a convex set.[37] That is, if X_1 and X_2 as just defined are in the set, then Y is also in the set. (Indeed, this is the main consequence of A7, although A7 is needed to derive several of the key geometric proper-ties of the forward sectors.)

Consider then a continuous curve in a UV diagram corresponding to the boundary of the set of states adiabatically accessible from a given state, which as expected in the L-Y scheme turns out to be a curve of constant entropy (see Fig. 1).[38]

For a standard thermodynamic system with positive pressure, U decreases on this curve with increasing deformation coordinate V. The dotted line between states X_1 and X_2 represents the locus of all convex combinations of these states obtained by ranging over the parameter t ($0 \leq t \leq 1$). Both X_1 and X_2 are adiabatically accessible from Y', and it follows from Axiom A7 that Y is too. In the figure, the adiabat is represented as a convex function, so $(\partial^2 U/\partial V^2)_S \geq 0$. So the forward sector defined relative to any state X on the adiabatic

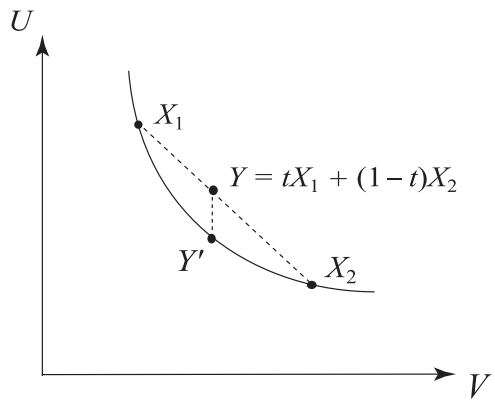

Fig. 1. A constant-entropy curve joining states X_1, Y', and X_2. State Y is a convex combination of X_1 and X_2 (see text).

is "upward pointing" in Lieb and Yngvason's terms: the projection on the energy axis of the normal to the tangent plane at X pointing to the interior of the forward section is positive.[39] This is a necessary condition for the Planck prin-ciple to hold. (Note in the figure that for the state Y there is a state Y' on the adiabat with the same deformation coordi-nates, and $U(Y) \geq U(Y')$.) However, if the adiabat is concave, then the forward sector will be downward pointing. So apart from the special case of a flat boundary, owing to Axiom A7 the shape of the adiabat will determine where the forward sector lies unambiguously.[40]

However, the shape of the adiabat, and hence the upward or downward pointing nature of the forward sector, are *not* determined by the L-Y axioms. The original energy ambiguity in Carathéodory's 1909 formulation has reappeared. Lieb and Yngvason are clearly aware, as we saw in Sec. V D, that two worlds are consistent with their axioms, but prefer to say that the choice of the familiar Kelvin world is one of "convention." In our opinion, Carathéodory's view of the matter is the correct one: it is nature, not the observer, that makes the choice, and given some background arrow of time, experiment is needed to see what nature prefers.

E. The flow of energy (heat)

It is perhaps worth comment that the convex combination Axiom A7 does not seem to be required to obtain important variants of the second law. Consider the simple derivation Lieb and Yngvason give of the proposition that energy (heat) spontaneously flows from hot to cold bodies, and not the converse (recall the related ambiguity in Sec. III).[41] A body A is defined to be hotter than another body B if the absolute temperature T_A is greater than T_B, where temperature $T_{A(B)}$ is defined as $(\partial S_{A(B)}/\partial U_{A(B)})^{-1}$, $S_{A(B)}$ being the entropy of $A(B)$. Note that this definition of "hotter than" is reasonable only if $T_{A(B)}$ is everywhere positive, and this constraint is a consequence[42] of the upward-pointing nature of the forward sectors and the choice of the function S and not $\tilde{S} \equiv -S$ in defining temperature in this way (see the end of Sec. V B). It is easy now for Lieb and Yngvason to obtain the desired irreversible flow of energy (heat) from A to B, given the conservation of total energy for the joint system, and the monotonicity of $U_{A(B)}$ with respect to $T_{A(B)}$. We won't repeat the details, other than to make two remarks.

