Gibbs, Boltzmann, and negative temperatures

Daan Frenkel, and Patrick B. Warren

Gibbs, Boltzmann, and negative temperatures

Daan Frenkel
Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom

Patrick B. Warren
Unilever R&D Port Sunlight, Quarry Road East, Bebington, Wirral CH63 3JW, United Kingdom

In a recent paper, Dunkel and Hilbert [Nat. Phys. **10**, 67–72 (2014)] use an entropy definition due to Gibbs to provide a "consistent thermostatistics" that forbids negative absolute temperatures. Here, we argue that the Gibbs entropy fails to satisfy a basic requirement of thermodynamics, namely, that when two bodies are in thermal equilibrium, they should be at the same temperature. The entropy definition due to Boltzmann does meet this test, and moreover, in the thermodynamic limit can be shown to satisfy Dunkel and Hilbert's consistency criterion. Thus, far from being forbidden, negative temperatures are inevitable, in systems with bounded energy spectra. © *2015 American Association of Physics Teachers.*

I. INTRODUCTION

The concept of negative temperature has often been discussed in the pages of this journal,[1–5] and there has recently been renewed interest in the area.[6–11] Formally, a negative temperature corresponds to a system under conditions in which the number of states $\omega(E)$ in the vicinity of energy E is a decreasing function, to wit $1/T \equiv \partial S/\partial E < 0$, where $S = k_B \ln \omega$ is the Boltzmann entropy. This situation usually arises in a system with a bounded energy spectrum. It has been realized experimentally in a famous early work on nuclear spin systems by Purcell and Pound,[12] and in recent experiments on ultra-cold quantum gases by Braun *et al.*[9] The existence of negative temperatures leads naturally to many questions. Prominent amongst these is whether a Carnot cycle can be constructed with a negative-temperature reservoir, and whether the consequent Carnot efficiency can be greater than one.[5,13–15]

In a recent article,[11] Dunkel and Hilbert argued that the use of Boltzmann's definition of entropy in the above expression for the temperature leads to unphysical predictions.[16] They propose instead to define entropy using Gibbs' notion of "extension in phase space" (see below).[17] They argue that this approach, which they attribute to Gibbs,[18] cannot give rise to negative temperatures nor can it predict the existence of Carnot cycles with an efficiency larger than one. Moreover, they show that the Boltzmann approach gives rise to unphysical predictions for systems with very few degrees of freedom. In what follows, we will ignore the latter point; as Gibbs himself stated on several occasions, it is unreasonable to expect a meaningful correspondence between statistical mechanics and thermodynamics for systems with only a few degrees of freedom. Our discussion will focus only on systems with very many degrees of freedom, i.e., those that are in the proverbial "thermodynamic limit."[19]

The point about negative temperatures and Carnot cycles is more interesting. We shall argue below that Dunkel and Hilbert are mistaken in their attempt to deprecate the Boltzmann entropy, and that instead it is the Gibbs entropy that fails to meet a basic criterion of thermodynamics. In contrast, the Boltzmann entropy does meet this basic test, and we also prove that it satisfies Dunkel and Hilbert's consistency criterion in the thermodynamic limit. With the Boltzmann entropy, negative temperatures are inevitable in

systems with bounded energy spectra, but this is not a problem if one pays attention to the details as explained by Ramsey over half a century ago.[20] To round off, we present a concrete, pedagogical example of a Carnot cycle connecting reservoirs of opposite temperatures, thereby exhibiting a Carnot efficiency greater than one.

II. CRITIQUE OF THE GIBBS ENTROPY

Let us briefly define the key quantities in modern notation. In particular, we will replace Gibbs' "extension in phase space" by $\Omega(E)$, the total number of quantum states of a system with an energy less than or equal to E. For classical systems with many degrees of freedom, $\Omega(E)$ is dominated by the number of states very close to E. To compute the number of states in a narrow interval $\Delta\epsilon$ around E, we simply differentiate $\Omega(E)$ with respect to E, to obtain

$$\frac{\partial \Omega(E)}{\partial E} \Delta\epsilon \equiv \omega(E)\,\Delta\epsilon. \tag{1}$$

Boltzmann's definition of entropy is

$$S_B(E) = k_B \ln \omega(E) + \text{constant}. \tag{2}$$

In contrast, Gibbs also considered other definitions of entropy, including

$$S_G(E) = k_B \ln \Omega(E) + \text{constant}. \tag{3}$$

Of course, Gibbs did not include k_B in his definition but, again, that is immaterial for the remainder of our argument. In what follows, we leave out the additive constant, as it is irrelevant for a discussion of thermal equilibrium and heat engines.

The definition of temperature in thermodynamics is

$$T = \left(\frac{\partial S(E)}{\partial E}\right)^{-1}. \tag{4}$$

For classical systems with many degrees of freedom, the difference in the value of the temperature based on S_B and S_G is negligible; the reason being that for such systems $\omega(E)$ increases very steeply with E, hence a constraint $\epsilon \leq E$ is

almost equivalent to $\epsilon = E$. However, for systems with an energy that is bounded from above, $\omega(E)$ may decrease for large energies, whereas $\Omega(E)$ is monotonically increasing. An example is a system of N non-interacting spins, which we shall discuss in more detail below. In the regime where $\omega(E)$ is not a monotonically increasing function of E, the two definitions of entropy lead to very different results for the temperature of a macroscopic system (one negative, the other positive). Only one can be right. It turns out that, in contrast to what is argued in Ref. 11, to meet a basic requirement of thermodynamics we must use the definition based on Boltzmann's entropy $S_B(E)$.

The nub of our argument turns on the behavior of systems, considered jointly, that are able to exchange energy. We shall argue that the condition for these to be in thermal equilibrium is that they all have the same temperature *à la* Boltzmann. This is essentially what is often termed "the zeroth law of thermodynamics." It implies that the Boltzmann entropy, *and only that*, can be used to construct a universal thermodynamic temperature scale. Without this basis, the whole edifice of classical thermodynamics is built on sand.

To see this, let us briefly restate the key properties that the statistical mechanical entropy and temperature should reproduce in order to correspond with the thermodynamic quantities defined by Clausius:

1. If a closed system is in equilibrium, its entropy must be at a maximum.
2. Heat never spontaneously flows from cold to hot.[21]
3. At the end of one cycle of a reversible heat engine the entropy of the system (i.e., engine plus reservoirs) has not changed.

