entropy

MDPI

Article

The "Real" Gibbs Paradox and a Composition-Based Resolution

Fabien Paillusson

School of Mathematics and Physics, University of Lincoln, Brayford Pool, Lincoln LN6 7TS, UK; fpaillusson@lincoln.ac.uk

Abstract: There is no documented evidence to suggest that J. W. Gibbs did not recognize the indistinguishable nature of states involving the permutation of identical particles or that he did not know how to justify on a priori grounds that the mixing entropy of two identical substances must be zero. However, there is documented evidence to suggest that Gibbs was puzzled by one of his theoretical findings, namely that the entropy change per particle would amount to $k_B \ln 2$ when equal amounts of any two different substances are mixed, no matter how similar these substances may be, and would drop straight to zero as soon as they become exactly identical. The present paper is concerned with this latter version of the Gibbs paradox and, to this end, develops a theory characterising real finite-size mixtures as realisations sampled from a probability distribution over a measurable attribute of the constituents of the substances. In this view, two substances are identical, relative to this measurable attribute, if they have the same underlying probability distribution. This implies that two identical mixtures do not need to have identical finite-size realisations of their compositions. By averaging over composition realisations, it is found that (1) fixed composition mixtures behave as homogeneous single-component substances and (2) in the limit of a large system size, the entropy of mixing per particle shows a continuous variation from $k_B \ln 2$ to 0, as two different substances are made more similar, thereby resolving the "real" Gibbs paradox.

Keywords: Gibbs paradox; mixtures; entropy

check for updates

Citation: Paillusson, F. The "Real" Gibbs Paradox and a Composition-Based Resolution. *Entropy* **2023**, *25*, 833. https://doi.org/10.3390/e25060833

Academic Editor: Dennis Dieks

Received: 4 April 2023
Revised: 12 May 2023
Accepted: 21 May 2023
Published: 23 May 2023

Copyright: © 2023 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

In most modern texts, the Gibbs paradox is referred to as the inability to ground the extensivity of thermodynamic potentials, such as the entropy or the Helmholtz free energy, within the framework of classical statistical mechanics developed by J. W. Gibbs in [1]. In his famous 1992 discussion of the Gibbs paradox, E. T. Jaynes summarises the situation well [2]:

> "For 60 years, textbooks and teachers (including, regrettably, the present writer) have impressed upon students how remarkable it was that Gibbs, already in 1902, had been able to hit upon this paradox which foretold—and had its resolution only in—quantum theory with its lore about indistinguishable particles, Bose and Fermi statistics, etc."

Jaynes, however, contends that Gibbs committed a mathematical mistake in his latest writings, which is ultimately responsible for the whole confusion around extensivity in statistical mechanics:

> "In particular, Gibbs failed to point out that an "integration constant" was not an arbitrary constant, but an arbitrary function. However, this has, as we shall see, nontrivial physical consequences. What is remarkable is not that Gibbs should have failed to stress a fine mathematical point in almost the last words he wrote; but that for 80 years thereafter all textbook writers (except possibly Pauli) failed to see it."

In the above passage, the mentioned integration constant would arise from a result derived by Gibbs in Chapter IV of his book on statistical mechanics. This point will be important for what we are going to discuss in a few paragraphs.

Whilst Jaynes' intention may have principally been to celebrate Gibbs' work and push against the necessity of a quantum foundation for classical statistical mechanics, these two quotes ironically end up further cementing a narrative that attributes a fault to Gibbs, which would have inadvertently led astray the physics and thermodynamics community for at least 80 years.

It is important to point out that such a narrative stems from a less-than-charitable, or even perhaps dishonest, reading of Gibbs' writing in his final book on statistical mechanics [1]. Jaynes even ventured to suggest that Gibbs was *older* and in *bad health*, which explains why his latest piece of work was lacking compared to previous ones. Interestingly, even biographies written just a few years after Gibbs' death, such as [3], do not appear to mention any purported intellectual decline or poor health in the later years of his life. However, one just needs to read the preface of *Elements of Statistical Mechanics* [1] to see that it was never Gibbs' intention to consider systems with varying particle numbers; thus, any analysis of extensivity for the results presented in the preceding chapters is automatically excluded until Chapter XV, which is the final chapter of his book:

> "Finally, in Chapter XV, we consider the modification of the preceding results which is necessary when we consider systems composed of a number of entirely similar particles of several kinds, all of which kind being entirely similar to each other, and when one of the numbers of variations to be considered is that of the numbers of the particles of the various kinds which are contained in the system."

At this point, the reader may be reminded that the alleged "integration constant" mistake (conflating a function of the particle number with a constant) that Gibbs made, according to Jaynes, is in Chapter IV.

A full discussion of how both the aforementioned extensivity issue and the purported mistakes attributed to Gibbs are actually not supported by any evidence in Gibbs' writings has been provided elsewhere [4]. In what follows, we shall assume that the few quotes provided above are sufficient to cast some suspicion about the historical, historiographical, and conceptual veracity of the main interpretation of Gibbs' paradox found in most textbooks. We shall now safely depart from this view and delve into what we term as the "real" Gibbs' paradox.

In his 1876 work *on the equilibrium of heterogeneous substances* [5], Gibbs found a somewhat intriguing result grounded in thermodynamics (but easily reproducible in statistical mechanics by later authors), where the change in entropy would amount to $k_B \ln 2$ per unit of mass/matter upon mixing equal amounts of any two different non-reactive substances. The surprising fact was that this result held no matter how similar these substances would be. Here, again, we may directly quote Gibbs [5]:

> "The fact is not less significant that the increase of entropy due to the mixture of gases of different kinds ... is independent of the nature of the gases ... and of the degree of similarity between them."

If the substances were considered to be exactly identical, however, one would find the entropy change to be exactly zero.

In [6], it is reported that Pierre Duhem was likely the first to mention the paradoxical nature of this result and highlight the "absurd consequences" that violated the continuity principle. To our knowledge, very few textbooks on statistical mechanics actually refer to the latter when discussing the Gibbs paradox. The only two such texts we are aware of are [7,8].

Van Kampen proposed a discussion of this version of the Paradox in [9]. In this discussion, he proposes the following reasoning to dispel the notion that there is any paradox at all

> "But suppose that [substances] A and B are so similar that the experimenter has no physical way of distinguishing between them. Then he does not have the semi-permeable walls needed for the second process, but on the other hand the first one will look perfectly reversible to him."

This leads to the conclusion that there should not be any change in entropy as long as the experimenter does not have any physical mean (or is not interested in using any such means) to separate two substances. Indeed, in the above quote, the second process mentioned refers to obtaining the final entropy by determining the work performed by a semi-permeable membrane letting A pass through and not B, for example.

This statement by van Kampen on the operational interpretation of the entropy of mixing was recently demonstrated experimentally with colloids in [10], but had already been mentioned by Gibbs one hundred years earlier in [5]:

> "when we say that two gas-masses of the same kind are mixed under similar circumstances there is no change of energy or entropy, we do not mean that the gases that have been mixed can be separated without change to external bodies. On the contrary, the separation of the gases is entirely impossible … because we do not recognise any difference in the substance of the two masses",

and was also likely known by Duhem as well.

The author remains uncertain that van Kampen's argument dispels the paradox pointed out by Gibbs and Duhem for all possible imaginable substances. For example, in his explanation, van Kampen assumes that either one can certainly discriminate between substance A and substance B, or one is incapable of discriminating them at all. However, what if, for example, substances A and B actually comprise the same molecules, but not in the same proportions? Surely a gas containing 80% dioxygen and 20% nitrogen is not "identical" to a gas containing 20% dioxygen and 80% nitrogen. How do we separate these two substances then, and what is the corresponding work needed to do so? As proposed in [4], it is possible to conceive of a generalised semi-permeable membrane that only has a given probability (less than or equal to unity) to bring one of the molecules of a given type from one side of the box to the other side, to separate substances A and B. A similar approach was used to experimentally test the Landauer bound in [11].

These new developments allow one to imagine *degrees of similarity* between substances rather than a binary identical/different view of substance similarity.

