Explanation of the Gibbs paradox within the framework of quantum thermodynamics

A. E. Allahverdyan[1] and Th. M. Nieuwenhuizen[2]

[1]*Yerevan Physics Institute, Alikhanian Brothers St. 2, Yerevan 375036, Armenia*
[2]*Institute for Theoretical Physics, University of Amsterdam, Valckenierstraat 65, 1018 XE Amsterdam, The Netherlands*

The issue of the Gibbs paradox is that when considering mixing of two gases within classical thermodynamics, the entropy of mixing appears to be a discontinuous function of the difference between the gases: it is finite for whatever small difference, but vanishes for identical gases. The resolution offered in the literature, with help of quantum mixing entropy, was later shown to be unsatisfactory precisely where it sought to resolve the paradox. Macroscopic thermodynamics, classical or quantum, is unsuitable for explaining the paradox, since it does not deal explicitly with the difference between the gases. The proper approach employs quantum thermodynamics, which deals with finite quantum systems coupled to a large bath and a macroscopic work source. Within quantum thermodynamics, entropy generally loses its dominant place and the target of the paradox is naturally shifted to a decrease of the maximally available work before and after mixing (mixing ergotropy). In contrast to entropy this is an unambiguous quantity. For almost identical gases the mixing ergotropy continuously goes to zero, thus resolving the paradox. In this approach the concept of "difference between the gases" gets a clear operational meaning related to the possibilities of controlling the involved quantum states. Difficulties which prevent resolutions of the paradox in its entropic formulation do not arise here. The mixing ergotropy has several counterintuitive features. It can increase when less precise operations are allowed. In the quantum situation (in contrast to the classical one) the mixing ergotropy can also increase when decreasing the degree of mixing between the gases or when decreasing their distinguishability. These points go against a direct association of physical irreversibility with lack of information.

I. INTRODUCTION

Studying mixtures and mixing processes is one of the oldest tasks of thermodynamics. Perhaps the most celebrated aspect of this task is the Gibbs paradox: the entropy increase upon mixing two different gases stays finite for an arbitrary small difference between the gases, but is zero for identical gases.

This paradox is discussed in many textbooks on thermodynamics and statistical physics—e.g., in [1,2]—and it created a vast amount of literature during the last hundred years until our day [3–17]. It was stated to be of a high principal and methodical value [4], since it displays the limits of applicability for classical (phenomenological) thermodynamics: the resolution of the paradox (if any) ought to lie outside this discipline. Already several times the paradox was claimed to be resolved, but each time it was reconsidered and seen as an open issue again.

The present status of the problem is somewhat controversial. The existing opinions can be roughly summarized as follows.

(i) The paradox is resolvable within the information theoretical approach already in classical statistical physics [11,13].

(ii) The most natural resolution of the paradox has been achieved within quantum statistical physics [4–9] thanks to the feature of partial distinguishability.

(iii) The quantum situation presents a natural setting for the resolution, but there is a specifically quantum peculiarity of the problem (induced by noncommutativity) which still prevents its ultimate resolution [14]. Thus, the Gibbs paradox in quantum statistical physics has so far not been resolved.

We share the last opinion. Our purpose is to present an explanation of this thermodynamical paradox starting from the first principles of quantum mechanics.

This is the program of *quantum thermodynamics*; see [18] for a short review. The crucial point in our discussion of the Gibbs paradox is to realize that it has to be formulated in terms of the available work, as was already realized by Landé in 1955 [6]. (Within the setup of classical thermodynamics this formulation is equivalent to the entropic one [6].) In contrast to entropy, the available work—by definition an ensemble average—is a well-defined quantity for any equilibrium or nonequilibrium state even of small quantum systems and it is a function of both the state and class of work sources employed for work extraction. Moreover, the features of work are grounded directly on the first principles of quantum mechanics. On top of that, the amount of available work adequately reflects intuition usually associated with entropy, such as being a measure of nonequilibrium or disorder.[1]

The above first-principles properties lead to the resolution of the paradox in terms of mixing work which shows perfectly continuous behavior when the difference between the gases goes to zero. Difficulties which prevent resolutions of the paradox in its entropic formulation do not arise here.

---

[1]In fact the priority of the available work over entropy was adequately understood already by Clausius; see discussions in [20]. Another example is Schrödinger, who in his famous book [21] gave importance to (neg)entropy for characterizing survival of organisms, but later on admitted that he should have spoken in this context about the available work rather than entropy.

Certain aspects of the proposed scheme—using work instead of entropy, time-scale separation, etc.—were already anticipated in literature—e.g., in [6,15]. However, these anticipations were conceived only in the framework of phenomenological thermodynamics, and this is why the resolution in terms of mixing work was not achieved.[2]

Our paper is organized as follows. In Sec. II we recall the classical formulation of the Gibbs paradox. The next section reviews the mixing entropy argument, an attempt to solve the paradox with help of quantum entropy. Section IV discusses in detail why this argument cannot be considered as a resolution of the paradox. Two basic reasons for this are outlined, and several pertinent issues are discussed. Section V presents the resolution of the paradox with help of mixing work. In Sec. VI we discuss the mixing work in the contexts of instruments available for work extraction. The analysis fully embodies the idea that the difference between two substances is first of all an operational notion and should not be given any absolute status [23]. Moreover, it appears that the dependence of the mixing work on the available instruments is nontrivial, since it can both increase or decrease upon introducing restrictions on those instruments. Though the mixing work is zero when mixing identical substances, in the quantum situation it can be a nonmonotonous function of the degree of mixing and of the (information-theoretic) distinguishability between the mixing substances. These are shown in Secs. VII and VIII, respectively. The last section presents our conclusions. Appendix A discusses definitions of entropy and their relations to the second law; Appendix B recollects several formulas.

II. CLASSICAL FORMULATION OF THE GIBBS PARADOX

Consider two reservoirs each one having volume V. They are separated by a wall and are filled with different ideal[3] Boltzmann gases—e.g., with two different isotopes of the same substance. The difference is not specified, but assumed to be tunable. The number of particles, N, pressure P, and temperature T in each reservoir are the same. The entropy S of each gas is [1][4]

$$S(N,V) = N \ln \frac{V}{N}. \tag{1}$$

Since the gases do not interact, the total entropy reads

$$S_{\mathrm{i}} = S_1(N,V) + S_2(N,V) = 2N \ln \frac{V}{N}. \tag{2}$$

Now remove the wall. The overall system of the two gases is assumed to be thermally isolated (the only influence of the external fields is in removing the wall).[5] The gases will mix, and after some transient time, a new equilibrium state is reached. Since in this state gases still do not interact, the final entropy S_{f} can be obtained again as a sum of two partial entropies, every component with N particles distributed in the volume $2V$,

$$S_{\mathrm{f}} = S_1(N,2V) + S_2(N,2V) = 2N \ln \frac{2V}{N}. \tag{3}$$

Thus the mixing entropy reads

$$\Delta S = S_{\mathrm{f}} - S_{\mathrm{i}} = 2N \ln 2. \tag{4}$$

The additional contribution $2N \ln 2$ arose due to the irreversible process of mixing, and it *does not depend* on any quantitative measure of the difference between the ideal gases.

Now consider the same process, but assume that initially the gases are identical. After removing the wall, Eq. (4) does not predict any entropy change. Indeed, in the final state we have a one-component gas with total number of particles, $2N$, in the volume $2V$. Thus, from Eq. (1),

$$S_{\mathrm{f}} = 2N \ln \frac{2V}{2N}$$

and this equals S_{i}, so $\Delta S = 0$. This is, of course, the expected and consistent result, since there is no irreversibility when mixing two identical gases in equilibrium; see in addition below and footnote 22. Thus we have arrived at the Gibbs paradox [3–5][6]:

When varying continuously the difference between the

---

[2]The to-be-presented resolution of the Gibbs paradox resembles the recent solution for the Maxwell demon problem presented by Scully and co-workers [22]: both find their basis in quantum thermodynamics—that is to say, the thermodynamics of small quantum systems coupled to a macroscopic bath and work source, the latter leading to a time-dependent Hamiltonian. Another recent result of quantum thermodynamics is our report on the breakdown of the Landauer inequality for the energy needed to erase one bit of information [19].

[3]For simplicity we choose to work with ideal gases. The ideality is not an issue for the Gibbs paradox: it exists for nonideal gases as well [9], and the resolution obtained below for ideal gases will be generalizable to the nonideal situation.

[4]In formula (8) we omitted a term $N f_m(T)$ with f being some function of temperature—e.g., $f_m(T) = \frac{5}{2} - \frac{3}{2}\ln[(mT)/(2\pi\hbar^2)]$ for a monoatomic gas. This term does not play any role in our discussion, since it drops out from the entropy difference. One also should not

be troubled by the presence of the dimension inside of the logarithm in Eq. (1), because it is canceled by the one of $f_m(T)$, while in our further discussion it drops out anyhow when calculating entropy differences.

[5]We shall focus on the mixing in the thermally isolated system. For ideal gases this coincides with the isothermal mixing, since the energy U of such a gas depends only on its temperature: $U/N = f_m(T) - T f'_m(T)$, where the function $f_m(T)$ is discussed in Footnote 4. In general (i.e., for nonideal gases), there will be a difference between the isothermal mixing, where the temperature is kept constant during the whole process with help of an external thermal bath, and the mixing in the thermally isolated system, where the final temperature is determined by the constancy of the overall energy. During the isothermal mixing the gases will exchange some energy with the bath (mixing heat).

[6]The paradox is not always formulated correctly in literature; see [9] for a detailed criticism. Some authors define entropy as $N \ln V$ and see the paradox in increasing the entropy when mixing two identical gases. Others think that the paradox is resolved by the very

gases, the entropy defined according to Eqs. (1), (2), and (4) changes discontinuously.

It is to be stressed that the existence of the Gibbs paradox is not connected with the thermodynamic limit $N \rightarrow \infty$. As discussed in [8,9], the finite-N situation does bring some differences in the expression for the entropy of mixing, but the paradox survives; see in addition footnote 22.

III. MIXING ENTROPY ARGUMENT

A. Assumptions of the argument

It was realized by many scholars that the origin of the paradox is that the difference between the gases is only assumed, but does not show up explicitly in Eqs. (1)–(4); i.e., the description that led to the discontinuity is *not sufficiently complete* [6–9]. In that respect the paradox demonstrates the limits of applicability of phenomenological thermodynamics.

It is expected that for two ideal gases the difference will be related to the internal states of their atoms [6–10]:

(a) Indeed, besides the translational motion which contributes to the entropy (8), the atoms of the gases also have internal states (e.g., spin states). These states are typically described by quantum mechanics. For Boltzmann gases the internal states of the atoms are decoupled from the translational motion. Returning to the above example of different gases in two reservoirs, let us assume that the first and second reservoirs contain atoms in internal states described by density matrices

$$\rho_1 \text{ and } \rho_2, \qquad (5)$$

respectively.[7]

One of the main points in taking the internal states into account is that now from the very beginning we can treat the two gases as identical, but being in different internal states ρ_1 and ρ_2 [8–10]. This is similar to what happens in nuclear physics, where the neutron and proton are considered as identical particles (nucleons) in different states distinguished by the value of the isotopic spin.