It should be stressed that Clausius' thermodynamics is based on experimental observations, not on axioms. The simplest way to construct statistical mechanics is to invoke the so-called *ergodic hypothesis*: "A system at a given energy E is equally likely to be found in any of its $\omega(E)$ quantum states."[22] The ergodic hypothesis does *not* apply to $\Omega(E)$. If a system is at a fixed energy E, it does not visit states with a lower energy. These lower-energy states are therefore irrelevant for counting the degeneracy of the system.

Let us consider a closed system with total energy E, comprised two subsystems with energies E_1 and E_2. We assume that the absolute value of the interaction energy between the two subsystems is much smaller than E. In that case, $E = E_1 + E_2$. The total number of states of this system is given by the product

$$\omega_T(E_1, E_2) = \omega_1(E_1) \times \omega_2(E_2). \qquad (5)$$

Of course, we can write the same expression for $\Omega(E_1, E_2)$ but, importantly, if we consider all the states for which $E_1 + E_2 \leq E$, we are *not* describing a system at fixed energy E, but rather a system with a fixed maximum energy E.[23] For classical systems, this distinction is irrelevant in the thermodynamic limit, but for a system with an energy that is bounded from above (say $E \leq E_M$), the difference is crucial. To illustrate this, let us consider an extreme (but perfectly legitimate) example: suppose that we regard $E_2 < E_M$ as merely a cap on the energy of system 2. Then, increasing E_2 by an amount ΔE does not necessarily correspond to an energy transfer of ΔE from system 1 to system 2: there is no relation between ΔE and heat transfer, and all relation with thermodynamics is lost. To establish a link with

thermodynamics, it is essential that the total energy E is equal to the sum $E_1 + E_2$. This condition can only be enforced if we work with subsystems that have well-defined energies, not just systems with an energy below a particular value.

Let us next consider what happens if we allow energy exchange between systems 1 and 2. In equilibrium, the entropy should be at a maximum with respect to a small energy transfer from system 1 to system 2. Hence

$$\frac{\partial \ln \omega_T}{\partial E_1} = 0. \qquad (6)$$

Using the fact that $dE_1 = -dE_2$, we obtain

$$\frac{\partial \ln \omega_1(E_1)}{\partial E_1} = \frac{\partial \ln \omega_2(E_2)}{\partial E_2} \qquad (7)$$

and hence

$$\frac{1}{T_1} = \frac{1}{T_2}. \qquad (8)$$

Note that we obtain this result only if we use the Boltzmann entropy $S = k_B \ln \omega$.

But, should the Gibbs entropy be so lightly dismissed? After all, as Dunkel and Hilbert point out, there is a mathematically rigorous equipartition theorem [Eq. (8) of Ref. 11] that features S_G and would appear to single-out the associated temperature $T_G = (\partial S_G / \partial E)^{-1}$ as a privileged quantity. However, it is dangerous to extrapolate from this that equality of T_G can be used to gauge whether systems are in thermal equilibrium; this can easily lead to absurd results. For example, consider the finite-size spin system in Fig. 1 of Ref. 11. In the population-inverted state (see below for the definition), the Boltzmann temperature is negative but T_G is positive and finite. One can easily construct a classical system (for example, a perfect gas) with the same value of T_G. Should one therefore conclude that the spin system with an inverted population can be in thermal equilibrium with a perfect gas? Of course not! The spin system would lose energy to the gas, raising the (Boltzmann) entropy of both. In the conventional picture, of course, a population-inverted state has a negative Boltzmann temperature and is always "hotter" than a normal system with a positive Boltzmann temperature, so one would always expect heat transfer to take place.

The only remaining point to discuss is Dunkel and Hilbert's consistency criterion. In Ref. 11, it is proved that S_G satisfies this requirement, independent of system size. Dunkel and Hilbert further argue that S_B must therefore fail this requirement because typically $S_B \neq S_G$ for small systems. However, it can be proved that in the thermodynamic limit S_B *does* satisfy the consistency criterion. In Appendix A, we present such a proof. Our conclusion therefore is that S_B meets all the requirements one would expect of a thermodynamic entropy, whereas S_G does not.

III. NEGATIVE TEMPERATURES AND CARNOT CYCLES

Now that we have convinced ourselves we should use the entropy *à la* Boltzmann, the implication is that, if $\omega(E)$ is not a monotonically increasing function, negative temperatures are inevitable. Are negative temperatures a problem? Not

really. First of all, heat still only flows from "hot" (low $1/T$) to "cold" (high $1/T$). But indeed, Carnot efficiencies can be larger than one.[24] This may seem strange, but it violates no known law of nature.[25]

Let us consider a Carnot cycle operating between two heat reservoirs (see Fig. 1). Following the usual sign conventions, let q_1 be the heat that flows *into* the engine from reservoir 1, and let q_2 be the heat that flows *out of* the engine to reservoir 2. Then the first law of thermodynamics (energy conservation) states that the work w done by the engine must equal the net heat absorbed, $q_1 - q_2$. Each stage in the cycle is reversible so that the entropy changes of the engine are respectively $\Delta S_1 = q_1/T_1$ and $\Delta S_2 = -q_2/T_2$. Now, since S is a state function, the total entropy of the engine does not change, and $\Delta S_1 + \Delta S_2 = 0$. This implies
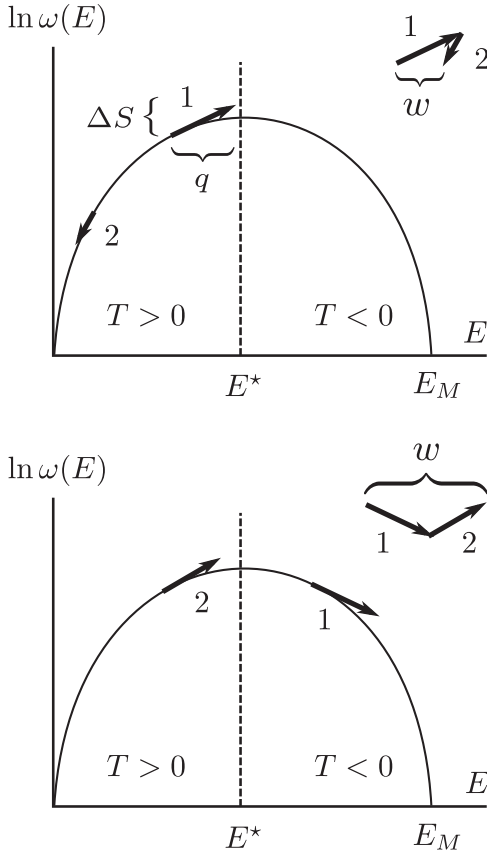

