

Article

# Rényi Entropy, Signed Probabilities, and the Qubit

Adam Brandenburger <sup>1,\*</sup>, Pierfrancesco La Mura <sup>2,†</sup> and Stuart Zoble <sup>3,†</sup>

<sup>1</sup> Stern School of Business, Tandon School of Engineering, NYU Shanghai, New York University, New York, NY 10012, USA

<sup>2</sup> HHL—Leipzig Graduate School of Management, 04109 Leipzig, Germany

<sup>3</sup> Signal Fox, Princeton, NJ 08542, USA

\* Correspondence: adam.brandenburger@stern.nyu.edu

† These authors contributed equally to this work.

**Abstract:** The states of the qubit, the basic unit of quantum information, are  $2 \times 2$  positive semi-definite Hermitian matrices with trace 1. We contribute to the program to axiomatize quantum mechanics by characterizing these states in terms of an entropic uncertainty principle formulated on an eight-point phase space. We do this by employing Rényi entropy (a generalization of Shannon entropy) suitably defined for the signed phase-space probability distributions that arise in representing quantum states.

**Keywords:** Rényi entropy; signed probability; uncertainty principle; qubit



**Citation:** Brandenburger, A.; La Mura, P.; Zoble, S. Rényi Entropy, Signed Probabilities, and the Qubit. *Entropy* **2022**, *24*, 1412. <https://doi.org/10.3390/e24101412>

Academic Editors: Yong Deng, Philip Broadbridge and Oleg Olendski

Received: 23 August 2022

Accepted: 30 September 2022

Published: 3 October 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The maximum entropy method was introduced into physics as a way of deriving the Boltzmann distribution of statistical mechanics (Jaynes [1]). In this paper, we apply entropy methods to characterize the basic unit of quantum information, namely, the qubit. Our work fits into the ongoing program to identify principles or axioms yielding quantum mechanics. This program goes back at least to Birkhoff and von Neumann [2] and their investigation of quantum mechanics as a non-classical logic. More recently, Hardy [3] reconstructed quantum theory from five axioms couched in terms of operations that can be conducted on a physical system. His work spurred many other axiomatizations based on communication complexity (Van Dam [4]), information causality (Pawlowski et al. [5]), information capacity (Dakić and Brukner [6]), and purification (Chiribella, et al. [7]), among other principles.

We aim to characterize the simplest quantum system, namely, a two-level system such as the spin of a particle. Empirically, the experimenter can observe a property such as spin in three arbitrarily chosen mutually orthogonal directions. In each direction, the outcome is binary (up or down). An empirical model gives the frequencies of these outcomes when identical copies of the same two-level system are prepared and one of the three measurements is performed on a given copy. We want to associate an entropy with an empirical model. This step is not immediate because entropy is a measure of the uncertainty in a single probability distribution, and an empirical model contains three distinct probability distributions (one for each direction). The solution is to move to phase space, where an empirical model is represented by a single probability distribution. Our phase space for a two-level system contains eight points, where each point specifies the outcome (up or down) of each of the three possible measurements. The possibility of a non-deterministic response to measurement—as in quantum mechanics—is allowed for by specifying a probability of each point in phase space. Later, we comment on the relationship between our phase-space framework and the finite-field representations due to Wootters and collaborators (e.g., Wootters [8]; Gibbons et al. [9]).

Our phase-space representation of an empirical model can be thought of as a particular (canonical) type of local hidden-variable model (Bell [10]), where the possible values of

the hidden variable are precisely the possible points in phase space. It follows from Bell's Theorem (Bell [10]) that there are empirical models which arise in quantum mechanics and which cannot be represented in phase space with ordinary probabilities. One answer is found in the Wigner distribution (Wigner [11]), which can take on negative values at certain locations in phase space. Dirac [12] and Feynman [13] also argued for admitting negative probabilities in quantum calculations. We emphasize that although we will allow phase-space probabilities to take negative values, the frequencies of all observable events remain non-negative. Further support for the use of negative probabilities comes from Abramsky and Brandenburger [14], who prove that the family of empirical models that can be represented in phase space this way is precisely the family of no-signaling theories (Popescu and Rohrlich [15]).

