Deriving the Qubit from Entropy Principles∗

Adam Brandenburger† Matthew Deutsch‡ Pierfrancesco La Mura§

Version 07/09/14

Abstract

We provide an axiomatization of the simplest quantum system, namely the qubit, based on entropic principles. Specifically, we show: The qubit can be derived from the set of maximum-entropy probabilities that satisfy an entropic version of the Heisenberg uncertainty principle. Our formulation is in phase space (following Wigner [41]) and makes use of Rényi [32] entropy (which includes Shannon [33] entropy as a special case) to measure the uncertainty of, or information contained in, probability distributions on phase space. We posit three axioms. The Information Reality Principle says that the entropy of a physical system, as a measure of the amount or quantity of information it contains, must be a real number. The Maximum Entropy Principle, well-established in information theory, says that the phase-space probabilities should be chosen to be entropy maximizing. The Minimum Entropy Principle is an entropic version of the Heisenberg uncertainty principle and is a deliberately chosen physical axiom. Our approach is thus a hybrid of information-theoretic (“entropic”) and physical (“uncertainty principle”) axioms.

1 Introduction

The Heisenberg uncertainty principle $\Delta x \cdot \Delta p \geq \hbar / 2$ ([23]) is one of the most famous features of quantum mechanics. It sets a fundamental limit to the extent to which certain pairs of physical variables (such as position $x$ and momentum $p$) can take definite values in microscopic systems.

The non-determinism implied by the Heisenberg uncertainty principle and other — perhaps even stranger — features of quantum mechanics such as superposition, entanglement, and nonlocality were viewed in the early days of the subject as almost-paradoxical. (Schlipp [34] is an account of the famous debates between Bohr and Einstein.) Nowadays, such features are viewed as valuable information resources and are at the heart of developments including quantum cryptography, algorithms, and computing.

Still, deep puzzles remain about the nature of quantum mechanics. In particular, the mathematical structure of the theory is not built up from physically interpretable principles, but, rather,
it is simply posited — and then works. This is very different from elsewhere in physics. Popescu and Rohrlich [30] described the contrast:

The special theory of relativity, we know, can be deduced in its entirety from two axioms: the equivalence of inertial reference frames, and the constancy of the speed of light. Aharanov {unpublished lecture notes} has proposed such a logical structure for quantum theory. ([30], p.380)

Popescu and Rohrlich went on to formulate their well-known no-signaling axiom for quantum mechanics. Interestingly, they showed that this axiom permits not just all quantum systems but also superquantum systems (the so-called PR boxes). It does not identify quantum mechanics.

In this paper, we provide an axiomatization of the simplest quantum system, namely the qubit. Our approach is to begin with the uncertainty principle and turn it into an axiom. We emphasize that we do not claim that the uncertainty principle constitutes an intuitive axiom. Our goal is not to use axioms to ‘tame’ quantum mechanics. In relativity theory, light speed invariance is not an intuitive axiom — the point is that it is physically intelligible. Our interest in the uncertainty principle as an axiom is that it passes this same test. It refers to a limit to definiteness in microscopic systems. This is a physically intelligible — if famously surprising — idea.

The next section puts our axiomatization into the context of the literature. The formal treatment begins in Section 3.

2 The Quantum Reconstruction Program

The search for axioms yielding quantum mechanics can be seen to go as far back as the work by Birkhoff and von Neumann [9] to understand quantum mechanics as a non-classical logic. With the rise of the field of quantum information, this search — often called the quantum reconstruction program — has experienced a notable renaissance.

A large part of the modern search for axioms takes the form of looking for information-theoretic principles that yield the mathematical structure of quantum mechanics. One goal is to provide insight into why the formal structure of the theory takes the shape it does. Another goal comes from the rise of quantum information and its discoveries of very surprising forms of information processing (such as the Grover [20] and Shor [36] results on quantum speedup of database search and prime factorization algorithms, respectively). Axiomatizations of quantum mechanics may yield insight into what kinds of tasks can and cannot be achieved by use of quantum information resources.

Much of the commonly used framework in this enterprise — involving abstractions of the ideas of preparation, transformation, and measurement of physical systems in the laboratory — was laid down in Hardy [22]. Since then, a large literature has developed and various information-theoretic principles have been proposed, including communication complexity (van Dam [39]), information causality (Pawlowski et al. [29]), information capacity (Dakić and Brukner [14]), and purification (Chiribella, D’Ariano, and Perinotti [12]).

The present paper uses the information-theoretic concept of entropy (Shannon [33]). We repeat, though, that we accept a physical principle, namely an uncertainty principle, in our formulation. Our approach is therefore a hybrid of information-theoretic (“entropic”) and physical (“uncertainty principle”) approaches. Fuchs [19] offers forceful support for the importance of finding “deep physical principles” [19] that yield quantum mechanics. Another hybrid paper is Oppenheim and Wehner [28] who also treat an uncertainty principle as an input rather than output of analysis.
Of course, one can debate the extent to which various approaches in the literature are pure or hybrid in the above sense. For example, the purification principle (Chiribella, D’Ariano, and Perinotti [12]) says of a system that “the best possible knowledge of a whole does not necessarily include the best possible knowledge of all its parts” (Schrödinger [35]). This is very mysterious at the everyday macroscopic level, but it is physically intelligible and evidently true of microscopic systems. In the end, what is most important about any axiomatization is not so much debating its classification as finding out where it leads. It remains to be seen whether or not our approach proves productive.

Our approach shares several features with Brukner and Zeilinger [11] (B-Z), who derive important features of quantum mechanics (though not a full axiomatization) from underlying principles. We follow B-Z in assuming that one can choose mutually complementary measurements in the system of interest (see Section 7). Both approaches are entropic, and both use a notion of 2-entropy. But we use Rényi entropy while B-Z use Tsallis [38] entropy, and our arguments for 2-entropy are very different. Finally, we adopt a phase-space framework (as we will see next) while B-Z work in the usual state-space setting.

3 Entropic Uncertainty Principles and Phase Space

We will follow a strand of the modern literature in quantum mechanics and state an uncertainty principle in terms of entropy (Shannon [33]). That quantum systems satisfy an entropic uncertainty principle was conjectured by Everett [17] and Hirschman [25], and proved by Beckner [5] and Bialynicki-Birula and Mycielski [8]. Bialynicki-Birula [7] advocates the use of entropy over the more traditional standard deviation as a measure of the ‘spread’ of probability distributions in quantum systems, and Wehner and Winter [40] survey various entropic uncertainty principles and their applications.