Fig. 1. Plots of entropy vs energy for a Carnot engine (or the working substance thereof) whose temperature can be either positive or negative. The heavy arrows indicate the reversible heat exchanges with reservoirs 1 and 2, respectively. Top panel: If both reservoirs have positive temperatures, then the condition that S is constant implies that because heat flows into the engine from reservoir 1, heat must flow out of the engine into reservoir 2. Since $\Delta S = q/T$ and $1/T = \partial S/\partial E$, it follows that $\partial S/\partial E = \Delta S/q$ and the arrows have a specific geometric interpretation: the horizontal displacement is the heat gain q and the vertical displacement is the entropy change ΔS. The Carnot cycle itself can then be represented by combining the arrows vectorially since in the adiabatic steps $q = \Delta S = 0$. The resultant vector should have no net vertical displacement since the net entropy change should vanish; the net horizontal displacement represents w, the amount of energy available to do useful work. This graphical construction is shown in the inset, demonstrating visually that $w < q_1$ so that the Carnot efficiency is less than one. Bottom panel: If reservoir 1 instead has a negative temperature, then the condition that S is constant implies that because heat still flows into the engine from reservoir 1, heat must also flow *into* the engine from reservoir 2. Here, a similar vector construction demonstrates that $w > q_1$ so that the Carnot efficiency is greater than one.

$$\frac{q_1}{T_1} = \frac{q_2}{T_2}. \tag{9}$$

Suppose now that the "hot" reservoir 1 has a negative temperature $T_1 < 0$, while the "cold" reservoir 2 has a positive temperature $T_2 > 0$ (Fig. 1, bottom panel). In this case, perhaps confusingly, $\Delta S_1 < 0$, but this must be so because when heat flows into the negative-temperature engine, its entropy decreases (and the entropy of the negative-temperature reservoir simultaneously increases as it gives up heat). Moreover, by Eq. (9), when T_1 is negative and T_2 is positive the signs of q_1 and q_2 must be opposite. Hence, if heat flows *into* the engine from reservoir 1 ($q_1 > 0$), then reversibility of the engine requires that heat *also flows into the engine from reservoir 2* ($q_2 < 0$). In such a situation, the work done is

$$w = q_1 - q_2 = q_1 + |q_2|. \tag{10}$$

The efficiency of the Carnot cycle is defined as the ratio

$$\eta = \frac{w}{q_1} = 1 + \frac{|q_2|}{q_1} > 1. \tag{11}$$

Hence, indeed, the Carnot efficiency is larger than one. However, no physical law forbids that.

There is one potential point of concern: it would seem that it is possible to run an engine in contact with a single, negative-temperature heat bath. Whilst this is true, negative-temperature heat baths do not occur naturally: they have to be prepared (in the case of lasers, we call this "pumping"). If the total heat and work budget include the preparation of the negative-temperature heat bath from a system at positive temperature, then it turns out that it is still not possible to run an engine sustainably by extracting heat from a single reservoir. Negative temperatures in real-world laboratory settings do not imply perpetual motion.

IV. SPIN SYSTEMS

Spin systems form an excellent arena to explore the issues raised here. Apart from the three-level system described below, much of this material can be found in the original literature,[20] and textbooks.[13,26] For an interesting mechanical analog of these systems, see Ref. 7.

A. Two-level system

The simplest case to consider is a system with N_0 spins in a ground state with energy zero and N_1 spins in an excited state at energy ϵ, such that the total number of spins $N = N_0 + N_1$ is fixed. Often one can visualize the N_0 spins as "down" and the N_1 spins as "up," but care is needed with this mental picture as we shall consider cases where $\epsilon < 0$ and $N_1 > N_0$ (population inversion). We shall suppose that the spins are distinguishable so that the number of microstates that are consistent with this arrangement of spins is $\omega = N!/(N_0!N_1!)$. For this system, the energy and (Boltzmann) entropy are therefore, respectively,

$$E = \epsilon N_1,$$
$$S = \ln\omega = N\ln N - N_0\ln N_0 - N_1\ln N_1. \tag{12}$$

We have supposed $N \gg 1$ in the entropy expression. For notational simplicity, we set $k_B = 1$ and drop the subscript

"B" from S_B since we shall be exclusively considering the Boltzmann entropy.

Since $N_0 = N - N_1$, and N is fixed, the (Boltzmann) temperature of the spin system is

$$\frac{1}{T} = \frac{\partial S}{\partial E} = \frac{(\ln N_0 + 1) - (\ln N_1 + 1)}{\epsilon} = \frac{1}{\epsilon} \ln \frac{N_0}{N_1}. \tag{13}$$

Therefore, one obtains the familiar Boltzmann result

$$N_1 = N_0 \, e^{-\epsilon/T}, \tag{14}$$

from which it follows that

$$\frac{N_1}{N} = \frac{e^{-\epsilon/T}}{1 + e^{-\epsilon/T}}, \quad \frac{N_0}{N} = \frac{1}{1 + e^{-\epsilon/T}}. \tag{15}$$

The free energy $F = E - TS$ has to be expressed in terms of its "natural" variables T and N. We first have

$$
\begin{aligned}
S &= N \ln N - N_0 \ln N_0 - N_1 \ln N_1 \\
&= (N_0 + N_1) \ln N - N_0 \ln N_0 - N_1 \ln N_1 \\
&= -N_0 \ln(N_0/N) - N_1 \ln(N_1/N),
\end{aligned} \tag{16}
$$

so consequently

$$F = \epsilon N_1 + T N_0 \ln \frac{N_0}{N} + T N_1 \ln \frac{N_1}{N}. \tag{17}$$

We substitute Eq. (15) into this expression to find

$$F = -TN \ln(1 + e^{-\epsilon/T}). \tag{18}$$

This rather neat result follows after a few lines of algebra, left as an exercise for the reader. It can also be derived from the partition function sum, which is left as a further exercise.

