

Article

Not peer-reviewed version

Entropy of the Canonical Occupancy (Macro) State in the Quantum Measurement Theory

[Arnaldo Spalvieri](#)*

Posted Date: 18 December 2023

doi: 10.20944/preprints202312.1296.v1

Keywords: Occupancy Numbers; Multivariate Hypergeometric Distribution; Multinomial Distribution; Canonical Typicality; Gibbs Correction Factor; Sackur-Tetrode Entropy Formula



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Entropy of the Canonical Occupancy (Macro) State in the Quantum Measurement Theory

Arnaldo Spalvieri

Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano

Abstract: The paper analyzes the probability distribution of the occupancy numbers and the entropy of a system at the equilibrium composed by an arbitrary number of non-interacting bosons. The probability distribution is derived both by tracing out the environment from a bosonic eigenstate of the union of environment and system of interest (the empirical approach) and by tracing out the environment from the mixed state of the union of environment and system of interest (the Bayesian approach). In the thermodynamic limit, the two coincide and are equal to the multinomial distribution. Furthermore, the paper proposes to identify the physical entropy of the bosonic system with the Shannon entropy of the occupancy numbers, fixing certain contradictions that arise in the classical analysis of thermodynamic entropy. Finally, by leveraging an information-theoretic inequality between the entropy of the multinomial distribution and the entropy of the multivariate hypergeometric distribution, Bayesianism of information theory and empiricism of statistical mechanics are integrated into a common "infomechanical" framework.

Keywords: occupancy numbers; multivariate hypergeometric distribution; multinomial distribution; canonical typicality; Gibbs Correction Factor; Sackur-Tetrode entropy formula

1. Introduction

The concept of physical entropy is controversial since the times of Boltzmann and Gibbs. The following passage of an interview with Shannon, whose fundamental contribution to the understanding of entropy in the information-theoretic sense is today widely recognized also in physics [1], can be found in [2]:

My greatest concern was what to call it. I thought of calling it 'information,' but the word was overly used, so I decided to call it 'uncertainty.' When I discussed it with John von Neumann, he had a better idea. Von Neumann told me, "You should call it entropy, for two reasons. In the first place your uncertainty function has been used in statistical mechanics under that name, so it already has a name. In the second place, and more important, no one really knows what entropy really is, so in a debate you will always have the advantage."

During the years, the number of different interpretations and definitions of entropy has grown, a recent collection of heterogeneous "entropies" being reported in [3].

In this fragmented and ambiguous context, one of the most debated points is the relationship between information and physical entropy. The controversial about the role of information in physics plunge its roots in the famous thought experiments of Maxwell and Szilard, and is still today object of discussion. From the one side, Landauer claims in [4] that

Information is a physical entity.

From the other side, Maroney writes in his thesis [5] that

The Szilard Engine is unsuccessful as a paradigm of the information-entropy link,

and Maroney and Timpson reiterate the same concept in [6]:

rejecting the claims that information is physical provides a better basis for understanding the fertile relationship between information theory and physics.

Independently of the different opinions and arguments, it is a matter of fact that entropy is the key quantity that allows to derive all the thermodynamic macro-properties of many-particle systems from the statistical properties of the constituent microscopic particles.

One standard approach to the entropy in the classical (non-quantum) regime is based on the Gibbs entropy of microstates which, for systems at the equilibrium with the environment, is the Shannon entropy of the probability distribution resulting from entropy maximization under the constraints imposed on the system by the environment that surrounds it. The principle of maximization of entropy of microstates, often referred to as MaxEnt, dates back to the seminal paper of Jaynes [7] and is inspired to Shannon's information-theoretic view of entropy. Regrettably, the probability distribution of microstates does not account for indistinguishability of particles. At the same time, indistinguishability of particles is essential, therefore it is imperative to incorporate it into the framework. This is typically done by subtracting from the entropy of microstates the Gibbsian correction term $\log(N!)$, where N is the number of system's particles.

The Gibbsian correction term enters the scene almost like a sudden intervention, akin to a *deus ex machina*, but the subtraction of $\log(N!)$ generates a glaring contradiction: when the entropy of microstates is smaller than $\log(N!)$, the difference between entropy of microstates and $\log(N!)$ becomes negative, as it actually happens to the Sackur-Tetrode entropy formula at low temperature-to-density ratio. Since the entropy of any random variable is guaranteed to be always non-negative, entropy of microstates minus $\log(N!)$ cannot be entropy. This pathology is not surprising because, after the subtraction of $\log(N!)$, it is no more specified the random variable whose "entropy" is calculated, so "entropy" becomes a mere formula that is attributed to an unspecified random variable that represents an unspecified physical system. In this controversial situation, after more than 100 years from its introduction, the Gibbsian $\log(N!)$ is still today object of research and debate, see e.g. [8,9].

This paper recovers coherency of the picture by considering, in place of the classical setting, the more general quantum setting, where entropy is the von Neumann entropy of the mixed state of the system. The mixed state of a system is commonly represented by a density operator living in the Hilbert space $\mathcal{H}^{\otimes N}$ obtained as the tensor product N times of the single-particle Hilbert space \mathcal{H} . Exactly as it happens with the entropy of microstates, the von Neumann entropy of this density operator is appropriate for systems of distinguishable particles, but it does not capture indistinguishability of particles. We observe that, to capture indistinguishability of particles of a bosonic system, i.e. a system whose state remains unchanged when particles' indexes are permuted, an eigenbasis that conveniently represents its quantum state is that of the vectors of the occupancy numbers, giving rise to the so-called second quantization formalism. The consideration of indistinguishability of particles and, with it, of the occupancy numbers, leads us to define the *bosonic* density operator i.e. the density operator in system's bosonic subspace. Due to the inherent capability of the bosonic density operator to capture indistinguishability of particles, its von Neuman entropy *is* the entropy of the system. This allows overcoming the standard approach to entropy based on distinguishable particles and resolves the puzzle of the Gibbsian $\log(N!)$, which, as shown in this paper, now enters the scene through the main entrance alongside other quantum correction terms ensuring non-negativity of entropy.