Entropy measures are well-suited to our axiomatic investigation because we want to derive the probabilities that arise in a quantum system, and entropy gives a well-established method of choosing probability distributions. This is the Maximum Entropy Principle, widely used in statistical physics, communication theory, computer science, and probability and statistics. (See, e.g., Cover and Thomas [13].) An early use in physics is by Jaynes [26, 27], who developed the information-theoretic interpretation of thermodynamics. The idea behind the Maximum Entropy Principle is that we should choose among probability distributions by making use of the knowledge that we have about the physical system in question, and nothing except this knowledge.

|                      | Prob(spin-up) | Prob(spin-down) |
|----------------------|---------------|-----------------|
| Spin in x-direction  | $q_x$         | $1 - q_x$       |
| Spin in y-direction  | $q_y$         | $1 - q_y$       |
| Spin in z-direction  | $q_z$         | $1 - q_z$       |

Figure 1: Empirical Model

We now specify our setting and the precise way we calculate entropies. Suppose an experimenter can undertake one of three possible measurements on a system, each of which has two possible outcomes. An associated empirical model gives the probabilities of the outcomes for each measurement. See Figure 1. This is the appropriate empirical-model description of a qubit, for an appropriate set of values for the probabilities $q_x, q_y, q_z$. To emphasize this (and for ease of exposition), the labels in Figure 1 refer to the particular case of the spin (up or down) in the $x$-, $y$-,
z-directions of a spin-$\frac{1}{2}$ particle. Obviously, we do not also put in the set of possible values for $q_x$, $q_y$, $q_z$, or there would be nothing left to derive. Our goal is to derive these probabilities from axioms.

The three-measurement feature of our set-up can itself be derived via the well-known parameter-counting argument of Wootters [43, 44] and Hardy [22]. Here is a brief sketch. Say a system has $N$ orthogonal states if a (non-degenerate) measurement on the system has $N$ possible outcomes. Let $g(N)$ be the number of parameters needed to specify the state of the system. Assuming no signaling (Popescu and Rohrlich [30]) and that measurements on the parts of a (composite) system are sufficient for determining the state of the whole, one obtains $g(N) = N^k - 1$, for some positive integer $k$. The classical case is $k = 1$. The quantum case is $k = 2$. For a two-outcome system, this gives $g(2) = 2^2 - 1 = 3$, as we assume. This step is by now routine in the axiomatic program, so we do not go into further detail here.

Next, we want to calculate the entropy of an empirical model. This poses a challenge since an empirical model contains more than one probability distribution — specifically, a separate distribution for each possible measurement on the system. One might try adding up the entropies of each of the distributions, to get an overall measure. But this would be conceptually questionable. Entropies can be added over statistically independent systems (see later), but here we are considering a number of probability distributions defined on a single system.

Our solution is to associate a phase-space model with an empirical model, where each point in the phase space is a complete specification of how the system responds to each of the possible measurements. Thus, a phase space for the empirical model in Figure 1 consists of eight points, as in Figure 2. We use 0 (resp. 1) to denote the spin-up (resp. spin-down) outcome.

| Spin in x-direction | Spin in y-direction | Spin in z-direction |
|---------------------|---------------------|---------------------|
| $\omega_1$         | 0                   | 0                   |
| $\omega_2$         | 0                   | 0                   |
| $\omega_3$         | 0                   | 1                   |
| $\omega_4$         | 0                   | 1                   |
| $\omega_5$         | 1                   | 0                   |
| $\omega_6$         | 1                   | 0                   |
| $\omega_7$         | 1                   | 1                   |
| $\omega_8$         | 1                   | 1                   |

Figure 2: Phase-Space Model

The possibility of a non-deterministic response to measurement — which is, of course, the case in quantum mechanics — is captured by specifying, in addition, a probability distribution over the points in phase space. Thus, the phase-space model of Figure 2 is completed by assigning a probability $p_i$ to each state $\omega_i$, for $i = 1, \ldots, 8$. Of course, we want these probabilities to relate correctly to the probabilities in the given empirical model. That is, we require:

$$p_1 + p_2 + p_3 + p_4 = q_x,$$

$$p_1 + p_2 + p_5 + p_6 = q_y,$$

$$p_1 + p_3 + p_5 + p_7 = q_z.$$

We will say that a phase-space model realizes a given empirical model when these conditions hold.

The key point is that we now have a single probability distribution $(p_1, \ldots, p_8)$ and we can think about associating an entropy with this distribution.
4 Negative Probabilities

The use of phase space in quantum mechanics goes back to Wigner [41], who gave a phase-space representation of quantum systems using quasi-probability distributions, i.e., distributions which may include negative weights on some points. Dirac [16] and Feynman [18] also promoted the use of negative probabilities in quantum mechanics.

This appearance of negative (signed) probabilities should not be a surprise. A phase-space model can be understood as a particular type of hidden-variable model, where the possible values of the hidden variable are precisely the possible states in phase space. Bell’s Theorem ([6]) tells us that empirical probabilities arise in quantum mechanics which cannot be reproduced by any hidden-variable model — where a hidden-variable model is understood to involve ordinary (unsigned) probabilities. If phase-space models are to reproduce all quantum systems, an assumption of Bell’s Theorem has to be weakened. The introduction of signed probabilities does the job. Abramsky and Brandenburger [1] prove that allowing signed probabilities on phase space makes it possible to reproduce precisely the set of empirical probabilities that arise under the no-signaling condition ([30]) — a set which we already noted includes the quantum set. (Abramsky and Brandenburger [2] offer an operational interpretation of negative probabilities.)

There is an important proviso. To make sense, an approach which allows negative probabilities on phase space must nevertheless insist that all observable events have non-negative probabilities. This is to ensure that the probabilities of observable events can be related, via statistical analysis, to observed frequencies — where the latter are non-negative by definition. In equations (1)-(3), some of the $p_i$’s may be negative, but the three sums must be equal to the non-negative numbers $q_x, q_y, q_z$, respectively.

5 Rényi Entropy

We are now ready to associate entropies with probability distributions on phase space. There are many different definitions of entropy, starting with Shannon entropy ([33]). We believe that one property of an entropy measure is very basic: It should be additive across statistically independent systems. We will view this as a fundamental structural requirement for entropy, and choose an entropy measure that seems least constrained beyond this. While it would be hard to argue that there is an unambiguous such choice, Rényi entropy ([32]) is a good candidate. It satisfies additivity across statistically independent systems. Moreover, together with a mean-value property (which is satisfied by many entropy measures) and some regularity conditions, additivity characterizes Rényi entropy. The Appendix provides a list of axioms yielding Rényi entropy. Shannon entropy is not ruled out by Rényi entropy, but now becomes a special case that satisfies a stronger requirement (in the form of a composition law).

Within quantum mechanics, the most common entropy measure is the von Neumann entropy $S(\rho) = -\text{Tr}(\rho \log \rho)$ (where $\rho$ is the density matrix). This is unsuitable for our purposes, for two reasons. First, it is defined within the quantum formalism, whereas we want to derive not assume the formalism. Second, von Neumann entropy captures only the uncertainty due to a mixture and not the uncertainty in a pure state ($S(\rho) = 0$ when $\rho$ is pure). The Rényi definition is already used in various ways in quantum mechanics; see Bialynicki-Birula [7] for a list of applications.

