Available online at:
www.4open-sciences.org

RESEARCH ARTICLE

OPEN ⌀ ACCESS

Logical entropy and negative probabilities in quantum mechanics

Giovanni Manfredi[*] (ORCID)

Université de Strasbourg, CNRS, Institut de Physique et Chimie des Matériaux de Strasbourg, UMR 7504, 67000 Strasbourg, France

Abstract − The concept of Logical Entropy, $S_\mathrm{L} = 1 - \sum_{i=1}^n p_i^2$, where the p_i are normalized probabilities, was introduced by David Ellerman in a series of recent papers. Although the mathematical formula itself is not new, Ellerman provided a sound probabilistic interpretation of S_L as a measure of the distinctions of a partition on a given set. The same formula comes across as a useful definition of entropy in quantum mechanics, where it is linked to the notion of purity of a quantum state. The quadratic form of the logical entropy lends itself to a generalization of the probabilities that include negative values, an idea that goes back to Feynman and Wigner. Here, we analyze and reinterpret negative probabilities in the light of the concept of logical entropy. Several intriguing quantum-like properties of the logical entropy are derived and discussed in finite dimensional spaces. For infinite-dimensional spaces (continuum), we show that, under the sole hypothesis that the logical entropy and the total probability are preserved in time, one obtains an evolution equation for the probability density that is basically identical to the quantum evolution of the Wigner function in phase space, at least when one considers only the momentum variable. This result suggest that the logical entropy plays a profound role in establishing the peculiar rules of quantum physics.

Keywords: Entropy, Negative Probabilities, Quantum Physics, Wigner Functions

1 Introduction

As its title suggests, this work sits at the crossroad of three different topics: (a) an alternative definition of entropy, (b) the extension of standard probabilities to negative values, and (c) the relevance of the first two items to our understanding of quantum mechanics. Here, we will introduce each topic separately, before bringing them together in the following sections.

1.1 Logical entropy

"Logical entropy" is a concept introduced by David Ellerman in a series of works spanning the last decade [1, 2]; see also Ellerman's paper in this Special Issue. Succinctly, logical entropy is based on the concept of *distinctions*. If a certain set U is partitioned into a number n of subsets B_i (such that $\cup_{i=1}^n B_i = U$), each endowed with a probability p_i of finding an element of U in that subset, then the probability that in two independent draws one will obtain elements in distinct subsets B_i and $B_{j\neq i}$ is: $p_i(1 - p_i)$. This is precisely the concept of *distinction*, i.e., the ability establish that two independent draws are different from one another.

Summing over all n subsets, we obtain the total number of distinctions, which is the definition of the logical entropy S_L:

$$S_\mathrm{L} = \sum_{i=1}^n p_i(1 - p_i) = 1 - \sum_{i=1}^n p_i^2, \qquad (1)$$

where we used the fact that $\sum_i p_i = 1$. The subsets B_i can possibly contain one single element, in which case S_L represents the probability that two consecutive draws yield different elements of U. In this work, we will mainly consider this case, unless otherwise stated. It is clear that $0 \leq S_\mathrm{L} \leq 1$. The lower bound is reached when one element has probability $p_i = 1$, while for all others $p_{j\neq i} = 0$. For equal probabilities ($p_i = \frac{1}{n}, \forall i$), one gets: $S_\mathrm{L} = 1 - \frac{1}{n} \to 1$, when $n \to \infty$.

*Corresponding author: giovanni.manfredi@ipcms.unistra.fr

Following [3, 4], one can also define the information I as the complement of the entropy to unity:

$$I = 1 - S_{\mathrm{L}} = \sum_{i=1}^{n} p_i^2. \tag{2}$$

This quantity reflects the knowledge we have of the state of a physical system, being maximum when we know its state with certainty, and minimum when all states are equally probable.[1] The information I has the nice property of being the square of a norm in $\mathbb{R}^n$, actually the Euclidean norm. This connection to Euclidean geometry allows one to use standard geometrical concepts when making use of the logical entropy. For instance, one can define the scalar product: $p \cdot q = \sum_{i=1}^{n} p_i q_i$ between two probability distributions $\{p_i\}$, $\{q_i\}$, and their Euclidean distance $d(p, q)$ as:

$$d^2(p,q) = \sum_{i=1}^{n} \left(p_i - q_i\right)^2. \tag{3}$$

Of course, the logical entropy definition (1) implies very different properties from the standard Shannon–Von Neumann entropy

$$S_{\mathrm{VN}} = - \sum_{i=1}^{n} p_i \log p_i. \tag{4}$$

In particular S_{VN} is additive, while S_{L} is not, at least not in the standard fashion, see [4, 5]. For a system known with certainty, both entropies yield, $S_{\mathrm{VN}} = S_{\mathrm{L}} = 0$, but for maximal uncertainty $S_{\mathrm{VN}} = \log n$, whereas $S_{\mathrm{L}} = 1 - \frac{1}{n}$.

Again we emphasize that, in contrast to the Shannon–Von Neumann entropy, the logical entropy S_{L} represents both a *probability* (of obtaining different results in two consecutive draws, as mentioned above) and a *norm* in the Euclidean space $\mathbb{R}^n$. These facts have important consequences, as we will see in the next section.

Although Ellerman [1, 2] provided a solid and fruitful probabilistic interpretation of this definition of entropy, the formulae (1) and (2) are not new. Quite the contrary, they have been discovered and rediscovered many times in the past, in very different areas of research. In biology and ecology, S_{L} is known as the Gini–Simpson index [6–8], which quantifies the diversity of species in an ecosystem. It was used by Polish mathematicians (and then by Alan Turing himself) to find patterns in messages generated by the Enigma machine during World War 2 [9]. In statistical mechanics, S_{L} is a special case of the Tsallis entropy [10] with index $q = 2$. In quantum physics, a version of S_{L} was used to quantify our knowledge of the state of a quantum system [3, 11]. It was also shown to be particularly adapted to the Wigner phase-space representation of quantum mechanics [4].

1.2 Negative probabilities

The very definitions of S_{L} and I lend themselves to the natural generalization whereby the probabilities p_i can take negative values. This is in analogy with vectors in $\mathbb{R}^n$, which can indeed have negative components, although their norm remains positive.