Our exploration of the quantum approach is situated within the framework of contemporary quantum thermodynamics, whose origins can be attributed to [10,11]. The two cited papers, which have deeply influenced the successive literature, e.g. [12–16], derive the mixed state of a system at the equilibrium with the environment by tracing out the environment from the universe, that, in the thermal case, is the union of system and heat bath that thermalizes the system. The analysis of [10] and [11] shows that, as the number of particles of the universe tends to infinity, the mixed state of the system is the same, at least in the weak sense, both if the universe is in a pure state or in a mixed state. Purity of the state of the universe makes unnecessary the introduction of a statistical ensemble of "universes," overcoming the subjectivism that is inherent in the Bayesian approach.

The outline of the paper is as follows. Section II introduces the bosonic Hilbert subspace and its eigenbasis. In Section III, we derive the bosonic density operator for the system under consideration by

tracing out the environment from the universe, assuming that the universe is in a bosonic eigenstate. With this assumption, we show that the probability distribution that weights the projectors of the bosonic Hilbert subspace of the system is the multivariate hypergeometric distribution. As the number of bosons of the universe tends to infinity, the multivariate hypergeometric distribution converges to the multinomial distribution, which we identify as the canonical distribution of the occupancy numbers. Section IV shows that, if, as in the Bayesian approach, the universe is assumed to be in a mixed bosonic state, then the distribution of the occupancy numbers of the system is multinomial provided that the occupancy numbers of the universe are multinomially distributed too. Section V discusses the application of the mentioned probability distributions to the entropy of physical systems and places our work within the framework of quantum information theory, unveiling an engaging connection between Bayesianism and empiricism in physics. To illustrate the intrinsic capacity of the canonical bosonic density operator to capture indistinguishability of particles in systems at the equilibrium, Section VI shows that its von Neumann entropy fits the entropy of the ideal gas in a container. Section VII sketches future application of our approach to the Szilard engine. Finally, in Section VIII we draw the conclusions.

2. The bosonic Hilbert subspace

Let $\{c \in \mathbb{C}\}$, $\mathbb{C} = \{1, 2, \dots, |\mathbb{C}|\}$, be the set of quantum numbers (the colors) allowed to a boson, let $\{\bar{c} = (c_1, c_2, \dots, c_N)\}$ (the overline denotes vectors and their size is not explicitly mentioned for brevity when it can be inferred from the context) be the set of *microstates* of a system made by N non-interacting bosons (the balls), and let $\{|\bar{c}\rangle\}$ be the complete set of eigenstates that span the Hilbert space $\mathcal{H}^{\otimes N}$ of the system. Let $\delta(\cdot)$ be the indicator function and let

$$n_c(c_1, c_2, \dots, c_N) \stackrel{\text{def}}{=} \sum_{i=1}^N \delta(c - c_i), \quad c = 1, 2, \dots, |\mathbb{C}|,$$

be the number of bosons whose quantum number is c . In the following, the dependency on (c_1, c_2, \dots, c_N) of $n_c(c_1, c_2, \dots, c_N)$ will be omitted when possible and the vector of the occupancy numbers $\bar{n} = (n_1, n_2, \dots, n_{|\mathbb{C}|})$ will be called *occupancy macrostate*. The quantum state of the system of indistinguishable bosons is

$$|\bar{n}\rangle = \sum_{\bar{c} \in \mathbb{C}^N(\bar{n})} \frac{|\bar{c}\rangle}{\sqrt{W(\bar{n})}}, \quad (1)$$

see e.g. [17] and chapter 7 of [18]. In the above equation, $\mathbb{C}^N(\bar{n})$ is the set of the $W(\bar{n})$ equiprobable microstates whose occupancy macrostate is \bar{n} and $W(\bar{n})$ is the multinomial coefficient,

$$W(\bar{n}) = \begin{cases} \frac{N!}{\prod_{c=1}^{|\mathbb{C}|} n_c!}, & n_c \geq 0, \forall c \in \mathbb{C}, \\ 0, & \text{elsewhere,} \end{cases}$$

i.e. the number of distinct permutations of the entries of any one of the elements of $\mathbb{C}^N(\bar{n})$.

The set $\{|\bar{n}\rangle, \bar{n} \in \mathbb{N}\}$, with

$$|\mathbb{N}| = \binom{N + |\mathbb{C}| - 1}{|\mathbb{C}| - 1}$$

see [19], is a complete set of bosonic eigenstates for the bosonic Hilbert subspace of the Hilbert space of the system.