First note that the spontaneity of this energy flow process (once thermal contact is achieved) leading to a common temperature can be secured by appeal to a special case of the equilibration principle.[43] Second, and more to the point, the strict monotonicity condition between $U_{A(B)}$ and $T_{A(B)}$ is a consequence of the concavity and differentiability of the entropy for simple systems. Concavity of entropy in general terms is established by Lieb and Yngvason's Theorem 2.8, which relies on the convex combination Axiom A7. However, what is actually required in the heat flow argument is the weaker claim that entropy is concave relative to the in-ternal energy, and Lieb and Yngvason had earlier established (p. 53) that this does not depend on Axiom A7.[11] It seems then that this axiom is not crucial to the argument, whereas the upwards pointing condition certainly is.

Finally, a word of caution about over-simplistic inferences as to what an anti-Kelvin world would be like according to the L-Y scheme. Suppose that the adiabat is strictly concave (not convex as depicted in Fig. 1), the Planck principle is false, and $(\partial^2 U/\partial V^2)_S < 0$. Now pressure P by definition

satisfies the equation $P = -(\partial U/\partial V)_S$.[44] Thus $(\partial P/\partial V)_S = -(\partial^2 U/\partial V^2) > 0$. Suppose a weight is placed on top of a piston with the pressure exactly set to balance the weight. It would seem that any bump involving the least increase of volume would increase the pressure and lead to a runaway process propelling the weight upwards; conversely a perturbation decreasing the volume would cause a continual collapse of the piston to zero volume. By Le Chatelier's principle, this means that all the states of the system are unstable equilibria, analogous to points on the upwards-sloping region of a van der Waals isotherm.[45] (This instability argument also holds for negative-pressure systems with U as an *increasing* function of V on the adiabats. In either case, whether any such instability threatens the validity of the equilibration principle is a moot point.) However, such a conclusion presupposes that the pressure P is interpreted as force per unit area, and that a force acting on a body at rest produces, as in our world, a motion in the same direction as the force. However, in a counterfactual anti-Kelvin world, this latter assumption may be false; how thermodynamical systems interact with mechanical systems is not established on the basis of the L-Y axioms alone.[46]

VI. CONCLUSIONS

We have argued that to understand aspects of Carathéodory's, or indeed any approach to the second law in thermodynamics, a background arrow of time needs to be specified, and we suggest defining it by way of the equilibration principle (minus first law). The L-Y approach, compared to Carathéodory's, requires weaker assumptions in order to derive a monotonic entropy function but needs considerably more axioms in order to establish a recognizable version of the second law. In part this reflects an admirable attention to detail and an attempt to make every step transparent while using less differential structure; one must also not overlook the ambitious program of deriving the comparison hypothesis. However, a strict analog of the energy ambiguity in Carathéodory's approach reappears in the L-Y scheme. The upward pointing nature of forward sectors (and hence the Planck principle) appears not to be a consequence of the L-Y axioms, nor is it a mere convention; it plays the role of an appeal to experience, over and above the axioms, of the kind Carathéodory needed in order to derive the standard Kelvin-Planck version of the second law.

ACKNOWLEDGMENTS

The authors thank Leah Henderson, Elliott Lieb, Jos Uffink, and especially Jakob Yngvason for helpful discussions. The authors are also grateful to the referees for extensive insightful comments, which led to a number of improvements in the Paper.

[a)]Electronic mail: marsland@mit.edu
[b)]Electronic mail: harvey.brown@philosophy.ox.ac.uk
[c]Electronic mail: valente@pitt.edu

[1]Accelerating clocks differ from inertial ones in the sense that their behavior generally depends on their constitution; see Harvey R. Brown, "The role of rods and clocks in general relativity and the meaning of the metric field," e-print arXiv:0911.4440v1 [gr-qc] (2009).

