# Canonical Typicality

Sheldon Goldstein,[1,*] Joel L. Lebowitz,[1,†] Roderich Tumulka,[2,‡] and Nino Zanghì[3,§]

[1]*Departments of Mathematics and Physics, Hill Center, Rutgers, The State University of New Jersey,
110 Frelinghuysen Road, Piscataway, New Jersey 08854-8019, USA*
[2]*Mathematisches Institut, Eberhard-Karls-Universität, Auf der Morgenstelle 10, 72076 Tübingen, Germany*
[3]*Dipartimento di Fisica dell'Università di Genova and INFN sezione di Genova, Via Dodecaneso 33, 16146 Genova, Italy*

It is well known that a system *S* weakly coupled to a heat bath *B* is described by the canonical ensemble when the composite *S + B* is described by the microcanonical ensemble corresponding to a suitable energy shell. This is true for both classical distributions on the phase space and quantum density matrices. Here we show that a much stronger statement holds for quantum systems. Even if the state of the composite corresponds to a single wave function rather than a mixture, the reduced density matrix of the system is canonical, for the overwhelming majority of wave functions in the subspace corresponding to the energy interval encompassed by the microcanonical ensemble. This clarifies, expands, and justifies remarks made by Schrödinger in 1952.

PACS numbers: 05.30.Ch, 03.65.−w, 05.20.Gg

A quantum system in thermal equilibrium at inverse temperature $\beta$ is described by the canonical density matrix

$$\rho_\beta = \frac{1}{Z} \exp(-\beta H^{(S)}), \qquad (1)$$

where $H^{(S)}$ is the system Hamiltonian and

$$Z = \text{tr} \exp(-\beta H^{(S)}).$$

The usual justification for (1) is that it is the reduced density matrix of the system when it is weakly coupled to a heat bath, and the composite system is described by the microcanonical density matrix at a suitable total energy $E$.

More explicitly, one assumes that it is permissible to neglect the relatively small interaction between the system and the bath, so that the total Hamiltonian of the composite $S + B$ is given by

$$H = H^{(S+B)} = H^{(S)} + H^{(B)}. \qquad (2)$$

The composite $S + B$ is then assumed to be represented by a microcanonical ensemble in some energy interval $[E, E + \delta]$, where $\delta$ is small on the macroscopic scale, $\delta \ll E$, but large enough for the interval to contain very many eigenvalues. The corresponding microcanonical density matrix is

$$\rho_{E,\delta} = (\dim \mathcal{H}_{[E,E+\delta]})^{-1} P_{\mathcal{H}_{[E,E+\delta]}}, \qquad (3)$$

with $P_{\mathcal{H}_{[E,E+\delta]}}$ the projection to $\mathcal{H}_{[E,E+\delta]}$, the spectral subspace for $H$ associated with energies in the interval $[E, E + \delta]$ in the Hilbert space $\mathcal{H} = \mathcal{H}^{(S+B)} = \mathcal{H}^{(S)} \otimes \mathcal{H}^{(B)}$. One readily proves (see below) that in the thermodynamic limit, when the size, i.e., the number of components $N$ of the heat bath, goes to infinity while $E/N \to e$, the reduced density matrix of the system $S$,

$$\rho^{(S)} = \text{tr}^{(B)} \rho_{E,\delta}, \qquad (4)$$

is equal to $\rho_\beta$. Here $\text{tr}^{(B)}$ denotes the partial trace over $\mathcal{H}^{(B)}$ and $\beta = \beta(e)$.

In this Letter we show how this result can be substantially strengthened: we prove that, in the thermodynamic limit, the reduced density matrices of the overwhelming majority of the wave functions of $S + B$ are canonical. We call this statement *canonical typicality*. As a consequence of canonical typicality, it follows that the canonical ensemble is even more inevitable in quantum mechanics— arising as it does there without the invocation of any genuine randomness—than it is classically. Results in this direction were first obtained by Schrödinger [1], and later by Gemmer and Mahler [2]; related results have been obtained by Tasaki [3] (see later).