A further issue that arises when thinking about the mixing of real substances is well illustrated by a thought experiment discussed in [12]: If two glasses of milk are poured from the same bottle, they will correspond to two different finite realisations of the same underlying substance (i.e., the milk coming from the bottle). Consequently, the two glasses are unlikely to contain the very same amounts of, say, fat globules or casein proteins, despite the fact that they are sampled from the same underlying substance. If one were to mix these two glasses into a larger container, would one expect a non-zero mixing entropy from a statistical mechanics or thermodynamics perspective? More importantly, for the purpose of the present paper, how can the experiment be reproduced in the first place?

Before moving on to what the present article seeks to address, it will be instructive to look first at a common approach that has been extensively used in the past two decades to describe mixtures, discrete or otherwise, either in the context of the Gibbs paradox (e.g., in s [4,12,13]) or in the context of multimer assemblies [14] and phase equilibria of polydisperse systems (e.g., [15–19]). One way to conceive this approach consists in considering a mixture comprising a fixed integral number N_i of particles of species i among, say, m species. If the system has a total of N particles, then we must have that $N = \sum_{i=1}^{m} N_i$. For an ideal gas model confined to a volume V, the corresponding free energy in the canonical ensemble is easily found to be in the large N_i limit (see, e.g., [12])

$$\beta F(N, \{N_i\}, \beta) \approx -N \ln \frac{V}{\Lambda^3} - N + \sum_{i=1}^{m} N_i \ln N_i, \tag{1}$$

where, for illustration purposes, we considered a gas of particles with no relevant internal degrees of freedom and possessing identical masses (thus distinguished by means other than their mass) and the thermal length scale Λ. The approach then goes on to make the following prescription:

$$N_i = Np(i), \tag{2}$$

where $p(i)$ is interpreted as being the probability of having a particle of species i in the system. Substituting the above prescription into Equation (1) gives

$$F(N, \{N_i\}, \beta) \approx Nk_BT\ln(n\Lambda^3) - N - Nk_BTs(p), \tag{3}$$

where $n = N/V$ is the particle number density and

$$s(p) = -\sum_{i=1}^{m} p_i \ln p_i. \tag{4}$$

At this stage, it is worth pointing out that some disagreement exists in the naming convention used to qualify the quantity $k_Bs(p)$ appearing in Equation (3) for the free energy of a single substance. For example, in [12], $k_Bs(p)$ is called the *mixing entropy*, while in [13], $k_Bs(p)$ is considered the characteristic of the composition of a substance (no matter how this substance is put together), as described by the probability $p(i)$, and is, therefore, called *composition entropy*. Given that the present work is a continuation of the work proposed in [13], we shall adopt the latter naming convention in what follows. On that latter view, refs. [4,13] have defined the mixing entropy as the change in the composition entropy between a system comprising N_A and N_B particles of two initially separated substances, A and B, and that of a substance C resulting from putting A and B together in the same volume. If we denote p_A, p_B, and $p_C = \frac{N_A}{N}p_A + \frac{N_B}{N}p_B$ as the composition probabilities of substances A, B, and C, respectively, then the (dimensionless) entropy change is found to be [13]

$$\Delta S/k_B = N_A D_{KL}(p_A|p_C) + N_B D_{KL}(p_B|p_C), \tag{5}$$

where

$$KL(p|q) \equiv -\sum_{i=1}^{m} p(i)\ln\frac{q(i)}{p(i)} \tag{6}$$

is the Kullback–Leibler divergence of the probability p from probability q. In the context of strongly asymmetric mixing, e.g., $N_A \ll N_B$, it was shown in [13] that Equation (5) reduces to a prescription by Sollich et al. for determining the phase equilibria of polydisperse systems, for instance in [18,19], where a dominant (parent) composition starts coexisting with a minority (incipient) composition. The traditional Gibbs mixing thought experiment is, however, more in line with a fully symmetric case, where $N_A = N_B = N/2$. In this case, Equation (5) reduces to [4,13]

$$\Delta S/k_B = ND_{JS}(p_A|p_B), \tag{7}$$

where

$$D_{JS}(p_A|p_B) \equiv \frac{1}{2}\sum_{i=1}^{m}\left(p_A(i)\ln\left(\frac{2p_A(i)}{p_A(i)+p_B(i)}\right) + p_B(i)\ln\left(\frac{2p_B(i)}{p_A(i)+p_B(i)}\right)\right), \tag{8}$$

is the Jensen–Shannon divergence between p_A and p_B, which corresponds to a square metric [20] between the probability distributions p_A and p_B.

The problem with the above approach is that it entirely relies on the prescription that $N_i = Np(i)$. However, upon inspection, this prescription does not appear justified for three reasons:

- Firstly, from a mathematical standpoint, if a mixture is characterised by a probability distribution $p(i)$, then each time a particle of that substance is added to the system, the species i is selected with probability $p(i)$. In a simpler case, where only two species are possible, i.e., $m = 2$, the problem becomes analogous to the flipping of N coins. In that case, the number of particles of a given species shall follow a binomial distribution, as studied in [21], and $Np(i)$ is the (binomial) expectation value of N_i. In practice,

however, a single realisation of N coin flips is not expected to give exactly $N_i = Np(i)$. Thus, the proposed prescription, albeit heuristically intuitive, conflates an instantiated value taken by a random variable with its expectation value; this is akin to a sort of 'mean field' approximation, especially given that N_i ends up in various logarithmic functions in statistical thermodynamics.

- Secondly, from a conceptual standpoint, there is a problem with the substitution of N_i by $Np(i)$; the latter is often not an integer. Gibbs' statistical mechanics, or its quantum extension, establishes a relationship between the dimensionality of state spaces to be explored and the number of particles in the system. These spaces possess an integral number of dimensions, not a fractional one.
- Finally, the prescription is often taken *after* the Stirling approximation has been used, which must assume that *each* N_i is large; this cannot be guaranteed for all system sizes for all composition probabilities.

Despite the above critique of the $N_i = Np(i)$ prescription, Equation (7) ultimately appears to make sense from a physical standpoint, as discussed in [4,13], and it can be obtained from a substantial body of existing works [12,15,19]. It is just that (a) the prescription is not mathematically consistent with the probability theory and (b) it is not conceptually satisfactory. More particularly, there is no notion of what happens if N cannot be assumed to be infinite. The aim of the present paper is, therefore, to extend the work on binary mixtures carried out in [21] to generalised mixtures, so as to more firmly ground both mathematically and conceptually the validity of Equation (7), and propose mixing entropy expressions for finite N.

In what follows, we propose a new formalism that enables one to address the above questions. More specifically, we consider that substances are ultimately defined by composition probabilities and that one can only ever mix two finite composition realisations of different substances. Given that the entropy change will depend on the specific realisations one is looking at, the reproducibility requirement of thermodynamics compels us to seek realisation-independent mixing entropy expressions. We obtain the latter by averaging over composition realisations in both substances to be mixed. We believe that this approach follows the operational view promoted by Gibbs and van Kampen, but when the two substances cannot be separated with certainty and have some overlap in their composition.

2. Materials and Methods

The method we introduce in this paper expands upon a preliminary work by the present author on binary mixtures [21]. This generalisation can be formulated in the following presuppositions:

1- Discrete mixtures comprising m possible *identifiable species*, labeled from 1 to m, are characterised—in a definitional sense—by an *ideal composition* corresponding to a set of fixed probabilities $\{p(i)\}_{i=1,\ldots,m}$, satisfying

$$\sum_{i=1}^{m} p(i) = 1. \tag{9}$$

2- Any real mixture with a finite number N of particles is but one of many *realisations* obtained from *independently* sampling each N particle identity from the sample space $\{1,\ldots,m\}$, with the corresponding probabilities $\{p(i)\}_{i=1,\ldots,m}$.