We are now ready to associate entropies with probability distributions on phase space. Within quantum mechanics, the most common entropy measure is the von Neumann entropy (von Neumann [16]). This is unsuitable for our purpose because it is defined within the quantum formalism, which we want to derive not assume. Shannon entropy (Shannon [17]) is also unsuitable when applied to probabilities in phase space, since, if the latter can be negative, then it would take complex values. Instead, we work with the more basic notion of Rényi entropy [18] and impose real-valuedness and smoothness. Rényi entropy satisfies the basic requirement of extensivity, i.e., it is additive across statistically independent systems. In fact, it is defined by this property together with some technical axioms (Daróczy [19]). Rényi entropy is used in various applications in quantum mechanics (Wehner and Winter [20]; Bialynicki-Birula and Rudnicki [21]; Coles et al. [22]).

The next step is to state an entropic uncertainty principle as an axiom on phase space. Note that different from other entropic uncertainty principles in quantum mechanics (Everett [23]; Hirschman [24]; Beckner [25]; Bialynicki-Birula and Mycielski [26]), our principle is formulated in phase space. Furthermore, we do not derive the principle from quantum mechanics but introduce it as an axiom. Our main result is that the set of probability distributions on phase space whose Rényi entropy exceeds a certain lower bound is exactly equal to the set of probability distributions that induce the qubit.

A paper that, broadly speaking, goes in the opposite direction to ours is Wootters and Sussman [27]. These authors work in a finite-field phase-space representation of discrete quantum systems and are able to identify certain minimum-uncertainty states. In particular, they show that a particular class of states (the "rotationally invariant states") minimize Rényi 2-entropy (a special case of our family of entropy functionals, as we shall see).

We consider our axiomatization of the qubit as in line with the program enunciated by Fuchs [28] to find "deep *physical* principles" that yield quantum mechanics. This said, we do not claim that our axiom is self-evident. In relativity theory, the principle of light speed invariance is not an intuitive axiom—the point is that it is physically intelligible. (This comparison between quantum theory and relativity theory is also made in Ongadinata et al. [29].) Our interest in an uncertainty principle is similar. Ever since the initial formulation by Heisenberg [30], uncertainty principles have been viewed as one of the notably unintuitive features of quantum mechanics. However, even though they are mysterious at the everyday macroscopic level, uncertainty principles are physically interpretable, and they are evidently true of microscopic systems.

## 2. Preliminaries

A basis for the space of  $2 \times 2$  Hermitian matrices is given by  $\{\sigma_0, \sigma_1, \sigma_2, \sigma_3\}$ , where  $\sigma_0 = \mathbf{I}$  is the  $2 \times 2$  identity matrix and  $\sigma_1, \sigma_2, \sigma_3$  are the Pauli matrices

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

A  $2 \times 2$  Hermitian matrix  $\mathbf{M}$  satisfies  $\text{Tr}(\mathbf{M}) = 1$  if and only if

$$\mathbf{M} = \frac{1}{2}(\mathbf{I} + r_1\sigma_1 + r_2\sigma_2 + r_3\sigma_3)$$

for some vector  $\mathbf{r} = (r_1, r_2, r_3) \in \mathbb{R}^3$ .

**Definition 1.** A  $2 \times 2$  Hermitian matrix  $\mathbf{M}$  with  $\text{Tr}(\mathbf{M}) = 1$  is called a potential quantum state. If, in addition,  $\mathbf{M}$  is positive semi-definite, then  $\mathbf{M}$  is a quantum state, or a state of the qubit. We also refer to the corresponding vectors  $\mathbf{r}$  as potential quantum states and quantum states.