(b) After removing the wall, the gases mix. We shall assume that the time scale on which the internal states of the gases change is much larger than the time scale related to mixing of the translational degrees of freedom.

(c) Thus after the mixing, the internal states will be described by the density matrix ($M=2$)

___
fact of not having any entropy increase when mixing identical gases. To avoid confusion, we stress that the paradox is in the discontinuous change of entropy when tuning the difference between the gases. A closely related point—which can also be viewed as paradoxical—is that the mixing entropy does not depend on the actual difference between the gases, provided this difference is not zero.

[7]Recall that the density matrix—as well as the wave function—refers to an ensemble of identically prepared systems; see, e.g., [2]. Thus by "state of a particle" we necessarily mean the density matrix of the ensemble to which this particle belongs.

$$\rho = \sum_{\alpha=1}^{M} \lambda_\alpha \rho_\alpha. \qquad (6)$$

Since two equal amounts of gases are mixed, the probability (weight) factors are equal, $\lambda_1 = \lambda_2 = \frac{1}{2}$.

The same equation (6) applies for the mixing of M gases with number of particles, $\{N_\alpha\}_{\alpha=1}^{M}$, and initial density matrices, $\{\rho_\alpha\}_{\alpha=1}^{M}$; the corresponding weights are

$$\lambda_\alpha = \frac{N_\alpha}{\sum_{\alpha=1}^{M} N_\alpha}, \quad \alpha = 1, \ldots, M. \qquad (7)$$

For the details of this generalization see Sec. III C.

B. Implementation of the argument [6–9]

Due to the above decoupling feature, the total entropy of the translational motions and the internal states of each gas is defined as (recall Footnote 4)

$$S_k(N,V) = N \ln \frac{V}{N} + N S_{vN}(\rho_k), \quad k = 1,2, \qquad (8)$$

$$S_{vN}(\rho) \equiv -\text{tr}[\rho \ln \rho], \qquad (9)$$

where ρ_k are given by (5), and where $S_{vN}(\rho)$ is the von Neumann entropy.

The initial entropy of the two gases is the sum of two contributions (recall that $N_1 = N_2 = N$)

$$S_i = 2N \ln \frac{V}{N} + N S_{vN}(\rho_1) + N S_{vN}(\rho_2), \qquad (10)$$

while the final entropy reads

$$S_f = 2N \ln \frac{2V}{2N} + 2N S_{vN}(\rho). \qquad (11)$$

Recall that we treat two gases as identical; so in the final state there is a single gas having $2N$ particles in volume $2V$. The mixing entropy $\Delta S = S_f - S_i$ thus reads

$$\Delta S = 2N \left[S_{vN}(\rho) - \frac{1}{2} S_{vN}(\rho_1) - \frac{1}{2} S_{vN}(\rho_2) \right]. \qquad (12)$$

Assume that the internal states were maximally different—i.e., orthogonal:

$$\rho_1 \rho_2 = 0. \qquad (13)$$

Such states can be distinguished by a single measurement, i.e.; if it is known that the state of a given single atom belongs to an ensemble described by either ρ_1 or ρ_2, then a single measurement suffices to establish the identity of the state. In this respect orthogonal states are similar to the classical case (perfect distinguishability). It is seen from definitions (6) and (9) that

$$S_{vN}(\rho) = -\text{tr}\left[\frac{\rho_1}{2} \ln \frac{\rho_1}{2} \right] - \text{tr}\left[\frac{\rho_2}{2} \ln \frac{\rho_2}{2} \right] \qquad (14)$$

and that the mixing entropy $\Delta S = 2N \ln 2$ agrees with the prediction (4) of classical thermodynamics.

The other extreme is when the states are identical,

$$\rho_1 = \rho_2, \tag{15}$$

which implies $\Delta S = 0$, again in agreement with the prediction of classical thermodynamics.

In general, if neither Eq. (13) nor (15) is true, the states ρ_1 and ρ_2 are only partially distinguishable; i.e., any finite number of measurements will distinguish these states with a finite error. Assume the states are pure:

$$\rho_1 = |a_1\rangle\langle a_1| \quad \text{and} \quad \rho_2 = |a_2\rangle\langle a_2|. \tag{16}$$

Noting the spectrum

$$\text{Spec}\left\{\frac{1}{2}|a_1\rangle\langle a_1| + \frac{1}{2}|a_2\rangle\langle a_2|\right\} = \frac{1}{2}(1 \pm |\langle a_1|a_2\rangle|), \tag{17}$$

we get, from Eq. (12),

$$\frac{\Delta S}{2N} = -\text{tr}[\rho \ln \rho] = h\left(\frac{1 - |\langle a_1|a_2\rangle|}{2}\right), \tag{18}$$

$$h(x) \equiv -x \ln x - (1-x)\ln(1-x). \tag{19}$$

This expression is minimal and equal to zero for identical gases $|\langle a_1|a_2\rangle| = 1$. It is maximal and equal to $2N \ln 2$ for totally distinguishable (orthogonal) states $|\langle a_1|a_2\rangle| = 0$. In the intermediate case $0 < |\langle a_1|a_2\rangle| < 1$, ΔS changes continuously, a conclusion that holds more generally [9]. This was seen as a resolution of the Gibbs paradox [6–10].[8] We shall recall counter arguments in Sec. IV.

C. Generalization to several mixing gases

We shall indicate how Eq. (12) changes for the mixing of two gases having an initially nonequal number of particles and nonequal volumes. The generalization to the mixing of several gases will be straightforward.

Let the first and second reservoirs contain, respectively, N_1 and N_2 particles in volumes V_1 and V_2. Since we are interested in irreversibilities coming due to mixing only, we should assume that the initial pressures P and temperatures T of the two gases are equal both initially and finally. The known ideal-gas relation $PV = NT$, applied for $V = V_1, V_2, V_1 + V_2$ and $N = N_1, N_2, N_1 + N_2$, implies

$$\frac{P}{T} = \frac{N_1}{V_1} = \frac{N_2}{V_2} = \frac{N_1 + N_2}{V_1 + V_2}. \tag{20}$$

Using Eq. (20) and proceeding along the same lines as when deriving Eq. (12), we get, for the mixing entropy ($M = 2$),

---

[8]Note that there are several differences between the positions undertaken by the authors of [7–9] versus the one of Landé in [6]. The detailed analysis carried out in [9] suggests that the approach by Landé contains errors, and his final formulas for the entropy of mixing are different from those in [7–9].

$$\frac{\Delta S}{\sum_{\gamma=1}^{M} N_\gamma} = S_{\text{vN}}\left(\sum_{\alpha=1}^{M} \lambda_\alpha \rho_\alpha\right) - \sum_{\alpha=1}^{M} \lambda_\alpha S_{\text{vN}}(\rho_\alpha), \tag{21}$$

where

$$\lambda_\alpha = \frac{N_\alpha}{\sum_{\gamma=1}^{M} N_\gamma}, \quad \alpha = 1, \ldots, M, \tag{22}$$

are the fractions of the two gases in the final density matrix. We already wrote Eqs. (21) and (22) such that they hold for any $M \geq 2$.

IV. CRITIQUE OF THE QUANTUM MIXING ENTROPY ARGUMENT

A. Thermodynamic entropy of mixing is ill defined in quantum mechanics

The above argument on the continuous change of ΔS was seen by many as the resolution of the Gibbs paradox—and it is often still believed to be. However, a more detailed analysis has shown that this explanation creates a new conceptual problem [14]. Let us recall the following features of the thermodynamical entropy.

(i) If two states A and B are connected by an irreversible process $A \rightarrow B$, then for defining thermodynamically the entropy change during this process, we should connect those states by a certain reversible process $A \Rightarrow B$—possibly by involving thermal baths and sources of work—and calculate the entropy change ΔS via the Clausius formula

$$\Delta S(A \rightarrow B) = \int_{A \Rightarrow B} \frac{dQ}{T}, \tag{23}$$

where dQ and T are, respectively, the differential heat (received from thermal baths) and the temperature.

Equation (23) provides entropy with an operational meaning and makes it observable via macroscopic measurements. Indeed, determining, e.g., the von Neumann entropy via its definition (9) implies knowledge of *all* eigenvalues of the corresponding density matrix ρ. This knowledge is not available for typical macroscopic or mesoscopic systems.

(ii) A reversible process $A \Rightarrow B$ is defined by requiring that it is possible to pass back along the same trajectory and to return to the same thermodynamical state,[9] such that, in particular, the entropy change during the resulting cyclic process $A \Rightarrow B \Rightarrow A$ is equal to zero:

$$\int_{A \Rightarrow B \Rightarrow A} \frac{dQ}{T} = 0.$$

Any statistical definition of entropy is expected to agree with the above thermodynamical one. An inspection shows, however, that this is not the case [14]: the partially distinguishable (i.e., nonorthogonal) states—which were supposed

---

[9]The thermodynamical state is defined by the values of certain macroscopic quantities, such as pressure, temperature, magnetization, entropy, etc.

to solve the paradox—create in this respect an inconsistency. It appears that for such states there is no reversible mixing process. Let us first of all note that when the internal states ρ_1 and ρ_2 are orthogonal—that is, they correspond to definite eigenvalues a_1 and a_2 of some physical observable (Hermitian operator) A—it is possible to separate the mixed gases and at least in principle to fulfill the requirement of a cyclic process. What one needs for this purpose is a suitable Hamiltonian [24]

$$H_{\text{sep}} = f(\vec{r}, A), \qquad (24)$$

which establishes strong correlations between the internal states of the atoms and their translational motion described by the position vector $\vec{r}$ [24]: the function $f(\vec{r}, a_i)$, with $i = 1, 2$, is very small for $\vec{r}$ being in, respectively, first and second reservoirs. The magnitude of H_{sep} has to be sufficiently large, so that all other terms in the overall Hamiltonian can be neglected. Together with a low-temperature bath, weakly coupled to the gases, the Hamiltonian H_{sep} will drive the system towards its minima and it will separate the mixed gases back into different reservoirs [24] without changing the internal states of the atoms (since $[H_{\text{sep}}, A] = 0$). There can be practical limitations on this procedure related, e.g., with limitations on the magnitude of H_{sep}, but in principle such a process is possible. Thus, one can apply Eq. (23) and recover of the usual thermodynamical formulas for entropy [24].

The problem is that once the gases described by initially *partially* distinguishable (nonorthogonal) density matrices ρ_1 and ρ_2 (state A) are mixed with weights λ_1 and λ_2, respectively (state B), then it is impossible to go back to the original state by any process such that the two gases return to their original states: There is *no* Hamiltonian similar to H_{sep} in Eq. (24) which can achieve such a separation [24], in particular because ρ_1 and ρ_2 do not form eigenstates of any Hermitian operator.

Are there, however, measurements which can help to achieve this separation? We need a careful discussion of this question, since the existing opinions—e.g., those presented in [14]—seem to us somewhat unclear.

First of all, we note that the procedure involving H_{sep} can be seen as a measurement, where the role of the measuring apparatus is played by the classical coordinate $\vec{r}$ of the atom.[10] The motion of this apparatus amounts to the separation of the gases. The above question can be thus reformulated as to concern other measuring apparatuses (not connected with the coordinates) and their role for separation of the gases. Our answer to this question is negative, and here is why.

