The Gibbs paradox revisited from the fluctuation theorem with absolute irreversibility

Yüto Murashita$^1$ and Masahito Ueda$^{1,2}$

$^1$Department of Physics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-0033, Japan
$^2$RIKEN Center for Emergent Matter Science (CEMS), Wako, Saitama 351-0198, Japan

(Dated: November 28, 2016)

The inclusion of the factor $\ln(1/N!)$ in the thermodynamic entropy proposed by Gibbs is shown to be equivalent to the validity of the fluctuation theorem with absolute irreversibility for gas mixing.

As analyzed by van Kampen [1], what is collectively called the Gibbs paradox actually involves the following three distinct issues (see Fig. 1):

GP-I: consistency within thermodynamics [1,3].
GP-II: consistency within classical statistical mechanics [1,4].
GP-III: consistency between thermodynamics and classical statistical mechanics [1,3,5].

As detailed below, all of them have been settled in the thermodynamic limit [1,5]. However, in view of the growing interest in small thermodynamic systems [3,5], it is worth revisiting the Gibbs paradox in this context. The conventional resolutions for GP-I [1,3] and GP-II [1] apply to small thermodynamic systems as well. However, the one for GP-III cannot be applied to small thermodynamic systems [3].

The essence of GP-III is to determine ambiguity function $f(N)$ which gives the difference between the thermodynamic entropy $S$ and the statistical-mechanical entropy $S_{\text{stat}}$ as

$$S(T,V,N) = S_{\text{stat}}(T,V,N) + k_B f(N). \quad (1)$$

In the thermodynamic limit, extensivity of the thermodynamic entropy leads to $f(N) = -N \ln N + cN$, where $c$ is a constant [3,9]. However, extensivity breaks down in small thermodynamic systems. In this Letter, we show that in small thermodynamic systems the fluctuation theorem [7,8,10,17] with absolute irreversibility [18] in place of extensivity leads to $f(N) = \ln(1/N!) + cN$.

A brief history. In the 1870s, Gibbs put forth the problem of GP-I [2] pp.227-229). Whereas identical-gas mixing at the same temperature and pressure does not produce an extensive amount of entropy, different-gas mixing does. The fact that the latter entropy production stays constant even when the difference between two species becomes infinitesimal seems paradoxical. Gibbs ascribes this to how we define thermodynamic states. This idea was further elaborated later. For example, van Kampen [1 Sec.2-4] argued that the Clausius equality defines only the difference in entropy of a system where the particle number $N$ is conserved. Therefore, to determine the $N$-dependence, we have to introduce a reversible process in which $N$ is varied. For identical gases, opening a channel between the two partitions is reversible. For different gases, this process is irreversible and consequently we have to invoke the familiar reversible procedure with semi-permeable walls. In this way, difference between identical- and different-gas mixings reduces the two different processes used to extend the definition of entropy. Which of the two processes we should use depends on whether we intend to use the semi-permeable walls or on how we operationally define the identical thermodynamic states. “Thus, whether such a process is reversible or not depends on how discriminating the observer is. The expression for the entropy [...] depends on whether or not he is able and willing to distinguish between the molecules A and B” [1 pp.306-307]. As Jaynes put it [3 Sec.5], the thermodynamic entropy has “anthropomorphic” nature in that its value depends on “human information.”

About a quarter century later, Gibbs discussed GP-III in his renowned textbook [5 Chap.XV] by introducing two types of phase spaces, “the generic phase” and “the specific phase.” In “the generic phase,” a set of points in phase space are identified when any point in the set can be obtained by permutation of particles of other points, whereas in “the specific phase,” they are not. Gibbs preferred “the generic phase” to “the specific phase” because the former results in no entropy production upon identical-gas mixing [5 pp.206-207]. Thus, Gibbs introduced the factor $\ln(1/N!)$ in classical statistical mechanics to ensure the consistency with thermodynamic observation. Van Kampen [1 Sec. 9] later claimed that this prescription is guaranteed by convention (namely, by assuming $f(N) = 0$ in Eq. [1] for the classical statistical-mechanical entropy in the generic phase). However, Jaynes [3 Sec.9] was able to determine...
be modified \[3, \text{Sec.9}\] (see \textit{Discussions}\). Start with open systems, the Clausius definition should be modified with temperature \(T\) and particle number \(N\) is taken along a quasi-static process. We note that the change in the entropy is not a real event \[20, \text{pp.61}\]. Now, in many standard textbooks (e.g., see Refs. \[21–26\]), the thermodynamic systems. The fluctuation theorem, especially the Jarzynski equality
\[\beta \langle W \rangle = 1\], determines the equilibrium free-energy difference \(\Delta F\) between two configurations from statistics of work \(W\) in a nonequilibrium process. Unfortunately, the Jarzynski equality cannot apply to gas mixing because the initial state is not a global thermodynamic system.