This is the model of the simplest quantum system, namely a two-level system such as the spin of a particle. Empirically, the experimenter can observe a property such as spin in three arbitrarily chosen mutually orthogonal directions  $x_1, x_2$ , and  $x_3$ . In each direction, the outcome of a measurement will be labeled  $+1$  or  $-1$ . The expectation of the outcome in direction  $i$  is (see, e.g., p.181 in Sakurai and Napolitano [31]).

$$\text{Tr}(\mathbf{M}\sigma_i) = r_i.$$

We want to associate an entropy with an empirical model. This step is not immediate because entropy is a measure of the uncertainty in a single probability distribution, and an empirical model contains three probability distributions (one for each direction). Our solution is to move to phase space, where an empirical model is represented by a single probability distribution. The *phase space* for a two-level system contains eight points,

$$\{+1, -1\}^3 = \{e_n \mid n = 1, \dots, 8\},$$

where  $e_n(i) = (-1)^{n_i}$  for  $(n_1, n_2, n_3)$  the base-2 digits of  $n - 1$ . Each point in phase space specifies the outcome of each of the three possible measurements. Non-deterministic responses to measurement are incorporated by specifying probabilities over the points in phase space. Let

$$Q = \{\mathbf{q} \in \mathbb{R}^8 \mid \sum_{i=1}^8 q_i = 1\}$$

denote the set of all signed probability distributions on phase space. That is, we do not require the probabilities to be positive, only that they sum to 1. We define a map  $\phi$  from  $Q$  to the set of potential quantum states by

$$\phi(\mathbf{q}) = \frac{1}{2}(\mathbf{I} + r_1\sigma_1 + r_2\sigma_2 + r_3\sigma_3),$$

where

$$r_i = \sum_{\{n \mid e_n(i)=+1\}} q_n \times (+1) + \sum_{\{n \mid e_n(i)=-1\}} q_n \times (-1).$$

The map  $\phi$  gives the correct transformation from phase space to the space of potential quantum states, in the sense of preserving the empirical probabilities. This map is linear and it will be helpful to fix some notation surrounding a matrix representation. Note we have folded the condition that  $\mathbf{q}$  is a probability distribution in as the last equation in the definition of representation below.

**Definition 2.** Let  $\mathbf{A}$  denote the matrix

$$\begin{pmatrix} 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}.$$

For  $\mathbf{r} \in \mathbb{R}^3$  define  $\hat{\mathbf{r}} = (r_1, r_2, r_3, 1) \in \mathbb{R}^4$ . For  $\mathbf{q} \in \mathbb{R}^8$  and  $\mathbf{r} \in \mathbb{R}^3$  we say  $\mathbf{q}$  represents  $\mathbf{r}$  if  $\mathbf{A}\mathbf{q} = \hat{\mathbf{r}}$ .

### 3. Rényi Entropy

We are going to use phase space to formulate an entropic uncertainty principle as an axiom, and derive the quantum states this way. In particular, we will allow only those potential quantum states  $r$  for which there is a phase-space representation  $q$  whose entropy exceeds a lower bound. The non-classicality of the qubit becomes apparent because there are quantum states for which the only representations with entropy exceeding the bound are signed probability distributions. The use of negative probabilities on phase space to represent quantum systems goes back to the Wigner quasi-probability distribution (Wigner [11]). The first task then is to choose a suitable definition of entropy for signed probabilities.

We extend Rényi entropy ([18]) to signed probabilities and impose a smoothness condition that identifies a particular family of entropy functionals. Fix a finite set  $X = \{x_1, \dots, x_n\}$  together with an ordinary (unsigned) probability distribution  $q$  on  $X$ . Rényi entropy is the family of functionals

$$H_\alpha(q) = -\frac{1}{\alpha - 1} \log_2\left(\sum_{i=1}^n q_i^\alpha\right),$$

where  $0 < \alpha < \infty$  is a free parameter. (Shannon entropy is the special case, via L'Hôpital's rule, when  $\alpha = 1$ .) We can preserve the real-valuedness of entropy under signed probabilities by taking absolute values

$$H_\alpha(q) = -\frac{1}{\alpha - 1} \log_2\left(\sum_{i=1}^n |q_i|^\alpha\right).$$