[2]Stephen Hawking, "The No Boundary Condition And The Arrow Of Time," in *Physical Origins of Time Asymmetry*, edited by J. J. Halliwell, J. Pérez-Mercador, and W. H. Zurek (Cambridge U.P., Cambridge 1994), pp. 346–357; see p. 348.

[3]This point was stressed by Hans Reichenbach in Chapter II, Sec. 5 of his 1956 book *The Direction of Time*, edited by Maria Reichenbach (Dover, Mineola, New York, 1999).

[4]Truesdell claims that all pioneers of thermodynamics regarded time as primitive, in Clifford A. Truesdell and Subramanyam Bharatha, *The Concepts and Logic of Classical Thermodynamics as a Theory of Heat Engines: Rigorously Constructed Upon the Foundation Laid by S. Carnot and F. Reech* (Springer-Verlag, New York, 1977), p. viii.

[5]Owen J. E. Maroney, "Does a computer have an arrow of time?," Found. Phys. **40**, 205–238 (2010).

[6]Ryan Smith, "Do brains have an arrow of time?," Philos. Sci. **81**, 265–275 (2014). An earlier identification of the psychological arrow of time with the thermodynamic arrow, based on the brain as (inter alia) a "registering instrument," is found in Hans Reichenbach, "Les Fondements logiques de la mechanic de quanta," Annales de l'Institut Henri Poincaré, Tome XIII. Fasc. II (Paris, 1953), pp. 156–157. English translations of relevant paragraphs are published in the Appendix to Reichenbach's 1956 book, Ref. 3. Although Reichenbach associated the thermodynamic arrow with non-decreasing entropy, it seems to us that the argument has more to do with the equilibration principle, which we are about to discuss.

[7]Jos Uffink, "Bluff your way in the second law of thermodynamics," Stud. Hist. Philos. Mod. Phys. **32**(3), 305–394 (2001).

[8]For the origin of both these titles, see Harvey R. Brown and Jos Uffink, "The origins of time asymmetry in thermodynamics: The minus first law," Stud. His. Philos. Mod. Phys. **32**(4), 525–538 (2001). This Paper contains a detailed account of the equilibration principle, the emphasis it has received in the literature, and its role in the foundations of thermodynamics.

[9]Meir Hemmo and Orly R. Shenker, *The Road to Maxwell's Demon: Conceptual Foundations of Statistical Mechanics* (Cambridge U.P., Cambridge, 2012), Chap. 2.

[10]See Ref. 4.

[11]Elliott H. Lieb and Jakob Yngvason, "The physics and mathematics of the second law of thermodynamics," Phys. Rep. **310**, 1–96 (1999). The 1999 online version of this Paper, to which our page numbers refer, is at arXiv:cond-mat/9708200v2. A shorter, more informal version of this Paper is found in Elliott H. Lieb and Jakob Yngvason, "A fresh look at entropy and the second law of thermodynamics," Phys. Today **53**(4), 32–37 (2000).

[12]Constantin Carathédory, "Untersuchungen über die Grundlagen der Thermodynamik," Math. Ann. **67**, 355–386 (1909). English translation by Joseph Kestin: "Investigation into the Foundations of Thermodynamics," in *the Second Law of Thermodynamics: Benchmark Papers on Energy*, edited by Joseph Kestin (Dowden, Hutchinson and Ross, Stroudsberg, Pennsylvania, 1976), Vol. 5, pp. 229–256. Page numbers refer to the translation.

[13]See Ref. 7, Sec. 11. A critique of this analysis and a related one in Ref. 8 is found in Leah Henderson, "Can the second law be compatible with time reversal invariant dynamics?," Stud. His. Philos. Mod. Phys. **47**, 90–98 (2014).

[14]See, for example, J. B. Boyling, "An axiomatic approach to classical thermodynamics," Proc. R. Soc. London, Sec A **329**, 35–70 (1972), pp. 42–43.