Negative probabilities have a long history of interest, especially among physicists struggling to make sense of some of the weird properties of quantum mechanics. Feynman [12] was one of the first to ponder the meaning of negative probabilities in a quantum context (although he published his ideas in 1987 in a volume in honor of David Bohm, he states there that he developed these reflections some twenty years earlier). For Feynman, negative probabilities should be considered as a useful bookkeeping tool just like negative numbers.[2] As an example, he mentions a man starting a day with five apples, giving away ten at midday and earning eight in the evening. The initial (5) and final (3) numbers of apples owned by the man are both positive and thus unambiguous to interpret. But if we take the numbers at face value, the man will have -5 apples some time in the afternoon, which does not quite make sense unless we postulate that one is allowed to count the number of apples only in the morning and in the evening, but not in the middle of the day. Hence, negative probabilities are allowed as long as they intervene in contexts where they cannot be observed directly. All this is reminiscent of the limitations on measuring some quantities, which are intrinsic to quantum physics [13, 14].

Of course, negative probabilities had appeared in quantum mechanics even earlier, when Wigner [15] introduced his celebrated pseudo-probability distribution in the classical phase space ("Wigner function"), which almost always takes negative values. Indeed, the negativity of a Wigner function can be used as a tool to quantify the degree of quantumness of a particular state, as was done even experimentally [16].

---

[1] As an aside, we note that the idea of information as distinctions (differences, distinguishability, and diversity) would take the higher logical entropy states as making more distinctions or showing more diversity and distinguishability between the outcomes. In that sense, higher logical entropy states may be thought as having *more*, rather than less, information. But here we stick to the definition of information as presented in the main text, which is the way it is usually interpreted in physics.

[2] The need for negative numbers can be circumvented through the trick of double-entry bookkeeping, see [33].

Negative probabilities have also been studied in a fundamental mathematical context [17–20] and for applications to financial modeling [21]. A thorough, if not very recent, review on the topic of negative probabilities in physics was published in 1986 [22], and contains quotations from several eminent scientists on this somewhat controversioal problem.

1.3 Quantum mechanics

The earliest relationship between negative probabilities and quantum mechanics dates back to Wigner [15], who in 1932 introduced a pseudo-probability distribution in the phase space (x, p) which possesses many of the properties of classical probability distributions (for instance, it can be used to compute averages using the classical formula), *except non-negativity*. The Wigner function $w(x, p, t)$ can describe both pure and mixed quantum states and evolves in time according to an integro-differential equation similar to the classical Liouville equation. Wigner functions have proven exceedingly useful in a variety of domains, ranging from condensed matter and nanophysics, to quantum plasmas and quantum optics (see [23] for a review).

The Wigner equation conserves in time not only the total probability $\int \int w(x, p, t)\mathrm{d}x\,\mathrm{d}p$, but also the integral of the square of the Wigner function: $\int \int w^2(x, p, t)\mathrm{d}x\,\mathrm{d}p$. Note that higher powers $\int \int w^r \mathrm{d}x\,\mathrm{d}p$, with $r > 2$, are not conserved, in contrast to the classical Liouville equation, for which the conservation property is valid for any value of r. Some time ago, the present author suggested that one uses

$$S_L = 1 - I = 1 - h \iint w^2 \mathrm{d}x\,\mathrm{d}p \tag{5}$$

as the definitions of entropy and information [4], where h is Planck's constant (this is necessary to render the integral term in the above expression non-dimensional). Equation (5) can be viewed as the continuous counterpart of equation (1), i.e., its extension to an infinite dimensional space. Also note that the logical entropy can be expressed in terms of the trace of the density operator, as $S_L = 1 - \mathrm{Tr}(\hat{\rho}^2)$.

More recently, negative probabilities have been explored in various quantum mechanical contexts, such as indistiguishability [24], quantum computation [25], and contextuality [26]. Besides, an operational interpretation of negative probabilities has been proposed by Abramsky and Brandenburger [27, 28]. In [28], they propose a simple scenario to illustrate pedagogically the use of negative probabilities in quantum mechanics, by considering a system comprising two-bit registers.

The rest of this work is devoted to the study of the properties of the logical entropy (1) and information (2) when one relaxes the requirement that $p_i \geq 0$, $\forall i$. It will be claimed that the logical entropy constitutes the natural framework for the introduction of negative probabilities. Interestingly, by combining the definition of logical entropy with negative probabilities, one can recover many properties that are typical of quantum systems.

The main result obtained here is that, simply by requiring the logical entropy to be conserved in time, one obtains an evolution equation for the probability density that is virtually identical to the evolution equation of the Wigner function in physics, at least when one considers only the momentum variable. This remarkable result suggest that the logical entropy plays a profound role in establishing the peculiar rules of quantum physics.

2 Finite-dimensional spaces

We consider a set of n outcomes, each endowed with probability p_i. The probabilities satisfy

$$\sum_{i=1}^{n} p_i = 1, \tag{6}$$

$$\sum_{i=1}^{n} p_i^2 = R^2, \tag{7}$$

where $0 \leq R \leq 1$. Then the logical entropy and information are, respectively, $S_L = 1 - R^2$ and $I = R^2$. The number R can be interpreted as the Euclidean norm of the vector $\boldsymbol{p} = (p_1, \dots p_n)$ in $\mathbb{R}^n$: $||\boldsymbol{p}|| = R$. Geometrically, equations (6) and (7) represent respectively a hyperplane and a hypersphere of radius R in $\mathbb{R}^n$, and their intersection yields the probability distributions $\{p_i\}$ satisfying those equations.