3. Empirical approach

We want to obtain the system's bosonic density operator $\hat{\nu}_{\bar{u}}$, where \bar{u} is the occupancy macrostate of the universe, tracing out the environment made by $U - N$ bosons from the projector $|\bar{u}\rangle \langle \bar{u}|$ of the universe made by U bosons. The first step is to write $|\bar{u}\rangle$ in the form of bosonic purification:

$$\begin{aligned} |\bar{u}\rangle &= \sum_{\bar{n} \in \mathbb{N}} \sum_{\bar{c}_1^N \in \mathbb{C}^N(\bar{n}), \bar{c}_{N+1}^U \in \mathbb{C}^{U-N}(\bar{u}-\bar{n})} \frac{|\bar{c}_1^N, \bar{c}_{N+1}^U\rangle}{\sqrt{W(\bar{u})}} \\ &= \sum_{\bar{n} \in \mathbb{N}} \sqrt{\frac{W(\bar{n})W(\bar{u}-\bar{n})}{W(\bar{u})}} |\bar{n}, \bar{u}-\bar{n}\rangle, \end{aligned}$$

where \bar{c}_i^j means windowing of vector \bar{c} between i and j and it is understood that the set $\mathbb{C}^N(\bar{n})$ ($\mathbb{C}^{U-N}(\bar{u}-\bar{n})$) is empty when one or more entries of \bar{n} ($\bar{u}-\bar{n}$) are negative. Observing that

$$\langle \bar{u}-\bar{n} | \bar{u}-\bar{n}' \rangle |\bar{n}\rangle \langle \bar{n}'| = \begin{cases} |\bar{n}\rangle \langle \bar{n}'|, & \bar{u}-\bar{n} = \bar{u}-\bar{n}', \\ 0, & \bar{u}-\bar{n} \neq \bar{u}-\bar{n}', \end{cases}$$

and that the condition $\bar{u}-\bar{n} = \bar{u}-\bar{n}'$ forces $\bar{n} = \bar{n}'$, we conclude that

$$\begin{aligned} \hat{\nu}_{\bar{u}} &= \text{Tr}_{\mathcal{H}^{\otimes(U-N)}} |\bar{u}\rangle \langle \bar{u}| \\ &= \sum_{\bar{n} \in \mathbb{N}} \frac{W(\bar{u}-\bar{n})W(\bar{n})}{W(\bar{u})} |\bar{n}\rangle \langle \bar{n}|, \end{aligned} \quad (2)$$

where $\text{Tr}_{\mathcal{H}^{\otimes(U-N)}}$ is the operator that traces out the Hilbert space of the environment. If the set $\mathbb{C}^N(\bar{n})$ ($\mathbb{C}^{U-N}(\bar{u}-\bar{n})$) is empty, then one or more entries of \bar{n} ($\bar{u}-\bar{n}$) are negative, in which case the multinomial coefficient $W(\bar{n})$ ($W(\bar{u}-\bar{n})$) is zero by definition. The fraction appearing in (2) is the multivariate hypergeometric distribution, which is the distribution of the occupancy numbers of colors in drawing without replacement N balls out of an urn containing U balls with color occupancy numbers \bar{u} . In many textbooks and papers, the multivariate hypergeometric distribution is expressed by binomial coefficients as

$$P_{\bar{u}}(\bar{n}) = \frac{W(\bar{u}-\bar{n})W(\bar{n})}{W(\bar{u})} = \frac{\prod_{c=1}^{|\mathbb{C}|} \binom{u_c}{n_c}}{\binom{U}{N}}, \quad (3)$$

where $P(\cdot)$ is the probability of the random variable inside the round brackets and \bar{u} in the subscript means that the probability distribution $\{P_{\bar{u}}(\bar{n})\}$ depends on the vector of known parameters \bar{u} .

In the thermodynamic limit the multivariate hypergeometric distribution converges to the multinomial distribution, i.e. the distribution of the occupancy numbers of colors in drawing with replacement N times a ball out of an urn containing colored balls with relative frequency of color c equal to $P(c)$:

$$\begin{aligned} \lim_{U \rightarrow \infty} P_{\bar{u}}(\bar{n}) &= W(\bar{n}) \lim_{U \rightarrow \infty} \frac{W(\bar{u}-\bar{n})}{W(\bar{u})} \\ &= W(\bar{n}) \lim_{U \rightarrow \infty} \prod_{c=1}^{|\mathbb{C}|} (U^{-1} u_c)^{n_c} \end{aligned} \quad (4)$$

$$= W(\bar{n}) \prod_{c=1}^{|\mathbb{C}|} (P(c))^{n_c}, \quad (5)$$

$$\lim_{U \rightarrow \infty} \hat{v}_{\bar{u}} = \hat{v} = \sum_{\bar{n} \in \mathbb{N}} W(\bar{n}) \prod_{c=1}^{|\mathbb{C}|} (P(c))^{n_c} |\bar{n}\rangle \langle \bar{n}|,$$

where (4) is Stirling's formula, while, by regarding \bar{u} as the result of a PVM operated on the universe, equation (5) is recognized to be the Law of Large Numbers (LLN) for the empirical one-particle distribution. *Concentration inequalities* that bound the probability of deviations of the empirical probability from its expectation, that is the probability of occurrence of *non-typical* bosonic eigenstates of the universe, can be found in [20] for hypergeometrically distributed random variables, in [21] for multinomially distributed random vectors. Paper [22] demonstrates that the multinomial distribution is the maximum entropy distribution of the occupancy numbers with constrained one-particle distribution. See also [23,24] for the multinomial distribution in statistical mechanics.