3- Let $\mathbf{N}_i$ denote the random variable representing the number of particles of species i in a given mixture and let $\mathbf{C} = \{\mathbf{N}_1,\ldots,\mathbf{N}_m\}$ represent the multivariate random variable characterising the *empirical composition*. If the mixture comprises N particles. We must have that

$$\sum_{i=1}^{m} \mathbf{N}_i = N. \tag{10}$$

4- We denote $\mathcal{C} \equiv \{N_1, \ldots, N_m\}$ the *composition realisation* of a given mixture, such that for any species index i, we have $\mathbf{N}_i = N_i$. Given that the indicator random variables for the N particles are considered independent and identically distributed with $\{p(i)\}_{i=1,\ldots,m}$, it follows that the probability distribution for $\mathcal{C}$ satisfies a multinomial distribution:

$$\mathcal{M}_{N,m,p}(\mathcal{C}) = \frac{N!}{\prod_{i=1}^{m} N_i!} \prod_{j=1}^{m} p(j)^{n_j}, \tag{11}$$

where it is understood that N_i must comply with Equation (10) as well.

5- The *Helmholtz free energy* $F(N, \beta = \frac{1}{k_B T}, \mathbf{C})$ of an N-particle system is also a random variable via the composition random variable $\mathbf{C}$. We seek the realisation-independent Helmholtz free energy of the system $\mathcal{F}(N, \beta)$ by averaging over all possible composition realisations:

$$\mathcal{F}(N, \beta) \equiv \langle\!\langle F(N, \beta, \mathbf{C}) \rangle\!\rangle \equiv \sum_{\mathcal{C}} \mathcal{M}_{N,m,p}(\mathcal{C}) F(N, \beta, \mathcal{C}). \tag{12}$$

To address Gibbs' paradox with the above formalism, we also need to specify an ensemble and a model to apply it to. We choose to work in the Gibbs canonical ensemble, i.e., fixed temperature and a fixed amount of matter. Given that the paradox arises already in the case of ideal gases, the very first step is, therefore, to discuss it in the context of N non-interacting, non-relativistic, and independent particles confined in a box of volume V. This is what shall be done in the rest of this paper.

For a system of N particles comprising N_i particles of species i of mass M_i corresponding to a given composition realisation $\mathcal{C}$, the Gibbs canonical partition function reads

$$Z(N, \beta, \mathcal{C}) = V^N \prod_{i=1}^{m} \left(\frac{z_i^{N_i}}{N_i! \Lambda_i^{3N_i}} \right), \tag{13}$$

where z_i is the internal dimensionless canonical partition function of a particle of type i, $\Lambda_i = \xi / \sqrt{2\pi M_i k_B T}$ is a characteristic length scale associated with the species i, and ξ is the area—preserved by the Hamiltonian flow—of an elementary two-dimensional face of a polygonal cell of a partition of the phase space. Of course, it is common in modern texts to identify Λ_i to the thermal de Broglie wavelength by setting $\xi = h$, the Planck constant, but this is not a necessity.

From Equation (11), we have that

$$\frac{1}{\prod_{i=1}^{m} N_i!} = \frac{\mathcal{M}_{N,m,p}(\mathcal{C})}{N! \prod_{j=1}^{m} p(j)^{N_j}}, \tag{14}$$

we find

$$\beta \mathcal{F}(N, \beta) = -N ln V + N \ln(\tilde{\Lambda}^3 / \tilde{z}) + \ln N! - N s(p) + H(\mathcal{M}_{N,m,p} | \mathcal{M}_{N,m,p}), \tag{15}$$

where $\tilde{\Lambda} = e^{\sum_{i=1}^{m} p(i) \ln \Lambda_i}$ is a single effective characteristic length scale for the substance, $\tilde{z} = e^{\sum_{i=1}^{m} p(i) \ln z_i}$ is an effective internal partition function for the substance, $s(p) \equiv -\sum_{i=1}^{m} p(i) \ln p(i)$ is the composition entropy, $H(\mathcal{M}_{N,m,p} | \mathcal{M}_{N,m,p}) \equiv -\sum_{\mathcal{C}} \mathcal{M}_{N,m,p}(\mathcal{C}) \ln \mathcal{M}_{N,m,p}(\mathcal{C})$ is the *composition realisation entropy*, and we use the fact that $\langle\!\langle \mathbf{N}_i \rangle\!\rangle = \sum_{\mathcal{C}} P(\mathcal{C}) N_i = N p(i)$ for the multinomial distribution (11).

One of the major difficulties in regard to using Equation (15) for analytical results for finite-sized systems is that the term $H(\mathcal{M}_{N,m,p} | \mathcal{M}_{N,m,p})$ corresponds to the entropy of the multinomial distribution and there is no known closed-form expression for it. We are, therefore, bound to explore different limiting regimes in the cases of different kinds of composition models, and propose conclusions on a case-by-case basis.

3. Results

Equation (15) is exact for any finite system consisting of N particles and characterised by a composition probability map p. This will serve as our starting point to elucidate the thermodynamic behaviours of general mixtures when they are considered on their own and also upon mixing any two mixtures. In particular, we shall also try to explore cases of so-called *continuous polydisperse* systems. In what follows, we shall use as a safeguarding strategy the principle that, in the large N limit, the realisation-independent free energy of mixture composition models should become extensive. Composition models that do not satisfy this requirement shall be considered provisionally inappropriate for the study of the thermodynamics of substances and would warrant further inspection outside of the scope of the present paper.

3.1. Case of a Single Mixture

As mentioned before, the potential utility of the proposed formalism is going to depend on one's ability to evaluate the entropy of the multinomial distribution $H(P(\mathcal{C})|\mathcal{M}_{N,m,p}(\mathcal{C}))$. In what follows, we will use the following convenient rewriting of this entropy:

$$H(\mathcal{M}_{N,m,p}|\mathcal{M}_{N,m,p}) = -\ln N! + Ns(p) + \sum_{i=1}^{m}\sum_{N_i=0}^{N} B_{N,p}(\mathbf{N}_i = N_i)\ln(N_i!), \tag{16}$$

where $B_{N,p}(\mathbf{N}_i = N_i)$ is the binomial probability distribution for finding $\mathbf{N}_i = N_i$ particles of type i in the composition realisation given the specified N and $p(i)$

$$B_{N,p}(\mathbf{N}_i = N_i) \equiv \frac{N!}{(N-N_i)!N_i!}p(i)^{N_i}(1-p(i))^{N-N_i}. \tag{17}$$

3.1.1. Finite Discrete Mixtures: m and Map p Are Fixed and $N > m$

The case of finite discrete mixtures corresponds to the "typical" situation that most thermodynamicists might have in mind for a mixture, i.e., there is a clearly defined set of m different chemical species and various probabilities assigned to each species, and N is often much larger than m. If N is large enough, such that $Np(i) \gtrapprox 10$ for any $i = 1, \ldots, m$, then one may use the Stirling approximation for $N_i!$ and take the normal distribution limit of $B_{N,p}(\mathbf{N}_i = N_i)$ giving [22] (see also Appendix A for numerical comparisons)

$$H(\mathcal{M}_{N,m,p}|\mathcal{M}_{N,m,p}) = \frac{m-1}{2}\ln(2\pi Ne) + \frac{1}{2}\sum_{i=1}^{m}\ln p(i) + \mathcal{O}\left(\frac{1}{N}\right). \tag{18}$$

One notes that the first term in Equation (18) is of order ~$\ln N$, the second term is, at most, of order ~$\ln m$, and the remainder of order ~$1/N$. This means that, in the large N limit, using the Stirling approximation, the realisation-free Helmholtz free energy per particle for finite discrete mixtures becomes

$$\frac{\mathcal{F}(N,T)}{N} \simeq k_B T(\ln(n_v\tilde{\Lambda}^3/\tilde{z}) - s(p) - 1), \tag{19}$$

where we introduce the particle density $n_v \equiv N/V$. Three important remarks follow from Equations (18) and (19):

- At fixed composition, i.e., at fixed m and probability map p on $\{1, \ldots, m\}$, $s(p)$ is a constant, and the mixture can be essentially conceived as a homogeneous substance. All thermodynamic properties of the system will be identical to that of a single component system with only the total substance particle density n_v (as opposed to partial densities) playing a role.
- In the large system size limit, the realisation-independent free energy becomes proportional to the total amount of matter N in the system, i.e., becomes extensive.