We now define Rényi entropy. Given a finite set $X = \{x_1, \ldots, x_n\}$ and a probability distribution $p$ on $X$, i.e., a tuple $p = (p_1, \ldots, p_n)$ where each $p_i \geq 0$ and $\sum_i p_i = 1$, the Rényi entropy of $p$ is
defined as a function of a free parameter $0 < \alpha < \infty$ by:

$$H_\alpha(p) = \begin{cases} 
\frac{-1}{\alpha - 1} \log_2 \left( \sum_{i=1}^{n} p_i^\alpha \right) & \text{if } \alpha \neq 1, \\
- \sum_{i=1}^{n} p_i \log_2 p_i & \text{if } \alpha = 1.
\end{cases}$$

(4)

Shannon entropy is the special case $\alpha = 1$. We can also obtain Shannon entropy $H_1(p)$ from $H_\alpha(p)$, for $\alpha \neq 1$, by letting $\alpha \to 1$ and using l’Hôpital’s Rule. Another limiting case is the min-entropy $H_\infty(p) := \lim_{\alpha \to \infty} H_\alpha(p) = -\log_2(\max_i p_i)$. Hartley entropy $H_0(p) := \log_2 n$ can be viewed as the limit as $\alpha \to 0$ (with the convention that $0^0 = 1$).

The Rényi entropy $H_\alpha(p)$ is a non-negative real number when $p$ is an ordinary (unsigned) probability distribution. But, as we saw in the previous section, we want to admit signed probability distributions, that is, we allow $p$ where each $p_i \geq 0$ and $\sum_i p_i = 1$. Once we do this, entropy may become complex-valued. (When $\sum_i p_i^\alpha < 0$, we get the log of a negative number.) This said, we impose:

**Information Reality Principle:** The entropy of a physical system must be real-valued.

The viewpoint here is that the information in a system is a physical concept. Entropy is the measure of the amount or quantity of this information in a particular system. As such, it must be a number in the usual sense, i.e., it must be real-valued. One might ask whether one should go further and insist that entropy is not just a real number but is non-negative, on the basis that there cannot be a negative amount of information. In fact, it will not be necessary to take a view on this here, since non-negativity will be implied by a subsequent axiom.

Our Information Reality Principle immediately imposes a restriction on the values that the parameter $\alpha$ in Rényi entropy can take: $H_\alpha(p)$ is a real number for all signed probability distributions $p$ is equivalent to the requirement that $\alpha$ is an even integer, i.e. $\alpha = 2k$, for $k = 1, 2, 3, \ldots$.

### 6 Maximum Entropy Distributions

The role of phase-space probabilities is to reproduce the probabilities which are observed by measurements on our physical system. In general, more than one phase-space distribution will give rise (via equations (1)-(3)) to the same observed probabilities. We next apply the Maximum Entropy Principle to enable us to select a phase-space distribution when what we know are the empirical probabilities in Figure 1.

It will be helpful for later to re-parameterize an empirical model in the way shown in Figure 3, where $-1 \leq r_x \leq 1$, $-1 \leq r_y \leq 1$, and $-1 \leq r_z \leq 1$. Thus, we now identify the set of empirical models with the set $[-1, 1]^3$.

| Direction    | Prob(spin-up) | Prob(spin-down) |
|--------------|---------------|-----------------|
| Spin in x-direction | $\frac{1}{2}(1 + r_x)$ | $\frac{1}{2}(1 - r_x)$ |
| Spin in y-direction | $\frac{1}{2}(1 + r_y)$ | $\frac{1}{2}(1 - r_y)$ |
| Spin in z-direction | $\frac{1}{2}(1 + r_z)$ | $\frac{1}{2}(1 - r_z)$ |

Figure 3: Re-parameterized Empirical Model
Given a signed probability distribution $p = (p_1, \ldots, p_8)$ on phase space, we can associate a triple of numbers $(r_x, r_y, r_z) \in \mathbb{R}^3$ analogously to equations (1)-(3):

\begin{align*}
    r_x &= 2(p_1 + p_2 + p_3 + p_4) - 1, \\
    r_y &= 2(p_1 + p_2 + p_5 + p_6) - 1, \\
    r_z &= 2(p_1 + p_3 + p_5 + p_7) - 1. \\
\end{align*}

(5) \hspace{1cm} (6) \hspace{1cm} (7)

Let $\Pi$ denote the set of signed probability distributions on phase space and let $f : \Pi \to \mathbb{R}^3$ denote the mapping defined by equations (5)-(7). Fix an empirical model $(r_x, r_y, r_z) \in [-1,1]^3$. In general, the pre-image $f^{-1}(r_x, r_y, r_z)$ will consist of more than one (signed) probability distribution $p$ in $\Pi$. To select from these distributions, we will appeal to:

**Maximum Entropy Principle:** A (signed) probability distribution $p$ corresponding to an empirical model will be one that maximizes the Rényi entropy $H_\alpha(p)$.

Algebraically, we will choose $p$ to solve:

$$
\max_{p \in f^{-1}(r_x, r_y, r_z)} H_\alpha(p),
$$

and we will write $h_\alpha(r_x, r_y, r_z)$ for the resulting maximum value of the entropy.

We next set up the Lagrangian program to find $h_{2k}(r_x, r_y, r_z)$ for $k = 1, 2, 3, \ldots$. Since \(\frac{1}{2k-1} \log_2(\cdot)\) is strictly increasing, we can solve the program:

$$
\max_{p_i; i=1,\ldots,8} - \sum_{i=1}^8 p_i^{2k}
$$

subject to

\begin{align*}
    \sum_{i=1}^8 p_i &= 1, \\
    p_1 + p_2 + p_3 + p_4 &= \frac{1}{2}(1 + r_x), \\
    p_1 + p_2 + p_5 + p_6 &= \frac{1}{2}(1 + r_y), \\
    p_1 + p_3 + p_5 + p_7 &= \frac{1}{2}(1 + r_z). \\
\end{align*}

(10) \hspace{1cm} (11) \hspace{1cm} (12) \hspace{1cm} (13)

The objective is strictly concave and the constraints are linear, so there is a unique maximum and the first-order conditions are necessary and sufficient. Introducing Lagrangian multipliers $\lambda$, $\mu$, $\nu$, and $\xi$ for the constraints (10), (11), (12), and (13), respectively, the first-order conditions for the Lagrangian $L$ are:

\begin{align*}
    \frac{\partial L}{\partial p_1} &= 0 \quad \Rightarrow \quad 2kp_1^{2k-1} = \lambda + \mu + \nu + \xi, \\
    \frac{\partial L}{\partial p_2} &= 0 \quad \Rightarrow \quad 2kp_2^{2k-1} = \lambda + \mu + \nu, \\
    \frac{\partial L}{\partial p_3} &= 0 \quad \Rightarrow \quad 2kp_3^{2k-1} = \lambda + \mu + \xi, \\
    \frac{\partial L}{\partial p_4} &= 0 \quad \Rightarrow \quad 2kp_4^{2k-1} = \lambda + \mu, \\
    \frac{\partial L}{\partial p_5} &= 0 \quad \Rightarrow \quad 2kp_5^{2k-1} = \lambda + \nu + \xi, \\
    \frac{\partial L}{\partial p_6} &= 0 \quad \Rightarrow \quad 2kp_6^{2k-1} = \lambda + \nu, \\
    \frac{\partial L}{\partial p_7} &= 0 \quad \Rightarrow \quad 2kp_7^{2k-1} = \lambda + \xi, \\
    \frac{\partial L}{\partial p_8} &= 0 \quad \Rightarrow \quad 2kp_8^{2k-1} = \lambda. \\
\end{align*}

(14) \hspace{1cm} (15) \hspace{1cm} (16) \hspace{1cm} (17)

We will be interested in solutions to equations (10)-(17) and the resulting maximal value $h_{2k}(r_x, r_y, r_z)$ of the entropy.
7 An Entropic Uncertainty Principle

In quantum mechanics, a set of measurements on a system is called mutually unbiased (or complementary) if complete knowledge of the measured value of the outcome of one of them implies no knowledge at all about the possible outcomes of the others. Measurements of position and momentum form the most famous example. We now introduce a physical principle that forms part of our axiomatization of the qubit, by assuming that the three observables in the empirical model of Figure 3 form a complementary set. This is a true statement about the $x$, $y$, and $z$-spins of a spin-$\frac{1}{2}$ particle, but, of course, it is a substantive assumption in an axiomatic approach (made also in e.g., Brukner and Zeilinger [11]). It will be embodied in the axiom we state next.