Typicality in quantum mechanics, as well as in classical mechanics, involves a probability distribution on the possible microstates of the system, the distribution in terms of which "overwhelming majority" is to be understood. In classical mechanics these microstates are points in the appropriate phase space and the distribution is then a measure on this phase space. To define typicality for quantum systems we shall take the microstates to be wave functions, i.e., points on the unit sphere of $\mathcal{H}$ (up to a phase). Even with this identification it may not be clear which distribution is appropriate for the (composite) system described by the density matrix $\rho_{E,\delta}$. Here we take that to be the probability distribution proposed long ago by Schrödinger [1,4] and Bloch [5]: it is the (normalized) uniform (surface area) measure $u_{E,\delta}$ on the unit sphere in the subspace $\mathcal{H}_{[E,E+\delta]}$, i.e., the uniform probability distribution over all normalized wave functions $\Psi$ with energies in $[E, E + \delta]$. If we expand $\Psi$ in terms of energy eigenfunctions $|E_\alpha\rangle$ of $H$, $\Psi = \sum c_\alpha |E_\alpha\rangle$, where the sum is restricted to levels with energies in the interval $[E, E + \delta]$, then $u_{E,\delta}$ corresponds to the uniform distribution on the surface of the sphere $\sum |c_\alpha|^2 = 1$. This measure was

shown in [5] to be, in a well-defined sense, the most appropriate distribution corresponding to the microcanonical density matrix, i.e., such that

$$\rho_{E,\delta} = \int u_{E,\delta}(d\Psi)|\Psi\rangle\langle\Psi|. \tag{5}$$

Let $\rho^{\Psi}$ denote the reduced density matrix of the system, given that $S + B$ is in a pure state $\Psi \in \mathcal{H}_{[E,E+\delta]}$,

$$\rho^{\Psi} = \mathrm{tr}^{(B)}|\Psi\rangle\langle\Psi|. \tag{6}$$

We may then ask this: for which wave functions $\Psi$ is $\rho^{\Psi}$ (approximately) of the canonical form (1)? We make the standard assumption that both $H^{(S)}$ and $H^{(B)}$ have pure point spectrum and are bounded from below. In $\mathcal{H}^{(S)}$ ($\mathcal{H}^{(B)}$) we choose an eigenbasis of $H^{(S)}$ ($H^{(B)}$), denoted $|E_1^{(S)}\rangle, |E_2^{(S)}\rangle, \ldots$ ($|E_1^{(B)}\rangle, |E_2^{(B)}\rangle, \ldots$), with corresponding eigenvalues $E_1^{(S)} \leq E_2^{(S)} \leq \cdots$ ($E_1^{(B)} \leq E_2^{(B)} \leq \cdots$).

First, we note the following: saying that for the majority of $\Psi$'s, $\rho^{\Psi}$ is close to $\rho_{\beta}$ is equivalent to saying that if $\Psi$ is *randomly* chosen with distribution $u_{E,\delta}$ then with overwhelming *probability* $\rho^{\Psi}$ is close to $\rho_{\beta}$. From now on we will thus regard $\Psi$ as a (Hilbert-space-valued) random variable.

We begin by recalling the standard derivation of the canonical ensemble from the microcanonical. One has for the reduced density matrix $\rho^{(S)}$ of the system that

$$\rho^{(S)} = (\dim\mathcal{H}_{[E,E+\delta]})^{-1}\sum_i \dim(\mathcal{H}_i^{(B)})|E_i^{(S)}\rangle\langle E_i^{(S)}|, \tag{7}$$

where

$$\mathcal{H}_i^{(B)} = \mathcal{H}_{[E-E_i^{(S)}, E-E_i^{(S)}+\delta]}^{(B)} \tag{8}$$

is the spectral subspace for $H^{(B)}$ associated with energies in the interval $[E - E_i^{(S)}, E - E_i^{(S)} + \delta]$. It is then more or less clear, and can be rigorously proven under suitable conditions, that when the bath is sufficiently large $\rho^{(S)} \approx \rho_{\beta}$, with $\beta = dS(E)/dE$ where $S(E)$ is the bath's entropy. This follows from the basic fact that $S(E) \approx \log \dim(\mathcal{H}_{[E,E+\delta]}^{(B)})$, so that

$$\dim(\mathcal{H}_i^{(B)}) \approx e^{S(E-E_i^{(S)})} \approx e^{S(E)-\beta E_i^{(S)}} \sim e^{-\beta E_i^{(S)}}. \tag{9}$$

More precisely, one proves that $\rho^{(S)} \to \rho_{\beta}$ in the thermodynamic limit with $\beta = ds(e)/de$, where $s(e) = \lim[S(E)/N]$ and $e = \lim[E/N]$.