**Conventional resolution to GP-III.** GP-III concerns the relation between the thermodynamic entropy \(S\) and the classical statistical-mechanical entropy \(S^{\text{stat}}\). While the latter is defined as the Gibbs entropy, namely the Shannon entropy of the canonical state, the former is defined through the Clausius equality as \(\Delta S = \int \delta Q / T\), where \(\delta Q\) is the heat transferred from the heat reservoir with temperature \(T\) to the system and the integration is taken along a quasi-static process. We note that the particle number \(N\) cannot be varied in this process; if we start with open systems, the Clausius definition should be modified \[5, \text{Sec.9}\] (see \textit{Discussions} for detail). Therefore, the thermodynamic entropy defined by the Clausius equality has ambiguity about \(N\). Consequently, the relation between the thermodynamic entropy and statistical-mechanical entropy should involve some ambiguity function \(f(N)\) as in Eq. \[1\] \[1\] Eq. (14) \[3\] Eq. (16)]. The theme of GP-III is to demonstrate \(f(N) = \ln(1/N!)\) (up to a constant per particle) for the classical statistical-mechanical entropy in the specific phase (or equivalently \(f(N) = 0\) for the one in the generic phase). Incidentally, we note that “exactly the same argument will apply in quantum statistical mechanics” \[3\] Sec.9]. Therefore, GP-III is not the issue of classical statistical mechanics in particular, but any statistical mechanics in general.

Jaynes \[3\] Sec.9] resolved this issue for free particles by analogy with the way by which Pauli phenomenologically determined the \(N\)-dependence of the thermodynamic entropy \[9\] pp.38-39]. By requiring that the thermodynamic entropy satisfy extensivity
\[S(T, qV, qN) = qS(T, V, N), \quad \forall q > 0, \quad (2)\]
and by invoking the formula of the statistical-mechanical entropy for an ideal gas \(S^{\text{stat}}(T, V, N) = Nk_B(3/2) \ln T + \ln V + \text{const}.)\), one can show
\[f(N) = N f(1) - N \ln N. \quad (3)\]
Here, the first term on the right-hand side of Eq. \[3\] represents an intrinsic entropy per particle, while the second term amounts to the factor \(- \ln N!\) in the large-\(N\) limit. Thus, the requirement of extensivity leads to the desired factor in the thermodynamic limit \((N \to \infty)\).

However, extensivity, in general, breaks down in small thermodynamic systems, and therefore the Pauli-Jaynes resolution cannot apply to them. Jaynes \[3\] Sec.9] expressed his concern about this point by stating, “The Pauli correction was an important step in the direction of getting “the bulk of things” right pragmatically; but it ignores the small deviations from extensivity that are essential for treatment of some effects; and in any event it is not a fundamental theoretical principle. A truly general and quantitatively accurate definition of entropy must appeal to a deeper principle which is hardly recognized in the current literature, [...]” Thus, we need a different guiding principle to determine \(f(N)\) in small thermodynamic systems.

**Method.** We use the fluctuation theorem \[8, 10–17\] as the guiding principle for small thermodynamic systems. The fluctuation theorem, especially the Jarzynski equality \(e^{-(W - \Delta F)} = 1\) \[12\], determines the equilibrium free-energy difference \(\Delta F\) between two configurations from statistics of work \(W\) in a nonequilibrium process. Unfortunately, the Jarzynski equality cannot apply to gas mixing because the initial state is not a global canonical equilibrium state of the entire box, which invalidates the prerequisite of the Jarzynski equality.