By equiprobability of the disjoint microstates belonging to the same occupancy macrostate, we see that, in (3),

$$\frac{W(\bar{u} - \bar{n}(c_1, c_2, \dots, c_N))}{W(\bar{u})} = P_{\bar{u}}(c_1, c_2, \dots, c_N).$$

Clearly

$$\begin{aligned} \lim_{U \rightarrow \infty} P_{\bar{u}}(c_1, c_2, \dots, c_N) &= \prod_{c=1}^{|\mathbb{C}|} (P(c))^{n_c(c_1, c_2, \dots, c_N)} \\ &= \prod_{i=1}^N P(c_i), \end{aligned}$$

which explicitly shows that, in the thermodynamic limit, microstates outcoming from the mentioned PVM becomes independent and identically distributed random variables. What happens is that the LLN cancels the dependencies inside the joint distribution of microstates induced by the constraints imposed on the colors of the N balls by the occupancy numbers \bar{u} of the universe in drawing without replacement. Furthermore, the probability distribution $\{\prod_{i=1}^N P(c_i)\}$ also becomes independent of the specific result \bar{u} of the PVM operated on the universe, in the sense that the empirical one-particle distribution $\{U^{-1}u_c\}$ for $U \rightarrow \infty$ is the same for *almost every bosonic eigenstate of the universe*, i.e., for the *typical* bosonic eigenstates of the universe. The one-particle distribution and, with it, the product $\{\prod_{i=1}^N P(c_i)\}$, will depend on the physical constraints imposed on the universe. When the constraint is the temperature, it is widely recognized that $\{\prod_{i=1}^N P(c_i)\}$ is the Boltzmann distribution, that can be found by Jaynes' constrained maximization of entropy [7].

Independency and identical distribution of the eigenstates of the individual bosons of the system lead to the following density operator in $\mathcal{H}^{\otimes N}$

$$\hat{\rho} = \sum_{\bar{c} \in \mathbb{C}^N} \prod_{i=1}^N P(c_i) |\bar{c}\rangle \langle \bar{c}|,$$

that we claim to be the canonical density operator in $\mathcal{H}^{\otimes N}$. Papers [10] and [11] claim that system's density operator in $\mathcal{H}^{\otimes N}$ converges to the canonical density operator as $U \rightarrow \infty$ for *almost every pure state of the universe*, i.e. for the *typical* pure states of the universe. For this reason convergence to the canonical state is called *canonical typicality* in [11]. Typicality, which can be intended in various senses, is often invoked in statistical mechanics. For instance, paper [25] uses the properties of the information-theoretic typical set to characterize weak convergence of microstates to equiprobability in the context of classical statistical mechanics. We point out that the claim of [10] and [11] is compatible with our claim of convergence to the bosonic canonical state if we do not pretend that typical bosonic eigenstates are typical states, as illustrated in Figure 1.

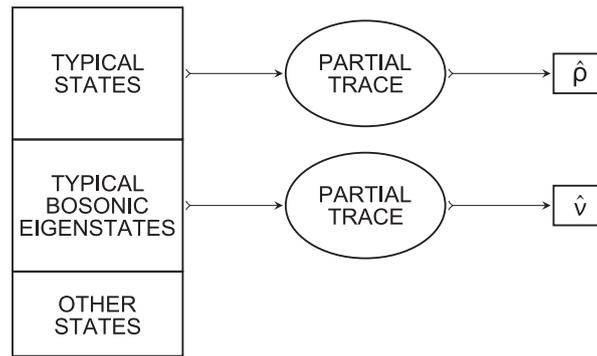


Figure 1. The box on the left represents the pure states of the universe. Although in the thermodynamic limit almost every pure state of the universe is typical, still there is space for infinitely many typical bosonic eigenstates and for infinitely many other pure states that are neither typical states nor typical bosonic eigenstates, among which are all the eigenstates of $\mathcal{H}^{\otimes U}$.

4. Bayesian approach

In the Bayesian approach, the vector of known parameters \bar{u} becomes a random vector whose probability distribution, the Bayesian *prior*, characterizes the bosonic density operator of the universe. In the Bayesian setting, the hypergeometric multivariate distribution of the density operator $\hat{v}_{\bar{u}}$ of (2) is the Bayesian *likelihood*, i.e. a conditional distribution where the random condition is the random vector of parameters \bar{u} . The probability distribution of the density operator of the system, the Bayesian *marginal*, is obtained by tracing out the environment from the bosonic density operator of the universe. The Bayesian approach is self-consistent when the prior is multinomial, because the marginal turns out to be multinomial too, whichever is the number of particles of the universe:

$$\begin{aligned}
 \hat{v} &= \sum_{\bar{u} \in \mathbb{U}} W(\bar{u}) \prod_{c=1}^{|\mathcal{C}|} (P(c))^{u_c} \text{Tr}_{\mathcal{H}^{\otimes (U-N)}} |\bar{u}\rangle \langle \bar{u}| \\
 &= \sum_{\bar{u} \in \mathbb{U}} W(\bar{u}) \prod_{c=1}^{|\mathcal{C}|} (P(c))^{u_c} \hat{v}_{\bar{u}} \\
 &= \sum_{\bar{n} \in \mathbb{N}} W(\bar{n}) \prod_{c=1}^{|\mathcal{C}|} (P(c))^{n_c} |\bar{n}\rangle \langle \bar{n}|, \tag{6}
 \end{aligned}$$

where the third equality is obtained by substituting (2) in the second line and manipulating. The comparison between the Bayesian (6) and the empirical (2) clarifies the relationship between the two approaches. The Bayesian approach captures the pre-measurement randomness of the mixed state of the universe by a mathematical procedure based on constrained entropy maximization whose result is the multinomial prior of the universe. Instead, the empirical approach renounces to mathematics and imports the result \bar{u} of the measurement of the occupancy numbers of the universe in the multivariate hypergeometric distribution of the occupancy numbers of a system that is separated from the universe after the measurement.