Thus to demonstrate canonical typicality it suffices to establish that (7) holds, at least approximately, when $\rho^{(S)}$ is replaced by $\rho^{\Psi}$ for typical $\Psi \in \mathcal{H}_{[E,E+\delta]}$. A key step of our argument is to note that the uniformly distributed random vector $\Psi$ can always be regarded as arising by normalization

$$\Psi = \frac{\Phi}{\parallel \Phi \parallel} \tag{10}$$

from a Gaussian random vector $\Phi \in \mathcal{H}_{[E,E+\delta]}$ with mean zero and covariance given by the identity operator on $\mathcal{H}_{[E,E+\delta]}$. This means that in the decomposition

$$\Phi = \sum_i \sum_j C_{ij}|E_i^{(S)}\rangle|E_j^{(B)}\rangle, \tag{11}$$

the real and imaginary parts, $\mathrm{Re}C_{ij}$ and $\mathrm{Im}C_{ij}$, of the coefficients are independent real Gaussian random variables with mean zero and variance $1/2$ for those $i$ and $j$ for which $E_i^{(S)} + E_j^{(B)} \in [E, E + \delta]$ (and $C_{ij} = 0$ otherwise). We obtain from (11) that

$$\Phi = \sum_i |E_i^{(S)}\rangle|\Phi_i\rangle \tag{12}$$

with

$$|\Phi_i\rangle = \sum_{j\in I_i} C_{ij}|E_j^{(B)}\rangle, \tag{13}$$

where $I_i$ is the set of bath levels $j$ such that $E_j^{(B)} \in [E - E_i^{(S)}, E - E_i^{(S)} + \delta]$, whence the reduced density matrix (6) is of the form

$$\rho^{\Psi} = \frac{1}{\parallel \Phi \parallel^2}\mathrm{tr}^{(B)}|\Phi\rangle\langle\Phi|$$

$$= \frac{1}{\parallel \Phi \parallel^2}\sum_{i,i'}\langle\Phi_i|\Phi_{i'}\rangle|E_i^{(S)}\rangle\langle E_{i'}^{(S)}|. \tag{14}$$

Now, if $\delta$ were so small that the system's energy spacings $\Delta E_i^{(S)} = E_{i+1}^{(S)} - E_i^{(S)}$ are all greater than $\delta$, then the relevant energy intervals $I_i$ for the heat bath would be pairwise disjoint and the $\Phi_i$ pairwise orthogonal,

$$\langle\Phi_i|\Phi_{i'}\rangle = \delta_{ii'} \parallel \Phi_i \parallel^2. \tag{15}$$

We argue now that Eq. (15) will continue to hold, at least approximately, even without the above assumption on $\Delta E_i^{(S)}$. This is so because, when $I_i$ and $I_{i'}$ have significant overlap, the contributions to $\Phi_i$ and $\Phi_{i'}$ corresponding to the sum over the terms in (13) belonging to both $I_i$ and $I_{i'}$ will typically be approximately orthogonal. To see this, note that these sums form two independent random vectors, with uniformly distributed directions, in a high-dimensional space. As such, they are, with probability approaching unity, nearly orthogonal since the expected value of the absolute square of the dot product of two independent random unit vectors in an $n$-dimensional space is, by symmetry, $1/n$.