This formula can also be derived axiomatically. (see Brandenburger and La Mura [32] who modify the original axioms for Rényi entropy in Rényi [18] and Daróczy [19].) We next impose a smoothness condition, requiring that  $H_\alpha$  be smooth on the space of signed probabilities. That is, we require  $H_\alpha(q_1, \dots, q_{n-1}, (1 - \sum_{i=1}^{n-1} q_i))$  to be  $C^\infty$  on  $\mathbb{R}^{n-1}$ . Now, if  $\alpha$  is not an integer let  $k$  be the least integer with  $k > \alpha$ . Then

$$\frac{\partial^k H_\alpha}{\partial q_i^k}(q) = \frac{f(q)}{g(q)},$$

where  $f(q) \neq 0$  and  $g(q) = 0$  for any  $q$  with  $q_i = 0$ . Thus,  $\alpha$  must be an integer. If  $\alpha$  is an odd integer then  $\partial^\alpha H_\alpha / \partial q_i^\alpha(q)$  is undefined for  $q$  with  $q_i = 0$ . Therefore, Rényi entropy takes the following form under our smoothness assumption.

**Definition 3.** Rényi entropy for signed probability distributions is the family of functionals

$$H_{2k}(q) = -\frac{1}{2k - 1} \log_2\left(\sum_{i=1}^n q_i^{2k}\right) = -\frac{2k}{2k - 1} \log_2(\|q\|_{2k}),$$

where  $k = 1, 2, \dots$  is a free parameter.

Finally in this section, we give an example of a quantum state such that the only representatives with Rényi entropy satisfying the lower bound are signed probabilities. Consider the quantum state  $(r_1, r_2, r_3) = (\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}})$ . Set  $k = 1$ . The (unique) maximum 2-entropy representation is

$$q = \frac{1}{8}(1 + \sqrt{3}, 1 + \frac{1}{\sqrt{3}}, 1 + \frac{1}{\sqrt{3}}, 1 - \frac{1}{\sqrt{3}}, 1 + \frac{1}{\sqrt{3}}, 1 - \frac{1}{\sqrt{3}}, 1 - \frac{1}{\sqrt{3}}, 1 - \sqrt{3}),$$

with negative final component. The 2-entropy of  $q$  is 2, which is the lower bound we impose below, so we cannot find a representation with all non-negative components with sufficiently high 2-entropy. In fact any state with  $|r_1| + |r_2| + |r_3| > 1$  will have this property.

#### 4. Main Theorem

We can now state an entropic uncertainty principle as an axiom on phase space. The axiom is inspired by the use of entropic uncertainty relations in quantum information (Wehner and Winter [20]; Bialynicki-Birula and Rudnicki [21]; Coles et al. [22]).

**Uncertainty Principle:** A potential quantum state  $r$  satisfies the Uncertainty Principle if for every  $k$ , there is a phase-space probability distribution  $q$  that represents  $r$  and satisfies  $H_{2k}(q) \geq 2$ .

This says that we allow as potential quantum states only those states  $r$  containing a minimum amount of uncertainty, as measured by the entropy of a corresponding probability distribution  $q$  on phase space. Note that our Uncertainty Principle is a sequence of conditions, one for each  $k$ . This is because Rényi entropy itself is not a single functional but a sequence of functionals (indexed by  $k$ ).

**Theorem 1.** *The potential quantum states satisfying the Uncertainty Principle are precisely the states of the qubit.*

**Proof.** We first show that the potential quantum states satisfying the Uncertainty Principle at  $k = 1$  are the states of the qubit. Note that

$$H_2(q) \geq 2 \text{ if and only if } \|q\|_2^2 \leq \frac{1}{4}.$$

For a general  $r$ , the representation  $q^*$  which maximizes 2-entropy is given by

$$q^* = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\hat{r}.$$

Using the fact that  $\mathbf{A}\mathbf{A}^T = 8\mathbf{I}$  we have

$$\|q^*\|_2^2 = \hat{r}^T(\mathbf{A}\mathbf{A}^T)^{-1}\hat{r} = \frac{1}{8}r^T r + \frac{1}{8} \leq \frac{1}{4}$$

if and only if

$$r_1^2 + r_2^2 + r_3^2 \leq 1,$$

and the result follows since the matrix  $\frac{1}{2}(\mathbf{I} + r_1\sigma_1 + r_2\sigma_2 + r_3\sigma_3)$  is positive semi-definite if and only if  $r_1^2 + r_2^2 + r_3^2 \leq 1$ .