In analogy with quantum physics, we shall call *pure states* the probability distributions for which $R = 1$ (corresponding to maximum information and minimum entropy) and *mixed states* those for which $R < 1$. Indeed, Wigner functions for pure and mixed quantum states satisfy precisely these properties, when the entropy is defined as in equation (5). If we request all probabilities to be nonnegative, then the only pure states are those for which $p_i = 1$ and $p_{j \neq i} = 0$, that is, the ith outcome can be predicted with certainty. However, if we admit negative probabilities, there exist other pure states with some $p_i < 0$ which still satisfy equations (6) and (7) with $R = 1$.
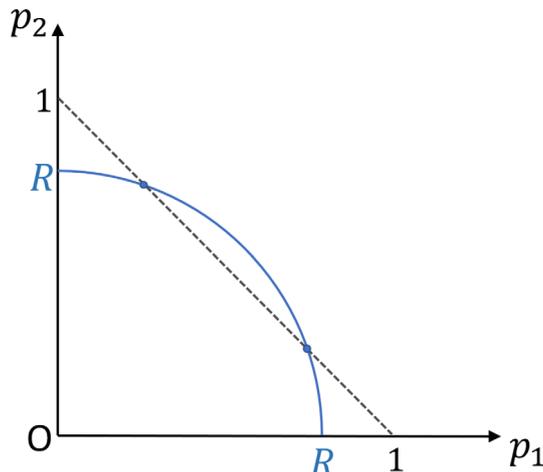
Figure 1. Schematic representation of the case $n = 2$. The total probability constraint (6) is represented by the dashed straight line, while the entropy constraint (7) is represented by the blue quarter circle of radius R. Solutions are given by their intersections.

To dissipate all ambiguities, here we are not dealing with "probability amplitudes" as in quantum mechanics. Probability amplitudes are complex quantities, while our p_is are real numbers, albeit potentially negative. Our approach is the same as the one based on Wigner functions (also real quantities), which represent quantum states with real, but signed, numbers.

In the rest of the present section, we will focus on the cases $n = 2$, which is trivial and does not admit negative probabilities, and $n = 3$, which is much richer. The infinite-dimensional case will be treated in Section 3.

2.1 General properties for $n = 2$ and $n = 3$

For $n = 2$, the solution is given by the intersection of the straight line and the circle shown in Figure 1. It is clear that for $R \leq 1$, only positive solutions are allowed. Solving equations (6) and (7) yields $p_{1,2} = \frac{1 \pm \sqrt{2R^2 - 1}}{2}$. No solutions exist for $R < \sqrt{2}/2$ (dashed straight line tangent to the circle). For this value of R, one obtains $p_1 = p_2 = 1/2$, which is the maximally mixed state (with largest entropy $S_{\mathrm{L}} = 1/2$).

The case $n = 3$ is depicted schematically in Figure 2a for the special case $R = 1$ (pure states). It is evident that there are three pure states with nonnegative probabilities: $(1, 0, 0)$, $(0, 1, 0)$ and $(0, 0, 1)$, which represent certainty for one of the three possible outcomes. These states form an orthonormal basis which we denote by e_i. However, there exist an infinity of other pure states with negative probabilities. These are the states that lie on the circle given by the intersection between the sphere of radius R and the plane π defined by the three vectors e_i. Actually, *all* pure states, except e_1, e_2 and e_3, feature some negative probabilities. A simple example is the state: $p = (2/3, 2/3, -1/3)$.

A view of the plane π is shown in Figure 2b. The circles represent the intersections of the plane and the sphere, for different values of the radius R. Points that lie outside the equilateral triangle ABC (with sides $a = \sqrt{2}$) have negative probabilities. The circumscribed circle, corresponding to $R = 1$, has radius $r_e = a/\sqrt{3} = \sqrt{2/3}$.

For information $I = R^2$ smaller than unity, i.e. for mixed states, there are some positive and negative solutions (thin red circle in Fig. 2b). Further decreasing R, we reach the situation of the inner circle of radius $r_i = r_e/2 = \sqrt{6}/6$, for which all probabilities are positive. To determine the value of R corresponding to r_i, we consider the cone of vertex O and base radius r_i (see Fig. 2c). The height k of the cone is the distance between the origin O and the plane π, which turns out to be $k = 1/\sqrt{3}$. From this, we deduce that the radius R corresponding to the inner circle in Figure 2b is $R = 1/\sqrt{2}$. Finally, for $R = k = 1/\sqrt{3}$, the sphere is tangent to the plane π, and the only solution is $p_1 = p_2 = p_3 = 1/3$, corresponding to maximum entropy $S_{\mathrm{L}} = 2/3$. For smaller R, there are no solutions.

In summary, defining the radii $R_{\max} = 1$, $R_{\mathrm{pos}} = 1/\sqrt{2}$, and $R_{\min} = 1/\sqrt{3}$, we obtain that:

- For $R_{\mathrm{pos}} < R \leq R_{\max}$, there exist some negative-probability solutions;
- For $R_{\min} \leq R \leq R_{\mathrm{pos}}$, there exist only positive-probability solutions;
- For $R = R_{\min}$: maximum entropy solution $p_1 = p_2 = p_3 = 1/3$;
- For $R < R_{\min}$, there exist no solutions.