[15]See, for example, Louis A. Turner, "Simplification of Carathéodory's treatment of thermodynamics. II," Am. J. Phys. **30**, 506–508 (1962), and Francis W. Sears, "Modified form of Carathéodory's second axiom," Am. J. Phys. **34**, 665–666 (1966).

[16]See Ref. 7, Sec. 9.

[17]See Ref. 7. A stronger analogy is with the geometric treatment of the conformal structure of special relativistic space-time found in A. A. Robb, *The Absolute Relations of Time and Space* (Cambridge U.P., Cambridge, 1921); see p. 144 of Jos Uffink, "Irreversibility and the second law of thermodynamics," in *Entropy*, edited by Andreas Greven, Gerhard Keller, and Gerald Warnecke (Princeton U.P., New Jersey, 2003), pp. 121–146.

[18]This is the way the postulate is commonly stated in the literature. Carathéodory's actual axioms was (p. 236) "In every arbitrarily close neighborhood of a given initial state there exist states that cannot be approached arbitrarily closely by adiabatic processes."[12]

[19]More specifically, quasi-static adiabatic processes are defined as those in which the difference between the work done externally and the limit of the energy change defined when the derivatives of the thermodynamic parameters converge uniformly to zero is less than experimental uncertainty. It is noteworthy that Carathéodory does not define quasi-static in the general case.

[20]Turner has argued that the zeroth law is not strictly necessary in Carathéodory's treatment. See Louis A. Turner, "Temperature and Carathéodory's treatment of thermodynamics," J. Chem. Phys. **38**, 1163–1167 (1963).

[21]A modern geometric treatment of Carathéodory's Theorem in the context of the Frobenius Theorem and nonholonomic constraints is found in Theodore Frankel, *The Geometry of Physics*, 2nd ed. (Cambridge U.P., New York, 2004); see Secs. 6.3d–6.3f.

[22]See, for example, Peter T. Landsberg, *Thermodynamics and Statistical Mechanics* (Dover Publications, Inc., New York, 1990), Sec. 5.2; Louis A. Turner, "Simplification of Carathéodory's treatment of thermodynamics," Am. J. Phys. **28**, 781–786 (1960); and Francis W. Sears, "A simplified simplification of Carathéodory's treatment of thermodynamics," Am. J. Phys. **31**(10), 747–752 (1963).

[23]For a discussion of these assumptions, see Peter T. Landsberg, "On suggested simplifications of Carathéodory's thermodynamics," Phys. Status Solidi B **1**(2), 120–126 (1961).

[24]Reference 12, Sec. 9. The recourse to experiment is further justified in Constantin Carathédory, "Über die Bestimmung der Energie und der absoluten Temperatur mit Hilfe von reversiblen Prozessen," Sitzungsberichte der Preussischen Akademie der Wissenschaften, Physikalisch-Mathematische Klasse 1 (1925), pp. 39–47; a discussion of historical responses to this feature of his theory is found in Ref. 7.

[25]See Peter T. Landsberg, "Foundations of thermodynamics," Rev. Mod. Phys. **28**, 363–392 (1956) and J. Dunning-Davies, "Carathéodory's principle and the Kelvin statement of the second law," Nature **208**, 576–577 (1965).

[26]See Arthur E. Ruark, "LXIII. The proof of the corollary of Carnot's Theorem," Philos. Mag. Series 6 **49**(291), 584–585 (1925).

[27]It is remarkable that the connection between the Kelvin-Planck formulation of the second law and Carathédory's inaccessibility principle was clarified only in the 1960s (independently) by B. Crawford, Jr. and I. Oppenheim, "The second law of thermodynamics," J. Chem. Phys. **34**(5), 1621–1623 (1961), and Peter T. Landsberg, "A deduction of Carathéodory's principle from Kelvin's principle," Nature **201**, 485–486 (1964); see also Dunning-Davies in Ref. 25.