We overcome this difficulty by applying the fluctuation theorem in the presence of absolute irreversibility
In particular, if the free-energy difference is calculated for gas mixing in the isothermal environment is equivalent to the sum of the probabilities of these events. Thus, absoluteness, irreversible events is larger due to the presence of the rightmost event which has no counterpart in (a).

\[ \langle e^{-\beta (W - \Delta F)} \rangle = 1 - \lambda \] \[ \text{(5)} \]

which allows an initial canonical state with spatial constraints. Here, the term \( \lambda \) is the absolutely irreversible probability which is mathematically defined as the integral of the singular part of the time-reversed probability measure with respect to the forward probability measure \[ \text{[13]} \]. For this singular part, the Crooks fluctuation theorem \[ \text{[13]} \] is ill-defined because the dissipative work is divergent with singularly strong irreversibility. This is why the Jarzynski equality \[ \text{[14]} \] should be modified in absolutely irreversible processes. We can calculate \( \lambda \) as the sum of the probabilities of the time-reversed events whose corresponding forward events vanish.

To demonstrate the existence of absolute irreversibility in gas mixing, we consider the time-reversed process of gas mixing, namely, wall insertion as illustrated in Fig. 2. The events indicated by blue arrows are absolutely irreversible because they have no counterparts in the forward process. The absolutely irreversible probability \( \lambda \) is given by the sum of the probabilities of these events. Thus, absolute irreversibility of gas mixing originates from fluctuations in the number of particles in each box after wall insertion and may be interpreted as an inevitable loss of information on the particle number in each box before gas mixing \[ \text{[29]} \].

**Main result.** We now show that the validity of the fluctuation theorem with absolute irreversibility

\[ \langle e^{-\beta (W - \Delta F)} \rangle = 1 - \lambda \]

for gas mixing in the isothermal environment is equivalent to the following choice of the ambiguity function:

\[ f(N) = N f(1) - \ln N! \]

In particular, if the free-energy difference is calculated from Eq. \[ \text{(4)} \], the thermodynamic entropy is defined and \( f(N) \) is given by Eq. \[ \text{(5)} \]. Recall that, in the thermodynamic limit, the validity of extensivity \[ \text{[2]} \] is equivalent to Eq. \[ \text{(3)} \]. Thus, the fluctuation theorem with absolute irreversibility \[ \text{(1)} \] in small thermodynamic systems takes the place of extensivity \[ \text{[2]} \] in the thermodynamic limit. Note that Eq. \[ \text{(4)} \] differs from Eq. \[ \text{(3)} \] at the subleading order, namely in the mesoscopic regime.