We now show that if a potential state  $r$  satisfies the Uncertainty Principle at  $k = 1$  then it satisfies the Uncertainty Principle at all  $k$ . This is the main mathematical argument. Fix  $k > 1$  and let  $r \in \mathbb{R}^3$  be a state of the qubit. Choose a  $q$  to maximize the  $2k$ -entropy of a representative of  $r$ . We want to show  $H_{2k}(q) \geq 2$  which is equivalent to  $\|q\|_{2k} \leq (\frac{1}{2})^{\frac{2k-1}{k}}$ .

Observe that  $q$  solves the norm minimization problem

$$\begin{aligned} &\min_{q \in \mathbb{R}^8} \|q\|_{2k} \\ &\text{subject to } \mathbf{A}q = \hat{r}. \end{aligned}$$

The dual problem is

$$\begin{aligned} &\max_{x \in \mathbb{R}^4} \hat{r}^T x \\ &\text{subject to } \|\mathbf{A}^T x\|_{\frac{2k}{2k-1}} \leq 1. \end{aligned}$$

(see pp.221–222 in Boyd and Vandenberghe [33].) Note that  $\|\cdot\|_{\frac{2k}{2k-1}}$  is the dual norm of  $\|\cdot\|_{2k}$ . Strong duality holds so the values of the primal and dual problems are equal. Let  $\mathbf{y}^1, \mathbf{y}^k$  be the maximizers of the dual problems for 2-entropy and  $2k$ -entropy, respectively. Let

$$C_1 = \{\mathbf{x} \in \mathbb{R}^4 \mid \|\mathbf{A}^T \mathbf{x}\|_2 \leq 1\}$$

and

$$C_k = \{\mathbf{x} \in \mathbb{R}^4 \mid \|\mathbf{A}^T \mathbf{x}\|_{\frac{2k}{2k-1}} \leq 1\}.$$

Note that  $C_k \subseteq C_1$  are both convex and, in fact,  $C_1$  is the ball of radius  $\frac{1}{\sqrt{8}}$ . Let

$$\mathbf{z}^k = (\hat{\mathbf{r}}^T \mathbf{y}^k / \|\hat{\mathbf{r}}\|_2^2) \hat{\mathbf{r}}$$

be the projection of  $\mathbf{y}^k$  onto  $\hat{\mathbf{r}}$ . Since  $\hat{\mathbf{r}}^T \mathbf{y}^k = \frac{\|\hat{\mathbf{r}}\|_2}{\sqrt{8}} \cos \theta$ , where  $\theta$  is the angle between them, we must have  $\theta = 0$  and so

$$\mathbf{y}^1 = (\hat{\mathbf{r}}^T \mathbf{y}^1 / \|\hat{\mathbf{r}}\|_2^2) \hat{\mathbf{r}}.$$

Since the values of the primal and dual problems are equal, these values are positive, so  $\frac{\|\mathbf{z}^k\|_2}{\|\mathbf{y}^1\|_2}$  is equal to the ratio of the value of the general  $k$  problem to the value of the  $k = 1$  problem. By assumption

$$\hat{\mathbf{r}}^T \mathbf{y}^1 \leq \frac{1}{2},$$

so it is enough to show

$$\frac{\|\mathbf{z}^k\|_2}{\|\mathbf{y}^1\|_2} \leq \left(\frac{1}{2}\right)^{\frac{k-1}{k}}.$$