We suppose that if there is probability 1 of observing spin-up (resp. spin-down) in the $x$-direction, then there will be probability 1/2 of observing spin-up and probability 1/2 of observing spin-down in the $y$-direction, and probability 1/2 of observing spin-up and probability 1/2 of observing spin-down in the $z$-direction. Algebraically, the condition is that if $r_x = 1$ (resp. $-1$), then $r_y = r_z = 0$. There is maximum uncertainty in the $y$- and $z$-directions. The same conditions hold with the $x$-, $y$-, and $z$-directions permuted.

We can see from the symmetry in equations (9)-(13) that for these six probability distributions, the maximum values of the associated entropies will be equal:

$$h_\alpha(1, 0, 0) = h_\alpha(-1, 0, 0) = h_\alpha(0, 1, 0) = h_\alpha(0, -1, 0) = h_\alpha(0, 0, 1) = h_\alpha(0, 0, -1). \quad (18)$$

Let us denote this common value by $h_\alpha^*$. We are now ready to capture the uncertainty principle of quantum mechanics within our framework via:

**Minimum Entropy Principle**: Empirical probabilities defined by $r_x$, $r_y$, and $r_z$ are admissible if and only if for every value of the parameter $\alpha$, the associated Rényi entropy $h_\alpha(r_x, r_y, r_z)$ is equal to or greater than the lower bound $h_\alpha^*$.

As expected, this is a quite different statement from the Heisenberg uncertainty principle. We use entropy not standard deviation, and we consider a finite-dimensional not infinite-dimensional system. This brings us closer to the entropic uncertainty relations of the quantum information literature (Wehner and Winter [40]). Our principle is still different in that it is stated on phase space. Most importantly, of course, we do not derive the principle but state it as an axiom. We now show:

**Theorem 7.1** The minimum entropy bound is given by $h_{2k}^* = 2$ for all $k = 1, 2, 3, \ldots$

**Proof.** Set $p_1 = p_2 = p_3 = p_4 = \frac{1}{2}$ and $p_5 = p_6 = p_7 = p_8 = 0$. (This implies $\lambda = \nu = \xi = 0$ and $\mu = k \times 2^{3-4k}$.) This gives a solution to equations (10)-(17) when $(r_x, r_y, r_z) = (1, 0, 0)$. As noted earlier, the first-order conditions are necessary and sufficient for the maximum, so these values for $p_1$ through $p_8$ indeed identify the maximum. We now calculate: $h_{2k}(1, 0, 0) = \frac{1}{2k-1} \log_2 (4 \times (\frac{1}{2})^{2k}) = 2$.

We can also check directly that for $(r_x, r_y, r_z) = (-1, 0, 0)$, setting $p_1 = p_2 = p_3 = p_4 = 0$ and $p_5 = p_6 = p_7 = p_8 = \frac{1}{2}$ solves equations (10)-(17), and that $h_{2k}(-1, 0, 0) = 2 = h_{2k}(1, 0, 0)$, as asserted in equation (18). We can similarly check the cases $(r_x, r_y, r_z) = (0, 1, 0), (0, -1, 0), (0, 0, 1),$ and $(0, 0, -1)$. ■
8 Deriving the Qubit

Summarizing so far, our three principles — Information Reality, Maximum Entropy, and Minimum Entropy — have identified the following set of empirical probabilities (the following “theory”):

\[ T = \{(r_x, r_y, r_z) \in [-1, 1]^3 : h_{2k}(r_x, r_y, r_z) \geq 2 \text{ for all } k \}. \] (19)

We now show that \( T \) is the unit ball in \( \mathbb{R}^3 \). Recalling our parameterization in Figure 3, we see that this says our axioms identify the Bloch sphere representation of the qubit.

The proof is in two steps. Define a sequence of sets

\[ T_{2k} = \{(r_x, r_y, r_z) \in [-1, 1]^3 : h_{2k}(r_x, r_y, r_z) \geq 2 \} \] (20)

for \( k = 1, 2, 3, \ldots \). Thus \( T = \bigcap_k T_{2k} \). The first step (Theorem 8.1) is to show that the \( T_{2k} \) form an increasing sequence, so that \( T = T_2 \). The second step (Theorem 8.2) is to show that \( T_2 \) is the unit ball in \( \mathbb{R}^3 \).

**Theorem 8.1** The sets \( T_{2k} \), for \( k = 1, 2, 3, \ldots \), are increasing.

**Proof.** First rewrite Rényi entropy \( H_{2k}(p) \) in terms of the \( 2k \)-norm:

\[ H_{2k}(p) = -\frac{2k}{2k - 1} \log_2(||p||_{2k}), \] (21)

for \( k = 1, 2, 3, \ldots \). We can then write the entropy maximization problem as the norm minimization problem:

\[
\min_{p \in \mathbb{R}^8} ||p||_{2k}
\]
subject to \( Ap = b \),

where

\[
A = \begin{pmatrix}
1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\
1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1
\end{pmatrix}, \quad b = \begin{pmatrix}
\frac{1}{2}(1 + r_x) \\
\frac{1}{2}(1 + r_y) \\
\frac{1}{2}(1 + r_z) \\
1
\end{pmatrix}. \] (24)

The dual problem (see Boyd and Vandenberghe [10], pp.221-222) is:

\[
\max_{x \in \mathbb{R}^4} b^T x
\]
subject to \( ||A^T x||_{\frac{2k}{2k+1}} \leq 1 \),

where the superscript \( T \) denotes transpose. Note that \( || \cdot ||_{\frac{2k}{2k+1}} \) is the dual norm of \( || \cdot ||_{2k} \). We will also use the fact that \( || \cdot ||_{\frac{2k+2}{2k+1}} \) is the dual norm of \( || \cdot ||_{2k+2} \).