5. Quantum entropy and quantum information

Since $\hat{v}_{\bar{u}}$ (\hat{v}) is an eigendecomposition, its von Neumann entropy is equal to the Shannon entropy of the multivariate hypergeometric (multinomial) distribution. For this reason, in the following we make no distinction between the von Neumann entropy and the Shannon entropy, calling it simply entropy. It is worth emphasizing that here we completely skip the notion of phase space, leading to the *exact* probability distribution of the quantum occupancy numbers, and, as a consequence, to the exact entropy. Conversely, the standard phase space approach inherently leads to approximations

to entropy, that ask for improvement at low temperature/density ratio, see e.g. [26], still remaining approximations.

The entropy of the multivariate hypergeometric distribution, which, owing to the distribution's symmetry, is equal to the entropy of both the system and the environment after their separation, is

$$\begin{aligned} S(\hat{v}_u) &= \log(W(\bar{u})) - E\{\log(W(\bar{u} - \bar{n}))\} - E\{\log(W(\bar{n}))\} \\ &= -E\{\log(P_{\bar{u}}(\bar{c}))\} - \log(N!) + \sum_{c=1}^{|\mathbb{C}|} E\{\log(n_c!)\}, \end{aligned} \quad (7)$$

where $S(\cdot)$ is the von Neuman entropy of the density operator inside the round brackets, $E\{\cdot\}$ is the classical expectation computed over the probability distribution of the random variable inside the expectation, the base of the logarithm is Euler's number, and the Boltzmann constant in front of the logarithm is omitted for brevity. The term $-E\{\log(P_{\bar{u}}(\bar{c}))\}$ is the Gibbs entropy of microstates, i.e. the entropy of the system of distinguishable particles, while the term $-E\{\log(W(\bar{n}))\}$ is due to indistinguishability of particles which, in the average, prevents the access to $E\{\log(W(\bar{n}))\}$ units of information. Since $(W(\bar{n}))^{-1}$ is the conditional probability of microstates given the macrostate, its expectation is the conditional Shannon entropy of microstates given the macrostate, so $S(\hat{v}_u)$ is the mutual information between microstates and macrostates. The term $\log(N!)$ in (7) was introduced by Gibbs to force compatibility between the non-quantized phase-space (differential) entropy of microstates and the physical entropy of systems of indistinguishable particles. We observe that, while the probability that two or more classical particles have the same position and momentum in the phase space is zero because position and momentum are dense variables, the probability that two or more quantum particles occupy the same quantum state is not zero. This non-zero probability is captured by the sum of expectations in (7). As the entropy of microstates becomes lower and lower, this sum becomes closer and closer to $\log(N!)$, till becoming equal to $\log(N!)$ when all the particles occupy the ground state. This prevents system's entropy to become negative also when the entropy of microstates becomes vanishingly small.

In the canonical case, the entropy of the bosonic density operator is the entropy of the multinomial distribution:

$$S(\hat{v}) = -NE\{\log(P(c))\} - E\{\log(W(\bar{n}))\}, \quad (8)$$

see [27] and [28] for the calculation of the above entropy, see also [24] for approximations to the entropy of the multinomial distribution in the context of statistical mechanics. Equation (8) is equation (11) of [24], where the authors call the entropy of the distribution of the occupancy numbers *entropy fluctuations*. Apart of certain exceptions, the authors of [24] consider these "entropy fluctuations" negligible compared to the "entropy" of the system, failing to recognize that the entropy of the occupancy numbers *is* the thermodynamic entropy of a system of indistinguishable particles.

The following inequalities sandwich the Boltzmann entropy $\log(W(E\{\bar{n}\}))$ between the two terms in the right hand side of (8):

$$\begin{aligned} -NE\{\log(P(c))\} &\geq \log(W(E\{\bar{n}\})) \\ &\geq E\{\log(W(\bar{n}))\}, \end{aligned} \quad (9)$$

where, with some abuse of notation, the factorials of the real numbers in the denominator of $W(E\{\bar{n}\})$ are intended as $x! = \Gamma(x + 1)$, where $\Gamma(\cdot)$ is the Gamma function. The first inequality is (11.22) of [19], the second inequality is obtained by applying the Jensen inequality

$$E\{f(n_c)\} \geq f(E\{n_c\}), \quad \forall c \in \mathbb{C},$$

to the convex (upward) function $f(n_c) = \log(n_c!)$. In statistical mechanics it is standard to derive from Stirling's formula an approximation between the two terms of (9).

Note that, if we pretend that entropy is a variable of state, then the probability distribution of the occupancy numbers must depend only on the state of the system. However, the multivariate hypergeometric distribution depends also on the state of the universe. The dependency becomes weaker and weaker as the number of particles of the universe tends to infinity, but it remains that this makes the empirical approach incompatible with the notion of entropy as variable of state. In conclusion, entropy can be a variable of state only if we accept the Bayesian approach.