From Eq. (18), the quantity conjugate to ϵ is

$$-\frac{\partial F}{\partial \epsilon}\bigg|_{T,N} = -\frac{N \, e^{-\epsilon/T}}{1 + e^{-\epsilon/T}} = -N_1. \tag{19}$$

This makes N_1 the natural variable to describe the arrangement of the spins. For this system, the N_1–ϵ plane is the analog of the p-V diagram encountered in textbooks.[27] The first expression in Eq. (15) serves as an "equation of state," providing isotherms in the N_1–ϵ plane.

For completeness, the entropy as a function of T and N follows most simply by solving $F = E - TS$. Substituting the relevant expressions, one finds (another exercise!)

$$\frac{S}{N} = \frac{\epsilon}{T} \frac{e^{-\epsilon/T}}{1 + e^{-\epsilon/T}} + \ln\left(1 + e^{-\epsilon/T}\right). \tag{20}$$

This function (see Fig. 2; cf. Fig. 14.23 in Ref. 13) depends only on the ratio ϵ/T and is in fact symmetric about $\epsilon/T = 0$, although this is not immediately obvious.

If we examine the derivations, we see that all of the above holds for positive and negative ϵ, and more crucially for positive and negative T as well. If $\epsilon > 0$ and $T > 0$ the spin population is "normal" in the sense that the majority of spins are in the ground state, whereas if $\epsilon > 0$ and $T < 0$ the spin population is "inverted" in the sense that the majority of spins are
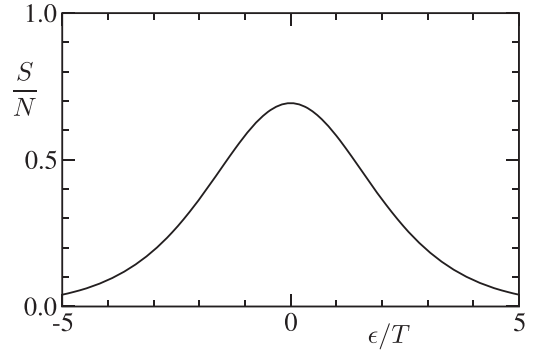

Fig. 2. The entropy function of Eq. (20).

in the excited state. An analogous situation holds, *mutatis mutandis*, if $\epsilon < 0$.

These inverted states are well defined and correspond to a total energy that is closer to the absolute maximum energy $E_M = \epsilon N$ than to $E = 0$, as we now demonstrate. Since $E = \epsilon N_1$, $E_M = \epsilon N$, and $N_0 = N - N_1$, the entropy can be written as

$$\frac{S}{N} = -\left(1 - \frac{E}{E_M}\right) \ln\left(1 - \frac{E}{E_M}\right) - \frac{E}{E_M} \ln \frac{E}{E_M}. \tag{21}$$

This function has the non-monotonic shape shown in Fig. 1 (cf. Fig. 14.22 in Ref. 13), with a maximum at $E^\star = E_M/2$. The spin system therefore has a negative temperature for $E > E_M/2$. The maximum in $S(E)$ corresponds to $1/T = 0$, which suggests that it should be impossible to pass adiabatically (i.e., at constant S) from positive to negative temperature through $1/T = 0$. This is in fact a general result.[2,14]

Isotherms for the spin system in the N_1–ϵ plane are illustrated in Fig. 3, for both positive and negative temperatures.
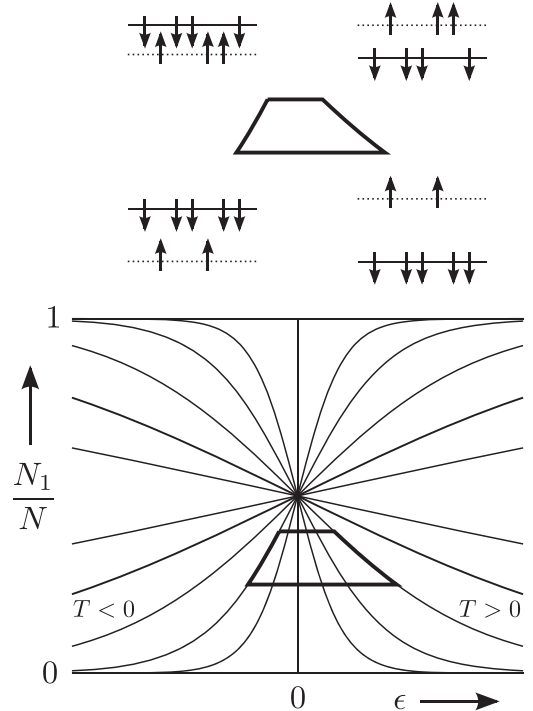

Fig. 3. A Carnot cycle connecting isotherms of opposite temperatures. The thin lines in the N_1-ϵ plane (lower plot) are isotherms, from the first expression in Eq. (15).

The spin system can be moved along an isotherm by connecting it to a thermal reservoir and changing ϵ. The thermal reservoir could be, for example, a much larger spin system, which can be at a positive or negative temperature.

Adiabatic changes in the spin system correspond to changing ϵ without changing the distribution of spins. One can see from Eqs. (12) that keeping N_0 and N_1 fixed leaves the entropy S unchanged. Adiabats in the N_1–ϵ plane in Fig. 3 are therefore horizontal lines. One further point can be made. Since N_0 and N_1 depend on the ratio ϵ/T, it follows that T remains strictly proportional to ϵ in an adiabatic change [cf. Eq. (20)]. This observation implies that we can invert the temperature of the spin system adiabatically by changing the sign of ϵ (but see below for a further discussion).

B. Three-level system

The three-level system is worth considering because it has some novel features not encountered in the two-level system, and it sheds further light on the problems associated with the Gibbs entropy. The main new feature is that a three-level spin system has an internal (macroscopic) degree of freedom, since specifying the energy and total number of spins is insufficient to fix all the population levels.