The above considerations can be easily extended to $n > 3$, yielding: $R_{\max} = 1$, $R_{\mathrm{pos}} = 1/\sqrt{n-1}$, and $R_{\min} = 1/\sqrt{n}$, with maximum entropy solution: $p_i = 1/n$, $\forall i$. We note that for large n, one has $R_{\min} \approx R_{\mathrm{pos}} \sim 1/\sqrt{n}$. Therefore, almost all existing solutions will display some negative values.
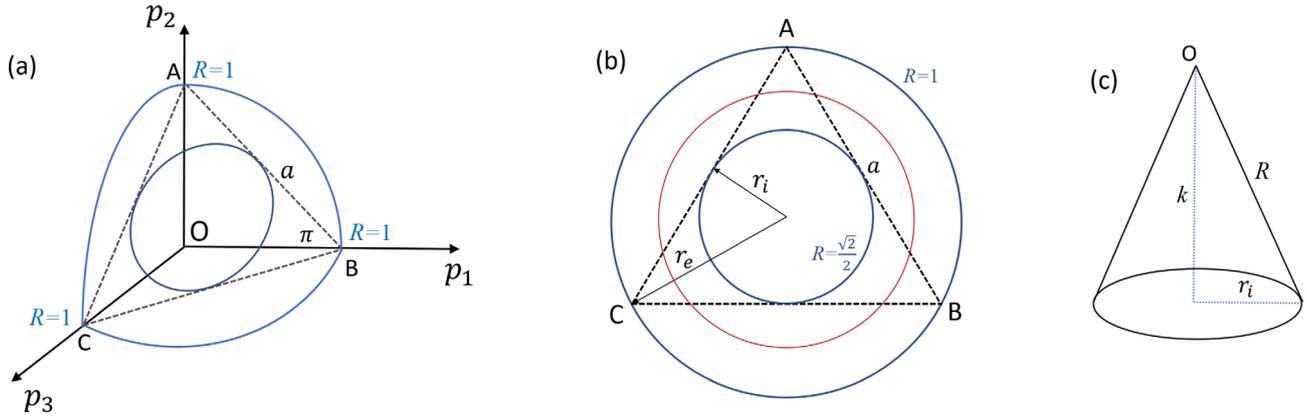
Figure 2. Schematic representation of the case $n = 3$. (a) The solutions of equations (6) and (7) lie on the circle given by the intersection of the sphere of radius R (here represented for the pure state with $R = 1$) and the plane π passing through the points A, B and C. The dashed triangle ABC, lying on the plane π, has all sides equal to $a = \sqrt{2}$. The inner circle is inscribed into the triangle. (b) View of the plane π with the circumscribed and inscribed circles of radii r_e and r_i, which correspond, respectively, to values $R = 1$ and $R = 1/\sqrt{2}$. The thin red circle is an intermediate case where positive and negative probabilities coexist (the latter lie outside the triangle). (c) Circular cone with apex at the origin O and basis circle of radius r_i. The height k of the cone is the distance between the origin and the plane π.

From Figure 2, it is evident that, for pure states, the most negative value of p_i is reached when two probabilities are identical and positive, and the third one is negative, i.e., $p_1 = p_2 = p$ and $p_2 = -q$. Direct computation yields the result $p = 2/3$ and $q = 1/3$. The three vectors: $\boldsymbol{u}_1 \equiv (2/3, 2/3, -1/3)$, $\boldsymbol{u}_2 \equiv (2/3, -1/3, 2/3)$ and $\boldsymbol{u}_3 \equiv (-1/3, 2/3, 2/3)$ also constitute an orthonormal basis in $\mathbb{R}^3$.

This reasoning can be extended to n dimensions, yielding $\mathrm{p}_1 = \ldots = p_{n-1} = 2/n$ and $p_n = (2 - n)/n$. From this, one can construct an orthonormal basis $\{\boldsymbol{u}_1 \ldots \boldsymbol{u}_n\}$. For instance, for $n = 4$ one gets $p_1 = p_2 = p_3 = 1/2$ and $p_4 = -1/2$.

Finally, we stress that, from this simple example with $n = 3$, negative probabilities arise very naturally if the p_i are requested to satisfy the two equations (6) and (7), which fix the total probability and total entropy (or information) of the system. Indeed, for a mixed state such as described by the thin red circle in Figure 2b, it would be odd to retain only the positive-probability solutions (inside the triangle) and discard the negative ones (outside the triangle). Hence, the entropy definition (7) calls for the acceptance of negative probabilities on the same footing as positive ones.

2.2 Maximization with constraints

We would like to maximize the entropy S_{L} (minimize the information) with a constraint. This is analogous to the statistical mechanics problem of finding the equilibrium probability distribution that maximizes entropy for given energy, which yields the Maxwellian distribution if one uses the Shannon–Von Neumann entropy. Let us call X our constraint, which has the mean value $m \equiv \langle X \rangle = \sum_i p_i X_i$. The functional F to be minimized is given by the information I augmented by two constraints on the total probability and the average of X:

$$F = \sum_i p_i^2 - \lambda \sum_i p_i + \mu \sum_i p_i X_i, \tag{8}$$

where λ and μ are Lagrange multipliers. Setting the variation of F to zero, i.e.:

$$\delta F = 2 \sum_i p_i \delta p_i - \lambda \sum_i \delta p_i + \mu \sum_i X_i \delta p_i = 0,$$

one gets

$$p_i = \frac{\lambda - \mu X_i}{2}. \tag{9}$$

The Lagrange multipliers are determined by using the constraints: $\sum_i p_i = 1$ and $\sum_i p_i X_i = m$.

As an example, we take again $n = 3$ and $X_i = (-1, 0, 1)$. This choice yields $\lambda = 2/3$, $\mu = -m$, and the "equilibrium" probability distribution:

$$p = \left(\frac{1}{3} - \frac{m}{2}, \frac{1}{3}, \frac{1}{3} + \frac{m}{2} \right). \tag{10}$$
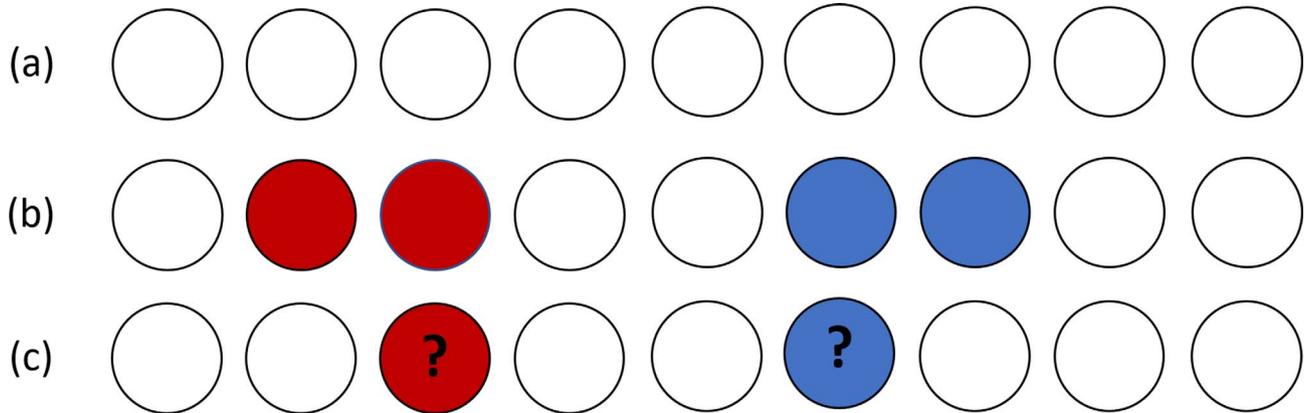
Figure 3. (a) Nine particles are drawn from a probability distribution p, corresponding to a pure state, but they are not yet observed. (b) We look at particles 2–3, which turn out to be both red, and then look at particles 6–7, which turn out to be both blue. (c) Had we drawn particles 3–6, we would have expected them to be of same color, but this is in contradiction with the "experiment" of row (b).