The geometry of the dual problem is depicted in Figure 4. Here, \( C_k \) labels the set of points \( x \) in \( \mathbb{R}^4 \) with \( ||A^T x||_{\frac{2k}{2k+1}} \leq 1 \). The point \( y^k \) is the maximizer of the dual problem, and the value of the dual problem is \( ||b||_2 \times ||z^k||_2 \) (where \( || \cdot ||_2 \) is the ordinary Euclidean norm). Figure 4 also depicts the dual problem for \( k + 1 \). Thus, \( C_{k+1} \) labels the set of points \( x \) in \( \mathbb{R}^4 \) with \( ||A^T x||_{\frac{2k+2}{2k+3}} \leq 1 \), \( y^{k+1} \) is the maximizer of the dual problem for \( k + 1 \), and \( ||b||_2 \times ||z^{k+1}||_2 \) is the corresponding value.
Referring back to equations (21) and (22), we see that \((r_x, r_y, r_z) \in T_{2k}\) iff the value of the primal problem for \(k\) is no more than \((1/2)^{\frac{k-1}{k+1}}\). Similarly, \((r_x, r_y, r_z) \in T_{2k+2}\) iff the value of the primal problem for \(k+1\) is no more than \((1/2)^{\frac{k+1}{k+2}}\). Strong duality holds, so the values of each primal and dual are equal. Therefore, the two conditions are, equivalently, that 
\[
|b|_2 \times ||z^k||_2 \leq (1/2)^{\frac{k-1}{k+1}}
\]
and 
\[
|b|_2 \times ||z^{k+1}||_2 \leq (1/2)^{\frac{k+1}{k+2}}.
\]

We want to show that \((r_x, r_y, r_z) \in T_{2k}\) implies \((r_x, r_y, r_z) \in T_{2k+2}\). Using
\[
\frac{2k-1}{k} + \frac{1}{k(k+1)} = \frac{2k+1}{k+1},
\]
it will suffice to show that
\[
\frac{||z^{k+1}||_2}{||z^k||_2} \leq \left(\frac{1}{2}\right)^{\frac{k+1}{k+2}}.
\]

By convexity of \(C_k\) and similar triangles:
\[
\frac{||z^{k+1}||_2}{||z^k||_2} \leq \frac{||y^{k+1}||_2}{||w^k||_2}.
\]

We now find the maximum value of the ratio \(||y^{k+1}||_2/||w^k||_2\) as \(w^k\) moves around the boundary of \(C_k\). Writing \(y^{k+1} = \lambda w^k\), for some \(\lambda\) between 0 and 1, we find:
\[
\lambda = \frac{1}{||A^T w^k||^{\frac{2k-1}{2k+2}}_{\frac{k-1}{k+1}}}
\]

In fact, we will write:
\[
\lambda = \frac{||A^T w^k||^{\frac{2k}{2k+2}}_{\frac{k-1}{k+1}}} {||A^T w^k||^{\frac{2k+2}{2k-1}}_{\frac{k-1}{k+1}}} = \frac{||A^T w||^{\frac{2k}{2k+1}}_{\frac{k-1}{k+1}}} {||A^T w||^{\frac{2k+2}{2k-1}}_{\frac{k-1}{k+1}}}
\]
for any $w \neq 0$ lying on the ray from 0 through $w^k$. (This uses homogeneity of norms.)

A calculation gives, for any numbers $m, n > 0$:

$$\frac{\partial}{\partial w_4} \left( \frac{||A^T w||_m^2}{||A^T w||_n^2} \right) =$$

$$\frac{1}{||A^T w||_n^2} \times \left[ ||A^T w||_n \times \left( \frac{||A^T w||_m}{||A^T w||_n} \right)^{m-1} - ||A^T w||_m \times \left( \frac{||A^T w||_m}{||A^T w||_n} \right)^{n-1} \right]. \quad (32)$$

Note that $||A^T w||_n \geq ||A^T w||_m$ whenever $m \geq n$, and $||A^T w||_{m-1} \geq ||A^T w||_m$ and $||A^T w||_{n-1} \geq ||A^T w||_n$ for all $m, n$. Setting $m = 2k/(2k-1)$ and $n = (2k + 2)/(2k + 1)$ we therefore find:

$$\frac{\partial}{\partial w_4} \left( \frac{||A^T w||_{2k}}{||A^T w||_{2k+2}} \right) \geq 0. \quad (33)$$

Again using the fact that the value of the ratio is constant on a given ray, we conclude that the ratio (31) is maximized on the ray from 0 through $(0, 0, 0, 1)$. Using $A^T w = (1, 1, 1, 1, 1, 1, 1)\top$ when $w = (0, 0, 0, 1)\top$, we find that this maximum value is $8^{\frac{2k-1}{2k+1}}/8^{\frac{2k+1}{2k+2}} = 8^{-\frac{1}{2k+1}} = [1/(2\sqrt{2})]^{\frac{1}{2k+1}} < (1/2)^{\frac{1}{2k+1}}$, as required. ■

**Theorem 8.2** The set $T_2$ is the unit ball in $\mathbb{R}^3$:

$$T_2 = \{(r_x, r_y, r_z) \in [-1, 1] : r_x^2 + r_y^2 + r_z^2 \leq 1\}. \quad (34)$$

**Proof.** We need to calculate $h_2(r_x, r_y, r_z)$ for each choice of $r_x, r_y, r_z$. For $k = 1$, we can solve equations (10)-(17) in closed form to obtain:

$$p_1 = \frac{1}{8}(1 + r_x + r_y + r_z), \quad p_2 = \frac{1}{8}(1 + r_x + r_y - r_z), \quad (35)$$

$$p_3 = \frac{1}{8}(1 + r_x - r_y + r_z), \quad p_4 = \frac{1}{8}(1 - r_x - r_y - r_z), \quad (36)$$

$$p_5 = \frac{1}{8}(1 - r_x + r_y + r_z), \quad p_6 = \frac{1}{8}(1 - r_x + r_y - r_z), \quad (37)$$

$$p_7 = \frac{1}{8}(1 - r_x - r_y + r_z), \quad p_8 = \frac{1}{8}(1 - r_x - r_y - r_z). \quad (38)$$

From equations (35)-(38) we can calculate the maximum Rényi entropy:

$$h_2(r_x, r_y, r_z) = -\log_2 \left[ \frac{1}{64} \left[ (1 + r_x + r_y + r_z)^2 + (1 + r_x + r_y - r_z)^2 + \right. \right.$$

$$(1 + r_x - r_y + r_z)^2 + (1 + r_x - r_y - r_z)^2 + (1 - r_x + r_y + r_z)^2 +$$

$$(1 - r_x + r_y - r_z)^2 + (1 - r_x - r_y + r_z)^2 + (1 - r_x - r_y - r_z)^2 \right], \quad (39)$$

which after some simplification yields:

$$h_2(r_x, r_y, r_z) = -\log_2 \left[ \frac{1 + r_x^2 + r_y^2 + r_z^2}{8} \right]. \quad (40)$$

It follows that:

$$T_2 = \{(r_x, r_y, r_z) \in [-1, 1] : -\log_2 \left[ \frac{1 + r_x^2 + r_y^2 + r_z^2}{8} \right] \geq 2\} \quad (41)$$

$$= \{(r_x, r_y, r_z) \in [-1, 1] : r_x^2 + r_y^2 + r_z^2 \leq 1\}, \quad (42)$$

as was to be shown. ■
9 Discrete Wigner function

A discrete Wigner function was introduced by Wootters [45] as an analog, for finite-dimensional systems, to the original Wigner [41] function defined for infinite-dimensional systems. The function is defined on the 4-point phase space shown in Figure 5. This is different from the 8-point phase space which we have used (Figure 2). It is important that we started with the full (unreduced) state space because our goal was to derive the structure of the qubit from the axioms we imposed. This is not necessary in the case of the discrete Wigner function, which is a representation not a derivation.