We start by noting the analog of Eq. (12)

$$
\begin{aligned}
N &= N_0 + N_1 + N_2, \\
E &= \epsilon_1 N_1 + \epsilon_2 N_2, \\
S &= N\ln N - N_0\ln N_0 - N_1\ln N_1 - N_2\ln N_2.
\end{aligned}
\tag{22}
$$

We have now explicitly included the constraint on the total number of spins. The entropy given above is Boltzmann's $S = \ln\omega$. There is absolutely no disagreement that the most likely macroscopic arrangement of spins is the one that maximizes ω, and hence S. The problem then is to maximize S in Eq. (22) with respect to N_0, N_1, and N_2, subject to the constraints on N and E. To solve this problem, the usual (textbook) method is to introduce Lagrange multipliers for the constraints (μ for N and β for E). However, the problem is sufficiently simple that one can also proceed, for example, by eliminating N_1 and N_2 from the expression for S in Eqs. (22) to find S as a function of N_0 alone, which is then maximized. This procedure is something that everyone should try, at least once! A slightly more elegant method is to differentiate all three of Eqs. (22) to get

$$
\begin{aligned}
dN &= dN_0 + dN_1 + dN_2 = 0, \\
dE &= \epsilon_1\,dN_1 + \epsilon_2\,dN_2 = 0, \\
dS &= -(\ln N_0 + 1)\,dN_0 - (\ln N_1 + 1)\,dN_1 \\
&\quad - (\ln N_2 + 1)\,dN_2.
\end{aligned}
\tag{23}
$$

We eliminate dN_1 and dN_2 between these as though they are algebraic quantities to find

$$
dS = -\left(\ln N_0 + \frac{\epsilon_2 \ln N_1}{\epsilon_1 - \epsilon_2} + \frac{\epsilon_1 \ln N_2}{\epsilon_2 - \epsilon_1}\right) dN_0.
\tag{24}
$$

The condition that S is an extremum corresponds to the requirement $dS = 0$ for an infinitesimal change $dN_0 \neq 0$. Hence, the factor multiplying dN_0 in Eq. (24) should vanish when $S(N_0)$ is maximized. After a small rearrangement, this leads to

$$
\frac{1}{\epsilon_1}\ln\frac{N_1}{N_0} = \frac{1}{\epsilon_2}\ln\frac{N_2}{N_0}.
\tag{25}
$$

Let us pause for a moment to see what this result is telling us. An experimentalist, for instance, would note that measurement of the relative population of one level (e.g., N_1/N_0) can be used to make a prediction of the relative population of the other level. Also, if there is a population inversion for one level (e.g., $N_1/N_0 > 1$ when $\epsilon_1 > 0$), there must be population inversion for the other level as well. Most likely, though, the experimentalist will complain that we have dressed a standard result up in rather unfamiliar language. In Eq. (25), let us call the quantity on either side of the equality "$-1/T$" (or "$-1/k_B T$" if we wish to retain the Boltzmann constant). Then, Eq. (25) can be written

$$
N_1 = N_0\, e^{-\epsilon_1/T}, \quad N_2 = N_0\, e^{-\epsilon_2/T}.
\tag{26}
$$

In other words, the populations satisfy the Boltzmann distribution. The actual value of T is determined by the constraints on N and E. The case where $T < 0$ corresponds to population inversion and will inevitably arise if E is made big enough at fixed N. Measuring the relative population of one level is the same as measuring T. When T is known, the relative population of the other level can be predicted. It is easy to see that these results generalize to multilevel systems (this is where the Lagrange multiplier method comes into its own), and it does not matter whether T is positive or negative.

But, what happens to the Gibbs entropy and T_G in this problem? Recall that we are concerned with large systems where in the normally populated state there is no difference between T_G and T. In the population-inverted state, however, T_G diverges exponentially with system size,[28] and has a complicated dependence on the Boltzmann temperature as a function of energy (see Appendix B). From an experimental point of view this means that T_G is difficult to access and provides little in the way of added value.

Note that in the discussion thus far, the three-level spin system has been kept isolated (i.e., in the microcanonical ensemble) and no reservoirs have been involved. We therefore cannot be accused of introducing an assumption of ensemble equivalence by sleight-of-hand. But we feel obliged to point to one more piece of evidence to support the ascendancy of the Boltzmann entropy. Specifically, once T is measured, the spin system is assured of being in thermal equilibrium with a heat reservoir at that same temperature, irrespective of the sign of T.

C. Carnot cycle with efficiency greater than one

We now return to the two-level system and use our knowledge of the thermodynamics outlined above to construct a Carnot cycle that operates between positive- and negative-temperature reservoirs. The cycle is shown as the thick solid line in the N_1–ϵ plane in Fig. 3. The corresponding changes in the spin population are illustrated above the main plot. Our construction can be viewed as a generalization of the adiabatic demagnetization procedure widely used as a refrigeration method in experimental low-temperature physics.[13,26]

Beginning at the upper-left corner and moving counter-clockwise, the cycle starts with a spin system in an inverted state with $\epsilon < 0$ in contact with a "hot" reservoir at some temperature $T_1 < 0$. The first step consists in isothermally increasing $|\epsilon|$. The entropy of the spin system *decreases* in

this step by some finite amount ΔS (cf. Fig. 2, noting that $\epsilon/T > 0$ increases). Since we are proceeding reversibly, the entropy decrease in the spin system is balanced by an entropy increase in the reservoir. Because of the peculiar properties of the negative-temperature "hot" reservoir, this means that the spin system withdraws an amount of heat $q_1 = |\Delta S/T_1|$ from this reservoir. Next, contact with the reservoir is removed and the spin population is inverted adiabatically by reversing the sign of ϵ to arrive at the lower right corner. In the third step, the now normally populated spin system is brought into contact with the "cold" reservoir at $T_2 > 0$, and ϵ is decreased isothermally (thus decreasing ϵ/T). This is continued until the entropy has increased by the exact same amount ΔS lost in the first step. Correspondingly the system also withdraws an amount of heat $q_2 = \Delta S/T_2$ from this reservoir. Finally, the cold reservoir is removed and the population inverted adiabatically once again to return to the upper-left corner. Since the system withdraws heat from both reservoirs, the conventionally defined Carnot efficiency $\eta = 1 - T_2/T_1$ is larger than one, as claimed in the general discussion.