**Proof.** First, we assume Eq. \[ \text{(4)} \] to derive Eq. \[ \text{(5)} \]. We consider mixing of identical gases and that of different gases as illustrated in Fig. 2 and assume that gases are initially in thermal equilibrium with inverse temperature \( \beta \). We assume that the gas particles have the same mass and the interaction potential between different particles is the same as that between two identical particles. Therefore the Hamiltonians for the two cases are identical. We also assume that the interactions do not break additivity. By the term additivity, we mean that the thermodynamic entropy of a system consisting of independent subsystems is equal to the sum of the entropies of the individual subsystems. The left (right) box is initially filled with an \( N \)-particle gas with volume \( M \pi \) and then the wall is removed. For identical and different gases, Eq. \[ \text{(4)} \] gives \( \langle e^{-\beta (W_{\text{id}} - \Delta F_{\text{id}})} \rangle = 1 - \lambda_{\text{id}} \) and \( \langle e^{-\beta (W_{\text{dif}} - \Delta F_{\text{dif}})} \rangle = 1 - \lambda_{\text{dif}} \), respectively. Because no work is performed upon gas mixing, \( \langle e^{-\beta W_{\text{dif}}} \rangle_{\text{id}} = 1 \) holds. Therefore, we obtain \( \Delta F_{\text{id}} - \Delta F_{\text{dif}} = k_B T \ln[(1 - \lambda_{\text{id}})/(1 - \lambda_{\text{dif}})] \). Now, we evaluate the left-hand side of this equality. Since \( \Delta F = \Delta U - T \Delta S \) and \( \Delta U_{\text{id}} = \Delta U_{\text{dif}} \) by the assumption of the same interaction potential, we obtain \( \Delta F_{\text{id}} - \Delta F_{\text{dif}} = -T(\Delta S_{\text{id}} - \Delta S_{\text{dif}}) \). Moreover, since the partition functions for the two processes calculated by statistical mechanics based on “the specific phase” are equal, we have \( \Delta S_{\text{id}} = \Delta S_{\text{dif}} \). It follows then from Eq. \[ \text{(4)} \] that the difference between the thermodynamic entropy productions can be evaluated only in terms of \( f(N) \) as \( \Delta S_{\text{id}} - \Delta S_{\text{dif}} = k_B \Delta f_{\text{id}} - \Delta f_{\text{dif}} \). For identical-gas mixing, the ambiguity function of the initial state is given by \( f_{\text{id}} = f(M + N) \) due to the assumption of additivity, whereas that of the final state is \( f_{\text{dif}} = f(M + N) \), resulting in \( \Delta f_{\text{id}} = f(M + N) - f(M) \). In contrast, for different-gas mixing, we can quasi-statically connect the initial and final states by using a semipermeable membrane. As a result, the entropy difference is uniquely determined by the Clausius equality, and therefore the ambiguity function remains unchanged, and hence \( \Delta f_{\text{dif}} = 0 \). As a consequence, we obtain

\[ f(M + N) - f(M) = \ln \frac{1 - \lambda_{\text{dif}}}{1 - \lambda_{\text{id}}} \]

This result implies that the ambiguity in the thermodynamic entropy represented by \( f(N) \) is directly related to and hence can be removed by the degree of absolute irreversibility in the gas mixing process.

To find \( \lambda \), let us consider the time-reversed process of
gas mixing, i.e., wall insertion as illustrated in Fig. 2. After the insertion, the state may or may not return to the original one. The events in which the state does not return to the original state are absolutely irreversible because they have no counterparts in the forward process; the event in which the state returns to the original state is the only one without absolute irreversibility. Therefore, $1 - \lambda$ is equal to the probability that the original state is restored after the wall insertion. Although this probability cannot be calculated explicitly in the presence of interactions, we can compare the probabilities for the two cases. The number of combinations for the $M + N$ particles in which the left box contains $M$ particles after the wall insertion is equal to the binomial coefficient $\binom{M+N}{M}$. These partitions can be realized with equal probability due to the assumption of identical interactions. For identical gases, all of them are identified with the original state, whereas there is only one such state for different gases. Therefore, we obtain $f(M + N) - f(M) - f(N) = -\ln \binom{M+N}{M}$. The solution of this functional equation is given by Eq. (5).

Next, we assume Eq. (5) to show Eq. (4). We first consider identical-gas mixing. Let $Z(N)$ denote the partition function of the gas with the particle number $N$ and the volume $N\nu$ calculated by statistical mechanics based on “the specific phase.” Then, from Eq. (5), the thermodynamic free energy is evaluated as $F(N) = -\beta^{-1} \ln Z(N) - \beta^{-1} f(N) = -\beta^{-1} \ln[Z(N)(f(1))^N/N!]$. Consequently, the left-hand side of Eq. (4) for identical-gas mixing reduces to $\langle e^{-\beta (W_{id} - \Delta F_{id})}\rangle_{id} = e^{\beta \Delta F_{id}} = M+N C_M \cdot Z(M)/Z(M+N)$. The right-hand side gives the probability to restore by the wall insertion the original state in which the number of particles in the left (right) partition is $M (N)$. This is by definition equal to $1 - \lambda_{id}$. Hence, Eq. (4) for identical-gas mixing is shown. A similar procedure vindicates Eq. (4) for different-gas mixing. Thus, the main result has been proven.