To continue the definition, one introduces a set of phase-point operators, i.e., a set of $2\times2$ matrices $A_{a_1a_2}$, for each point $(a_1, a_2) = (0, 0), (1, 0), (0, 1), (1, 1)$:

$$A_{a_1a_2} = \frac{1}{2}[I + (-1)^{a_2}\sigma_x + (-1)^{a_1+a_2}\sigma_y + (-1)^{a_1}\sigma_z], \quad (43)$$

where $I$ is the identity matrix and $\sigma_x, \sigma_y, \sigma_z$ are the Pauli matrices. Given a density matrix $\rho$, the associated (discrete) Wigner function $W$ (Wootters [45]) is then given by:

$$W_{a_1a_2} = \frac{1}{2}\text{tr}(\rho A_{a_1a_2}), \quad (44)$$

for each point $(a_1, a_2)$.

One can check that the Wigner function is a signed probability distribution: $W_{00} + W_{10} + W_{01} + W_{11} = 1$. It yields empirical probabilities by marginalization. For example, the probability that a measurement in the $x$-direction yields spin-up is given by the sum $W_{00} + W_{01}$.

The Wigner function can be brought into a more useful form for us by using the well-known fact that any density matrix $\rho$ can be written in the form:

$$\rho = \frac{1}{2}(I + r_x\sigma_x + r_y\sigma_y + r_z\sigma_z), \quad (45)$$

where $(r_x, r_y, r_z)$ lies in the Bloch sphere. We now calculate:

$$W_{00} = \frac{1}{2}\text{tr}(\rho A_{00}) = \frac{1}{8}\text{tr}[(I + r_x\sigma_x + r_y\sigma_y + r_z\sigma_z)(I + \sigma_x + \sigma_y + \sigma_z)] = \frac{1}{8}\text{tr}(I + r_xI + r_yI + r_zI) = \frac{1}{4}(1 + r_x + r_y + r_z), \quad (46)$$
where we rely on the usual properties of the Pauli matrices. We can similarly calculate:

\begin{align*}
W_{10} &= \frac{1}{4}(1 - r_x + r_y - r_z), \\
W_{01} &= \frac{1}{4}(1 + r_x - r_y - r_z), \\
W_{11} &= \frac{1}{4}(1 - r_x - r_y + r_z).
\end{align*}

Comparing equations (46)-(49) with equations (35)-(38), we find the following relationships:

\begin{align*}
p_1 + p_2 + p_3 + p_4 &= W_{00} + W_{10}, \\
p_1 + p_2 + p_5 + p_6 &= W_{00} + W_{01}, \\
p_1 + p_3 + p_5 + p_7 &= W_{00} + W_{11}.
\end{align*}

from which (using \(W_{00} + W_{10} + W_{01} + W_{11} = 1\)) we can derive the \(W_{a_1a_2}\) from the \(p_i\).

In this way, we can also view our axiomatization as directly yielding a representation of the qubit in terms of the discrete Wigner function. There is some precedent for axiomatizing quantum mechanics via the Wigner function (in the infinite-dimensional case). See Wigner [42] and Hillery et al. [24], although the axioms there are not based on entropy as in this paper.

## 10 Discussion

There is a subtle interplay between the role of negative probabilities and the role of the uncertainty principle in our analysis. Suppose, for a moment, we do not impose our Minimum Entropy Principle. Then any empirical model \((r_x, r_y, r_z) \in [-1,1]^3\) can be realized by non-negative probabilities. An easy way to see this is to imagine creating an urn with billiard balls which come in two possible colors, two possible weights, and two possible diameters. By choosing the proportions of each of the eight types of billiard ball appropriately, any triple \((r_x, r_y, r_z)\) can be obtained as the probabilities of getting the first color, weight, and diameter, respectively.

Now, (re-)impose our Minimum Entropy Principle but disallow negative probabilities. We can see from equations (35)-(38) that only those \((r_x, r_y, r_z)\) satisfying \(|r_x| + |r_y| + |r_z| \leq 1\) can now be obtained. This condition defines the octahedron with vertices \((1,0,0), (-1,0,0), (0,1,0), (0,-1,0), (0,0,1),\) and \((0,0,-1)\). The octahedron is strictly contained in the unit ball \(T_2\). For example, it does not include the point \((\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}})\). This point can be realized by the non-negative probabilities \(p_1 = \frac{1}{2}(1 + \frac{1}{\sqrt{3}}), p_8 = \frac{1}{2}(1 - \frac{1}{\sqrt{3}}), p_i = 0\) otherwise (among other choices), but then \(H_2(p) = -\log_2(\frac{2}{3}) < 2\) and the Minimum Entropy Principle is violated. Of course, we know this point can be realized by signed probabilities which satisfy the Minimum Entropy Principle. Such probabilities are given by equations (35)-(38) and, in particular, we find \(p_8 = \frac{1}{2}(1 - \sqrt{3}) < 0\). In a manner of speaking, then, Nature disallows certain physical systems (all those lying outside the octahedron) by virtue of the uncertainty principle. But then Nature gives back some systems (all those lying outside the octahedron and within the unit ball) by allowing negative probabilities.

Our entropic approach to deriving the qubit may shed some light on the debate over epistemic vs. ontic interpretations of quantum mechanics. Prominent on the epistemic side is Spekkens [37] (who introduced the epistemic vs. ontic terminology). A prominent recent argument on the ontic side was given by Pusey, Barrett, and Rudolph [31]. The maximum-entropy method which we adopt may offer at least a degree of reconciliation of the two sides. Under this approach, the probabilities in a quantum system can be thought of as representing the point of view of the observer (an epistemic
notion), but they cannot be fully personalistic in that they must respect the probabilities given empirically and be maximally noncommittal beyond this (more of an ontic view).

Finally, a historical note: Jaynes ([26], p.623) observed that quadratic entropy (i.e., Rényi entropy with $\alpha = 2$) works well in various applications of information theory, but that a drawback is that maximum-entropy calculations via Lagrangian multipliers may then yield negative probabilities. In the quantum case, we have seen that $\alpha = 2$ is the correct entropy to use and that negative probabilities are, in fact, fully expected.

A Appendix

We show here how Rényi entropy ([32]) can be derived from information-theoretic axioms. Rényi himself showed that his definition satisfied a list of axioms and he conjectured (but did not prove) that this list characterized his definition. Daróczy [15] supplied the missing proof. We have not found a convenient self-contained treatment in the literature, so we supply one here. (Daróczy [15] is in German. Aczél and Daróczy [3] is a comprehensive treatment, but at the level of ‘master’ theorems.)