We will bound this expression by a function that can be explicitly maximized. Note that for every nonzero vector  $\mathbf{w}$  there are unique  $\lambda < \nu$  such that

$$\|\mathbf{A}^T \nu \mathbf{w}\|_2 = 1$$

and

$$\|\mathbf{A}^T \lambda \mathbf{w}\|_{\frac{2k}{2k-1}} = 1.$$

This follows immediately from linearity, homogeneity, the fact that  $\mathbf{A}$  has full rank, and the fact that  $\frac{2k}{2k-1} < 2$ . Now let

$$f(\mathbf{w}) = \frac{\|\mathbf{A}^T \mathbf{w}\|_2}{\|\mathbf{A}^T \mathbf{w}\|_{\frac{2k}{2k-1}}}.$$

By the previous observation and the fact that  $f(\lambda \mathbf{w}) = f(\mathbf{w})$  for any nonzero scalar  $\lambda$ , we see that  $f(\mathbf{w})$  is the ratio of the distance to the boundary of  $C_1$  along the ray through  $\mathbf{w}$  to the distance to the boundary of  $C_k$ . Let  $\mathbf{w}^1 = \nu \mathbf{y}^k$  belong to the boundary of  $C_1$ . Figure 1 depicts the situation in the plane containing  $\hat{\mathbf{r}}$  and  $\mathbf{y}^k$ .

**Claim 1.**  $\frac{\|\mathbf{z}^k\|_2}{\|\mathbf{y}^1\|_2}$  is bounded by a value of  $f$ .

**Proof.** We claim that

$$\frac{\|\mathbf{z}^k\|_2}{\|\mathbf{y}^1\|_2} \leq \frac{\|\mathbf{y}^k\|_2}{\|\mathbf{w}^1\|_2}.$$

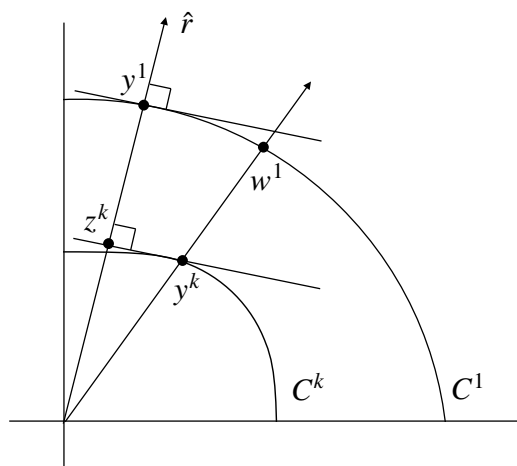


Figure 1. Comparing the maximizers of the dual problem.

Note that

$$\hat{r}^T w^1 \leq \hat{r}^T y^1,$$

so the length of the projection of  $w^1$  onto  $\hat{r}$  (call this vector  $v$ ) cannot exceed the length of  $y^1$ . By similar triangles then

$$\frac{\|y^k\|_2}{\|z^k\|_2} = \frac{\|w^1\|_2}{\|v\|_2} \geq \frac{\|w^1\|_2}{\|y^1\|_2},$$

so  $\frac{\|z^k\|_2}{\|y^k\|_2} \leq \frac{\|y^1\|_2}{\|w^1\|_2} = f(w^1)$ . □

To complete the proof of Theorem 1, it suffices to show that

$$\max\{f(w) \mid w \in \mathbb{R}^4\} = \left(\frac{1}{2}\right)^{\frac{k-1}{k}},$$

which we do in the Appendix A. □

### 5. Conclusions

We have shown that an entropic Uncertainty Principle formulated on an eight-point phase space characterizes the states of the qubit. We see our result as contributing to the program that aims to reconstruct quantum mechanics from physically interpretable axioms. Of course, our derivation is only for the simplest, two-level quantum system. We anticipate that to characterize an  $n$ -qubit system, methods will be needed that go beyond those in this paper. In particular, it may be necessary not only to extend our entropic Uncertainty Principle to the  $n$ -qubit case, but to identify new axioms. The Wootters and Sussman [27] analysis may be an important guide in this direction in that they are able to identify certain minimum-uncertainty states in an  $n$ -qubit system. A full characterization may be possible combining techniques across the two papers.