Discussions. We have demonstrated that the validity of Eq. (4) is equivalent to that of Eq. (5). Therefore, mathematically speaking, requiring either of them produces the same results. However, from a physical point of view, thermodynamics should have an operational basis. In this sense, it is noteworthy that Eq. (4) defines the equilibrium free-energy difference from statistics of the work that we can operationally obtain by repeating a fixed nonequilibrium processes. Hence, we think that requiring Eq. (4) as the starting point is more appropriate for thermodynamics as an operational theory.

As clarified in the context of GP-I [1], the thermodynamic entropy depends on the operational capability of the beholder. Therefore, to deal with the thermodynamic entropy, we must define what we mean by the same and different states operationally. This is the conventional wisdom to resolve the Gibbs paradox. We do so by introducing different probabilities of absolute irreversibility for the two mixing processes because they are defined as the probability of time-reversed events that do not have the counterparts in the forward process. In this sense, we follow the conventional wisdom and claim no novelty on this point. Rather, we believe that any resolution of the Gibbs paradox is destined to follow it.

As we have discussed above the fluctuation theorem with absolute irreversibility (4) is closely related to extensivity (2). For simplicity, we consider mixing of identical ideal gases with the same temperature and volume: $(T, V, N) \rightarrow (T, 2V, 2N)$. In the thermodynamic limit, we can explicitly show from Eq. (4) that the thermodynamic entropy production upon gas mixing is sub-extensive, and such a sub-extensive correction is thermodynamically ignored, giving $S(T, V, N)|T, V, N) = S(T, 2V, 2N)$. Therefore, assuming additivity $S(T, V, N) = S(T, V) + S(T, V) = 2S(T, V, N)$, we obtain $S(T, 2V, 2N) = 2S(T, V, N)$, which is extensivity (2) with $q = 2$. Thus, Eq. (4) reduces to extensivity (2) in the thermodynamic limit.

We have dealt with interacting gases by assuming that the interaction does not break additivity. In the Pauli-Jaynes method [3, 9], only noninteracting gases were discussed. Although it is possible to extend their method to interacting gases such as van der Waals gases, a general discussion seems intractable because we have to invoke the specific form of the statistical-mechanical entropy. In contrast, our method can consider general interacting gases. We consider both identical- and different-gas mixing and compare their entropy production to cancel out the effects of interactions. We stress that to do so Eq. (4) plays a crucial role. We translate the free-energy difference $\Delta F$, which is not explicitly computable, into the absolutely irreversible probability $\lambda$, which is not either. However, $\Delta F_{id} - \Delta F_{dif}$ is computable because the right-hand side of Eq. (6) can be evaluated by combinatorics. Thus, the fluctuation theorem with absolute irreversibility (4) is crucial to remove the ambiguity $f(N)$ in the presence of interparticle interactions, although the fluctuations of work do not exist in gas mixing.

One may wonder why we do not consider the thermodynamic entropy of an open system from the beginning. We could extend the Clausius definition to an open system as $\Delta S = \int_{q,s} (\delta Q + \mu \delta N)/T$. However, to define the thermodynamic entropy in this way, we must define the chemical potential $\mu$ beforehand. The crucial point is that the chemical potential is defined from the very $N$-dependence of the thermodynamic entropy as $\mu = T \partial S/\partial N$. To avoid falling into a circular argument, we cannot start from an open thermodynamic system.

Finally, we discuss why we consider isothermal systems instead of isolated systems. First of all, the original discussion by Gibbs himself [2] considered the isothermal setup at a constant temperature. Furthermore, in small isolated systems, the relation between thermodynamics
and statistical mechanics is not completely understood. On the other hand, in isothermal systems, thermodynamic responses are considered to be reproduced by statistical mechanics as long as the heat bath is infinitely large. Thus, GP-III is more well-defined in isothermal systems than in isolated systems.

Conclusion. We have demonstrated that the validity of the fluctuation theorem with absolute irreversibility \[] is equivalent to the specific form \[] of the ambiguity function of the thermodynamic entropy in Eq. (1). Requiring Eq. (4) is more appropriate than requiring Eq. (5) from an operational point of view.