Using the example given by Eqs. (16), it is seen that there is not any measurement which would discriminate unambiguously—and without disturbing the initial states—between $|a_1\rangle$ and $|a_2\rangle$, if $\langle a_2 | a_1 \rangle$ is neither zero nor one [24]. Thus, it is impossible to separate the gases without disturbing the states of their atoms. However, requiring cyclic

changes of every single-atom state is too much for a thermodynamical reasoning. It suffices to require cyclic change of all collective (macroscopic variables) of the gases. In particular, the (final) internal states of the atoms in each reservoir are to be described by the density matrices $\rho_1 = |a_1\rangle\langle a_1|$ and $\rho_2 = |a_2\rangle\langle a_2|$, respectively. Such (generalized) measurements do exist.[11] Assume for simplicity that the internal state is a spin-$\frac{1}{2}$ state represented by the Pauli matrices $\vec{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$. One comes with another set of particles carrying spin $\frac{1}{2}$ described by the Pauli matrices $\vec{s} = (s_1, s_2, s_3)$. The spins $\vec{\sigma}$ and $\vec{s}$ undergo a controlled unitary evolution, after which one measures, e.g., s_3 with the help of a suitable macroscopic measurement apparatus. After selecting measurement results (i.e., the eigenvalues ± 1 of s_3), the initial mixed ensemble $\rho = \lambda_1 |a_1\rangle\langle a_1| + \lambda_2 |a_2\rangle\langle a_2|$ of the $\vec{\sigma}$ spin is separated into two subensembles $\rho_1 = |a_1\rangle\langle a_1|$ and $\rho_2 = |a_2\rangle\langle a_2|$, with the probabilities (weights) λ_1 and λ_2, respectively [26,27]. This is the desired separation.

However, quantum measurements are by their very nature *noncyclic*, since dissipative processes are connected with the motion of the pointer variable. In the above example both the spin $\vec{s}$ and the apparatus measuring s_3 have undergone such noncyclic processes. These certainly do generate an independent (and sizable) amount of entropy which is not taken into account in Eq. (23).

(iii) In summary, the possibility to define a cyclic process is a necessary condition for the thermodynamical meaning of entropy. When mixing gases that have non-orthogonal states, there are no suitable cyclic unmixing processes. This precludes entropy from having the proper thermodynamical meaning. Thus, trying to solve the problem in one place the quantum mixing entropy argument creates a new problem almost at the same time. The Gibbs paradox thus remains unexplained.

1. Why it is impossible simply to define entropy via the von Neumann formula

In the context of the above objection to the thermodynamical meaning of the mixing entropy in the quantum situation, one can ask why it is not possible simply to *define* entropy via the von Neumann formula (9) without worrying about its precise relation to other thermodynamical notions. If desired, such a definition may be motivated, e.g., via information-theoretic arguments [2].

In our opinion this is not possible to do, since entropy in statistical physics is never defined as an independent macroscopic observable; note again that the calculation of the von Neumann entropy via Eq. (9) requires knowledge of the full spectrum of the density matrix ρ, which is microscopic information normally not available for statistical systems. For internal states, it may be available, though. More generally, entropy cannot be defined from first principles without taking into account the corresponding formulations of the sec-

---

[10]A closely related quantum mechanical model for quantum (and classical) measurements was recently analyzed in detail [25]

[11]This is a known fact in the physics of quantum ensembles; see, e.g., Ref. [26]. The described procedure amounts to POVM (positive operator measured values). Recently we discussed in detail its implications for defining fluctuations of work [27].

ond law of thermodynamics, a fact that strictly speaking precludes any really noncircular derivation of these entropic formulations from first principles [1]. In contrast, formulations of the second law that operate with work instead of entropy normally do have first-principles derivations; see, e.g., [28–30]. The nonunique character of entropy is recalled and illustrated in Appendix A.

B. The employed notion of "difference between gases" does not have a clear operational status

Another difficulty with the above argument is that this attempted resolution of the paradox does not depend on the available experimental instruments and tools to be employed in control of the internal states of the atoms. As it stands within the entropic argument, the resolution depends on the difference between the states which is determined by their initial preparations via density matrices ρ_1 and ρ_2. However, preparation and control are different things and in general cannot be combined in a single density matrix. As an example, consider the preparation of a Gibbsian state with density matrix $\rho \propto \exp[-H/T]$, where H and T are, respectively, the Hamiltonian and temperature. This preparation needs only a weak interaction between the system and a thermal bath at temperature T; it does not contain any information on what we can measure or control in this system.

On general grounds, it was argued in [23] that the resolution of the Gibbs paradox *has to be* operational, since there are situations when two objects are formally different, but no computable (i.e., solvable by algorithms) operation can establish this difference. Worse, we cannot exclude unknown laws of physics that in the future would force us to distinguish (states of) atoms or particles which in our present understanding are considered as identical.

This operational aspect is also important, because, as we see below for the approach that takes this properly into account, the dependence on the available instruments is nontrivial: less refined instruments can—depending on the situation—indicate less or more irreversibility of mixing.

One may perhaps counterargue the above criticism by noting that the operational meaning and the dependence on the available instruments might be provided by the information-theoretic approach to statistical physics; see, e.g., [11,13]. We, however, should simply note that information-theoretic constructions are not at all guaranteed to have the proper physical meaning, as we saw for the above reversibility problem. Moreover, uncritical use of information-theoretic concepts may by itself lead to problems; see, e.g., [19], where the first-principles derivation of the Landauer bound for information erasure was found in conflict with the information-theoretical one.

V. RESOLUTION OF THE PARADOX VIA THE CONCEPT OF MAXIMAL MIXING WORK

The main point of the present paper is to employ quantum thermodynamics—the thermodynamics of finite systems coupled to a macroscopic worksource and possibly to a macroscopic bath. As realized in earlier works [18,19,31], this

approach generally acknowledges that one should study work instead of entropy—in the absence of a thermodynamic limit the latter has no firm meaning and each definition leads to a new value. This shift of paradigm will allow us to resolve the Gibbs paradox without the difficulties and ill-defined meaning of the mixing entropy argument. The reason for this solution lies in the fact that work and its properties are deduced from the first principles of quantum mechanics without any need of thermodynamic postulates (such as reversibility or existence of cyclic processes); see in this context footnotes 14 and 15. In other words, the resolution of the paradox is sought by going to the first principles of quantum mechanics alone and *without involving any thermodynamic argument.*

We start by recalling the definition of available work for a general, thermally isolated process done on a quantum system.

A. Definition of work

A quantum system is described at the initial time $t=0$ by a density matrix $\rho(0)$ and interacts with an external macroscopic work source. The resulting evolution of the system is generated by (an effective) Hamiltonian $H(t) = H\{R(t)\}$, which is time dependent via classical (c-number) parameters $R(t)$ (control fields).

We shall be concerned with processes where the change of the Hamiltonian is cyclic:

$$H(\tau) = H(0) = H. \qquad (25)$$

The situation where the work source interacts with the system for a finite time belongs to this class of processes, since the corresponding system-work-source interaction Hamiltonian is zero both initially and finally.[12] Note that processes with a cyclic Hamiltonian are obviously different from the processes that are cyclic in the sense of various macroscopic quantities. However, it is necessary to have a cyclic change of the Hamiltonian for the process to be cyclic in the sense of macroscopic quantities.

Thus, the process is assumed to be thermally isolated and the Hamiltonian $H(t)$ generates a unitary evolution:

$$i\hbar \frac{\mathrm{d}}{\mathrm{d}t}\rho(t) = [H(t), \rho(t)], \qquad (26)$$

$$\rho(t) = U_t \rho(0) U_t^\dagger, \quad U_t = \overleftarrow{\mathrm{exp}}\left[-\frac{i}{\hbar}\int_0^t \mathrm{d}s\, H(s)\right], \qquad (27)$$

where $\overleftarrow{\mathrm{exp}}$ denotes the time-ordered exponent. It is well known that, in general, a Hamiltonian evolution for two coupled systems does not reduce to a Hamiltonian evolution for one of them. However, in the present case the evolution

---

[12]All constructions below generalize to processes, where the initial and final Hamiltonians are different. In the context of the Gibbs paradox this more general setting may provide some advantages, though it does not give any conceptual novelty as compared to the cyclic-Hamiltonian case.

of the system is Hamiltonian owing, in particular, to the *macroscopic* character of the work source, as discussed in [2,32].[13]

(i) The work W done by the external source between times 0 and τ in the thermally isolated process[14] is identified with the average energy change of the system [1,2]:

$$W = \mathrm{tr}[\rho(\tau)H(\tau) - \rho(0)H(0)]. \qquad (28)$$

(ii) Due to conservation of the (average) energy, W is equal to the average energy decrease of the work source.

(iii) This is a classical, mechanical energy that can be transferred with 100% efficiency to other macroscopic work sources, and, in particular, it can transferred to another mechanical degree of freedom performing classical deterministic motion.

(iv) W is typically observed via suitable (classical) measurements done on the macroscopic work source or, alternatively, by measuring the initial and final average energies on the ensemble of (many) identically prepared systems. Both these ways are routinely employed in practice, e.g., in NMR and electron spin resonance (ESR) physics, where the system corresponds to spin $\frac{1}{2}$ under the influence of external magnetic fields (work source).

(v) The definition of work and its features are based purely on the first principles of quantum mechanics. They do not depend on thermodynamical concepts, such as reversibility.[15] In contrast, the work as it is known in thermodynamics can be deduced from the first principles of quantum mechanics.

---

[13]The Appendix of Ref. [32] contains a clear discussion of certain additional conditions that have to be satisfied for the time-dependent Hamiltonian evolution and for the proper identification of the work source.

[14]From the viewpoint of work exchange every process can be completed to a thermally isolated one by including in the system its environment (e.g., thermal baths). Then the work (28) for this thermally isolated process coincides with the usual definition of work for an arbitrary process: $W = \int_0^\tau dt\, \mathrm{tr}(\rho_S(t) \partial_t H(t))$, where $\rho_S(t)$ is the time-dependent density matrix of the system. Indeed, let H_E and H_I be, respectively, the Hamiltonian of the environment and the system-environment interaction. Recall that the work sources act only on the system; thus the total Hamiltonian $\mathcal{H}(t)$ of the system +environment is $\mathcal{H}(t) = H(t) + H_E + H_I$, where only the system Hamiltonian $H(t)$ is time dependent. To prove the desired statement we have to write down expression (28), $W = \mathrm{tr}[\rho_{SE}(\tau)\mathcal{H}(\tau) - \rho_{SE}(0)\mathcal{H}(0)]$, where $\rho_{SE}(t)$ is the time-dependent density matrix of the system+environment, apply the von Neumann equation of motion for the thermally isolated process, $i\dot{\rho}_{SE} = [\mathcal{H}, \rho_{SE}]$, and transform $W = \int_0^\tau dt\, \mathrm{tr}(\rho_S(t)(\partial_t H(t)) = \int_0^\tau dt\, \mathrm{tr}(\rho_{SE}(t)(\partial_t \mathcal{H}(t))$ with the help of integration by parts.