We hereafter introduce the *empirical* information, sketching a new engaging connection between the Bayesian and the empirical approaches to information in quantum measurement theory. Let us regard the PVM operated on the universe as a POVM operated on the system. The difference between the entropy of the multinomial Bayesian marginal and the expectation over the multinomial Bayesian prior $\{P(\bar{u})\}$ of the entropy of the multivariate hypergeometric Bayesian likelihood of the system is equal to the Holevo upper bound χ above the accessible quantum information that any POVM can achieve [29]:

$$\chi = S(\hat{v}) - \sum_{\bar{u} \in U} P(\bar{u})S(\hat{v}_{\bar{u}}) \geq 0.$$

Whichever is U , in place of the above Bayesian information one could consistently consider the following empirical information, that does not need the definition of a prior:

$$S(\hat{v}_{\text{emp}}) - S(\hat{v}_{\bar{u}}) \geq 0, \quad (10)$$

where $S(\hat{v}_{\text{emp}})$ is the Shannon entropy of the empirical multinomial distribution $\{P_{\text{emp}}(\bar{n})\}$ of the occupancy numbers of the system,

$$P_{\text{emp}}(\bar{n}) = W(\bar{n}) \prod_{c=1}^{|\mathcal{C}|} (U^{-1}u_c)^{n_c}.$$

In (10), the one-particle probability distribution $\{U^{-1}u_c\}$ is the same for the multivariate hypergeometric distribution and for the empirical multinomial distribution, therefore the inequality is guaranteed by the maximum entropy property of the multinomial distribution demonstrated in [22]. The difference (10) is the "empirical information" brought by PVM operated on the universe about the system. As $U \rightarrow \infty$, both $S(\hat{v}_{\bar{u}})$ and $S(\hat{v}_{\text{emp}})$ tend to $S(\hat{v})$ and both the Bayesian information and the empirical information tend to zero. If, after the POVM, a PVM is operated on the system, the total information brought by the two measurements is $S(\hat{v})$ in the Bayesian approach, $S(\hat{v}_{\text{emp}})$ in the empirical approach. As $U \rightarrow \infty$, the total empirical information becomes equal to the total Bayesian information.

6. Entropy of the ideal gas in a container

In the case of an ideal monoatomic gas in a cubic container of side L , one particle of the gas is modelled as a quantum "particle in a box" with three degrees of freedom, whose energy eigenvalues with aperiodic boundary conditions are

$$\epsilon_c = (c_x^2 + c_y^2 + c_z^2) \frac{h^2}{8mL^2},$$

where c consists of the three quantum numbers (c_x, c_y, c_z) , m is the mass of the particle and $h = 6.626 \cdot 10^{-34} \text{ J} \cdot \text{s}$ is the Planck constant. The one-particle Boltzmann distribution for a gas at the thermal equilibrium at temperature T Kelvin degrees is

$$P(c) = Z^{-1} e^{-\epsilon_c/k_B T},$$

and the associated multinomial distribution of the occupancy numbers for a gas of N particles is

$$P(\bar{n}) = W(\bar{n}) Z^{-N} \prod_{c=1}^{|\mathcal{C}|} e^{-n_c \epsilon_c/k_B T},$$

where $k_B = 1.38 \cdot 10^{-23}$ J/K is the Boltzmann constant and Z is the one-particle partition function:

$$Z = \sum_{c=1}^{|\mathcal{C}|} e^{-\epsilon_c/k_B T}.$$

When the temperature-to-density ratio is high, it becomes possible to employ two approximations. In the first one, the partition function is approximated to an integral, see eqn. 19.54 of [30], leading to

$$-NE\{\log(P(c))\} \approx \frac{3N}{2} \left(1 + \log \left(\frac{2\pi m k_B T L^2}{h^2} \right) \right).$$

In the second one, with the idea that the probability that two or more particles occupy the same state is negligible, the denominator of the multinomial coefficient is ignored and, for large number of particles, the logarithm of the numerator $\log(N!)$ is approximated to $N \log(N) - N$, leading to

$$-E\{\log(W(\bar{n}))\} \approx -N \log(N) + N.$$

Plugging the two approximations in (8) one gets textbook Sackur-Tetrode formula:

$$S(\hat{v}) \approx N \left(\log \left(\frac{L^3}{N} \left(\frac{2\pi m k_B T}{h^2} \right)^{\frac{3}{2}} \right) + \frac{5}{2} \right).$$

Note that, as already mentioned in the introduction, the exact entropy of the multinomial distribution (8) is guaranteed to be non-negative, while the Sackur-Tetrode formula becomes negative at low temperature-to-density ratio, where the two mentioned approximations do not hold, see [31,32] for details.

7. Future work

We hereafter sketch the application of our proposed approach to the state of Szilard engine after the insertion of the piston, letting the complete analysis of the Szilard cycle to future work. In the case of N particles, after the insertion of the piston the number b of particles that are found in one of the two sub-container of volume V' is a binomial random variable $(N, V'/V)$, where V is the total volume. The probability distribution of the occupancy numbers is

$$P(\bar{n}', \bar{n}'') = \sum_{b=0}^N P(\bar{n}'|b) P(\bar{n}''|b) P(b),$$

where $\{P(\bar{n}'|b)\}$ ($\{P(\bar{n}''|b)\}$) is the probability distribution of the occupancy numbers of a gas with b ($N - b$) particles in the sub-container of volume V' ($V - V'$). For $N = 1$ and $V' = V/2$, the entropy after the insertion of the wall is

$$\log(2) - \sum_{c=1}^{|\mathcal{C}|} P(c) \log(P(c)),$$

where the famous $\log(2)$ of Landauer [33] comes from the binary equiprobable random variable b and $\{P(c)\}$ is the probability distribution of one particle in a box of volume V' . We have evaluated the partition function of the Boltzmann distribution with the parametrization of [34], that is mass of the particle $m = 9.11 \cdot 10^{-31}$ kg, temperature $T = 300$ K, and one-dimensional box of size $L = 20 \cdot 10^{-9}$ m. We obtain that the entropy of the single particle with one degree of freedom before the insertion of the piston is 1.988 in k_B units, while with size of the one-dimensional box equal to $10 \cdot 10^{-9}$ m, that is, after the insertion of the piston, the entropy in k_B units is 1.243, leading to the difference $1.988 - 0.693 - 1.243 = 0.052$, in excellent agreement with the entropy fall shown in Figure 3 of [34], where the result is derived by the phase space approach.