Acknowledgments. We appreciate useful comments by Naoyuki Sakumichi, Takashi Mori, Kiyoshi Kanazawa, Ken Funo, Yuto Ashida and Ruyusuke Hamazaki. This work was supported by KAKENHI Grant No. JP26287088 from the Japan Society for the Promotion of Science, a Grant-in-Aid for Scientific Research on Innovative Areas “Topological Materials Science (KAKENHI Grant No. JP15H05855), the Photon Frontier Network Program from MEXT of Japan, and the Mitsubushi Foundation. YM was supported by the Japan Society for the Promotion of Science through the Program for Leading Graduate Schools (MERIT) and JSPS Fellowship (JSPS KAKENHI Grant Number JP15J00410).

[1] N. G. van Kampen, *The Gibbs Paradox*, edited by W. E. Parry (Pergamon, 1984).
[2] J. W. Gibbs, *On the Equilibrium of Heterogeneous Substances* (Connecticut Acad. Sci., 1875-78).
[3] E. T. Jaynes, *The Gibbs Paradox*, edited by C. R. Smith, G. J. Erickson, and P. O. Neudorfer (Kluwer Academic Publishers, 1992).
[4] P. Ehrenfest and V. Trkal, Ann. Physik 65, 609 (1921).
[5] J. W. Gibbs, *Elementary Principles in Statistical Mechanics* (Yale University Press, 1902).
[6] K. Sekimoto, *Stochastic Enegetics* (Springer, 2010).
[7] R. Klages, W. Just, and C. Jarzynski, eds., *Nonequilibrium Statistical Physics of Small Systems* (Wiley-VCH, 2013).
[8] U. Seifert, Rep. Prog. Phys. 75, 126001 (2012).
[9] W. Pauli, *Thermodynamics and the Kinetic Theory of Gases* (MIT Press, 1973).
[10] D. J. Evans, E. G. D. Cohen, and G. P. Morriss, Phys. Rev. Lett. 71, 2401 (1993).
[11] G. Gallavotti and E. G. D. Cohen, Phys. Rev. Lett. 74, 2694 (1995).
[12] C. Jarzynski, Phys. Rev. Lett. 78, 2690 (1997).
[13] G. E. Crooks, Phys. Rev. E 60, 2721 (1999).
[14] T. Hatano and S. I. Sasa, Phys. Rev. Lett. 86, 3463 (2001).
[15] U. Seifert, Phys. Rev. Lett. 95, 040602 (2005).
[16] C. Jarzynski, Annual Review of Condensed Matter Physics 2, 329 (2011).
[17] C. Van den Broeck and M. Esposito, *Physica A* 418, 6 (2015).
[18] Y. Murashita, K. Funo, and M. Ueda, Phys. Rev. E 90, 042110 (2014).
[19] M. Planck, S.-B. Preuß. Akad. Wiss. , 49 (1925).
[20] E. Schrödinger, *Statistical Thermodynamics* (Cambridge University Press, 1946).
[21] R. P. Feynman, *Statistical Mechanics: A Set of Lectures* (Westview Press, 1972).
[22] D. Zubarev, *Nonequilibrium Statistical Thermodynamics* (Plenum Pub. Corp., 1974).
[23] L. D. Landau and E. M. Lifshitz, *Statistical Physics* (Butterworth-Heinemann, 1975).
[24] M. Toda, R. Kubo, and N. Saito, *Statistical Physics I: Equilibrium Statistical Mechanics* (Springer, 1978).
[25] H. B. Callen, *Thermodynamics and an Introduction to Thermostatistics* (John Wiley & Sons, 1985).
[26] D. Kondepudi and I. Prigogine, *Modern Thermodynamics: From Heat Engines to Dissipative Structures* (John Wiley & Sons, 1998).
[27] D. Frenkel, Mol. Phys. 112, 2325 (2014).
[28] M. E. Cates and V. N. Manoharan, Soft Matter 11, 6538 (2015).
[29] Y. Ashida, K. Funo, Y. Murashita, and M. Ueda, Phys. Rev. E 90, 052125 (2014).