[28]B. Bernstein, "Proof of Carathéodory's local Theorem and its global application to thermostatics," J. Math. Phys. **1**, 222–224 (1960). See also Ref. 7, pp. 367–368.

[29]This is related to the local nature of Frobenius' Theorem. See Ref. 21, pp. 183, 184. Note that the energy ambiguity is again resolved by appeal to experience on p. 186.

[30]Uffink has argued that the definition of adiabatic accessibility therein is subtly different from most definitions in the literature—and in particular that of Carathéodory; see Ref. 7, p. 381 *ff*. We need not dwell on this point.

[31]The first incentive strikes us as unnecessary, given that Carathéodory had shown that heat is not fundamental, and that it is perfectly well-defined given his first law and the conservation of energy, both of which feature in the L-Y approach. Their claim (p. 6) that no one "has ever seen heat, nor will it ever be seen, smelled or touched"[11] can also be said to apply to entropy! The second incentive also strikes us as curious, particularly because the authors themselves stress (Sec. G) that if irreversible processes did not exist, "it would mean that nothing is forbidden, and hence there would be no second law."[11]

[32]Here we part company with Ref. 11, p. 24, where it is claimed that if all states are adiabatically equivalent, entropy is constant.

[33]Reference 11, p. 44. There are two notions of universality involved here, and Lieb and Yngvason are anticipating two theorems. The first is Theorem 3.3, which states that the forward sectors *of all states* in the state space of a simple system (i.e., the set of states adiabatically accessible from the given state) point the same way energy-wise. The second is Theorem 4.2, which states that the forward sectors *of all simple systems* point the same way. It is worth emphasizing that the proofs of these theo-rems depend on several axioms additional to the set A1-A6 needed for the entropy principle. The content of one of these additional axioms, A7, is discussed in Sec. V D below.

[34]This is Theorem 3.4, Ref. 11, p. 45.

[35]See, for example, both Papers in Ref. 15.

[36]For many systems, the axiom is a direct consequence of the equilibration principle, which implies that for any two samples of a gas in states X_1 and X_2, when you partition the samples as just described and then connect the two partitioned regions, the resulting system will eventually reach an equilibrium state Y. This process can be done adiabatically, and by conservation of energy state Y will have energy $U = tU_1 + (1-t)U_2$. By assumption, the state also has volume $V = tV_1 + (1-t)V_2$. Compare with p. 31 in Ref. 11.

[37]From Sec. III onwards in Ref. 11, the state space for a simple system is taken to be a convex subset of $\mathbf{R}^{(n+1)}$, n being the number of deformation coordinates.

[38]For a discussion of the continuity of such an adiabat, see Ref. 11, p. 42.

[39]Figure 1 is a special case of Fig. 3 in Ref. 11, pp. 32 and 96. Orthogonality is defined with respect to the canonical scalar product on $\mathbf{R}^{(n+1)}$.

[40]The role of A7 in determining the directionality of processes in the L-Y approach is also pointed out and explained by Henderson in Ref. 13, though in a different way to our own.

[41]See Theorem 5.4, Ref. 11, pp. 66–67.

[42]See Ref. 11, pp. 44 and 62.

[43]Compare the discussion of the axiom T_1 of thermal contact in Ref. 11, p. 52.

[44]This equation is a consequence of the definition of pressure and a number of continuity assumptions in the L-Y formulation; see pp. 40–41 in Ref. 11.

[45]Note that the minus sign is omitted in Eq. (3.6) therein; this is not the case in Elliott H. Lieb and Jakob Yngvason, "A guide to entropy and the second law of thermodynamics," e-print arXiv:math-ph/9805005v1 (1998): see Eq. (14).

For an introductory explanation of thermodynamic stability conditions, see Frederick Reif, *Fundamentals of Statistical and Thermal Physics* (McGraw-Hill, Boston, 1965).

We are grateful to Jakob Yngvason for clarification of this point.