The total information is $I = R^2 = \frac{1}{3} + \frac{m^2}{2}$. As it must be smaller or equal to unity, we have a constraint on the maximum mean value allowed for the variable X: $m \leq \frac{2}{\sqrt{3}} \equiv m_{\max} \approx 1.15$. For $m = m_{\max}$, we obtain the pure state

$$p = \left(1 - \frac{\sqrt{3}}{3}, \frac{1}{3}, \frac{1+\sqrt{3}}{3}\right), \tag{11}$$

for which $p_1 < 0$. Indeed, p_1 is negative whenever $2/3 < m < m_{\max}$. For $m < 2/3$ all probabilities become positive and for $m = 0$ we recover the maximally mixed state with all probabilities equal to $1/3$. Similar considerations apply for the symmetric cases with negative m.

The above situation can be viewed as that of a die with three faces. For $m = 0$ the die is even, and all faces are equally probable. Hence, $|m|$ may be interpreted as an index of unevenness of the die. Classically, i.e. only allowing positive probabilities, the most uneven die is obtained for $m = 2/3$, yielding the state $p = (0, 1/3, 2/3)$ (for $m = -2/3$ the roles of p_1 and p_3 are interchanged), which has information $I = 5/9$. But if we admit negative probabilities, m can be increased up to $m_{\max} = \frac{2}{\sqrt{3}}$, which gives the state of equation (11), with information $I = 1$.

2.3 Interpretation

The existence of negative probabilities induces some nonstandard properties that are reminiscent of the paradoxes encountered in quantum physics. For example, let us consider a pure state p with $n = 3$ and assimilate the three possible outcomes to the colors of marbles drawn from a bag: red (R), blue (B) and green (G). Like for all pure states, the probability to get the same color in two consecutive draws is $I = 1$, while the probability to get different colors is $S_L = 0$. Let us suppose that we draw a number of marbles, but do not look at their colors for the moment (Fig. 3a). Then we look at the second and third marble and observe that they have the same color (as they should), namely red. Subsequently, we look at the sixth and seventh marble and notice they are both blue (Fig. 3b).

So far, all is in agreement with our expectations. But what would have happened if we had first looked at marbles number 3 and 6 (Fig. 3c)? According to the previous "experiment", they should be of different colors (red and blue), but this is not allowed by the probability distribution of a pure state. Hence, we should find that they have the same color, which is in contradiction with the experiment (b) on the figure. We are forced to conclude that the marbles do not have a predefined color prior to the observation, something that is typical for quantum objects [29–31].

As a second example, let us consider two probability distributions $p = \left(\frac{2}{3}, \frac{2}{3}, -\frac{1}{3}\right)$ and $q = \left(-\frac{1}{3}, \frac{2}{3}, \frac{2}{3}\right)$, which we can be visualized as two different bags containing, respectively, red (R), blue (B) and green (G) marbles in different proportions. They are both pure states and orthogonal to each other, $p \cdot q = \sum_i p_i q_i = 0$. The latter property means that the outcomes of the two bags are perfectly anticorrelated, i.e. if the outcome of the first bag is R then that of the second bag must be not R (denoted $\overline{R}$). We draw pairs of marbles from each bag. From the second bag, the probability of drawing a pair of red marbles is: $\text{Prob}_q(RR) = q_1^2 = \frac{1}{9}$. Since the outcome of the first bag is perfectly anticorrelated with that of the second bag, this number should also represent the probability of *not* drawing a pair of red marbles from the first bag. However, if we compute the same probability using the distribution p of the first bag, we obtain: $\text{Prob}_p(\overline{RR}) = \text{Prob}_p(BB) + \text{Prob}_p(GG) = p_2^2 + p_3^2 = \frac{5}{9}$, which is manifestly different.

This example shows that the following two procedures are mutually exclusive: (i) drawing one marble from bag 1 and another from bag 2, which gives perfectly anticorrelated results; (ii) drawing two marbles from either bag, which yields perfectly correlated results. If two experimentalists draw a marble from each bag and then communicate their results, they always observe anticorrelation. However, once they have done so, they cannot use this knowledge to predict their next draw by using the correlation property of each bag, because the latter is valid only if pairs of marbles are observed together. (Remember that the logical entropy quantifies *distinctions* between two draws, but says nothing about single draws. Indeed the outcome of a single draw is meaningless, as its probability can be negative; only pairs of consecutive draws are meaningful.) Similarly, if one experimenter observes BB in one bag and communicates this result to the second experimentalists, the latter cannot use it to predict that her next draw will be $\overline{\text{BB}}$, because the anticorrelation property holds only as long as both elements of the draw are still unknown.

2.4 Dynamics

The probability distribution $p(t)$ should evolve in a way that preserves both the total probability (of course) and the total information or entropy. In 3D this is possible only if the vector $\boldsymbol{p}$ performs a rotation around the axis perpendicular to the plane π and going through the origin O (see Fig. 2a). This can be viewed as a rotation around the vector $v = (1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3})$, which yields the evolution equation

$$\frac{\mathrm{d}p}{\mathrm{d}t} = v \times p(t),$$

where $\times$ denotes the standard 3D cross product. However, the representation using the vector product cannot be readily extended to dimensions $n > 3$, so it is more useful to write the above equation in matrix form:

$$\frac{\mathrm{d}p_i}{\mathrm{d}t} = \frac{\sqrt{3}}{3} \sum_{j=1}^{n} M_{ij} p_j, \tag{12}$$

where $\mathbf{M} = \{M_{ij}\}$ is the antisymmetric matrix

$$\mathbf{M} = \begin{pmatrix} 0 & -1 & +1 \\ +1 & 0 & -1 \\ -1 & +1 & 0, \end{pmatrix} \tag{13}$$

satisfying $M_{ij} = -M_{ji}$ and $\sum_i M_{i\,j} = \sum_j M_{ij} = 0$. The latter conditions guarantee that the total probability and the total information are indeed conserved during the evolution.