[15]In particular, the definition of a reversible process can be based on the notion of work [33]. A process is reversible if (i) it can be supplemented by its mirror reflection that goes back along the same trajectory and (ii) the work done on this completed process is zero. It is also clear that the definition of heat need not supersede the definition of work. The reason for this is that from the viewpoint of work exchange any process can be completed to a thermally isolated one, where the work is uniquely related to the energy; see footnote 14 in this respect.

B. Maximally available work

One of the fundamental tasks of thermodynamics is to determine the maximal amount of work which can be extracted from a given (nonequilibrium) system in the initial state ρ under cyclic-Hamiltonian (sufficiently smooth) processes (25). The latter condition is imposed, since otherwise there may not be any limit in the extracted work (e.g., for the final Hamiltonian being negative and very large by the absolute value). It is via this task posed by Clausius and solved within phenomenological thermodynamics that entropy acquires its physical meaning as a measure of order related to high-graded energy (work) [1,34]. While the standard solution of this task is well known and based on the notion of reversible process (in the same way as the definition of entropy is), it was recently shown that the problem can, and should, be solved from the first principles of quantum mechanics *without* invoking any thermodynamical axiom [32,34]. The solution differs from the standard one, though the latter provides a correct bound for the maximal work $W_{\max}$ in Eq. (32) below and is expected to agree with it when phenomenological thermodynamics is supposed to apply— i.e., for weakly nonequilibrium states of generic macroscopic systems.

To describe the solution to the maximal work extraction problem, we denote the eigenresolutions of the Hamiltonian H and of the density matrix ρ as, respectively,

$$H = \sum_{k=1}^n \varepsilon_k |\varepsilon_k\rangle\langle\varepsilon_k|, \quad \rho = \sum_{k=1}^n p_k |p_k\rangle\langle p_k|, \qquad (29)$$

where $\{|\varepsilon_k\rangle\}_{k=1}^n$ and $\{|p_k\rangle\}_{k=1}^n$ with $\langle\varepsilon_k|\varepsilon_l\rangle = \langle p_k|p_l\rangle = \delta_{kl}$ are the eigenvectors of H and ρ, respectively, and where ε_k and p_k are the corresponding eigenvalues. We shall assume that

$$\varepsilon_1 \leq \varepsilon_2 \leq \cdots. \qquad (30)$$

The non-increasing ordering of $\{|p_k\rangle\}_{k=1}^n$ is denoted as:

$$p_1^\downarrow \geq p_2^\downarrow \geq \cdots. \qquad (31)$$

Then the maximal available work is defined as [34]

$$W_{\max} \equiv -\mathcal{W} = \min_U \{\mathrm{tr} H[\rho(\tau) - \rho(0)]\}, \qquad (32)$$

where $\mathcal{W}$ is the non-negative absolute value of the maximal work and where the minimization in $\min_U$ is taken over *all* smooth, cyclic Hamiltonians[16]:

---

[16]Note that for an n-level system the minimization over all Hamiltonians (33) can be carried out by minimizing over Hamiltonians of the form $H(t) = H + \sum_{i=1}^m b_i(t) X_i$, where $b_i(t)$ are time-dependent c functions and where X_i are operators such that any generator of the group $SU(n)$ can be obtained via linear combinations of $H, X_1, X_2, \ldots, X_m$ and their multiple commutators [35]. For $n = 2$ and $H = \sigma_3$ this Hamiltonian is $H(t) = \sigma_3 + b(t)\sigma_1$, with σ_1 and σ_3 being the corresponding Pauli matrices. For $n = 3$ the analogous Hamiltonian is $H(t) = \lambda_3 + b_1(t)\lambda_1 + b_4(t)\lambda_4$, where λ_k are the Gell-Mann matrices [generators of $SU(3)$]. If the minimization is carried out via Hamiltonians $\lambda_3 + b_1(t)\lambda_1$, the unitary transformations act only on the upper left 2×2 sector of the 3×3 density matrix ρ.

$$H(t) = H + V(t), \quad V(0) = V(\tau) = 0, \qquad (33)$$

where τ is the cycle time of the Hamiltonian.[17] Minimizing over the Hamiltonians in Eqs. (33) is equivalent to minimizing over all unitary operators U [34]; this is why we denoted this minimization as $\min_U$. An explicit formula for the optimal Hamiltonian is given in [34], while the result of the minimization in Eq. (32) yields the *ergotropy* [34,32]

$$\mathcal{W} = \text{tr}(\rho H) - \sum_{k=1}^{n} p_k^{\downarrow} \varepsilon_k \geqslant 0. \qquad (34)$$

This is a difference between the final and initial average energies of the system, as it should be for the work extracted in a thermally isolated system. It has a simple interpretation: since in quantum mechanics the eigenvalues of ρ are conserved under the unitary evolution caused by macroscopic external sources, the lowest final energy is reached when the largest eigenvalue of ρ becomes the ground-state occupation, the one but largest eigenvalue occupies the first excited state, and so on. Various features of $\mathcal{W}$, in particular those contrasting the thermodynamical intuition, were studied in [32,34]. We suggested to call Eq. (34) the *ergotropy* of the state ρ.

It is seen from Eq. (34) that no work extraction is possible (i.e., $\mathcal{W}=0$) if ρ is a monotonically decreasing function of H:

$$\rho = f(H), \quad f'(x) \leqslant 0. \qquad (35)$$

This, in particular, includes Gibbs equilibrium states $\rho \propto e^{-H/T}$, where $T > 0$ is the temperature. This confirms Thomson's formulation of the second law: no work extraction from an equilibrium state by means of cyclic-Hamiltonian processes [28,29].

C. Operational meaning of the available work

The concept of maximal work takes into account the notion of available instruments. Indeed, in Eq. (32) we optimized the extracted work over all cyclic-Hamiltonian thermally isolated processes, which assumes that the optimal one is available. If there are restrictions on the availability of sources of work, the amount of extractable work will, in general, be smaller than $\mathcal{W}$. It is even possible that no work at all can be extracted by some restricted class of work sources.[18]

To make this point clear, let us assume that the possible unitary evolutions U in Eq. (32) are restricted to permutations of the diagonal elements,

---

[17]We note that there are no restrictions on the product of τ with the typical magnitude of $V(t)$ (i.e., on the dimensionless coupling constant characterizing the sources of work). It is also assumed that the initial state ρ is known. Limitations on this knowledge will, in general, lower the value of the maximal work.

[18]The class of employed work sources corresponds to what in [15] was called a thermodynamical construction: a set of nonrelaxed mechanical degrees of freedom that define the very meaning of various thermodynamical quantities.

$$\pi_k = \langle \varepsilon_k | \rho | \varepsilon_k \rangle, \qquad (36)$$

of the density matrix ρ in the energy representation.[19] Then instead of Eqs. (34) we will have

$$W'_{\max} \equiv -\mathcal{W}'(\rho, H) = \sum_{k=1}^{n} \pi_k^{\downarrow} \varepsilon_k - \text{tr}(\rho H) \qquad (37)$$

$$= \sum_{k=1}^{n} \varepsilon_k [\pi_k^{\downarrow} - \pi_k]. \qquad (38)$$

In general, we have for the ergotropy

$$\mathcal{W} \geqslant \mathcal{W}', \qquad (39)$$

where the equality sign is realized for $[\rho, H] = 0$. It is now possible that $\pi_k^{\downarrow} = \pi_k$ and thus $\mathcal{W}' = 0$, though $\mathcal{W} > 0$ due to the nondiagonal elements of ρ.

D. Explanation of the paradox

We shall now immediately deal with M gases with arbitrary weights $\lambda_\alpha = (N_\alpha)/(\mathcal{N})$, and the total of particles

$$\mathcal{N} = \sum_{\alpha=1}^{M} N_\alpha. \qquad (40)$$

Let us return to the assumptions presented in Sec. III A and list them again: (i) the necessity of taking into account the internal states and (ii) decoupling of the internal and translational degrees of freedom: the total Hamiltonian $H_{\text{tot}}^{(\alpha)}$ of each gas contained in the corresponding reservoir is

$$H_{\text{tot}}^{(\alpha)} = H_0^{(\alpha)} + \sum_{i=1}^{N_\alpha} H^{(\alpha,i)}, \quad \alpha = 1, \dots, M, \qquad (41)$$

where H_0 is the sum of kinetic energies of all N gas particles plus the potential generated by the walls of the reservoir and where $H^{(\alpha,i)}$ is the Hamiltonian of internal motions of the atom with index i belonging to the gas with index α. Since we assume that all atoms in both reservoirs are identical and differ by their states only, we shall assume that all atoms have the same internal Hamiltonian:

$$H^{(\alpha,i)} = H. \qquad (42)$$

(iii) Time-scale separation between the translational and internal degrees of freedom during the mixing; thus Eq. (6), $\rho = \Sigma_{\alpha=1}^{M} \lambda_\alpha \rho_\alpha$, holds for the post-mixed density-matrix for M gases with the initial internal states ρ_α and arbitrary weights λ_α.

In our opinion, these assumptions are physically sound; it is only their implementation within the mixing entropy argument that is problematic. We shall avoid that argument by using work (more precisely, its maximum in absolute value, ergotropy) instead of entropy.

---

[19]There are, of course, many other ways to introduce limitations on the available unitary evolutions. For more examples, see Ref. [36], as well as, the last part of Footnote 16 and Footnote 17.

Before mixing, how much work can be extracted from the total system containing M separate gases? The answer depends on the specification of the interaction between the gases and the sources of work. These interactions are chosen under the following assumptions.

(1) Since the gases are ideal, it is natural to assume that the sources act on each particle separately; i.e., the sources by themselves do not introduce interparticle interactions.

(2) Work sources act on the internal degrees of freedom only. This is because the internal and the translational degrees of freedom are decoupled and because the translational degrees of freedom are in (local) equilibrium, so it is useless to try to extract any work from them; recall our discussion around (35).

(3) We allow different sources of work to act on different gases. This is again reasonable, since the gases start out perfectly separated from each other.

Given the above assumptions we are led to the following time-dependent, internal Hamiltonian for each gas:

$$H^{(\alpha)}(t) = \sum_{i=1}^{N} [H^{(i)} + V^{(\alpha,i)}(t)]. \qquad (43)$$

Since all particles within the given reservoir are equivalent, we have

$$V^{(\alpha,i)}(t) = V^{(\alpha)}(t), \qquad (44)$$

where

$$V^{(\alpha)}(0) = V^{(\alpha)}(\tau) = 0, \qquad (45)$$

as required by the cyclic-Hamiltonian feature [compare with Eqs. (33)].

It is now seen that the maximal work extractable from the premixed state reads

$$\mathcal{W}_{\mathrm{i}} = \mathcal{N} \sum_{\alpha=1}^{M} \lambda_\alpha \mathcal{W}(\rho_\alpha, H), \qquad (46)$$

where $\mathcal{W}(\rho_\alpha, H)$ defined in Eq. (34) is the maximal work extracted from the initial state ρ_α with the initial (and final) Hamiltonian H. Note that $\mathcal{W}_{\mathrm{i}}$ is proportional to the total number of particles $\mathcal{N} = \Sigma_{\alpha=1}^{M} N_\alpha$ thanks to the above assumptions respecting the ideal gas structure of the problem.