Given a finite set $X = \{x_1, \ldots, x_n\}$, a generalized probability distribution (gpd) $Q$ on $X$ is a tuple $Q = (q_1, \ldots, q_n)$ where each $q_i \geq 0$ and $0 < \sum_{i=1}^n q_i \leq 1$. (Rényi [32] suggests interpreting gpd’s as reflecting possible unobservability of certain outcomes. It seems best to think of this kind of unobservability as a conceptual thought experiment which is useful in developing axioms, and not to confuse it with any physical constraints on observability.) The quantity $w(Q) = \sum_{i=1}^n q_i$ will be called the weight of $Q$. Given two gpd’s $P = (p_1, \ldots, p_m)$ and $Q = (q_1, \ldots, q_n)$, we denote by $P \ast Q$ the gpd which is the product $(p_1q_1, \ldots, p_1q_n, \ldots, p_mq_1, \ldots, p_mq_n)$. Also, we denote by $P \cup Q$ the gpd $(p_1, \ldots, p_m, q_1, \ldots, q_n)$ whenever it is well defined, i.e., whenever $w(P) + w(Q) \leq 1$. We impose the following axioms on entropy $H$:

Axiom A.1 (Symmetry) $H(Q)$ is a symmetric function of the elements of $Q$.

Axiom A.2 (Continuity) $H((q))$ is continuous in the interval $(0,1]$. (Note that $(q)$ is the gpd consisting of the single probability $q$.)

Axiom A.3 (Calibration) $H((\frac{1}{2})) = 1$.

Axiom A.4 (Additivity) $H(P \ast Q) = H(P) + H(Q)$.

Axiom A.5 (Mean-Value Property) There is a strictly monotone and continuous function $g$ such that, for any $P, Q$ with $w(P) + w(Q) \leq 1$:

$$H(P \cup Q) = g^{-1}\left[\frac{w(P)g(H(P)) + w(Q)g(H(Q))}{w(P) + w(Q)}\right].$$

Axiom A.6 (Certainty) $H((q, 1-q)) \downarrow 0$ as $q \downarrow 0$.

Note that Axiom A.1 is built into the set-up. The value $H((q))$ could in principle differ according to which of the $x_1, \ldots, x_n$ gets weight $q$. Under symmetry, only the quantity $q$ matters and so $H((q))$ is well defined.

Theorem A.1 Axioms A.1-A.6 hold if and only if:
\[ H(Q) := H_{\alpha}(Q) = \begin{cases} \frac{1}{1-\alpha} \log_2 \left( \frac{\sum_i q_i^\alpha}{\sum_i q_i} \right) & \text{if } \alpha \neq 1, \\ -\sum_i q_i \log_2 q_i & \text{if } \alpha = 1. \end{cases} \]

where \( \alpha \) is a free parameter with \( 0 < \alpha < \infty \).

Two lemmas are key to proving the theorem.

**Lemma A.2** Under Axioms A.2, A.3, and A.4, if \( q \neq 0 \), then \( H((q)) = -\log_2 q \).

**Proof.** Let \( h(q) := H((q)) \). Then \( h(pq) = h(p) + h(q) \) by Axiom A.4. This is Cauchy’s logarithmic functional equation (Aczél and Dhombres [4], Chapter 5). Using Axioms A.2 and A.3, we find the unique solution \( h(q) = -\log_2 q \).

**Lemma A.3** Under Axioms A.4 and A.5, we have \( g(x) = -ax + b \) (linear) or \( g(x) = a^2(1-a)x + b \) (exponential), where \( a \neq 0, b, \) and \( \alpha \neq 1 \) are arbitrary constants.

**Proof.** We reproduce the argument in Daróczy [15]. Let \( P, Q \) be gpd’s. From Axiom A.5 and Lemma A.2, we obtain:

\[ H(Q) = H((q_1) \cup \cdots \cup (q_n)) = g^{-1}\left[ \sum_j w((q_j)) g(H((q_j))) \right] = g^{-1}\left[ \sum_j q_j g(-\log_2 q_j) \right]. \]

We also have, using Axiom A.4:

\[ g^{-1}\left[ \sum_{i,j} p_i q_j g(-\log_2 p_i q_j) \right] = g^{-1}\left[ \sum_i p_i g(-\log_2 p_i) \right] + g^{-1}\left[ \sum_j q_j g(-\log_2 q_j) \right]. \]

Now let \( f(t) = g(-\log_2 t) \). Substituting, we get:

\[ f^{-1}\left[ \sum_{i,j} p_i q_j f(p_i q_j) \right] = f^{-1}\left[ \sum_i p_i f(p_i) \right] \times f^{-1}\left[ \sum_j q_j f(q_j) \right]. \]

Setting \( Q = ((q)) \) (so that \( q \neq 0 \)), this becomes:

\[ \frac{1}{q} f^{-1}\left[ \sum_i p_i f(p_i q) \right] = f^{-1}\left[ \sum_i p_i f(p_i) \right]. \]

Define \( h_q(t) = f(qt) \). Then:

\[ h_q^{-1}\left[ \sum_i p_i h_q(p_i) \right] = f^{-1}\left[ \sum_i p_i f(p_i) \right]. \]

This shows that \( h_q \) and \( f \) generate the same means. By a theorem on mean values (Hardy, Littlewood, and Pólya [21], Theorem 84), and using Axiom A.2, this implies that:

\[ h_q(t) = a(q)f(t) + b(t), \]

where \( a(q) \) and \( b(q) \) are constants (i.e., do not depend on \( z \)) and \( a(q) \neq 0 \). Substituting, we get:

\[ f(qt) = \alpha(q)f(t) + b(q). \]
Again using Hardy, Littlewood, and Pólya [21] (this step also appears in Rényi [32], pp.557-558), this functional equation has the solution:

\[ f(t) = a \log_2 t + b, \]

or:

\[ f(t) = at^{\alpha-1} + b, \]

where \( a \neq 0, b, \) and \( \alpha \neq 1 \) are arbitrary constants. Recalling the definition of \( f, \) we then find that either:

\[ g(x) = -ax + b, \]

or

\[ g(x) = a2^{(1-\alpha)x} + b, \]

as required. ■

**Proof of Theorem A.1.** To complete the proof of sufficiency of Axioms A.1-A.6, first observe that if, per Lemma A.3, \( g \) is linear, then by induction on Axiom A.5 we obtain Shannon entropy:

\[ H(Q) = \sum_i q_i \log_2 q_i \sum_i q_i. \]

Likewise, if, per Lemma A.3, \( g \) is exponential, then again by induction on Axiom A.5 we obtain Rényi entropy:

\[ H(Q) = \frac{1}{1-\alpha} \log_2 \left( \frac{\sum q_i^\alpha}{\sum q_i} \right). \]

Finally, Axiom A.6 is easily seen to rule out rule out \( \alpha \leq 0. \)

The proof of necessity of Axioms A.1-A.6 is a straightforward calculation. ■

This axiomatization applies to non-negative (unsigned) probability distributions. In the text, we (boldly?) applied the formula for Rényi entropy to signed probability distributions as well. More satisfactory would be to re-derive the formula by modifying (as needed) the axioms to apply to signed probability distributions. This remains to be done.

**References**

[1] Abramsky, S., and A. Brandenburger, “The Sheaf-Theoretic Structure of Non-Locality and Contextuality,” *New Journal of Physics*, 13, 2011, 113036.

[2] Abramsky, S., and A. Brandenburger, “An Operational Interpretation of Negative Probabilities and No-Signalling Models,” unpublished, 2014, at [http://arxiv.org/abs/1401.2561](http://arxiv.org/abs/1401.2561).