Consequently, given the safeguarding criterion we chose, the finite discrete mixture model described above constitutes a valid composition model of general mixtures.

- By applying the defining relation for the thermodynamic entropy $\mathcal{S}(N,T) \equiv -\frac{\partial \mathcal{F}}{\partial T}$ to Equation (19), we obtain

$$\mathcal{S}(N,T) \simeq Nk_B \ln V - \frac{3}{2} Nk_B \ln T + Nk_B T \frac{\partial \ln \tilde{z}}{\partial T} + K(p), \tag{20}$$

where $K(p)$ is a function that depends only on the composition characterised by the map p. This result for the entropy expression of a substance was already anticipated by Gibbs in 1876 in [5] in the case where $\tilde{z} = 1$ (and with slightly different notations).

> "[Equation (20)] applies to all gases of constant composition for which the matter is entirely determined by a single variable [N]"

We should note that Equation (19) has already been obtained by the author in [4], but within the heuristic theoretical framework described in the introduction section, whereby whole particle numbers for each species were directly replaced by their composition averages into the expression of the canonical partition function. This amounts to a sort of "mean-field" approximation, which is somewhat unjustified from a mathematical standpoint and has the conceptual disadvantage of invoking fractional particle numbers, which is hardly satisfactory.

On the contrary, the formalism proposed in Equations (9)–(15) leading to Equation (19) is free from these shortcomings.

3.1.2. Infinite Discrete Mixtures: $m >> N$ and $N^2 p(i) \ll 1$

It is tempting to imagine that so-called continuous mixtures or continuous polydisperse mixtures would correspond to the composition model we term as *infinite discrete mixtures*, where $m \gg N$, no matter the system size, including the thermodynamic limit, and where the values of all the probabilities $p(i)$ are monotonously decreasing functions of m with zero as the asymptotic limit for $N^2 p(i)$. A trivial example of this model would be the uniform probability model $p(i) = 1/m$ when $m \gg N^2 \gg 1$.

For infinite discrete mixture models, obtaining a particle of a given species becomes incredibly rare and the number N of particles in the system is too small compared to m for any single realisation of the composition $\mathcal{C} = \{N_1, \ldots, N_m\}$ to provide any faithful representation of the composition probabilities $p(i)$. Consequently, the third term of Equation (16) is dominated by the value $\mathbf{N}_i = 2$ i.e.,

$$\sum_{i=1}^{m} \sum_{N_i=0}^{N} B_{N,p}(\mathbf{N}_i = N_i) \ln N_i! \simeq \frac{N(N-1)\ln 2}{2} \sum_{i=1}^{m} p(i)^2 \sim \mathcal{O}(N^2 p(i)) \ll 1. \tag{21}$$

It follows that for infinite discrete mixtures, including cases where m tends to infinity at finite N, as well as in thermodynamic limit cases where the m limit is taken first, and only then the limit of infinite N, we obtain the entropy of multinomial distribution (see Appendix A for numerical comparisons)

$$H(\mathcal{M}_{N,m,p} | \mathcal{M}_{N,m,p}) \approx -\ln N! + Ns(p), \tag{22}$$

which, upon being substituted in Equation (15), gives

$$\beta \mathcal{F}(N,\beta) \approx -N \ln V + N \ln(\tilde{\Lambda}^3/\tilde{z}). \tag{23}$$

It is quite straightforward to see that Equation (23) does not provide an extensive form for the realisation-independent free energy $\mathcal{F}$. Therefore, given our extensivity criterion aimed at ensuring consistency with thermodynamics, we shall consider the infinite discrete mixture model described in this sub-section as being inadequate to model a typical thermodynamic situation. In particular, it does not appear to capture the thermodynamics displayed by continuous polydisperse systems, such as colloids.

3.1.3. Finite Continuous Mixtures: $p(i) = \rho(i)\Delta(i)$ and $N > m$

The inadequacy of the infinite continuous mixture model to characterise an expected thermodynamic behaviour of a single mixture stems from adopting a value for m, which is much larger than N. In that case, composition realisations with 0 or 1 particle per species are overwhelmingly more likely than those with at least one species having two particles. This pathological behaviour need not be rooted in the physics of the problem. Rather, it may arise from an inadequate categorisation of the different species in the system relative to the available number of particles N. In this situation, one may need to resort to a coarse-graining procedure, effectively grouping together different species into a much smaller number of effective particle species. With this in mind, we introduce *finite continuous mixture* composition models for which the probabilities $p(i)$ can be expressed as $p(i) = \rho(i)\Delta(i)$ for any $i = 1, \ldots, m$, where ρ is the probability density of a continuous random variable X and where $\Delta(i)$ is an interval size of X associated to the species index value i. This is, in fact, a special case of finite discrete mixtures, where we require the intermediate probability density ρ to be a smooth function and to not wildly change as m is increased. In fact, in this model, we expect the experimental ρ to converge to some theoretical one as m is increased.

There are two distinct, but non-mutually exclusive, situations that would give rise to such models:

- Mathematical limit: The model for map $p(i)$ may depend on m in a manner such that, as m becomes large "enough", $p(i)$ is very well-approximated by the distribution function of a continuous variable and is identical to it in the infinite m limit. Let us investigate in more detail the typical example of a binomial model for map $p(i)$ with parameters m and w, i.e.,

$$p(i) = \frac{m!}{i!(m-i)!} w^i (1-w)^{m-i}, \tag{24}$$

  where w is a new parameter characteristic of the composition model. Within this model, the most probable type of particle is that with the integer value, the closest to mw. Any particle index $i \neq mw$ will have a lower probability that decays rapidly to approach very small values as m becomes large enough. Indeed, in the large m limit at a fixed w, we have that the binomial distribution tends to a normal distribution (see Appendix B for a numerical comparison)

$$p(i) \simeq \frac{e^{-\frac{(i-mw)^2}{2mw(1-w)}}}{\sqrt{2\pi mw(1-w)}}. \tag{25}$$

  Consider now the map between the species label i taking values in $\{1, \ldots, m\}$ and a new variable $r \equiv i/m$ taking values in the finite interval of the rationals $\{1/m, \ldots, 1\}$. From Equation (25) and the bijective character between i and r, it follows that

$$\text{Prob}(r) \simeq \frac{e^{-\frac{(r-w)^2}{2w(1-w)/m}}}{\sqrt{2\pi w(1-w)/m}} \frac{1}{m}. \tag{26}$$

  In expression (26), r must be limited to take values in the set $\{1/m, \ldots, 1\}$. However, we really do have that $\text{Prob}(r) = \rho(r)\Delta(r)$, where $\rho(x)$ is the probability density function of a normally distributed *continuous* random variable, and where we can identify $\Delta(r) = 1/m$. (cf. Figure A2 in Appendix B).
- Experimental considerations: Whether the composition is characterised by the preparation protocol or by a measurement technique, any experimental process is accompanied by a corresponding finite precision. Thus, in practice, even if one were to use a particle attribute that takes continuous values and is associated with a theoretical underlying ρ to characterise the composition of a mixture, it is bound to be expressed

in terms of a discrete set $\{\Delta(1), \ldots, \Delta(m)\}$ of a finite number m of continuous intervals effectively corresponding to the m identifiable 'species' of the system with the provided resolution and particle number. If m were to be increased, this would just improve the "granularity" of these intervals without making the set actually continuous. If this granularity is fine enough, one may *define* an empirical probability density $\rho_{\text{emp}}(i) \equiv p(i)/\Delta(i)$ and fit it to a corresponding continuous probability density model. Convergence to a stable continuous model is expected as m (i.e., the resolution) is increased. This experimental resolution aspect can also be justified under the assumption that the excess free energy of the mixture solely depends on a subset of moments of the polydisperse distribution [19], where these moments are related to the precision of the measurement.