8. Conclusion

Entropy is a macroscopic property of a physical system and, at the same time, it is a mathematical property of randomness. As such, it *must* be a property of the randomness of system's macrostates. However, despite the necessity of introducing the somewhat ad hoc term $-\log(N!)$ as a kind of *deus ex machina*, from Boltzmann to the present day the "entropy" of a system at the equilibrium is commonly intended as the entropy of microstates, be it the Gibbs entropy or, to some extents, the Boltzmann entropy $\log(W(E\{\bar{n}\}))$. This author has come to the conclusion that the misunderstanding arises from the following two conceptual errors.

The first error consists in regarding $\log(W(\bar{n}))$ as the "entropy of a macrostate." Even an exceptionally deep author as Jaynes writes in [35]: *To emphasize this, note that a "thermodynamic state" denoted by $X = \{X_1, \dots, X_n\}$ defines a large class $C(X)$ of microstates compatible with X . Boltzmann, Planck, and Einstein showed that we may interpret the entropy of a macrostate as $S(X) = k \log(W(C))$, where $W(C)$ is the phase volume occupied by all the microstates in the chosen reference class C .* This misconception is pervasive in the entire statistical mechanics. The influential authors of [36] attribute to Einstein the idea that an individual macrostate can have entropy: *In fact, already Einstein (1914, Eq. (4a)) argued that the entropy of a macro state should be proportional to the log of the "number of elementary quantum states" compatible with that macro state... .* But entropy is a property of a statistical ensemble, therefore one individual macrostate cannot have entropy. Moreover, if by entropy of the macrostate we mean the entropy of the microstates that constitute it, then we must acknowledge that this entropy has no physical significance because the *elementary quantum state compatible with that macro state* are physically meaningless when particles are indistinguishable. This misconception is so widespread that can be found also in standard textbooks. We hereafter quote a passage from the introduction to chapter 16 of [30]: *specification of a macrostate constitutes incomplete information.* In absence of a formal definition of information, this statement risks to become misleading. Actually, we have shown that the entropy of the occupancy numbers *is* the complete information about the system because the entropy of microstates belonging to the same macrostate, due to indistinguishability of particles, is not informative, as it is subtracted in (8) to the entropy of microstates.

The second error is the lack of consideration of the absolute randomness of macrostates. Navigating through the literature of statistical mechanics, you may come across passages like the following as the one of [23] that is hereafter reported: *A crucial observation in statistical mechanics is that the distribution of all macrostate variables gets sharply peaked and narrow as system size N increases. ... In the limit $N \rightarrow \infty$ the probability of measuring a macrostate becomes a Dirac delta...* Clearly, the quoted statement is wrong, because the *absolute* width of the probability distribution of the occupancy numbers, and, more generally, of any macroscopic observable as for instance system's energy, becomes broader and broader as the number of particles grows, while, by the LLN, its *relative* width, i.e. the width compared to the mean value, sharpens. Certainly, it is apparent that

$$\lim_{N \rightarrow \infty} N^{-1} S(\hat{v}) = 0,$$

which shows that the relative randomness, i.e. the randomness per particle, becomes vanishingly small as $N \rightarrow \infty$. Meanwhile, the absolute randomness, represented by $S(\hat{\nu})$, increases as N becomes larger and larger.

In the end, the lack of formal specification of the physical role of microstates, together with the lack of consideration of the absolute randomness of macrostates, lead to the questionable belief that all the properties of a system of a large numbers of particles, including "entropy," depend on the macrostate $E\{\bar{n}\}$, because it contains "the overwhelming majority" of microstates. From this standpoint, the subtraction of $\log(N!)$ is seen as a technical maneuver that becomes necessary to circumvent the challenge posed by the indistinguishability of particles. Note that the idea itself that one macrostate can contain the overwhelming majority of microstates is inherently questionable, because the ratio $|\mathbb{C}^N(\bar{n})|/|\mathbb{C}^N|$ tends to zero whichever is \bar{n} as $N \rightarrow \infty$.

In conclusion, this paper made clear that, since the quantum state of a system of indistinguishable particles (bosons) is completely specified by the occupancy numbers of the quantum states allowed to system's particles, the entropy the physical system is the Shannon entropy of the random occupancy numbers \bar{n} , which is obtained by subtracting the expectation of $\log(W(\bar{n}))$ to the entropy of microstates. Recognizing that $E\{\log(W(\bar{n}))\}$ is the conditional Shannon entropy of microstates given the macrostate, we equivalently express the above concept by saying that the entropy of the physical system is equal to the mutual information between microstates and macrostates.