The above matrix form of the evolution equation (13) is readily adapted to higher dimensions, and will be generalized to infinite dimensional systems (continuum) in the next section.

3 Infinite-dimensional spaces (continuum)

3.1 Generalities

The logical entropy and information can be generalized to an infinite-dimensional system, i.e. in the continuum. We define the probability density $f(z)$, with $z \in \mathbb{R}$, normalized so that $\int_{-\infty}^{\infty} f(z)\,\mathrm{d}z = 1$. Then the logical entropy and the information are defined as follows [4]:

$$S_L = 1 - I = 1 - h \int_{-\infty}^{\infty} f^2(z)\mathrm{d}z, \tag{14}$$

where the constant h has the same dimensions as z, and f has the dimensions of h^{-1}. The so-defined information is basically the L^2 norm in the space of real square-integrable functions.

Given the arbitrariness of the constant h, it is not automatic that $0 \leq S_L \leq 1$: some very peaked functions of z may yield an entropy that is negative, or equivalently an information greater than unity. Hence, we *require* that $0 \leq S_L \leq 1$, and restrict the space of allowed probability densities to those whose entropy satisfies this condition.

A useful bound on $f(z)$, which is reminiscent of the bound on Wigner functions [23], can be obtained as follows. Let us consider pure states ($I = 1$) and write

$$h \int f^2(z)\mathrm{d}z = 1 = \left(\int f(z)\mathrm{d}z \right)^2 = \iint f(x)f(y)\mathrm{d}x\mathrm{d}y = \iint f\left(z - \frac{\lambda}{2}\right)f\left(z + \frac{\lambda}{2}\right)\mathrm{d}z\mathrm{d}\lambda.$$

This can be reformulated as

$$\int dz \left[h f^2(z) - \int f\left(z - \frac{\lambda}{2}\right) f\left(z + \frac{\lambda}{2}\right) d\lambda \right] = 0.$$

Setting the integrand equal to zero yields:

$$h f^2(z) = \int f\left(z - \frac{\lambda}{2}\right) f\left(z + \frac{\lambda}{2}\right) d\lambda,$$

which is an integral equation for $f(z)$. Finally, using the Cauchy–Schwartz inequality, we get

$$h f^2 \leq \left(\int \left| f\left(z - \frac{\lambda}{2}\right) \right|^2 d\lambda \right)^{1/2} \left(\int \left| f\left(z + \frac{\lambda}{2}\right) \right|^2 d\lambda \right)^{1/2} = 2 \int f^2(z) dz = \frac{2}{h},$$

from which we deduce the bound

$$\max_z |f(z)| = \frac{\sqrt{2}}{h}. \tag{15}$$

Obviously, the above bound limits the peakedness of $f(z)$ for a given value of h. For a mixed state with information $I < 1$, the bound becomes: $\max |f| = \sqrt{2I}/h$.

For instance, if the probability density is a Gaussian with standard deviation σ: $f(z) = e^{-z^2/2\sigma^2}/(\sqrt{2\pi}\sigma)$ and we require that $I = 1$, we obtain

$$\sigma = \frac{h}{2\sqrt{\pi}}.$$

This value yields exactly the maximum of equation (15), showing that the bound is saturated for a Gaussian distribution of unit information (pure state). For $\sigma > \frac{h}{2\sqrt{\pi}}$, we have $I < 1$, i.e. a mixed state.

All this is similar to a bound that can be obtained on the quantum Wigner function $w(x, p)$ [23], where x and p are respectively position and momentum: $\max_{x,p} |w(x, p)| = 2/h$, where here h is Planck's constant. The additional factor $\sqrt{2}$ is due to the fact that the maximization is done in the 2D phase space (x, p) instead of the 1D space (z) considered above. These considerations establish a suggestive link between the present results and the properties of quantum mechanics, on which we will further elaborate in the forthcoming subsections.

3.2 Dynamics

The time evolution of the probability density $f(z, t)$ must preserve both the total probability and the entropy, hence it has to be a rotation in the appropriate functional space. In analogy with the finite-dimensional case, see equations (12) and (13), we write the general evolution equation for $f(z, t)$ as

$$\frac{\partial f}{\partial t} = \frac{1}{h} \int M(z, z') f(z', t) dz', \tag{16}$$

where M must be antisymmetric: $M(z, z') = -M(z', z)$. In order to preserve the total probability in time, one should also have: $\int M(z, z') dz = 0 = \int M(z, z') dz'$, which follows immediately upon integrating (16) over z. Further, by multiplying equation (16) by $f(z, t)$ and integrating, we obtain

$$\frac{d}{dt} \int f^2(z, t) dz = \frac{2}{h} \iint f(z, t) M(z, z') f(z', t) dz' dz = 0.$$

The last equality follows because the function $\varphi(z, z') \equiv f(z, t) M(z, z') f(z', t)$ is such that $\varphi(z, z') = -\varphi(z', z)$, hence it is odd with respect to the diagonal of the (z, z') plane, and integration over all such planes yields zero.

The two-variable function $M(z, z')$ can be conveniently written as $M(z, z') = m(z - z')$, where $m(\zeta)$ is a single-variable odd function: $m(\zeta) = -m(-\zeta)$. The so-constructed $M(z, z')$ satisfies all the properties mentioned in the preceding paragraph. Hence, we rewrite:

$$\frac{\partial f}{\partial t} = \frac{1}{h} \int m(z - z') f(z', t) dz'. \tag{17}$$

We have included explicitly the constant h in the evolution equation for further comparison with Wigner functions. With this choice, m has the dimensions of an inverse time.