Regardless of whether the use of a continuous random variable and its associated probability density is justified by a mathematical limit of the composition model or inevitable experimental precision considerations, or both, ultimately, these preconditions allow one to write $p(i) = \rho(i)\Delta(i)$, where ρ is the probability density of a continuous random variable and for which $N > m$.

Given that $N > m$, we can use the approximate expression for the entropy of the multinomial distribution from Equation (18) and apply it to finite continuous mixtures. Substituting it into Equation (15), we obtain the following result in the limit of large N:

$$\frac{\mathcal{F}(N,T)}{N} \simeq k_B T(\ln(n_v \tilde{\Lambda}^3/\tilde{z}) - s_\Delta(\rho) - 1), \tag{27}$$

where we use the Taylor–McLaurin formula to the first order to replace the discrete sum for the composition entropy $s(p)$ by $s_\Delta(\rho) \equiv -\int_{-\infty}^{+\infty} \rho(x)\ln(\rho(x)\Delta(x))dx$.

We note from Equation (27) that $\mathcal{F}$ is extensive and, thus, finite continuous mixtures serve as valid composition models for probing the thermodynamic behaviors of general mixtures in finite and large N system sizes. This is not too surprising given that, as stated above, what we call finite continuous models are special cases of finite discrete models with some additional regularity and smoothness requirements. Therefore, Equation (20) would hold as well for such mixture models.

3.2. Mixing of Two Substances

We will now utilize the results obtained in the previous section to calculate the entropy change that occurs when two finite discrete mixtures, substances A and B, are mixed. We denote m the number of distinct possible chemical species in substances A and B so that all particle types of substances A and B have a label i, taking values in the same sample space $\{1, \ldots, m\}$. We denote $p_A(i)$ (resp. $p_B(i)$) as the probability for substance A (resp. B) to have a particle of type i.

As is common in the literature, we consider a mixing scenario comprising two different equilibrium situations:

- Firstly, let us consider a situation where a box of volume V is separated into two equally sized compartments by a removable wall with $N/2$ particles of substance A in, say, the left-hand side compartment, and $N/2$ particles of substance B in the right-hand-side compartment.
- Secondly, the wall separating the substances is removed so as to let them intermix until equilibrium is reached.

This scenario is the usual one within which the Gibbs paradox of mixing tends to be discussed. The quantitative aspects may be affected if one is not mixing an equal amount of substances A and B, or further contributions to the entropy changes can be expected if the densities in the two compartments are not initially the same.

3.2.1. Free Energy before Removing the Wall

Before the wall is removed, each substance is separated into its own compartment and at equilibrium at the same temperature. Introducing the composition of multivariate

random variables $\mathbf{C}_A$ and $\mathbf{C}_B$, the free energy in each compartment is also a random variable, which is a function of either $\mathbf{C}_A$ or $\mathbf{C}_B$, which, upon averaging over composition realisations in each compartment, gives

$$\beta\mathcal{F}_{A/B}\left(\frac{N}{2},\beta\right) = \frac{N}{2}\left(\ln\left(\frac{\tilde{\Lambda}^3_{A/B}}{\tilde{z}_{A/B}}\right) - \ln\frac{V}{2} - s(p_{A/B})\right) + \ln\left(\frac{N}{2}!\right) + H(\mathcal{M}_{\frac{N}{2},m,p_{A/B}}|\mathcal{M}_{\frac{N}{2},m,p_{A/B}}), \tag{28}$$

where we have simply reproduced Equation (15) with indices A or B to specify whether the substance being looked at is characterised by p_A or p_B.

In the end, from the additivity of the free energy, we obtain that

$$\mathcal{F}_{\text{unmix}}(N,\beta) \equiv \mathcal{F}_A\left(\frac{N}{2},\beta\right) + \mathcal{F}_B\left(\frac{N}{2},\beta\right). \tag{29}$$

3.2.2. Free Energy after Having Removed the Wall

Once the separating wall has been removed, substances A and B will mix and eventually reach equilibrium. After mixing, the number of particles of type i is going to be associated with the random variable $\mathbf{N}_i^{\text{mix}} = \mathbf{N}_i^A + \mathbf{N}_i^B$, where $\mathbf{N}_i^{A/B}$ is the random variable corresponding to the number of particles of type i in substance A (resp. B). Given that each substance composition follows a multinomial distribution $\mathcal{M}_{\frac{N}{2},m,p_{A/B}}$, it follows that

$$\langle\!\langle \mathbf{N}_i^{\text{mix}} \rangle\!\rangle = \frac{N}{2}(p_A(i) + p_B(i)) \equiv N p_{\text{mix}}(i), \tag{30}$$

where we identify p_{mix} the ideal composition probability distribution of the mixture of A and B. In Equation (30), the averaging over composition realisations is done by using the joint probability measure $\text{Prob}(\mathcal{C}_A,\mathcal{C}_B) = \mathcal{M}_{\frac{N}{2},m,p_A}(\mathcal{C}_A)\mathcal{M}_{\frac{N}{2},m,p_B}(\mathcal{C}_B)$, where the substances are considered completely independent.

The repeated mixing of $N/2$ particles from different composition realisations of substances A and B can be conceived as an actual protocol to sample N particles from the effective composition p_{mix}. The probability of obtaining a given composition $\mathcal{C}_{\text{mix}} = \{N_1^{\text{mix}},\ldots,N_m^{\text{mix}}\}$ then follows the multinomial distribution

$$\mathcal{M}_{N,m,p_{\text{mix}}}(\mathcal{C}_{\text{mix}}) = \frac{N!}{\prod_{i=1}^m N_i^{\text{mix}}!}\prod_{j=1}^m p_{\text{mix}}^{N_i^{\text{mix}}} = \frac{N!}{\prod_{i=1}^m (N_i^A + N_i^B)!}\prod_{j=1}^m p_{\text{mix}}^{N_j^A+N_j^B}, \tag{31}$$

where the last equality makes the connection between N^{mix} and $N^{A/B}$ stemming from the aforementioned mixing protocol used to sample p_{mix}.

For two given composition realisations, $\mathcal{C}_A$ and $\mathcal{C}_B$, each comprising $N/2$ particles, we also have that, after mixing, the canonical partition function reads

$$Z_{\text{mix}}(N,\beta,\mathcal{C}_A,\mathcal{C}_B) = \frac{V^N}{\prod_{i=1}^m (N_i^A + N_i^B)!}\prod_{j=1}^m\left(\frac{z_j}{\Lambda_j^3}\right)^{N_j^A+N_j^B}. \tag{32}$$

Substituting Equation (31) into Equation (32), we have

$$Z_{\text{mix}}(N\beta,\mathcal{C}_A,\mathcal{C}_B) = \frac{V^N \mathcal{M}_{N,m,p_{\text{mix}}}(\mathcal{C}_A+\mathcal{C}_B)}{N!\prod_{i=1}^m p_{\text{mix}}^{N_i^A+N_i^B}}\prod_{j=1}^m\left(\frac{z_j}{\Lambda_j^3}\right)^{N_j^A+N_j^B}, \tag{33}$$

where $\mathcal{C}_A + \mathcal{C}_B = \{N_1^A + N_1^B,\ldots,N_m^A + N_m^B\}$.

In the end, upon averaging over composition realisations $\mathcal{C}_A$ and $\mathcal{C}_B$, we have

$$\beta\mathcal{F}_{\text{mix}}(N,\beta) = -N\ln V + \ln N! - Ns(p_{\text{mix}}) + \frac{N}{2}\ln\left(\frac{\tilde{\Lambda}_A^3\tilde{\Lambda}_B^3}{\tilde{z}_A^3\tilde{z}_B^3}\right) + H(\mathcal{M}_{\frac{N}{2},m,p_A}\mathcal{M}_{\frac{N}{2},m,p_B}|\mathcal{M}_{N,m,p_{\text{mix}}}), \quad (34)$$

where it is understood that the last term involves sums over values taken by N_i^A and N_i^B.