Let us now determine how much work we can extract after the M gases have mixed. The above conditions for a system-work-source interaction remain valid except the last one.

(3′) Since the gases now form a single homogeneous system with the density matrix $\rho = \Sigma_{\alpha=1}^{M} \lambda_\alpha \rho_\alpha$, we cannot enforce the different particles (atoms) to couple to different sources of work. At best we can couple the $\mathcal{N} = \Sigma_{\alpha=1}^{M} N_\alpha$ particles with the same type of work sources. Thus, the physically acceptable cyclic Hamiltonians has the form (33).

The resulting maximal work reads, from Eqs. (6) and (34),

$$\mathcal{W}_{\mathrm{f}} = \mathcal{N} \mathcal{W}\left(\sum_{\alpha=1}^{M} \lambda_\alpha \rho_\alpha, H \right). \qquad (47)$$

The difference between Eqs. (46) and (47) is defined to be the *maximal mixing work* or *mixing ergotropy* $\Delta \mathcal{W}$:

$$\Delta \mathcal{W} \equiv \mathcal{W}_{\mathrm{i}} - \mathcal{W}_{\mathrm{f}} \qquad (48)$$

$$= \mathcal{N}\left[\sum_{\alpha=1}^{m} \lambda_\alpha \mathcal{W}(\rho_\alpha, H) - \mathcal{W}\left(\sum_{\alpha=1}^{M} \lambda_\alpha \rho_\alpha, H \right) \right], \qquad (49)$$

$$= \mathcal{N} \sum_{k=1}^{n} \varepsilon_k \left(p_k^{\downarrow} - \sum_{\alpha=1}^{M} \lambda_\alpha p_{k,\alpha}^{\downarrow} \right), \qquad (50)$$

where we employed Eq. (34) and where $p_k^{\downarrow}$ and $p_{k,\alpha}^{\downarrow}$ are nonincreasingly ordered eigenvalues of ρ and ρ_α, respectively.

The fact that maximal work cannot increase upon mixing,

$$\Delta \mathcal{W} \geqslant 0, \qquad (51)$$

should be obvious from the very construction. Here is, however, the formal proof. Recall (32) and note that

$$\max_{U_\alpha}(\mathrm{tr}[H\rho_\alpha] - \mathrm{tr}[HU_\alpha \rho_\alpha U_\alpha^\dagger]) \qquad (52)$$

$$= \mathrm{tr}[H\rho_\alpha] - \mathrm{tr}[H\tilde{U}_\alpha \rho_\alpha \tilde{U}_\alpha^\dagger] \qquad (53)$$

$$\geqslant \mathrm{tr}[H\rho_\alpha] - \mathrm{tr}[HU\rho_\alpha U^\dagger], \qquad (54)$$

where $\tilde{U}_\alpha$ is the optimal unitary operator which maximizes (52) and where U is any other unitary operator, including the one which maximizes $\mathrm{tr}[H\rho - HU\rho U^\dagger]$. The desired (51) is now recovered via multiplying (52)–(54) by λ_α and summing over α.

The very same argument applies if the maximization in the definition of $\mathcal{W}$ is carried out over a restricted class of unitary operators or cyclic Hamiltonians (we assume, of course, that this is the same class initially and finally). Analogous to (51), we then deduce from (37) that $\Delta \mathcal{W}' \geqslant 0$.

Turning to the conceptual implications of the mixing work $\Delta \mathcal{W}$, we note that, of course, $\Delta \mathcal{W} = 0$ for $\rho_\alpha = \rho$, when identical gases are mixed. Moreover, it goes to zero continuously with $\rho_\alpha \to \rho$.

We therefore consider this continuity of maximally extractable work as the resolution of the Gibbs paradox within quantum thermodynamics.

The first objection for the entropic argument—see our discussion around (23) and Ref. [14]—is now harmless, since now the concept of thermodynamical reversibility is not employed anywhere; the machinery of the maximal work extraction is based completely on quantum mechanics alone. As we stressed repeatedly, work is a first-principles concept, more fundamental than entropy [30].[20]

Note that when ρ_1 and ρ_2 are pure states, the converse of the above statement appears to be valid: if ρ_1 and ρ_2 are different, then $\Delta \mathcal{W} > 0$. This is because the only pure state

---

[20] In phenomenological thermodynamics, the problem of the maximal work extraction is treated by employing the reversibility concept and features of entropy [1]. In our opinion, this is the reason

that cannot provide work is the ground state of the Hamiltonian H. If, however, at least one of the two density matrices is mixed, there are different states ρ_1 and ρ_2 such that $\Delta \mathcal{W} = 0$. For the simplest example recall (35) and take as ρ_1 and ρ_2 two equilibrium states with different temperatures T_1 and T_2.

To illustrate the above statements in more detail, we turn to the density matrices given by (16)–(19), where the Hamiltonian H has two energy levels 0 and $\varepsilon > 0$. Recalling (17) we get from (49) that $\Delta \mathcal{W}$ is a simple function of the overlap:

$$\Delta \mathcal{W} = \frac{\mathcal{N} \varepsilon}{2} [1 - |\langle a_1 | a_2 \rangle|]. \tag{55}$$

For completely distinguishable, classical states $|\langle a_1 | a_2 \rangle| = 0$ this gives $\Delta \mathcal{W} = \varepsilon / 2$, while for identical states $|\langle a_1 | a_2 \rangle| = 1$, $\Delta \mathcal{W} = 0$. The classical argument describes *only* these extremes (i.e., completely different or identical) and, thus, creates the paradox.

VI. HOW THE MIXING WORK DEPENDS ON THE AVAILABLE INSTRUMENTS

Let us now turn to the second objection against the entropic argument. We recall from Sec. IV B that once the difference between two states is recognized to be an operational notion—two states may not differ under inspection by some instruments, but turn out to be different if more refined ones are used—we should expect that this feature is reflected in a satisfactory resolution of the Gibbs paradox.

As we stressed repeatedly, the notion of available work is operational in the above sense. So is the mixing work defined in Eq. (49). Moreover, the situation is nontrivial, since $\Delta \mathcal{W}$ can both increase or decrease under restricting the available instruments (i.e., system-work-source interactions), as we show now.

To illustrate this fact, let us take the internal Hilbert space of all particles having two dimensions (e.g., spin $\frac{1}{2}$):

$$\rho_\alpha = \frac{1}{2}(1 + \vec{n}_\alpha \vec{\sigma}), \quad \alpha = 1, \ldots, M, \tag{56}$$

where $\vec{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$ are the Pauli (2×2) matrices and where

$$\vec{n}_\alpha = (n_{1,\alpha}, n_{2,\alpha}, n_{3,\alpha}), \quad |\vec{n}_\alpha| \leq 1, \tag{57}$$

is the Bloch c vector. Recalling the spectrum

$$\text{Spec}\{\rho_\alpha\} = \frac{1}{2}(1 \pm |\vec{n}_\alpha|), \tag{58}$$

we get, from Eq. (49),

$$\Delta \mathcal{W} = \frac{\mathcal{N} \varepsilon}{2} \left(\sum_{\alpha=1}^{M} \lambda_\alpha |\vec{n}_\alpha| - \left| \sum_{\alpha=1}^{M} \lambda_\alpha \vec{n}_\alpha \right| \right). \tag{59}$$

On the other hand, if for the Hamiltonian

$$H = \frac{\varepsilon(1 + \sigma_3)}{2} \tag{60}$$

the maximization over the uninary operators in Eq. (34) is carried out only over those unitary operators which permute the diagonal elements of the corresponding density matrices in the energy representation [compare with (37) and (39)], the mixing work will read

$$\Delta \mathcal{W}' = \frac{\mathcal{N} \varepsilon}{2} \left(\sum_{\alpha=1}^{M} \lambda_\alpha |n_{3,\alpha}| - \left| \sum_{\alpha=1}^{M} \lambda_\alpha n_{3,\alpha} \right| \right), \tag{61}$$

where $n_{3,\alpha}$ is the third component of the vector $\vec{n}$.

It is obvious that there are cases where

$$\Delta \mathcal{W} > \Delta \mathcal{W}', \tag{62}$$

e.g., choose $n_{3,\alpha}$ all having the same sign which leads to $\Delta \mathcal{W}' = 0$. It is, however, less expected that there can also be situations where

$$\Delta \mathcal{W}' > \Delta \mathcal{W}. \tag{63}$$

This means that *use of less precise instruments can increase the amount of mixing work.*

To show this, let us choose the case

$$\sum_{\alpha=1}^{M} \lambda_\alpha n_{3,\alpha} = 0 \tag{64}$$

and write, from Eqs. (59) and (61),

$$\frac{\Delta \mathcal{W}' - \Delta \mathcal{W}}{\mathcal{N}(\varepsilon/2)} = \sum_{\alpha=1}^{M} (\sqrt{\lambda_\alpha^2 n_{3,\alpha}^2} - \sqrt{\lambda_\alpha^2 [n_{3,\alpha}^2 + n_{1,\alpha}^2 + n_{2,\alpha}^2]}) \tag{65}$$

$$+ \sqrt{\left(\sum_{\alpha=1}^{M} \lambda_\alpha n_{1,\alpha} \right)^2 + \left(\sum_{\alpha=1}^{M} \lambda_\alpha n_{2,\alpha} \right)^2}. \tag{66}$$

In Eq. (65) we use the inequality[21]

$$\sqrt{x} - \sqrt{x+y} \geq -\frac{y}{2\sqrt{x}}. \tag{67}$$

Taking for simplicity $\lambda_\alpha n_{1,\alpha} = \lambda_\alpha n_{2,\alpha} = b$ and $|\lambda_\alpha n_{3,\alpha}| = a$—and thus M should be even to satisfy Eq. (64)—we get

$$\Delta \mathcal{W}' - \Delta \mathcal{W} \geq \sqrt{2} M \mathcal{N} |b| \left(1 - \frac{|b|}{a\sqrt{2}} \right). \tag{68}$$

By suitable choice of a and b, one can make the right-hand side (RHS) of (68) positive, thus proving the desired statement (63).

---

why the concept of work—though mentioned as a helpful one for interpreting the Gibbs paradox [6,15]—was never seriously employed for resolving the paradox.

[21]To prove (67) make an incomplete Taylor expansion for $f(x) = \sqrt{x}$: $f(x+y) = f(x) + yf'(x) + (y^2)/(2(f''(\xi)$, where $x \leq \xi \leq x+y$, and disregard $(y^2)/(2)f''(\xi) \leq 0$.

VII. MIXING WORK AND THE DEGREE OF MIXING

As we saw, the mixing work is zero when there is no true mixing—i.e., when the internal states of the mixed gases are identical.[22] It is expected that the mixing work will decrease together with the degree of mixing.