One point needs addressing, since $T \propto \epsilon$ and population inversion is achieved by reversing the sign of ϵ. In a practical procedure, this reversal would be accomplished by reversing the external magnetic field. Clearly, this avoids the aforementioned technical problem of adiabatically connecting regions of opposite temperature through $1/T = 0$, but it means that the system passes through the state where $T = 0$. However, this should also be impossible to achieve, at least in a reversible adiabatic process. The reason is that, according to the third law of thermodynamics, the entropy $S \to 0$ as $T \to 0$ so one cannot connect states with entropies $S > 0$ by an adiabatic (constant S) process that passes through $T = 0$.

This means that any practical procedure can only work by allowing some irreversibility to creep in.[10,12] For example, one can adiabatically decrease the external magnetic field to an arbitrary small value, then suddenly switch the sign of the (now small) field. In this way, the effects of the irreversibility can be controllably confined to the vicinity of $T = 0$. If we incorporate this controlled irreversibility into the Carnot cycle, the efficiency will fall below that of the ideal, reversible cycle, but it seems to us there is no fundamental reason why the efficiency could not still be bigger than one, even if it falls short of the Carnot ideal.

From Eq. (20), the entropy of our spin system satisfies $S = \ln 2$ at $\epsilon = 0$ (cf. Fig. 2). This holds even if subsequently $T \to 0$, in apparent violation of the third law of thermodynamics. The discrepancy arises, of course, because an idealized system of non-interacting spins does not exist in reality. Remnant interactions provide the necessary means to enforce the third law in the real world.[26]

V. DISCUSSION

In summary, our claim is that the Gibbs entropy fails to meet a very basic expectation—the second law of thermodynamics. The Boltzmann entropy does meet this requirement. Moreover, in the thermodynamic limit the Boltzmann entropy also satisfies the consistency criterion demanded by Dunkel and Hilbert. Therefore, we see no reason why the Boltzmann entropy should be displaced from its position as the linchpin connecting statistical mechanics with thermodynamics.

Proponents of the Gibbs entropy may claim that we have taken it out of context, and that it should be discussed only

in relation to isolated systems (i.e., in the microcanonical ensemble). This would be a singularly narrow definition of temperature since it would mean we are not allowed to use it in the context of thermal equilibrium.[29] But, as we have seen for the three-level spin system, even with such an extreme position, the Gibbs entropy provides little in the way of added value.

What are we to make of the fact that the Gibbs entropy satisfies certain exact mathematical theorems, as adduced by Dunkel and Hilbert? Of course we do not dispute these. Rather, we say that for small systems they are evidence of the well-known inequivalence of ensembles, and the difficulty in finding a suitable entropy definition.[30] For large systems, the theorems can be used to prove that the Boltzmann entropy acquires certain desired features. But if the exact theorems lead towards nonsensical conclusions, for example that the temperature diverges exponentially with system size, then what they are telling us is that the *interpretation* is wrong. The original position is seen to become untenable.

With our viewpoint, negative temperatures are inevitable in systems with bounded energy spectra.[31] The necessary extension of the formalism of thermodynamics to treat this was described by Ramsey in 1956.[20] One must abandon the postulate that the entropy must be an increasing function of energy. But as so many times in the history of science, an abandoned postulate opens the door to an enriched formalism, capable of describing phenomena that were not envisaged in the foundational era of the subject.

APPENDIX A: THERMODYNAMIC CONSISTENCY OF S_B

The derivations in this appendix (and the next) use rather more advanced mathematics than we have employed in the main text. We do not claim much originality and indeed similar material can be found in many textbooks.[32,33] In this first appendix, we prove that in the thermodynamic limit S_B satisfies the consistency criterion specified by Dunkel and Hilbert in Ref. 11.

We first prove that in the *canonical* ensemble[33]

$$-\frac{\partial F}{\partial A_\mu}\bigg|_T = -\left\langle \frac{\partial H}{\partial A_\mu} \right\rangle_T. \tag{A1}$$

Here, A_μ is some parameter in the Hamiltonian H, for example, the position of a wall. The left-hand-side (LHS) of this equation is the generalized *thermodynamic* force, corresponding to this parameter. The right-hand-side (RHS) is the canonical ensemble average of the generalized *mechanical* force, corresponding to the same parameter. These forces are said to be conjugate to A_μ. For example, in standard thermodynamics, the quantities p and V form a well-known conjugate pair, and in our spin system the quantities $-N_1$ and ϵ

form another conjugate pair.[34] Equation (A1) is well known and forms the basis of a wide variety of Monte-Carlo free energy sampling methods.

The ensemble average in Eq. (A1) is given by

$$\langle \cdots \rangle_T = \frac{\text{Tr}\left[(\cdots) e^{-\beta H} \right]}{\text{Tr}[e^{-\beta H}]},$$ (A2)

where "Tr" can be read as "sum over states" and $\beta = 1/k_B T$. Provided there is a bounded energy spectrum, the sums remain well defined for $\beta < 0$.

The proof of Eq. (A1) is quite easy and starts from the expression that defines the free energy

$$e^{-\beta F} = \text{Tr}[e^{-\beta H}].$$ (A3)

Differentiating both sides with respect to A_μ gives

$$-\beta \left. \frac{\partial F}{\partial A_\mu} \right|_T e^{-\beta F} = -\beta \, \text{Tr}\left[\frac{\partial H}{\partial A_\mu} e^{-\beta H} \right].$$ (A4)

Equation (A1) is then obtained by dividing Eq. (A4) by Eq. (A3), and canceling β.

In order to recover the consistency criterion of Dunkel and Hilbert, what we need to do is transfer these results to the *microcanonical* ensemble. To do this, we first define the microcanonical ensemble average

$$\langle \cdots \rangle_E = \frac{\text{Tr}\left[(\cdots) \delta(E - H) \right]}{\text{Tr}[\delta(E - H)]},$$ (A5)

and the microcanonical Boltzmann entropy *via*

$$\text{Tr}[\delta(E - H)] = e^{S_B/k_B}.$$ (A6)

In these formulas, the Dirac δ-function selects only those states with energy E.