3.2.3. Entropy Change

Given that the energy in the system does not change upon mixing (because the species are non-reactive and non-interacting), we now define the *entropy change upon mixing* as being $\Delta\mathcal{S}_{\text{mix}} \equiv (-\mathcal{F}_{\text{mix}} + \mathcal{F}_{\text{unmix}})/T$, which gives:

$$\Delta\mathcal{S}_{\text{mix}}(N,\beta) = k_B\ln\left(\frac{2^N(N/2!)^2}{N!}\right) + Nk_BD_{JS}(p_A|p_B) + k_B\Delta H, \quad (35)$$

where $\Delta H = -H(\mathcal{M}_{\frac{N}{2},m,p_A}\mathcal{M}_{\frac{N}{2},m,p_B}|\mathcal{M}_{N,m,p_{\text{mix}}}) + \sum_{i=A,B} H(\mathcal{M}_{\frac{N}{2},m,p_i}|\mathcal{M}_{\frac{N}{2},m,p_i})$ and where $D_{JS}(p_A|p_B)$ is the Jensen–Shannon divergence introduced in Equation (8).

Equation (35) constitutes the main result of this paper and gives an *exact* realisation-independent expression of the classical mixing entropy of equal amounts of any two ideal-gas-like mixtures, A and B, for any system size N. A similar expression has already been found for binary mixtures in [21], but its generalisation to any extensive mixture model is, to our knowledge, new. It is worth going through the meaning of each of the terms in the entropy change expression.

- $\ln\left(\frac{2^N(N/2!)^2}{N!}\right)$ corresponds to the *partitioning entropy*, i.e., the entropy gained by releasing initially confined particles into double the initial volume. Note that this term is oblivious to the particle type and is, therefore, always positive, even if the substances are identical.
- $D_{JS}(p_A|p_B)$ measures a specific, bounded, square distance between substances p_A and p_B. We shall propose an additional interpretation for this term a bit later.
- ΔH represents the entropy change owing to composition realisations, i.e., to the fact that, upon repeating experiments, one is bound to have different sampled empirical compositions from the ideal compositions expressed by the probability distributions p_A and p_B.

Unfortunately, there is no closed-form expression for ΔH, which involves the entropy and cross-entropy of multinomial distributions. In practice, each of these terms in ΔH needs to be evaluated numerically for given composition models. For the large N, however, it can be shown (cf. Appendix C) that $\Delta H \sim \mathcal{O}(\ln N)$.

For the large N, we also have that the first term in Equation (35) is actually of the form $\sim \ln\left(\left(\frac{N}{2}^{N/2}\right)^2 2^N N^{-N}\right) + \mathcal{O}(\ln N) \sim \mathcal{O}(\ln N)$. Consequently, for the large N, i.e., for the system sizes initially considered by Gibbs, the only contribution (which happens to be extensive) left for the entropy of mixing is

$$\Delta\mathcal{S}_{\text{mix}}^{\text{Gibbs}} = Nk_BD_{JS}(p_A|p_B). \quad (36)$$

Note that the Jensen–Shannon divergence $D_{JS}(p_A|p_B)$ is positive definite and is bounded from above by $\ln 2$. Figure 1 illustrates the value of the Gibbs entropy of mixing for two substances, A and B, corresponding to a finite discrete model of the kind described in Equation (24). Contrary to the sentence by Gibbs quoted in Section 1, we see that the Gibbs mixing entropy varies *continuously* from $\ln 2$ to 0 as the substance compositions become more similar.

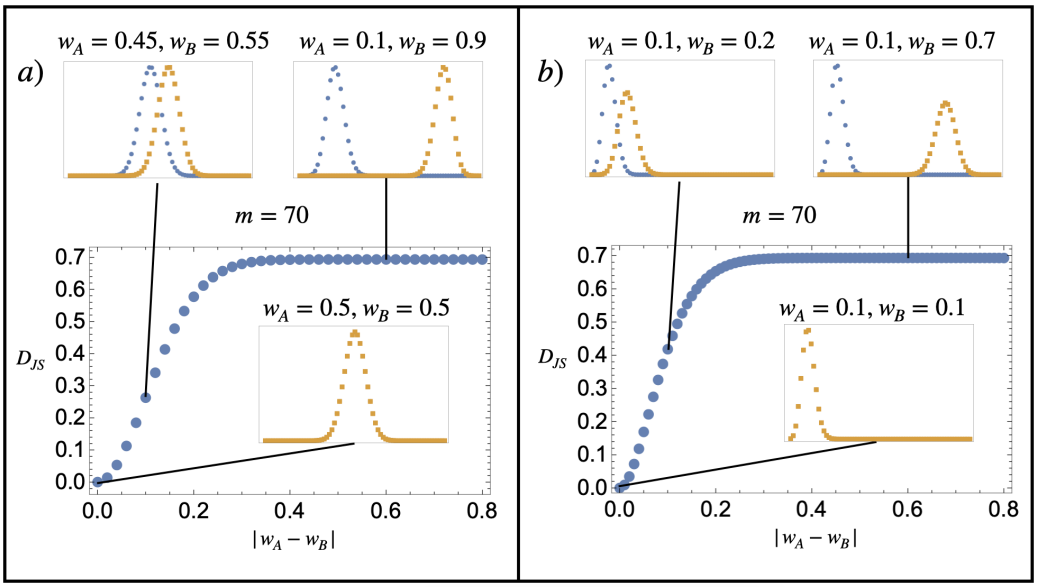


Figure 1. *Dimensionless Gibbs mixing entropy per particle.* The main plots represent $D_{JS}(p_A, p_B)$ for two probability distributions p_A (blue) and p_B (orange), which follow the composition model of Equation (24), with $m = 70$ and various values of w_A and w_B indicated in the figure. The similarity between the distribution is characterised by the absolute value in the difference between the w parameters $|w_A - w_B|$. When it is zero (the probability graphs are on top of one another), the substances are identical in the sense that they have the same composition probabilities. The figures show that the dimensionless Gibbs mixing entropy per particle varies continuously from 0 to $\ln 2 \approx 0.69$ as the similarity between the compositions p_A and p_B is decreased. The exact details of how it does so depend on how $|w_A - w_B|$ is varied. (**a**) Having both distributions moving closer to $w = 0.5$ in a symmetric fashion and (**b**) fixing $w_A = 0.1$, with w_B varying from 0.1 to 0.9.

4. Discussion

In 1876, Gibbs derived results that are generally undisputed; he showed that the mixing entropy of two substances would amount to $k_B \ln 2$ per particle, irrespective of the degree of similarity between the substances, but would discontinuously drop to 0 as soon as they were *sensibly* exactly identical.

In this article, we have developed an approach to address the thermodynamic behaviour of general mixtures. This approach relies on the characterisation of the composition of a substance by an *a priori* probability distribution p acting on the space of all identifiable species. For a given composition probability p, any real system can just sample from that distribution and provide finite-size realisations of it. Thermodynamic quantities are retrieved from averaging over composition realisations of the Helmholtz free energy, the Massieu potential for fixed temperature, and system size ensembles. We found that a restricted class of composition models—that of finite discrete mixtures—was suitable for describing the extensive behaviour of a given substance. These models also encompass the possibility of characterising the particle attribute with a continuous variable. On the contrary, it was found that attempting to characterise a continuous mixture by setting the number of particle types to be much larger than the number of particles N in the system was inconsistent with thermodynamics. One intuitive explanation of this failure is that such infinite continuous mixture models do not allow for any finite-size empirical realisation of size N to actually be representative of the underlying distribution they are sampled from, to the extent that they all appear to be completely different, and therefore incommensurate with, any of the other realisations. It must be noted that using intervals of a continuous random variable to group many species into a similar effective species is not the only way of dealing with the thermodynamically invalid behaviour of what we call infinite continuous mixture models. The crucial aspect is mostly to reduce the dimensionality of the space within which one characterises the composition. This can also be done via a dimensional

reduction approach that relies on determining a few moments of the underlying probability distribution of a specific attribute. The readers interested in such an approach may look at the work developed in [18,19], for instance.

For the selected class of models, it was found that, at a fixed composition, the thermodynamic behaviour is indistinguishable from that of a homogeneous one-component system, as already anticipated by Gibbs in the 1870s.