Consider the mixing work $\Delta\mathcal{W}(\vec{\lambda})$ as a function of the weights $\vec{\lambda} = \{\lambda_\alpha\}_{\alpha=1}^M$. For fixed states $\{\rho_\alpha\}_{\alpha=1}^M$, we expect that if $\vec{\lambda}$ is more inhomogeneous than $\vec{\mu}$, then

$$\Delta\mathcal{W}(\vec{\mu}) \geqslant \Delta\mathcal{W}(\vec{\lambda}). \qquad (69)$$

Here is an exaggerated example illustrating (69): for two species the degree of mixing is expected to be higher when having 100 particles of each type than when having 199 and 1, respectively. The weights for this example are, respectively, $\mu_1 = \mu_2 = 1/2$ and $\lambda_1 = 199/200$, $\lambda_2 = 1/200$.

Below we clarify in which sense the intuitive expectation (69) is correct.

A. Majorization

First we need the proper formalization for the notion of "inhomogeneous." This is provided by the concept of majorization [37] which we shortly recall below.

For two sets of probabilities $\vec{\lambda} = \{\lambda_\alpha\}_{\alpha=1}^M$ and $\vec{\mu} = \{\mu_\alpha\}_{\alpha=1}^M$, $\vec{\lambda}$ majorizes $\vec{\mu}$ (i.e., $\vec{\lambda}$ is more inhomogeneous than $\vec{\mu}$), denoted as

$$\vec{\mu} \prec \vec{\lambda}, \qquad (70)$$

if, for all $1 \leqslant m \leqslant M$,

$$\sum_{\alpha=1}^m \lambda_\alpha^\downarrow \geqslant \sum_{\alpha=1}^m \mu_\alpha^\downarrow, \qquad (71)$$

where $\vec{\lambda}^\downarrow$ means nonincreasing ordering of $\vec{\lambda}$ [recall (31)].

To illustrate (71), the uniform vector $(1/M, \ldots, 1/M)$ is majorized by all other probability vectors, while any deterministic vector—e.g., $(1, 0, \ldots, 0)$—majorizes all others. It follows from (71) that $\sum_{\alpha=1}^M f(\lambda_\alpha) \leqslant \sum_{\alpha=1}^M f(\mu_\alpha)$ for any concave function $f(x)$ [37]—e.g., $f(x) = -x \ln x$ (entropy).

The majorization property is transitive: $\vec{\mu} \prec \vec{\lambda}$ and $\vec{\lambda} \prec \vec{\nu}$ imply $\vec{\mu} \prec \vec{\nu}$. Also $\vec{\lambda} \prec \vec{\mu}$ and $\vec{\mu} \prec \vec{\lambda}$ imply $\vec{\lambda}^\downarrow = \vec{\mu}^\downarrow$. However, this property is incomplete: for $n \geqslant 3$ there are vectors $\vec{\lambda}$ and $\vec{\mu}$ for which neither $\vec{\lambda}$ majorizes $\vec{\mu}$ nor does $\vec{\mu}$ majorize $\vec{\lambda}$ [37].

---

[22]Note that when the overall numbers of particles N_α in each reservoir is not very large, even the mixing of completely identical gases brings about changes in their final state [9]. This is due to different fluctuation characteristics of the translational motion [9]; e.g., before mixing the number of particles in the volume V_α is precisely N_α, while after mixing this number of particles will fluctuate being equal to N_α only on average. We shall neglect this effect assuming N_α to be sufficiently large.

B. Quasiclassical situation

Let the initial states of the gases be M pure, orthonormal states

$$\rho_\alpha = |\psi_\alpha\rangle\langle\psi_\alpha|, \quad \langle\psi_\alpha|\psi_\beta\rangle = \delta_{\alpha\beta}. \qquad (72)$$

We call this situation quasiclassical, since following the original formulation of the Gibbs paradox within classical thermodynamics, the internal states are completely distinguishable and provide definite values for any observable that has $\{|\psi_\alpha\rangle\}_{\alpha=1}^M$ as its eigenfunctions.

Let us now prove that if $\vec{\lambda}$ is more inhomogeneous than $\vec{\mu}$—i.e., if (71) holds—then inequality (69) is valid. To this end we first employ summation by parts:

$$\sum_{k=1}^n \varepsilon_k \lambda_k^\downarrow = \varepsilon_n - (\varepsilon_2 - \varepsilon_1)\lambda_1^\downarrow - (\varepsilon_3 - \varepsilon_2)(\lambda_1^\downarrow + \lambda_2^\downarrow) - (\varepsilon_4 - \varepsilon_3)$$
$$\times (\lambda_1^\downarrow + \lambda_2^\downarrow + \lambda_3^\downarrow) - \cdots,$$

and then recalling (50) we get

$$\frac{\Delta\mathcal{W}(\vec{\mu}) - \Delta\mathcal{W}(\vec{\lambda})}{\mathcal{N}} = \sum_{k=1}^n \varepsilon_k[\mu_k^\downarrow - \lambda_k^\downarrow] = (\varepsilon_2 - \varepsilon_1)(\lambda_1^\downarrow - \mu_1^\downarrow)$$
$$+ (\varepsilon_3 - \varepsilon_2)(\lambda_1^\downarrow + \lambda_2^\downarrow - \mu_1^\downarrow - \mu_2^\downarrow) + \cdots \geqslant 0. \qquad (73)$$

Here each separate term is non-negative due to (30) and (71).

For this quasiclassical situation the above intuition (more mixing means larger mixing work) is correct.

C. Quantum situation

Let us assume that the initial states $\{\rho_\alpha\}_{\alpha=1}^M$ are not orthogonal. For simplicity we shall work with the simplest nontrivial situation

$$n = M = 2, \qquad (74)$$

i.e., two-dimensional internal state and two mixed gases. $\Delta\mathcal{W}$ is now given by (59). We assume that $\vec{\lambda}$ is more ordered than $\vec{\mu}$ in the sense of majorization, which for $M=2$ implies

$$\lambda_1 \geqslant \lambda_2, \quad \mu_1 \geqslant \mu_2, \quad \lambda_1 \geqslant \mu_1. \qquad (75)$$

Note that for the considered two-dimensional situation, $n=2$, the majorization order coincides, e.g., with the entropic order: Eq. (75) implies $-\lambda_1 \ln \lambda_1 - \lambda_2 \ln \lambda_2 \leqslant -\mu_1 \ln \mu_1 - \mu_2 \ln \mu_2$.

We now intend to clarify under which conditions inequality (69) holds. Recalling (59) this inequality is equivalent to

$$(\mu_1 - \lambda_1)(|\vec{n}_1| - |\vec{n}_2|)$$
$$\geqslant \sqrt{(\mu_1|\vec{n}_1| + \mu_2|\vec{n}_2|)^2 - 2\mu_1\mu_2|\vec{n}_1||\vec{n}_2|(1 - \cos\phi)}$$
$$- \sqrt{(\lambda_1|\vec{n}_1| + \lambda_2|\vec{n}_2|)^2 - 2\lambda_1\lambda_2|\vec{n}_1||\vec{n}_2|(1 - \cos\phi)}, \qquad (76)$$

where $\cos\phi$ is defined as

$$\vec{n}_1 \cdot \vec{n}_2 = |\vec{n}_1||\vec{n}_2|\cos\phi. \tag{77}$$

When both states are pure, $|\vec{n}_1|=|\vec{n}_2|=1$, inequality (76) reduces to $\mu_1\mu_2 \geqslant \lambda_1\lambda_2$ or

$$\lambda_1 + \mu_1 \geqslant 1, \tag{78}$$

a condition which is always satisfied in view of (75).

Assume in (76) that ϕ is small and expand (76) to first order of $1-\cos\phi$. After algebraic steps we get a generalization of (78):

$$\lambda_1 + \mu_1 \geqslant 1 + \lambda_1\mu_1\left[1 - \frac{|\vec{n}_1|}{|\vec{n}_2|}\right]. \tag{79}$$

This inequality is already not always satisfied. When $|\vec{n}_1|/|\vec{n}_2|$ is sufficiently small—i.e., one of the states is considerably more mixed—Eq. (79) may be violated; take, e.g., $\lambda_1=0.8$ and $\mu_1=0.7$. We conclude that *in the quantum situation the mixing work may be a nonmonotonous function of the degree of mixing, though it goes to zero continuously when the substances become identical.*

VIII. DISTINGUISHABILITY AND MIXING

Another way to control the mixing is to keep the weights equal, but make the internal states ρ_1 and ρ_2 closer to each other. It is natural to ask whether the mixing work is a monotonic function of the difference between these substances—i.e., whether decreasing this difference always makes the mixing work smaller. Below we are going to show that this is not always the case, though the mixing work, of course, goes to zero in the limit of identical substances.

First of all we need a clear understanding of the proper distance (closeness) between two density matrices ρ_1 and ρ_2. The answer is trivial for pure states as in (16): any monotonic function of the overlap

$$\mathrm{tr}(\rho_1\rho_2) = |\langle a_1|a_2\rangle|^2 \tag{80}$$

can be taken as the proper degree of closeness.

The generalization of the overlap (80) to mixed states is also well known and was derived from several different perspectives [38,39]. This "distinguishability" reads

$$d(\rho_1,\rho_2) = [\mathrm{tr}(\sqrt{\rho_1^{1/2}\rho_2\rho_1^{1/2}})]^2. \tag{81}$$

Let us note that $d(\rho_1,\rho_2)$ is symmetric,

$$d(\rho_1,\rho_2) = d(\rho_2,\rho_1), \tag{82}$$

concave,

$$d(\rho, x\rho_1 + (1-x)\rho_2) \geqslant xd(\rho,\rho_1) + (1-x)d(\rho,\rho_2), \tag{83}$$

and varies between 0 and 1,

$$0 \leqslant d(\rho_1,\rho_2) \leqslant 1, \tag{84}$$

being equal to 1 if and only if $\rho_1=\rho_2$. It is also multiplicative,

$$d(\rho_1 \otimes \rho_3, \rho_2 \otimes \rho_4) = d(\rho_2,\rho_1)d(\rho_3,\rho_4), \tag{85}$$

invariant under unitary transformations,

$$d(\rho_1,\rho_2) = d(U\rho_1 U^\dagger, U\rho_2 U^\dagger), \quad U^\dagger U = 1, \tag{86}$$

increases under completely positive evolution, and reduces to $\mathrm{tr}(\rho_1\rho_2)$ if ρ_1 or ρ_2 is pure.

In particular, $d(\rho_1,\rho_2)$ has the proper information-theoretic meaning as arising from the statistical distance between the data acquired by optimal measurements carried out for distinguishing between ρ_1 and ρ_2 [38].

In Appendix B we determine $d(\rho_1,\rho_2)$ for two spin-$\frac{1}{2}$ density matrices ρ_1 and ρ_2, given as in Eq. (56), with Bloch vectors $\vec{n}_1$ and $\vec{n}_2$, respectively[23]:

$$2d(\rho_1,\rho_2) - 1 = \vec{n}_1 \cdot \vec{n}_2 + \sqrt{1-|\vec{n}_1|^2}\sqrt{1-|\vec{n}_2|^2}. \tag{87}$$

For pure states $|\vec{n}_\alpha|=1$, and we expectedly obtain from Eq. (87) proportionality between the overlap and the scalar product of the two Bloch vectors.