We now write $m(\zeta)$ in terms of its Fourier transform

$$m(\zeta) = i \int d\lambda \, \hat{m}(\lambda) \exp\left(\frac{2\pi i \zeta \lambda}{h}\right). \tag{18}$$

If $\hat{m}(\lambda) = -\hat{m}(-\lambda)$ then it follows that $m(\zeta)$ is indeed an odd function. As the Fourier transform of an odd real function is purely imaginary, we also have that $m(\zeta)$ is real, as intended. Note that λ is dimensionless. Let us now write the odd function $\hat{m}(\lambda)$ as follows, without loss of generality:

$$\hat{m}(\lambda) = \Omega\left(a + \frac{\lambda}{2}\right) - \Omega\left(a - \frac{\lambda}{2}\right),$$

where a is a constant. Inserting all these definitions into the evolution equation (17), one gets

$$\frac{\partial f}{\partial t} = \frac{i}{h} \iint \left[\Omega\left(a + \frac{\lambda}{2}\right) - \Omega\left(a - \frac{\lambda}{2}\right)\right] \exp\left(\frac{2\pi i (z - z')\lambda}{h}\right) f(z', t) dz' d\lambda. \tag{19}$$

In the next section we will show that this equation is basically identical to the quantum evolution equation of the Wigner function.

4 Relationship to quantum mechanics

Equation (19) was built purely on the two assumptions that the total probability and the logical entropy should be conserved in time. It is therefore striking that this equation bears a close resemblance to the evolution equation for the Wigner function w in quantum mechanics [15, 23], as will be discussed shortly.

The Wigner formalism is a representation of quantum mechanics in the classical position-momentum phase space (x, p), which is strictly equivalent to the more usual Schrödinger or Heisenberg pictures. The state of a quantum system, either pure or mixed, is defined by a real function $w(x, p, t)$. The Wigner function is constructed from the wave function for a pure quantum state or from the density matrix for a mixed state. The Wigner function possesses many of the properties of standard probability distributions. For instance, it can be used to compute the average of a phase-space variable $A(x, p)$ as: $\langle A \rangle = \iint w(x, p) A(x, p) dx \, dp$, where we have assumed the normalization $\iint w(x, p) dx \, dp = 1$. However, w can take negative values, which precludes the possibility of interpreting it as a true probability density.

The Wigner function evolves in time according to an integro-differential equation that reads as:

$$\frac{\partial w}{\partial t} + \frac{p}{m} \frac{\partial w}{\partial x} = \frac{2\pi i}{h^2} \iint \left[V\left(x + \frac{\lambda}{2}\right) - V\left(x - \frac{\lambda}{2}\right)\right] \exp\left(\frac{2\pi i (p - p')\lambda}{h}\right) w(x, p', t) \, dp' d\lambda, \tag{20}$$

where $V(x)$ is the potential energy. Interestingly, the above evolution equation preserves in time both $\iint w \, dx \, dp$ and $\iint w^2 dx \, dp$, but not higher powers of w. This fact has motivated choosing the logical entropy as the natural definition of entropy in Wigner's quantum mechanics [4].

Now, we consider a Wigner function concentrated near a position $x = a$ and write: $w(x, p, t) = \overline{w}(p, t)\delta(x - a)$, where δ is the Dirac delta function. We also define $\Omega(x) \equiv 2\pi V(x)/h$, which has the dimensions of an inverse time. Substituting into equation (20) and integrating over x yields

$$\frac{\partial \overline{w}}{\partial t} = \frac{i}{h} \iint \left[\Omega\left(a + \frac{\lambda}{2}\right) - \Omega\left(a - \frac{\lambda}{2}\right)\right] \exp\left(\frac{2\pi i (p - p')\lambda}{h}\right) \overline{w}(p', t) \, dp' d\lambda, \tag{21}$$

which is identical to equation (19) with the correspondence $z \leftrightarrow p$.

It is quite remarkable that, based on the sole assumption that the probability density $f(z, t)$ preserves the total probability and the information (or entropy), we were able to construct an evolution equation (19) that is identical to the evolution equation of the Wigner function. In other words, the quantum evolution appears to stem uniquely from the property of conservation of the logical entropy (apart from the trivial conservation of total probability). This fundamental role played by the quantity $\iint w^2 dx \, dp$ had already been noticed in earlier works [4, 32].

An important caveat is that the probability density $f(z, t)$ depends only the only variable z (plus time), whereas the Wigner function depends on the two phase-space variables x and p. For that reason, we had to consider a Wigner function that is localized in space (Dirac delta function) in order to establish the equivalence with the Wigner evolution equation. This is a significant difference, because it means overlooking a crucial feature of quantum physics, namely the existence

of conjugate variables like position and momentum, whose simultaneous measurement is forbidden by the Heisenberg uncertainty principle.

In order to recover the full Wigner equation, we should work with probability distributions which, in the finite-dimensional case, depend on *two* indexes, such as p_{ij}, i.e., a matrix or tensor. The appropriate norm here appears to be the Frobenius norm $||p|| = \sqrt{\sum_{i,j} p_{ij}^2}$, with the information defined as $I = ||p||^2$. Then, in order to establish an evolution equation that preserves the norm, one would need to define a rotation of the tensor p_{ij} in the appropriate space. The generalization to an infinite dimensional space should lead to an evolution equation for a two-variable probability density $f(z_1, z_2, t)$, which will have to be compared to the full Wigner equation (20) for $w(x, p, t)$. This extension is left for future work.

5 Conclusions

In this work, we made use of the definition of logical entropy and information to extend the notion of probability to negative values. Although negative probabilities have been considered extensively in the past (and often dismissed as unphysical), we argued that they fit nicely within the framework of the logical entropy. Indeed, rejecting negative probabilities would appear as rather arbitrary and odd if one trusts the definition of logical entropy.

Our strategy was to posit that all normalized probability distributions $\{p_i\}$ for which the logical entropy lies in the interval [0, 1] are allowed, irrespective of the sign of the p_is. Of course, the constraint on the entropy limits the absolute negative values that can be taken by the probabilities.

We also pointed out that the logical information has a straightforward interpretation as the square of the Euclidean norm of the probability vector in $\mathbb{R}^n$, or the L^2 norm in the case of a continuous probability density. This simple geometric property is extremely fruitful to derive various interesting properties. In particular, the set of allowed probability distributions may be seen as the intersection of a hypersphere and a hyperplane in $\mathbb{R}^n$.