Applying this new formalism to a mixing scenario of equal amounts of two different substances, a general exact expression was derived for the classical mixing entropy of ideal substances. It was found that (a) even in cases where the substances were identical (same underlying probability distribution), some contributions to the mixing entropy would be non-zero, namely the partitioning entropy and the composition realisation entropy, and (b) in the large system size limit, this exact expression converges to a universal quantity that we call the Gibbs mixing entropy, which is extensive, and proportional to the square distance between the two ideal probability distributions characterising the substances being mixed. This extensive entropy of mixing was shown to vary continuously from 0 to $k_B \ln 2$ per particle as the degree of dissimilarity was increased between the substances. This allowed us to provide a resolution to the paradoxical violation of the continuity principle pointed out by Duhem in the finding of Gibbs.

There are various open questions left to answer with regard to the view developed in the present paper and which will be left for future works: (1) while it is mostly straightforward to extend the proposed formalism to the semi-classical limit of quantum statistical mechanics of independent particles (essentially replacing discrete sums by integrals), it is much less obvious to the author as to how to implement it for quantum statistical situations, where the fermionic or bosonic characters of the particles matter. By this, we mean that, as discussed in [4], it is unclear how to make the fermion/boson character of elementary and composite particles compatible with both the existing body of works on polydisperse systems and the present one. (2) Many real substances are also (at least partially) reactive substances and, therefore, a given particle may not maintain a specific identity at all times. Some stationary compositions may be obtained but, in some systems with reactive processes occurring over time scales larger than thermodynamic experiments, some ergodicity considerations with regard to the present approach may be warranted.

In conclusion, in this paper, we have shown that what we have referred to as "real" Gibbs paradox could find a resolution within classical statistical mechanics by employing a composition-based description of general substances and examining averages over finite-size composition realisations.

Funding: This research was partially funded by the Leverhulme Trust project grant RPG-2021-039.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The author declares no conflict of interest.

Appendix A

This section focuses on numerical testing to determine the extent to which Equations (18) and (22) can serve as useful approximate expressions for the entropy of the multinomial distribution $H(\mathcal{M}_{N,m,p}|\mathcal{M}_{N,m,p})$. In Figure A1, we test it on a simple uniform composition model, i.e., $p(i) = 1/m$, and propose various cases that depend on the values of m and N.

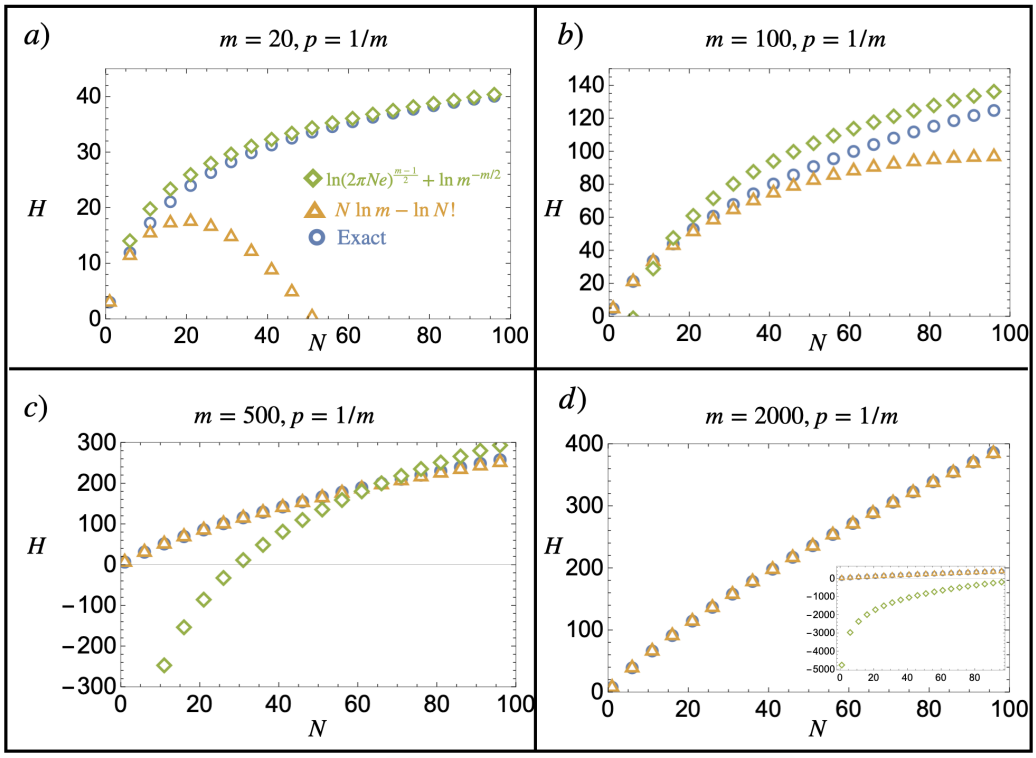


Figure A1. Entropy of the binomial distribution. Numerical comparison of the exact entropy of the binomial distribution (blue open circles) with Equation (18) (green open diamonds) and Equation (22) (orange open triangles), focusing on a uniform composition model, where $p(i) = 1/m$, and exploring various values of m and a function of N. (**a**) $m = 20$ shows good agreement between Equation (18) and the exact entropy from $N \gtrapprox 60 \simeq 3m$. (**b**) For the case of $m = 100$ and $N \leq 100$, neither Equation (18) nor Equation (22) provide a good approximation to the exact entropy. However, both expressions appear to provide reliable upper and lower bounds, respectively. (**c**,**d**) When the value of m is larger than N, Equation (22) is a good approximation for the entropy of the multinomial distribution. In (**c**), the deviation from good fit starts to become apparent when $N \simeq 100 \simeq m/5$.

Appendix B

This section illustrates the notorious normal distribution limit of the binomial distribution in the context of a composition model following Equation (24). In particular, it shows the main idea underlying the mathematical justification of the introduced notion of finite continuous mixtures by comparing Equations (24) and (26). A numerical implementation of different values of m and w is given in Figure A2.

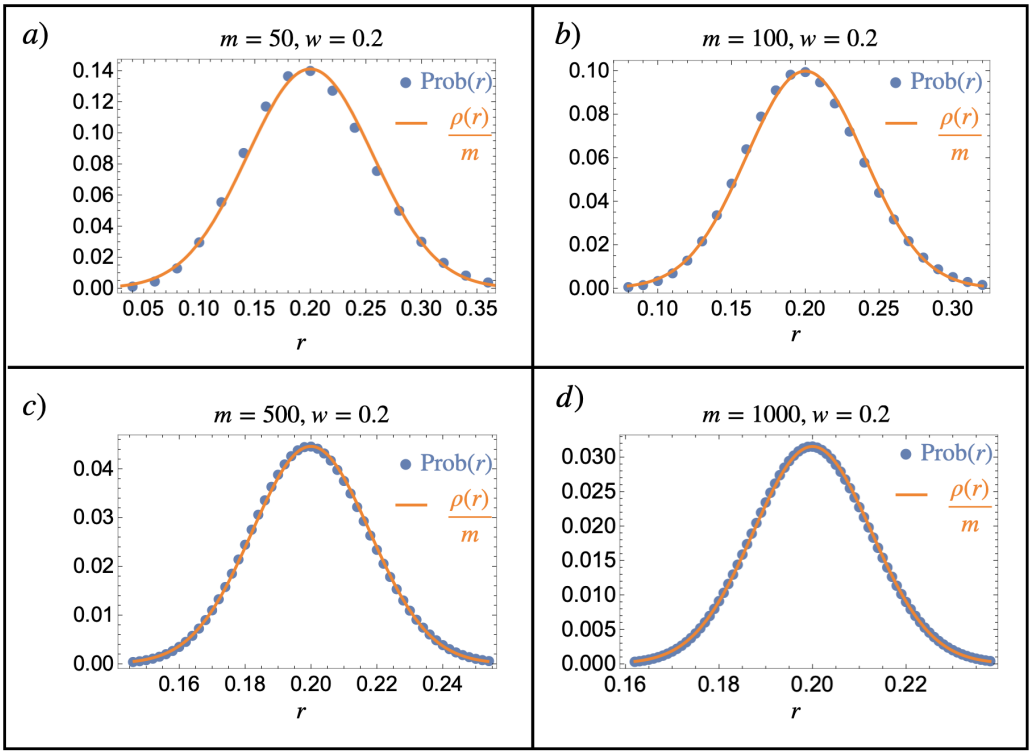


Figure A2. Normal limit of the binomial distribution. Comparison of Equations (24) (blue dots) and (26) (orange line) for various parameter values of m and $w = 0.2$: (**a**) $m = 50$, (**b**) $m = 100$, (**c**) $m = 500$, and (**d**) $m = 1000$. We see that even for $m = 50$, the normal distribution is already a good approximation of the binomial probability distribution, which becomes better as m is increased.