With these definitions in hand, we can write

$$\text{Tr}[(\cdots) e^{-\beta H}] = \int_0^\infty dE \, e^{-\beta E} \text{Tr}[(\cdots) \delta(E - H)],$$ (A7)

and therefore

$$e^{-\beta F} \langle \cdots \rangle_T = \int_0^\infty dE \, e^{-\beta E + S_B/k_B} \langle \cdots \rangle_E.$$ (A8)

A special case is obtained when $(\cdots) = 1$, namely,

$$e^{-\beta F} = \int_0^\infty dE \, e^{-\beta E + S_B/k_B}.$$ (A9)

Equations (A8) and (A9) have the structure of Laplace transforms, with (E, β) being the transform variables. They are valid irrespective of system size.[35] As the system size increases, though, the integrand in both cases becomes dominated by the peak in the exponential. We can therefore evaluate the integrals by the saddle-point method.[32] Taking Eq. (A9) first, one has

$$-\beta F = -\beta E + \frac{S_B}{k_B}, \quad \text{where} \quad -\beta + \frac{\partial}{\partial E}\left(\frac{S_B}{k_B} \right) = 0.$$ (A10)

The second expression is the condition that the integrand is a maximum, as a function of E. We now recall that $\beta = 1/k_B T$ and so the above can be rearranged to

$$F = E - TS_B, \quad \text{where} \quad \frac{1}{T} = \frac{\partial S_B}{\partial E}.$$ (A11)

Thus the saddle-point method has converted the Laplace transform into a Legendre transform. Equation (A11) looks very familiar but we should emphasize that T is the temperature in the canonical ensemble. So we have also proved that, in the thermodynamic limit

$$T = T_B.$$ (A12)

This can be viewed as an expression of the zeroth law of thermodynamics.

Turning now to Eq. (A8) and proceeding in the same way, the thermodynamic functions cancel, leaving

$$\langle \cdots \rangle_T = \langle \cdots \rangle_E.$$ (A13)

This confirms that ensemble averages are equivalent in the thermodynamic limit.

The next step in the proof consists in differentiating the first of Eqs. (A11) with respect to A_μ, paying careful attention to the dependent variables. One finds

$$\left. \frac{\partial F}{\partial A_\mu} \right|_T = \left(1 - T \frac{\partial S_B}{\partial E} \right) \frac{\partial E}{\partial A_\mu} - T \left. \frac{\partial S_B}{\partial A_\mu} \right|_E.$$ (A14)

The first term on the RHS vanishes, by virtue of the second of Eqs. (A11), and therefore

$$\left. \frac{\partial F}{\partial A_\mu} \right|_T = -T \left. \frac{\partial S_B}{\partial A_\mu} \right|_E.$$ (A15)

Recalling Eqs. (A1), (A12), and (A13), we have now proved

$$T_B \left. \frac{\partial S_B}{\partial A_\mu} \right|_E = -\left\langle \frac{\partial H}{\partial A_\mu} \right\rangle_E.$$ (A16)

This is exactly Dunkel and Hilbert's consistency criterion [cf. Eq. (7) in Ref. 11]. It holds in the thermodynamic limit and is our desired result.[36]

APPENDIX B: SOME PROPERTIES OF T_G

In this second appendix, we analyze the behavior of T_G in the limit of large system size. Starting from the definitions, $k_B T_G = (\partial \ln \Omega / \partial E)^{-1}$ and $\partial \Omega / \partial E = \omega$, we find $k_B T_G = \Omega/\omega$, as in Ref. 11. From this, we obtain

$$T_G = k_B^{-1} \int_0^E dE' \, e^{[S_B(E') - S_B(E)]/k_B}.$$ (B1)

This compact expression essentially contains everything we need to know about the behavior of T_G. The argument of the exponential is proportional to the system size (is extensive). Therefore, for large systems, the value of T_G is determined by the location of the maximum.

There are two cases. If S_B is monotonically increasing up to E, the integral is determined by the behavior at the upper limit of integration. Setting $\Delta E = E - E'$, we first have

$$S_B(E') - S_B(E) = -\frac{\Delta E}{T_B} + O(\Delta E^2), \tag{B2}$$

since $1/T_B = \partial S_B /\partial E$. Changing the integration variable to ΔE we then have

$$T_G = k_B^{-1} \int_0^E d(\Delta E)\, e^{-\Delta E/k_B T_B}. \tag{B3}$$

The upper integration limit can be replaced by ∞ since the correction becomes vanishingly small in the limit of a large system size. The integral can then be done

$$T_G = k_B^{-1} \int_0^\infty d(\Delta E)\, e^{-\Delta E/k_B T_B} = T_B. \tag{B4}$$

Therefore, we conclude that as long as S_B is monotonically increasing, $T_G = T_B$ in the thermodynamic limit.

In the second case, the maximum of the argument of the exponential in Eq. (B1) occurs at $E' = E^\star$, somewhere between the integration limits [note that $S_B(E)$ is just an offset]. Approximating the integral by the maximum value of the integrand, we find

$$T_G = k_B^{-1} e^{[S_B(E^\star) - S_B(E)]/k_B}. \tag{B5}$$

This demonstrates that to leading order T_G grows exponentially with system size.[28] Again recalling that $1/T_B = \partial S_B /\partial E$, we have finally

$$T_G = k_B^{-1} \exp\left(-\int_{E^\star}^E \frac{dE'}{k_B T_B}\right). \tag{B6}$$

Hence, to leading order T_G acquires a complicated dependence on the Boltzmann temperature between $E^\star$ and E. Note that $T_B < 0$ in Eq. (B6) so T_G is still an increasing function of E.

[1] R. J. Tykodi, "Negative Kelvin temperatures: Some anomalies and a speculation," Am. J. Phys. **43**, 271–273 (1975).

[2] A.-M. Tremblay, "Comment on: Negative Kelvin temperatures: Some anomalies and a speculation," Am. J. Phys. **44**, 994–995 (1976).

[3] R. J. Tykodi, "Negative Kelvin temperatures," Am. J. Phys. **44**, 997–998 (1976).

[4] A. Danielian, "Remarks on Tykodi's note on negative Kelvin temperatures," Am. J. Phys. **44**, 995 (1976).

[5] R. J. Tykodi, "Quasi-Carnot cycles, negative Kelvin temperatures, and the laws of thermodynamics," Am. J. Phys. **46**, 354–359 (1978).

[6] A. P. Mosk, "Atomic gases at negative kinetic temperature," Phys. Rev. Lett. **95**, 040403 (2005).