[3] Aczéľ, J., and Z. Daróczy, *On Measures of Information and Their Characterizations*, Academic Press, 1975.

[4] Aczéľ, J., and J. Dhombres, *Functional Equations in Several Variables*, Cambridge University Press, 1989.
[5] Beckner, W., “Inequalities in Fourier Analysis,” Annals of Mathematics, 102, 1975, 159-182.

[6] Bell, J., “On the Einstein-Podolsky-Rosen Paradox,” Physics, 1, 1964, 195-200.

[7] Bialynicki-Birula, I., “Rényi Entropy and the Uncertainty Relations,” in Adenier, G., C. Fuchs, and A. Khrennikov (eds.), Foundations of Probability and Physics, AIP Conference Proceedings 889, American Institute of Physics, Melville, 2007, 52.

[8] Bialynicki-Birula, I., and J. Mycielski, “Uncertainty Relations for Information Entropy in Wave Mechanics,” Communications in Mathematical Physics, 44, 1975, 129-132.

[9] Birkhoff, G., and J. von Neumann, “The Logic of Quantum Mechanics,” Annals of Mathematics, 37, 1936, 823-843.

[10] Boyd, S., and L. Vandenberghe, Convex Optimization, Cambridge University Press, 2004.

[11] Brukner, C., and A. Zeilinger, “Information Invariance and Quantum Probabilities,” Foundations of Physics, 39, 2009, 677-689.

[12] Chiribella, G., G. D’Ariano, and P. Perinotti, “Informational Derivation of Quantum Theory,” Physical Review A, 84, 2011, 012311.

[13] Cover, T., and J. Thomas, Elements of Information Theory, 2nd edition, Wiley-Interscience, 2006.

[14] Dakić, B., and C. Brukner, “Quantum Theory and Beyond: Is Entanglement Special?” in Halvorson, H. (ed.), Deep Beauty: Understanding the Quantum World through Mathematical Innovation, Cambridge University Press, 2011, 365-392.

[15] Daroczy, Z., “Über die gemeinsame Charakterisierung der zu den nicht vollständigen Verteilungen gehörigen Entropien von Shannon und von Rényi,” Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, 1, 1963, 381-388.

[16] Dirac, P., “The Physical Interpretation of Quantum Mechanics,” Proceedings of the Royal Society A, 180, 1942, 1-40.

[17] Everett, H., On the Foundations of Quantum Mechanics, doctoral dissertation, Princeton University, 1957.

[18] Feynman, R., “Negative Probability,” in Hiley, B., and F.D. Peat, Quantum Implications: Essays in Honour of David Bohm, Routledge, 1987, 235-248.

[19] Fuchs, C., “Quantum Mechanics as Quantum Information (and only a little more),” in Khrennikov, A., (ed.), Quantum Theory: Reconsideration of Foundations, Växjö University Press, 2002, 463-543. Also available at http://arxiv.org/abs/quant-ph/0205039

[20] Grover, L., “A Fast Quantum Mechanical Algorithm for Database Search,” Proceedings of the Twenty-Eighth Annual ACM Symposium on Theory of Computing, 1996, 212-219.

[21] Hardy, G., J. Littlewood, and G. Pólya, Inequalities, 2nd edition, Cambridge University Press, 1952.

[22] Hardy, L., “Quantum Mechanics from Five Reasonable Axioms,” 2001, at http://arxiv.org/abs/quant-ph/0101012
[23] Heisenberg, W., “Uber den anschulichen Inhalt der quantentheoretischen Kinematik und Mechanik,” *Zeitschrift fur Physik*, 43, 1927, 172-198.

[24] Hillery, M., R. O’Connell, M. Scully, and E. Wigner, “Distribution Functions in Physics: Fundamentals,” *Physics Reports*, 106, 1984, 121-167.

[25] Hirschman, I., “A Note on Entropy,” *American Journal of Mathematics*, 79, 1957, 152-156.

[26] Jaynes, E., “Information Theory and Statistical Mechanics,” *The Physical Review*, 106, 1957, 620-630.

[27] Jaynes, E., “Information Theory and Statistical Mechanics II,” *The Physical Review*, 108, 1957, 171-190.

[28] Oppenheim, J., and S. Wehner, “The Uncertainty Principle Determines the Nonlocality of Quantum Mechanics,” *Science*, 330, 2010, 1072-1074.

[29] Pawlowski M., T. Paterek, D. Kaszlikowski, V. Scarani, A. Winter, and M. Zukowski, “Information Causality as a Physical Principle,” *Nature*, 461, 2009, 1101-1104.

[30] Popescu, S., and D. Rohrlich, “Quantum Nonlocality as an Axiom,” *Foundations of Physics*, 4, 1994, 379-385.

[31] Pusey, M., J. Barrett, and T. Rudolph, “On the Reality of the Quantum State,” *Nature Physics*, 8, 2012, 475478.

[32] Rényi, A., “On Measures of Information and Entropy,” in Neymann, J., (ed.), *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1961, 547-561.

[33] Shannon, C., “A Mathematical Theory of Communication,” *Bell System Technical Journal*, 27, 1948, 379-423 and 623-656.

[34] Schlipp, P. *Albert Einstein, Philosopher-Scientist: The Library of Living Philosophers Volume VII*, 3rd edition, Open Court, 1998.

[35] Schrödinger, E., “Discussion of Probability Relations between Separated Systems,” *Proceedings of the Cambridge Philosophical Society.*, 31, 1935, 555-563.

[36] Shor, P., “Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer,” *SIAM Journal of Computing*, 26, 1997, 1484-1509.

[37] Spekkens, R., “Evidence for the Epistemic View of Quantum States: A Toy Theory,” *Physical Review A*, 75, 2007, 032110.

[38] Tsallis, C., “Possible Generalizations of the Boltzmann-Gibbs Statistics,” *Journal of Statistical Physics*, 52, 1988, 479-487.

[39] Van Dam, W., “Implausible Consequences of Superstrong Nonlocality,” 2005, at [http://arxiv.org/abs/quant-ph/0501159](http://arxiv.org/abs/quant-ph/0501159).

[40] Wehner, S., and A. Winter, “Entropic Uncertainty Relations — A Survey,” *New Journal of Physics*, 12, 2010, 025009.
[41] Wigner, E., “On the Quantum Correction For Thermodynamic Equilibrium,” Physical Review, 40, 1932, 749-759.

[42] Wigner, E., “Quantum Mechanical Distribution Functions Revisited,” in Yourgrau, W., and A. van der Merwe (eds.), Perspectives in Quantum Theory, MIT Press, 1971, 25-36.

[43] Wootters, W., “Quantum Mechanics without Probability Amplitudes,” Foundations of Physics, 16, 1986, 391-405.

[44] Wootters, W., “Local Accessibility of Quantum States,” in Zurek, W. (ed.), Complexity, Entropy and the Physics of Information, Volume VIII, Santa Fe Institute Studies in the Science of Complexity, Westview Press, 1990, 39-46.

[45] Wootters, W., “A Wigner-Function Formulation of Finite-State Quantum Mechanics,” Annals of Physics, 176, 1987, 1-21.