Appendix C

Given that there is no closed-form expression for ΔH in Equation (35), we are bound to find limiting cases for relatively large N values. This can be done by exploiting a few properties of ΔH:

- From Jensen's inequality [23] $\langle\!\langle f(\mathbf{X})\rangle\!\rangle \geq f(\langle\!\langle \mathbf{X}\rangle\!\rangle)$, it follows that, by choosing $f(x) = \ln x$, and
$$\mathbf{X} = \frac{\mathcal{M}_{N,m,p_{\text{mix}}}(\mathbf{C}_A + \mathbf{C}_B)}{\mathcal{M}_{\frac{N}{2},m,p_A}(\mathbf{C}_A)\mathcal{M}_{\frac{N}{2},m,p_B}(\mathbf{C}_B)}, \tag{A1}$$
we have

$$-H(\mathcal{M}_{\frac{N}{2},m,p_A}\mathcal{M}_{\frac{N}{2},m,p_B}|\mathcal{M}_{N,m,p_{\text{mix}}}) + H(\mathcal{M}_{\frac{N}{2},m,p_A}\mathcal{M}_{\frac{N}{2},m,p_B}|\mathcal{M}_{\frac{N}{2},m,p_A}\mathcal{M}_{\frac{N}{2},m,p_B}) \geq 0$$
$$\therefore \Delta H \geq 0 \tag{A2}$$

- It is possible to estimate the order of magnitude of ΔH in the large N limit. Indeed, given that the entropy of a multinomial distribution is positive, it automatically follows that
$$\begin{aligned}\Delta H &\leq H(\mathcal{M}_{\frac{N}{2},m,p_A}\mathcal{M}_{\frac{N}{2},m,p_B}|\mathcal{M}_{N,m,p_{\text{mix}}}) + \sum_{i=A,B} H(\mathcal{M}_{\frac{N}{2},m,p_i}|\mathcal{M}_{\frac{N}{2},m,p_i}) \\ &\leq 2\sum_{i=A,B} H(\mathcal{M}_{\frac{N}{2},m,p_i}|\mathcal{M}_{\frac{N}{2},m,p_i}),\end{aligned} \tag{A3}$$
where we use Equation (A2) and the fact that $H(\mathcal{M}_{\frac{N}{2},m,p_A}\mathcal{M}_{\frac{N}{2},m,p_B}|\mathcal{M}_{\frac{N}{2},m,p_A}\mathcal{M}_{\frac{N}{2},m,p_B}) = \sum_{i=A,B} H(\mathcal{M}_{\frac{N}{2},m,p_i}|\mathcal{M}_{\frac{N}{2},m,p_i})$ for the latter inequality. From Equation (18), it follows that $\Delta H \lessapprox \mathcal{O}(\ln N)$ for large N.

References

1. Gibbs, J.W. *Elementary Principles in Statistical Mechanics*; Ox Bow Press: Woodbridge, CT, USA, 1981.
2. Jaynes, E.T. The Gibbs' paradox. In *Proceedings of the Maximum Entropy and Bayesian Methods*; Smith, C., Erickson, G., Neudorfer, P., Eds.; Kluwer Academic: Norwell, MA, USA, 1992; pp. 1–22.
3. Hastings, C.S. Bibliographical Memoir of Josiah Willard Gibbs. In *Proceedings of the Bibliographical Memoirs, Part of Volume VI*; National Academy of Sciences: Washington, DC, USA, 1909.
4. Paillusson, F. Gibbs' paradox according to Gibbs and slightly beyond. *Mol. Phys.* **2018**, *116*, 3196. [CrossRef]
5. Gibbs, J.W. *On the Equilibrium of Heterogenous Substances*; Connecticut Academy of Arts and Sciences: New Haven, CT, USA, 1876; p. 108.
6. Darrigol, O. The Gibbs paradox: Early history and solutions. *Entropy* **2018**, *20*, 443. [CrossRef] [PubMed]
7. Kondepudi, D.; Prigogine, I. *Modern Thermodynamics*, 2nd ed.; John Wiley & Sons, Ltd.: Hoboken, NJ, USA, 2014.
8. Sator, N.; Pavloff, N. *Physique Statistique*, 1st ed.; Vuibert: Paris, France, 2016.
9. Van Kampen, N.G. The Gibbs' paradox. In *Proceedings of the Essays in Theoretical Physics: in Honor of Dirk ter Haar*; Parry, W., Ed.; Pergamon: Oxford, UK, 1984.
10. Cates, M.E.; Manoharan, V.N. Testing the Foundations of Classical Entropy: Colloid Experiments. *Soft. Matt.* **2015**, *15*, 6538. [CrossRef] [PubMed]
11. Bérut, A.; Arakelyan, A.; Petrosyan, A.; Ciliberto, S.; Dillenschneider, R.; Lutz, E. Experimental verification of Landauer's principle linking information and thermodynamics. *Nature* **2012**, *5*, 183. [CrossRef] [PubMed]
12. Frenkel, D. Why Colloidal Systems Can be described by Statistical Mechanics: Some not very original comments on the Gibbs' paradox. *Mol. Phys.* **2014**, *112*, 2325. [CrossRef]
13. Paillusson, F.; Pagonabarraga, I. On the role of compositions entropies in the statistical mechanics of polydisperse systems. *J. Stat. Mech.* **2014**, *2014*, P10038. [CrossRef]
14. Flory, P.J. *Principles of Polymer Chemistry*; Cornell University Press: Ithaca, NY, USA, 1953.
15. Salacuse, J.J. Random systems of particles: An approach to polydisperse systems. *J. Chem. Phys.* **1984**, *81*, 2468. [CrossRef]
16. Sollich, P. Projected free energy for polydisperse phase equilibria. *Phys. Rev. Lett.* **1998**, *80*, 1365. [CrossRef]
17. Warren, P.B. Combinatorial entropy and the statistical mechanics of polydispersity. *Phys. Rev. Lett.* **1998**, *80*, 1369. [CrossRef]
18. Sollich, P. Predicting phase equilibria in polydisperse systems. *J. Phys. Condens. Matter* **2002**, *14*, 79–117. [CrossRef]
19. Sollich, P.; Warren, P.B.; Cates, M.E. Moment Free Energies for Polydisperse Systems. In *Advances in Chemical Physics*; John Wiley & Sons, Ltd.: Hoboken, NJ, USA, 2001; Volume 116, pp. 265–336.
20. Endres, D.M.; Schindelin, J.E. A new metric for probability distributions. *IEEE Trans. Inf. Theory* **2003**, *49*, 1858. [CrossRef]
21. Paillusson, F. On the Logic of a Prior-Based Statistical Mechanics of Polydisperse Systems: The Case of Binary Mixtures. *Entropy* **2019**, *21*, 599. [CrossRef] [PubMed]
22. Cichón, J.; Golebiewski, Z. On Bernouilli Sums and Bernstein Polynomials. *Discret. Math. Theor. Comput. Sci.* **2012**, 179–190. [CrossRef]
23. Jensen, J.L.W.V. Sur les fonctions convexes et les inégalités entre les valeurs moyennes. *Acta Math.* **1906**, *30*, 175–193. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.