From the representation (13) we have that

$$\parallel \Phi_i \parallel^2 = \sum_j |C_{ij}|^2, \tag{16}$$

so that $\parallel \Phi_i \parallel^2$ is the sum of the $N_i = \dim(\mathcal{H}_i^{(B)})$ inde-

pendent, identically distributed random variables $|C_{ij}|^2$ with mean 1. It thus follows, by the law of large numbers, that typically

$$\| \Phi_i \|^2 \approx \dim(\mathcal{H}_i^{(B)}). \qquad (17)$$

We thus have for the reduced density matrix (6), using (14), that typically

$$\rho^\Psi \approx (\dim \mathcal{H}_{[E,E+\delta]})^{-1} \sum_i \dim(\mathcal{H}_i^{(B)}) |E_i^{(S)}\rangle\langle E_i^{(S)}|, \qquad (18)$$

which is what we needed to show.

Concerning (18), we remark that it follows merely from the fact that the reduced density matrix $\rho^\Psi$ does not depend upon $\Psi$ for typical $\Psi \in \mathcal{H}_{E,\delta}$, that whenever the reduced microcanonical density matrix $\rho^{(S)} \approx \rho_\beta$, the same is true for $\rho^\Psi$ for typical $\Psi$: applying the partial trace $\text{tr}^{(B)}$ to (5), one obtains that $\rho^{(S)} = \int u_{E,\delta}(d\Psi)\rho^\Psi \approx \rho^\Psi$, for typical $\Psi$.

Some essential parts of the argument we have presented here have already been described by Schrödinger in an appendix, written in 1952, to his book on statistical thermodynamics [1]. We note that Schrödinger in [1] made the assumption that "in a big system ... the amplitude-squares ... are on the average [in time] equal for ... eigenfunctions belonging to the same [energy] eigenvalue." He uses this assumption for the combined system $S + B$ to derive "exactly the same canonical distribution between the amplitude-squares, as is in the customary treatment said to indicate the probability of the system being on this or that energy level." However, Schrödinger neither connects the assumption with typicality, nor his conclusion with the reduced density matrix $\rho^\Psi$, which he does not even mention. His concern is rather with showing that one need not regard a system in thermal equilibrium as being in an energy eigenstate. As he states in his preface, "To ascribe to every system always one of its sharp energy values is an indefensible attitude."

Tasaki [3] has studied, as we do here, the reduced density matrix of a system coupled to a heat bath when $S + B$ is described by a wave function $\Psi$. He shows that for a special form of the coupling Hamiltonian the long-time average of $\rho^{\Psi(t)}$ is canonical. He then argues in a heuristic way that also for typical large times, $\rho^{\Psi(t)} \approx \rho_\beta$. [In this argument, there is a hidden typicality assumption

on the initial wave function $\Psi(0)$, namely, that all of the energy expansion coefficients are small, which is true of most wave functions.] His argument does not yield our stronger statement that $\rho^{\Psi(t)} \approx \rho_\beta$ even at $t = 0$ for typical wave functions. Tasaki also includes some examples that he studied rigorously, concerning energy eigenstates of the composite or states that are initially product states $\Psi^{(S)} \otimes \Psi^{(B)}$ with $\Psi^{(S)}$ an eigenstate of $H^{(S)}$.

More recently, Gemmer and Mahler [2] have established canonical typicality under the assumption of very large degeneracy for energy eigenstates by computing appropriate Hilbert space volumes.

*Note added.*—In a very recent paper, Popescu, Short, and Winter have established canonical typicality under great generality by invoking Levy's Lemma [6].

---

*Electronic address: oldstein@math.rutgers.edu
†Electronic address: lebowitz@math.rutgers.edu
‡Electronic address: tumulka@everest.mathematik.uni-tuebingen.de
§Electronic address: zanghi@ge.infn.it

[1] E. Schrödinger, *Statistical Thermodynamics* (Cambridge University Press, Cambridge, England, 1952).
[2] J. Gemmer and G. Mahler, Eur. Phys. J. B **31**, 249 (2003).
[3] H. Tasaki, Phys. Rev. Lett. **80**, 1373 (1998).
[4] E. Schrödinger, Ann. Phys. (Berlin) **83**, 956 (1927).
[5] John Dirk Walecka, *Fundamentals of Statistical Mechanics. Manuscript and Notes of Felix Bloch* (Stanford University Press, Stanford, CA, 1989).
[6] S. Popescu, A. J. Short, and A. Winter, quant-ph/0511225.