For the mixing work we have, from Eq. (59),

$$\frac{\Delta\mathcal{W}}{\mathcal{N}} = \frac{\varepsilon}{2}(\lambda_1|\vec{n}_1| + \lambda_2|\vec{n}_2| - \sqrt{\lambda_1^2|\vec{n}_1|^2 + \lambda_2^2|\vec{n}_2|^2 + 2\lambda_1\lambda_2\vec{n}_1 \cdot \vec{n}_2}). \tag{88}$$

When comparing Eq. (88) with Eq. (87) we see that if only the scalar product $\vec{n}_1 \cdot \vec{n}_2$ is varied—with the modules $|\vec{n}_1|$ and $|\vec{n}_2|$ being fixed—making the two states closer, the mixing work $\Delta\mathcal{W}$ indeed monotonically decreases. In particular, this is the case for pure states ρ_1 and ρ_2. However, as seen from Eq. (87), for mixed states ρ_1 and ρ_2 the scalar product between the corresponding Bloch vectors is only one aspect of closeness. To look at another setup, vary $|\vec{n}_1|$ with $|\vec{n}_2|$ while keeping their mutual angle ϕ fixed [see Eq. (77) for the definition of ϕ]. Note that $\Delta\mathcal{W}$ always increases with $|\vec{n}_1|$:

$$\frac{\partial\Delta\mathcal{W}}{\partial|\vec{n}_1|} = \frac{\varepsilon}{2|\lambda_1\vec{n}_1 + \lambda_2\vec{n}_2|}(|\lambda_1\vec{n}_1 + \lambda_2\vec{n}_2| - \lambda_1|\vec{n}_1| - \lambda_2|\vec{n}_2|\cos\phi)$$

$$\geqslant 0. \tag{89}$$

On the other hand, we have, from Eq. (87),

$$\frac{\partial d(\rho_1,\rho_2)}{\partial|\vec{n}_1|} = \frac{1}{2\sqrt{1-|\vec{n}_1|^2}}(\cos\phi|\vec{n}_2|\sqrt{1-|\vec{n}_1|^2} - |\vec{n}_1|\sqrt{1-|\vec{n}_2|^2}). \tag{90}$$

When the scalar product is positive, $\vec{n}_1 \cdot \vec{n}_2 = \cos\phi > 0$, Eq. (90) can be positive; i.e., the states can get closer with increasing $|\vec{n}_1|$, if $|\vec{n}_1|$ is sufficiently small or if $|\vec{n}_2|$ is sufficiently close to 1. Comparing with Eq. (89) we conclude that *it is possible to make the two states of the mixing substances closer to each other and simultaneously increase the mixing work.*[24]

We stress that all conclusions of the present section are valid under other reasonable measures of distance between

---

[23]Note the difference with $2\mathrm{tr}(\rho_1\rho_2)-1=\vec{n}_1\cdot\vec{n}_2$.

[24]Note that Ref. [16] discusses a similar situation in classical chemical physics. The analogy, however, appears to be superficial, since the author of Ref. [16] bases his conclusions on the nonadditive classical formula $S_{cl}(N,V)=N\ln V$ for entropy.

ρ_1 and ρ_2—e.g., $\text{tr}[(\rho_1 - \rho_2)^2]$. Indeed, it amounts to a simple check that the qualitative conclusion we got after Eq. (90) is valid as well for this measure of closeness.

IX. CONCLUSION

Since its formulation in the late 1870s, the Gibbs paradox has, lacking a simple solution, become a quest for the understanding of phenomenological thermodynamics from a more fundamental theory. This attempt to go to a deeper level is the reason for its importance [4–9]. Its understanding happens to have several layers. First, it was realized that it is necessary to take into account explicitly the difference between the particles, which drives the classical formulation of the paradox, but how much they differ shows up nowhere in formulas. Together with the separation of characteristic relaxation times and the von Neumann definition of entropy, this brought about the quantum mixing entropy argument which for many years was seen as the resolution of the Gibbs paradox [6–10]. It was, however, pointed out that the argument introduces a new conceptual difficulty precisely when it claims to solve the paradox [14]. The details being presented in Sec. III, we simply recall that this difficulty has to do with the features of entropy, more precisely, with the fact that the entropy is not a sufficiently primitive (first-order) quantity in the situation at hand. So a deeper reduction level has to be involved for the resolution of the Gibbs paradox.

In our opinion, the basic reason why classical thermodynamics fails for the understanding of mixing entropy is that the difference between an A atom and a B atom is not dealt with properly. We are thus left with the search for a more fundamental approach. Such a possibility is offered by the field of quantum thermodynamics, which has been considered in recent years by several groups; see, e.g., [18,19,22,27,30–32,34,36,40–44].

In the current paper we have presented an explanation of the Gibbs paradox within this field. Here the notion of entropy is known to be easily blurrred, and a paradigm shift is called for towards the more ancient concept of work (energy transferred to macroscopic work sources), which still plays a clear and empirical role. In particular, quantum thermodynamics applies to finite systems; e.g., the basic formulations of the second law are well defined both conceptually and operationally [30]. Indeed, it could be shown that Landauer's principle that connects the minimal energy dispersion to erase one bit of information $\Delta Q \geqslant kT \ln 2$ may lose its validity in the domain of quantum thermodynamics [19], while the Maxwell demon problem just found new viewpoints there [22,45].

It was further shown that the maximally extractable work (which we called "ergotropy" in an earlier paper with R. Balian [34]) can be clearly defined before and after the mixing process. The difference between them defines the maximal mixing work or mixing ergotropy, a non-negative quantity which smoothly goes to zero when the substances become more and more equal to each other, as it is for a single substance, thus solving the Gibbs paradox in the work formulation. (As should be clear from our presentation, we consider that the Gibbs paradox in its entropic formulation

has not been properly solved so far and that we even do not believe that it is consistently resolvable in that form.)

In contrast to entropy, the features of work can be directly based on the first principles of quantum mechanics and are well defined for any (equilibrium or nonequilibrium) state of a system interacting with macroscopic sources of work. In particular, there is no need to involve features of thermodynamical reversibility for defining and interpreting the mixing work; see in this context footnotes 14 and 15. On top of that, the concept of maximal work has a well-defined operational character, because it is always defined with respect to a definite class of work sources acting on the system of interest. The features of work and entropy are recalled and contrasted in Secs. V A and IV A 1 and Appendix A. Recall in this context that the concept of work was already employed in the literature devoted to the Gibbs paradox [6,15], but its potential applications were conceived in the framework of phenomenological thermodynamics. In that way, they encounter almost all objections raised against the mixing entropy argument. Only after the problem of maximal work extraction was solved from first principles [34] did it became possible to approach an explanation of the Gibbs paradox with the help of the mixing work. This explanation is free of the difficulties which plagued the quantum entropy argument.

To keep our approach as natural as possible, we have supposed that, after allowing the gases to mix, the translational degrees of freedom equilibrate rather quickly, while their spin degrees of freedom do not equilibrate at all at the time scales for which our discussion applies because their dynamics are supposed to take place on a much larger time scale. For this reason, these degrees of freedom can be considered as not coupled to the bath, which saves us from discussing the more complicated situation where heat exchange of the spins with the bath would also matter.

The consistent resolution of the paradox presents features that might not have been anticipated before. It appears that less precise control can, depending on the situation, bring a larger or a smaller amount of mixing work. We have also seen that a naive intuition relating the degree of mixing and the distinguishability with the mixing ergotropy may not always be correct: sometimes making the initial states of the mixed substances closer to each other (in the proper information-theoretic sense) can make the amount of mixing work larger. These are warnings against a direct association of physical irreversibility (i.e., mixing work) with lack of information: while the amount of mixing work is nonzero due to less information on the identity of atoms in the post-mixed state, the relation of this lack of information to the physical irreversibility can be nontrivial and counterintuitive.

APPENDIX A: THE FORMULATION OF THE SECOND LAW SETS THE CHOICE OF THE ENTROPY

Here we shall recall why entropy is not uniquely defined and why its possible definitions depend on the formulation of the second law of thermodynamics.

Consider an adiabatically isolated process done on a quantum system described by density matrix $\rho(t)$. The process is realized via a time-dependent Hamiltonian, with the cyclic feature defined according to Eq. (25). The evolution of the system starts from some Gibbsian equilibrium state at a positive temperature. It is well known from thermodynamics [1,2] and can be derived from the first principles of quantum mechanics—see [28,29] and our discussion in Sec. V B—that in this process the system consumes positive work which is the statement of the second law in Thomson's formulation. It is natural to look for the counterpart of this formulation in terms of entropy. Since the dynamics of the thermally isolated system is unitary, the von Neumann entropy $S_{\mathrm{vN}}(\rho(t))$ is constant in time, so it is not suitable for being the counterpart of the Thomson's formulation. This argument is sometimes dismissed on the grounds that the unitary dynamics is reversible and thus the constant behavior of the von Neumann entropy is reasonable. In the present context this seems incorrect, in particular because a positive amount of work is put into the system in accordance with Thomson's formulation of the second law.

In the spirit of the relevant entropy approach [2] (there are many entropies, each one for its own situation and its own use) we can regard as physical another entropy

$$S_{\mathrm{T}}(t) = -\sum_k \pi_k(t)\ln \pi_k(t), \qquad (A1)$$

with π_k being the time-dependent probabilities of various values of the system's energy (given by the time-dependent Hamiltonian) in the state $\rho(t)$. This definition of entropy was proposed and advocated by Tolman [46]. For the considered process, S_{T} does have several reasonable properties.

(i) At the end of the cyclic-Hamiltonian process, S_{T} is larger than in the initial equilibrium state [28,29].

(ii) Under conditions specified in [30], the change of S_{T} is minimal for the adiabatically slow process, again as required by thermodynamics.

(iii) S_{T} is maximal in equilibrium.

Each of these three features corresponds to a specific formulation of the second law. Features (i) and (ii) will not be valid when using the von Neumann entropy.

Thus, we are led to employ the Tolman definition of entropy following to the requirements of the second law.

Let us now consider an isothermal process, where the system (e.g., a spin or a Brownian particle) weakly interacts with an equilibrium thermal bath at temperature T. The bath being in equilibrium means for the present context two things. First, it starts in the equilibrium state at temperature T, and, second, its relevant characteristic times are much larger than those of the system. (An additional feature of weak interactions was stressed by us above.) It is again well known from phenomenological thermodynamics and is derived from the first principles of quantum mechanics that during the relaxation of the system to equilibrium, the (non-equilibrium) free energy decays, a statement known as the H theorem [1,2,29]:

$$\frac{dF}{dt} = \frac{d}{dt}[E(t) - TS_{\mathrm{vN}}(t)] \equiv \frac{d}{dt}\{\mathrm{tr}(\rho(t)H) + T\mathrm{tr}[\rho(t)\ln \rho(t)]\}$$

$$\leq 0, \qquad (A2)$$

where $\rho(t)$ is the density matrix of the system and where H is its time-independent Hamiltonian. Note especially that the H theorem will in general not be valid if instead of the von Neumann entropy we shall use in (A2) the Tolman entropy S_{T}. Thus, here for isothermal processes we had to return to the von Neumann definition of entropy. What is the proper definition of entropy when the process is neither isothermal nor thermally isolated is in general not known [31].