[7] J. Arnaud, L. Chusseau, and F. Philippe, "Mechanical equivalent of quantum heat engines," Phys. Rev. E **77** 061102 (2008).

[8] A. Rapp, S. Mandt, and A. Rosch, "Equilibration rates and negative absolute temperatures for ultra-cold atoms in optical lattices," Phys. Rev. Lett. **105**, 220405 (2010).

[9] S. Braun *et al.*, "Negative absolute temperature for motional degrees of freedom," Science **339**, 52–55 (2013).

[10] V. Romero-Rochín, "Nonexistence of equilibrium states at absolute negative temperatures," Phys. Rev. E **88**, 022144 (2013).

[11] J. Dunkel and S. Hilbert, "Consistent thermostatistics forbids negative absolute temperatures," Nat. Phys. **10**, 67–72 (2014). This paper rediscovers arguments given in V. Berdichevsky, I. Kunin, and F. Hussain, "Negative temperature of vortex motion," Phys. Rev. A **43**, 2050–2051 (1991).

[12] E. M. Purcell and R. V. Pound, "A nuclear spin system at negative temperature," Phys. Rev. **81**, 279–280 (1951).

[13] M. W. Zemansky, *Heat and Thermodynamics*, 5th ed. (McGraw-Hill, New York, 1957).

[14] E. C. G. Stueckelberg and P. B. Scheurer, *Thermocinétique Phénoménologique Galiléenne* (Birkhäuser, Basel, 1974).

[15] P. T. Landsberg, "Heat engines and heat pumps at positive and negative temperatures," J. Phys. A: Math. Gen. **10**, 1773–1780 (1977).

[16] Strictly speaking, the definition is due to Planck rather than Boltzmann; see M. Planck, "Ueber das gesetz der energieverteilung im normal spectrum," Ann. Phys. (Leipzig) **309**, 553–563 (1901).

[17] J. W. Gibbs, *Elementary Principles of Statistical Mechanics* (Charles Scribner's Sons, New York, 1902).

[18] This claim is not fully substantiated by Chapter XIV of Ref. 17; see pp. 171–172 and Eq. (488) on p. 174.

[19] A. L. Kuzemsky, "Thermodynamic limit in statistical physics," Int. J. Mod. Phys. B **28**, 1430004 (2014).

[20] N. F. Ramsey, "Thermodynamics and statistical mechanics at negative absolute temperatures," Phys. Rev. **103**, 20–28 (1956).

[21] The crucial word here is "never."

[22] This statement makes more sense if we introduce the width of the energy shell $\Delta\epsilon$. However, the ergodic hypothesis requires only that $\Delta\epsilon$ is so small that $\omega(E)$ is constant between E and $E + \Delta\epsilon$. We note that the ergodic hypothesis is not disputed by Dunkel and Hilbert in Ref. 11.

[23] A topical example may clarify this point. Suppose that there is social unrest in the banking sector because bank employees working in branch offices (the "proletariat") find that their wages are too low, and that those of the executives (the "bourgeoisie") are too high. This problem could be solved by raising the average salary per employee, while decreasing the average salary (including bonuses) for executives, such that the total sum spent on salaries remains the same. That would be the "Boltzmann" solution. The "Gibbs" solution according to Ref. 11 would be to increase the *maximum* salary of bank employees, for example, to $1M/yr, a measure from which very few, if any, employees would benefit, whilst decreasing the maximum salary for executives accordingly (say from $10bn/yr to $1bn/yr), such that the maximum amount that could be spent on salaries remains the same. Most likely, very few executives would protest. But more importantly, this measure would not result in any substantial transfer of money (our equivalent of "heat") from one reservoir to the other. In Marxist terminology, there is no redistribution of wealth from the bourgeoisie to the proletariat. Hence, it is not a strategy that is likely to satisfy the bank employees.

[24] We use throughout the standard definition of Carnot efficiency. As we explain in the text, with reservoirs of opposite temperatures, heat is also extracted from the cold reservoir. If one defines the efficiency to be the ratio of the work done to the heat extracted from *both* reservoirs, the efficiency can never be greater than one.[7,15]

[25] Excepting that a minor improvement is required in the Kelvin-Planck formulation of the second law.[13,15]

[26] J. R. Waldram, *The Theory of Thermodynamics* (Cambridge U.P., Cambridge, 1985), pp. 186–190.

[27] Writing $dF = (N_1/N)\, dE_M$ makes this precise: $-N_1/N$ is the analogue of p, and $E_M = \epsilon N$ is the analogue of V.

[28] J. M. G. Vilar and J. Miguel Rubi, "Communication: System-size scaling of Boltzmann and alternate Gibbs entropies," J. Chem. Phys. **140**, 201101 (2014).

[29] With this standpoint, it is then surely illogical to discuss the Gibbs entropy in the context of heat engines.

[30] J. Naudts, "Boltzmann entropy and the micro-canonical ensemble," Europhys. Lett. **69**, 719–724 (2005).

[31] For a related perspective see U. Schneider *et al.*, "Comment on 'Consistent thermostatistics forbids negative absolute temperatures,'" arXiv:1407.4127.

[32] E. Schrödinger, *Statistical Thermodynamics* (Cambridge U.P., Cambridge, 1952).

[33] T. L. Hill, *Statistical Mechanics: Principles and Selected Applications* (McGraw-Hill, New York, 1956).

[34] In the spin system, the Hamiltonian is $H = \epsilon \sum_{i=1}^M s_i$, where s_i is zero or one according to whether the ith spin is in the ground state or excited state. Equation (A1) is then satisfied almost trivially since $\langle \partial H/\partial\epsilon \rangle = \langle \sum_{i=1}^N s_i \rangle = N_1$.

[35] This demonstrates the well-known ensemble inequivalence for small systems, since in that case $-\partial F/\partial T \neq S_B$.

[36] Another proof proceeds by considering in addition to the Gibbs entropy $S_G = k_B \ln \mathrm{Tr}[\Theta(E - H)]$ the complementary entropy $\bar{S}_G = k_B \ln \mathrm{Tr}[\Theta(H - E)]$. By adapting Eq. (7) in Ref. 11, one can easily show that $\bar{S}_G$ *also* meets the consistency criterion. Then, from S_G and $\bar{S}_G$, a piecewise approximation to S_B can be constructed which becomes exact in the thermodynamic limit.