In order for the total probability and entropy to be conserved in time, the probability vector must *rotate* in the appropriate space, and this rotation is defined by an antisymmetric matrix. We next generalized this rotation to the infinite dimensional case (continuum). Quite remarkably, this leads to an evolution equation for the probability density $f(z, t)$ that is virtually identical to the Wigner equation for a quantum system, at least when one considers only the momentum variable. These findings highlight the fundamental role played by the logical entropy in the mathematical structure of quantum mechanics.

Our future program is to prove that the full Wigner formulation of nonrelativistic quantum mechanics may be deduced from just two simple postulates: (i) conservation of the total probability $\int\int w(x, p, t) \mathrm{d}x \, \mathrm{d}p$ and (ii) conservation of the logical information $h \int\int w^2(x, p, t) \mathrm{d}x \, \mathrm{d}p$. For this, one should extend the present derivation to probability densities that depend on two variables, namely position and momentum. Once realized, this program would establish an alternative axiomatic foundation to nonrelativistic quantum mechanics.

Conflict of interest

The author declares no conflict of interest.

Acknowledgments

References

1. Ellerman D (2009), Counting distinctions: On the conceptual foundations of Shannon's information theory, Synthese 168, 1, 119–149.
2. Ellerman D (2018), Logical entropy: Introduction to classical and quantum logical information theory. Entropy 20, 9, 679.
3. Brukner Č, Zeilinger A (1999), Operationally invariant information in quantum measurements. Phys Rev Lett 83, 3354–3357.
4. Manfredi G, Feix MR (2000), Entropy and Wigner functions. Phys Rev E 62, 4665–4674.
5. Wehrl A (1978), General properties of entropy. Rev Mod Phys 50, 221–260.
6. Simpson EH (1949), Measurement of diversity. Nature 163, 4148, 688.
7. Hunter PR, Gaston MA (1988), Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. J Clin Microbiol 26, 11, 2465–2466.
8. Crupi V (2019), Measures of biological diversity: Overview and unified framework, in: E Casetta, J Marques da Silva, D Vecchi (Eds), From Assessing to Conserving Biodiversity. History, Philosophy and Theory of the Life Sciences, vol. 24, Springer, Cham, pp. 123–136.

9. Christensen C (2007), Polish mathematicians finding patterns in enigma messages, Math Mag 80, 4, 247–273.
10. Tsallis C (1988), Possible generalization of Boltzmann-Gibbs statistics. J Stat Phys 52, 1, 479–487.
11. Brukner Č, Zeilinger A (2003), Information and fundamental elements of the structure of quantum theory, in: L Castell, O Ischebeck (Eds), Time, Quantum, Information, Springer, Berlin, Heidelberg, pp. 323–354.
12. Feynman RP (1987), Negative probability, in B Hiley, FD Peat (Eds), Quantum implications: Essays in honour of David Bohm, Routledge, London, pp. 235–248.
13. Scully MO, Walther H, Schleich W (1994), Feynman's approach to negative probability in quantum mechanics. Phys Rev A 49, 1562–1566.
14. Curtright T, Zachos C (2001), Negative probabilities and uncertainty relations, Mod Phys Lett A 16, 37, 2381–2385.
15. Wigner E (1932), On the quantum correction for thermodynamic equilibrium. Phys Rev 40, 749–759.
16. Deléglise S, Dotsenko I, Sayrin C, Bernu J, Brune M, Raimond J-M, Haroche S (2008), Reconstruction of non-classical cavity field states with snapshots of their decoherence. Nature 455, 7212, 510–514.
17. Bartlett MS (1945), Negative probability, Math Proc Camb Philos Soc 41, 1, 71–73.
18. Khrennikov AY (2008), EPR-Bohm experiment and Bell's inequality: Quantum physics meets probability theory. Theor Math Phys 157, 1, 1448–1460.
19. Khrennikov A (2009) Interpretations of probability, de Gruyter, Berlin, New York.
20. Burgin M (2010), Interpretations of negative probabilities. arXiv preprint arXiv:1008.1287.
21. Burgin M, Meissner G (2012), Negative probabilities in financial modeling. *Wilmott* 2012, 58, 60–65.
22. Mückenheim W, Ludwig G, Dewdney C, Holland PR, Kyprianidis A, Vigier JP, Cufaro Petroni N, Bartlett MS, Jaynes ET (1986), A review of extended probabilities. Phys Rep 133, 6, 337–401.
23. Hillery M, O'Connell RF, Scully MO, Wigner EP (1984,) Distribution functions in physics: Fundamentals. Phys Rep 106, 3, 121–167.
24. de Barros JA, Holik F (2020), Indistinguishability and negative probabilities, *Entropy* 22, 8, 829.
25. Veitch V, Ferrie C, Gross D, Emerson J (2012), Negative quasi-probability as a resource for quantum computation. New J Phys 14, 11, 113011.
26. Spekkens RW (2008), Negativity and contextuality are equivalent notions of nonclassicality. Phys Rev Lett 101, 020401.
27. Abramsky S, Brandenburger A (2011), The sheaf-theoretic structure of non-locality and contextuality. New J Phys 13, 11, 113036.
28. Abramsky S, Brandenburger A (2014), An operational interpretation of negative probabilities and no-signalling models, in: F van Breugel, E Kashefi, C Palamidessi, J Rutten (Eds.), Horizons of the Mind: A Tribute to Prakash Panagaden, Springer International Publishing, Cham, pp. 59–75.
29. Bell JS (1966), On the problem of hidden variables in quantum mechanics. Rev Mod Phys 38, 447–452.
30. Kochen S, Specker EP (1968), The problem of hidden variables in quantum mechanics. Indiana Univ Math J 17, 59–87.
31. Kochen S, Specker EP (1975), The problem of hidden variables in quantum mechanics, in: CA Hooker (Ed), The Logico-Algebraic Approach to Quantum Mechanics, Springer, Heidelberg, pp. 293–328.
32. Baker GA (1958), Formulation of quantum mechanics based on the quasi-probability distribution induced on phase space. Phys Rev 109, 2198–2206.
33. Ellerman D (1985), The mathematics of double entry bookkeeping, *Math Mag* 58, 4, 226–233.