In short, in statistical physics the definitions of entropy are contextual, since they already depend on various formulations of the second law. It is, therefore, questionable whether arguments based on entropies are able to resolve thermodynamical paradoxes.

APPENDIX B: A USEFUL IDENTITY

Here we outline how to calculate the overlap defined in (87) for two spin-$\frac{1}{2}$ density matrices

$$\rho = \frac{1}{2}(1 + \vec{n}_\alpha \cdot \vec{\sigma}), \quad \alpha = 1,2. \qquad (B1)$$

We need the following facts. First, note that the square root of ρ is most conveniently calculated when representing ρ as

$$\rho = \frac{1}{2}(1 + \sin \theta \vec{\chi} \cdot \vec{\sigma}), \qquad (B2)$$

where $0 \leq \theta \leq \pi/2$ and where $\vec{\chi}$ is a unit vector $|\vec{\chi}| = 1$. Then,

$$\sqrt{\rho} = \sqrt{\frac{1}{2}}\left(\cos \frac{\theta}{2} + \sin \frac{\theta}{2}\vec{\chi} \cdot \vec{\sigma}\right) \qquad (B3)$$

and

$$\mathrm{tr}\sqrt{\rho} = \sqrt{2} \cos \frac{\theta}{2} = \sqrt{1 + \sqrt{1 - |\vec{n}|^2}}. \qquad (B4)$$

Next, we need the known identity for Pauli matrices,

$$(\vec{n}_1 \cdot \vec{\sigma})(\vec{n}_2 \cdot \vec{\sigma}) = (\vec{n}_1 \cdot \vec{n}_2) + i\vec{\sigma} \cdot [\vec{n}_1 \times \vec{n}_2], \qquad (B5)$$

where $[\vec{n}_1 \times \vec{n}_2]$ is the vector product. And, finally, the last ingredient is given by

$$(\vec{\chi} \cdot \vec{\sigma})(\vec{\xi} \cdot \vec{\sigma})(\vec{\chi} \cdot \vec{\sigma}) = [2(\vec{\chi} \cdot \vec{\xi})\vec{\chi} - \vec{\xi}] \cdot \vec{\sigma}, \qquad (B6)$$

where χ and ξ are unit vectors. Equation (B6) follows from Eq. (B5) and the double-vector-product identity

$$[\vec{n}_3 \times [\vec{n}_1 \times \vec{n}_2]] = \vec{n}_1(\vec{n}_2 \cdot \vec{n}_3) - \vec{n}_2(\vec{n}_1 \cdot \vec{n}_3). \qquad (B7)$$

[1] L. D. Landau and E. M. Lifshitz, *Statistical Physics, I* (Pergamon Press, Oxford, 1978).

[2] R. Balian, *From Microphysics to Macrophysics* (Springer, Berlin, 1992), Vol. I.

[3] J. Willard Gibbs, *On the Equilibrium of Heterogeneous Substances* (Connecticut Academy of Science, 1875–1878) pp. 108–248; J. Willard Gibbs, *The Scientific Papers*, (Longman, Green and Co., New York, 1906), pp. 165–168. *Selected Papers on Thermodynamics and Statistical Physics* (Cambridge University Press, London, 1951).

[4] E. Schrödinger, *Statistical Thermodynamics* (Cambridge University Press, London, 1952).

[5] I. E. Tamm, Usp. Fiz. Nauk **6**, 112 (1926).

[6] A. Landé, *New Foundations of Quantum Mechanics* (Cambridge University Press, Cambridge, England, 1965); *Foundations of Quantum Theory* (Yale University Press, Yale, 1955).

[7] M. J. Klein, Am. J. Phys. **26**, 80 (1958).

[8] V. L. Luboshitz and M. I. Podgoretskii, Sov. Phys. Usp. **14**, 662 (1972).

[9] Ya. M. Gelfer, V. L. Luboshitz, and M. I. Podgoretskii, *Gibbs Paradox and Identity of Particles in Quantum Mechanics* (Nauka, Moscow, 1975) (in Russian).

[10] A. M. Lesk, J. Phys. A **13**, L111 (1980).

[11] Yu. S. Varshavskii and A. B. Sheinin, Dokl. Akad. Nauk SSSR **148**, 1099 (1963).

[12] N. G. van Kampen, in *Essays in Theoretical Physics in Honor of Dirk ter Haar*, edited by W. E. Parry (Pergamon, Oxford, 1984).

[13] C. Y. Tseng and A. Caticha, e-print cond-mat/0109324.

[14] D. Dieks and V. van Dijk, Am. J. Phys. **56**, 430 (1988).

[15] L. A. Blumenfeld and A. Yu. Grosberg, Biophysics (Engl. Transl.) **40**, 653 (1995).

[16] Shu-Kun Lin, Electron. J. Theor. Chem. **1**, 135 (1996).

[17] E. T. Jaynes, in *Maximum Entropy and Bayesian Methods*, edited by C. R. Smith, G. J. Erickson, and P. O. Neudorfer (Kluwer, Dordrecht, 1992).

[18] A. E. Allahverdyan, R. Balian, and Th. M. Nieuwenhuizen, J. Mod. Opt. **51**, 2703 (2004).

[19] A. E. Allahverdyan, and Th. M. Nieuwenhuizen, Phys. Rev. E **64**, 056117 (2001).

[20] P. A. Corning and S. J. Kline, Syst. Res. Behav. Sci. **15**, 273 (1998).

[21] E. Schrödinger, *What is Life?* (Cambridge University Press, Cambridge, England, 1944).

[22] M. O. Scully, Y. Rostovtsev, Z. Sariyanni, and M. S. Zubairy, Physica E (Amsterdam) **29**, 29 (2005); Y. Rostovtsev, Z. Sariyanni, M. S. Zubairy, and M. O. Scully, *ibid.* **29**, 40 (2005); Z. Sariyanni, Y. Rostovtsev, M. S. Zubairy, and M. O. Scully, *ibid.* **29**, 47 (2005).

[23] R. Rosen, Philos. Sci. **31**, 232 (1964).

[24] A. Peres, *Quantum Theory: Concepts and Methods* (Kluwer Academic, Dordrecht, 1995).

[25] A. E. Allahverdyan, R. Balian, and Th. M. Nieuwenhuizen, Europhys. Lett. **61**, 452 (2003); e-print cond-mat/0408316.

[26] E. Schrödinger, Proc. Cambridge Philos. Soc. **32**, 446 (1936); L. P. Hughston, R. Jozsa, and W. K. Wootters, Phys. Lett. A **183**, 14 (1993).

[27] A. E. Allahverdyan and Th. M. Nieuwenhuizen, Phys. Rev. E **71**, 066102 (2005).

[28] G. N. Bochkov and Yu. E. Kuzovlev, Sov. Phys. JETP **45**, 125 (1977); I. M. Bassett, Phys. Rev. A **18**, 2356 (1978); W. Pusz and S. L. Woronowicz, Commun. Math. Phys. **58**, 273 (1978); A. Lenard, J. Stat. Phys. **19**, 575 (1978); W. Thirring, *A Course in Mathematical Physics 4: Quantum Mechanics of Large Systems* (Springer, Vienna, 1983); H. Tasaki, e-print cond-mat/0009244; J. Kurchan, e-print cond-mat/0007360; A. E. Allahverdyan and Th. M. Nieuwenhuizen, Physica A **305**, 542 (2002).

[29] G. Lindblad, *Non-Equilibrium Entropy and Irreversibility* (Reidel, Dordrecht, 1983).

[30] A. E. Allahverdyan and Th. M. Nieuwenhuizen, Phys. Rev. E **71**, 046107 (2005).

[31] A. E. Allahverdyan and Th. M. Nieuwenhuizen, Phys. Rev. Lett. **85**, 1799 (2000); Th. M. Nieuwenhuizen and A. E. Allahverdyan, Phys. Rev. E **66**, 036102 (2002).

[32] D. Janzing, J. Stat. Phys. (to be published), e-print quant-ph/0502019.

[33] P. Perrot, *A to Z of Thermodynamics* (Oxford University Press, New York, 1998).

[34] A. E. Allahverdyan, R. Balian and Th. M. Nieuwenhuizen, Europhys. Lett. **66**, 419 (2004); e-print cond-mat/0401574.

[35] V. Jurdjevic and H. J. Sussmann, J. Differ. Equations **12**, 313 (1972); A. G. Butkovskiy and Yu. I. Samoilenko, *Control of Quantum Mechanical Processes and Systems* (Kluwer Academic, Dordrecht, 1990).

[36] D. Janzing, e-print quant-ph/0511271.

[37] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and its Applications* (Academic Press, New York, 1979).

[38] S. L. Braunstein and C. M. Caves, Phys. Rev. Lett. **72**, 3439 (1994); C. A. Fuchs and C. M. Caves, Open Syst. Inf. Dyn. **3**, 1 (1995); e-print quant-ph/9604001.

[39] D. J. C. Bures, Trans. Am. Math. Soc. **135**, 199 (1969); A. Uhlmann, Rep. Math. Phys. **9**, 273 (1976); M. Hübner, Phys. Lett. A **163**, 239 (1992); R. Jozsa, J. Mod. Opt. **41**, 2315 (1994).

[40] M. O. Scully, Phys. Rev. Lett. **87**, 220601 (2001); 0031-9007Phys. Rev. Lett. **88**, 050602 (2002); Y. V. Rostovtsev, A. B. Matsko, N. Nayak, M. S. Zubairy, and M. O. Scully, Phys. Rev. A **67**, 053811 (2003); A. E. Hill, Yu. V. Rostovtsev, and M. O. Scully, Phys. Rev. A **72**, 043802 (2005).

[41] T. Feldmann and R. Kosloff, Phys. Rev. E **61**, 4774 (2000); R. Kosloff and T. Feldmann, *ibid.* **65**, 055102(R) (2002); Y. Rezek and R. Kosloff, e-print quant-ph/0601006; J. P. Palao, R. Kosloff, and J. M. Gordon, e-print quant-ph/0106048.

[42] J. Gemmer, A. Otte, and G. Mahler, Phys. Rev. Lett. **86**, 1927 (2001); J. Gemmer and G. Mahler, Eur. Phys. J. D **17**, 385 (2001); **31**, 249 (2003). J. Gemmer, M. Michel, and G. Mahler, *Quantum Thermodynamics*, Vol. 657 of *Lecture Notes in Physics* (Springer, Berlin, 2004).

[43] A. E. Allahverdyan, R. Serral Gracia, and Th. M. Nieuwenhuizen, Phys. Rev. Lett. **93**, 260404 (2004); Phys. Rev. E **71**, 046106 (2005).

[44] D. Janzing, P. Wocjan, R. Zeier, R. Geiss, and Th. Beth, Int. J. Theor. Phys. **39**, 2217 (2000).

[45] Th. M. Nieuwenhuizen and A. E. Allahverdyan, in *Quantum Limits to the Second Law*, edited by D. P. Sheenan, AIP Conf. Proc. (AIP, Melville, NY, 2002), p. 436.

[46] R. C. Tolman, *The Principles of Statistical Mechanics* (Oxford University Press, London, 1938).