References

1. "The bit juggler." *nature physics* 19.919 (2023): 919.
2. Tribus, Myron, and Edward C. McIrvine. "Energy and information." *Scientific American* 225.3 (1971): 179-190.
3. Safranek, Dominik, et al. "A brief introduction to observational entropy." *Foundations of Physics* 51 (2021): 1-20.
4. Landauer, Rolf. "Information is a physical entity." *Physica A: Statistical Mechanics and its applications* 263.1-4 (1999): 63-67.
5. Maroney, Owen Jack Ernest. "Information and entropy in quantum theory." arXiv preprint quant-ph/0411172 (2004).
6. Maroney, Owen JE, and Christopher G. Timpson. "How is there a physics of information? On characterizing physical evolution as information processing." *Physical perspectives on computation, computational perspectives on physics* (2018): 103-126.
7. Jaynes, E. T. "Information theory and statistical mechanics." *Physical review*, 1957, 4, 620–630.
8. Sasa, Shin-ichi, et al. "Quasi-static Decomposition and the Gibbs Factorial in Small Thermodynamic Systems." *Journal of Statistical Physics* 189.2 (2022): 31.
9. Tasaki, Hal. "The best answer to the puzzle of Gibbs about $N!!$: A note on the paper by Sasa, Hiura, Nakagawa, and Yoshida." arXiv preprint arXiv:2206.05513 (2022).
10. Popescu, Sandu, Anthony J. Short, and Andreas Winter. "Entanglement and the foundations of statistical mechanics." *Nature Physics* 2.11 (2006): 754-758.
11. Goldstein, Sheldon, et al. "Canonical typicality." *Physical review letters* 96.5 (2006): 050403.
12. Bartsch, Christian, and Jochen Gemmer. "Dynamical typicality of quantum expectation values." *Physical review letters* 102.11 (2009): 110403.
13. Von Spakovsky, Michael R., and Jochen Gemmer. "Some trends in quantum thermodynamics." *Entropy* 16.6 (2014): 3434-3470.
14. Goold, John, et al. "The role of quantum information in thermodynamics - a topical review." *Journal of Physics A: Mathematical and Theoretical* 49.14 (2016): 143001.
15. Parrondo, Juan MR, Jordan M. Horowitz, and Takahiro Sagawa. "Thermodynamics of information." *Nature physics* 11.2 (2015): 131-139.
16. Deffner, Sebastian, and Steve Campbell. *Quantum Thermodynamics: An introduction to the thermodynamics of quantum information*. Morgan and Claypool Publishers, 2019.
17. Beggi, Andrea, et al. "Back and forth from Fock space to Hilbert space: a guide for commutators." *European Journal of Physics* 39.6 (2018): 065401.

18. Kardar, M. "Statistical physics of particles." Cambridge University Press, 2007.
19. Cover, T. M., and Thomas, J. A. "Elements of information theory, 2nd edition." Wiley series in telecommunications and signal processing, 2006.
20. Greene, Evan, and Jon A. Wellner. "Exponential bounds for the hypergeometric distribution." *Bernoulli: official journal of the Bernoulli Society for Mathematical Statistics and Probability* 23.3 (2017): 1911.
21. Agrawal, Rohit. "Finite-sample concentration of the multinomial in relative entropy." *IEEE Transactions on Information Theory* 66.10 (2020): 6297-6302.
22. Stern, Hal, and Thomas M. Cover. "Maximum entropy and the lottery." *Journal of the American Statistical Association* 84.408 (1989): 980-985.
23. Hanel, Rudolf, Stefan Thurner, and Murray Gell-Mann. "How multiplicity determines entropy and the derivation of the maximum entropy principle for complex systems." *Proceedings of the National Academy of Sciences* 111.19 (2014): 6905-6910.
24. Županović, Paško, and Domagoj Kuić. "Relation between Boltzmann and Gibbs entropy and example with multinomial distribution." *Journal of Physics Communications* 2.4 (2018): 045002.
25. Spalvieri, Arnaldo. "The Shannon-McMillan theorem proves convergence to equiprobability of Boltzmann's microstates." *Entropy* 2021, 23, 899.
26. Qiu, Tian, et al. "Quantum corrections to the entropy and its application in the study of quantum Carnot engines." *Physical Review E* 101.3 (2020): 032113.
27. Kaji, Yuichi. "Bounds on the entropy of multinomial distribution." 2015 IEEE International Symposium on Information Theory (ISIT). IEEE, 2015.
28. Cheraghchi, Mahdi. "Expressions for the entropy of binomial-type distributions." 2018 IEEE International Symposium on Information Theory (ISIT). IEEE, 2018.
29. Nielsen, Michael A., and Isaac L. Chuang. "Quantum Computation and Quantum Information." (2010).
30. Sekerka, R. F. "Thermal Physics: Thermodynamics and Statistical Mechanics for Scientists and Engineers." Elsevier, 2015.
31. Spalvieri, Arnaldo. "Infomechanics of Independent and Identically Distributed Particles." arXiv preprint arXiv:2204.06952 (2022).
32. Spalvieri, Arnaldo. "The Exact Entropy Formula of the Ideal Gas and its Information-Theoretic Interpretation." arXiv preprint arXiv:2305.04652 (2023).
33. Landauer, Rolf. "Irreversibility and heat generation in the computing process." *IBM journal of research and development* 5.3 (1961): 183-191.
34. Aydin, Alhun, Altug Sisman, and Ronnie Kosloff. "Landauer's Principle in a quantum Szilard engine without Maxwell's Demon." *Entropy* 22.3 (2020): 294.
35. Jaynes, Edwin T. "The Gibbs paradox." *Maximum entropy and bayesian methods*. Springer, Dordrecht, 1992. 1-21.
36. Goldstein, Sheldon, et al. "Gibbs and Boltzmann entropy in classical and quantum mechanics." *Statistical mechanics and scientific explanation: Determinism, indeterminism and laws of nature*. 2020. 519-581.

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.