We re-emphasize that our paper is aimed at a derivation not a representation of the qubit. This explains why our phase space contains eight points while in Wootters [8] and Gibbons et al. [9] the phase space for a single qubit comprises four points. A four-point space is the appropriate domain for the discrete Wigner function, but we do not assume a Wigner representation.

A related derivation of the qubit is Onggadinata et al. [29]. Similar to our paper, they employ Rényi entropy, but fix  $\alpha = 2$ . This instance of Rényi entropy is often called collision entropy. Their postulate is that the collision entropy is constant under any dynamics (not necessarily deterministic) on a finite one-dimensional lattice. From this, they recover the qubit with its full dynamics as defined on the Bloch sphere.

Both Wootters and Sussman [27] and Onggadinata et al. [29] work with Rényi 2-entropy. By contrast, we are able to obtain our results not just for 2-entropy, but for the entire family of  $2k$ -entropies, which we derived from basic principles. It would be interesting to see if our methods could be employed to generalize the results in these papers.

**Author Contributions:** A.B., P.L.M. and S.Z. did the research and prepared the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by NYU Stern School of Business, NYU Abu Dhabi, NYU Shanghai, J.P. Valles, the HHL—Leipzig Graduate School of Management, and the Deutsche Bundesbank.

**Institutional Review Board Statement:** Not applicable.

**Data Availability Statement:** No new data were created or analyzed in this study. Data sharing is not applicable to this article.

**Acknowledgments:** We are grateful to three referees and two associate editors for suggestions that importantly improved this paper. Samson Abramsky, Časlav Brukner, Matthew Deutsch, Christopher Fuchs, Leslie Greengard, Michael J.W. Hall, Jerry Keisler, Elliot Lipnowski, Kelvin Onggadinata, David Pine, Rui Soares Barbosa, Blake Stacey, Yuri Tschinkel, and Noson Yanofsky provided valuable input.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the writing of the manuscript; or in the decision to publish the results.

### Appendix A

Let  $w \in \mathbb{R}^4$  and  $v = wA$ . Let  $t \in \mathbb{R}^8$  be defined by  $t_i = v_i^{\frac{1}{2k-1}}$ . Note that the critical points of  $f$  are the same as the critical points of

$$\frac{f^2(w)}{8} = \frac{\|v\|_2^2}{8\|v\|_{2k/2k-1}^2} = \frac{wAA^T w^T}{8\|v\|_{2k/2k-1}^2} = \frac{\|w\|_2^2}{\|v\|_{2k/2k-1}^2},$$

which are the solutions of the system of first-order conditions

$$w_i = h(w)r_i t^T \quad i = 1, 2, 3, 4,$$

where

$$h(w) = \frac{\|w\|_2^2}{\|v\|_{2k/2k-1}^2} > 0$$

and  $r_i$  is the  $i$ th row of the matrix  $A$ . It is helpful to write out the system with  $\gamma$  denoting  $\frac{1}{2k-1}$  for readability:

$$w_1 = h(w)[(w_1 + w_2 + w_3 + w_4)^\gamma - (-w_1 + w_2 + w_3 + w_4)^\gamma + (w_1 - w_2 + w_3 + w_4)^\gamma - (-w_1 - w_2 + w_3 + w_4)^\gamma + (w_1 + w_2 - w_3 + w_4)^\gamma - (-w_1 + w_2 - w_3 + w_4)^\gamma + (w_1 - w_2 - w_3 + w_4)^\gamma - (-w_1 - w_2 - w_3 + w_4)^\gamma],$$

$$w_2 = h(w)[(w_1 + w_2 + w_3 + w_4)^\gamma + (-w_1 + w_2 + w_3 + w_4)^\gamma - (w_1 - w_2 + w_3 + w_4)^\gamma - (-w_1 - w_2 + w_3 + w_4)^\gamma + (w_1 + w_2 - w_3 + w_4)^\gamma + (-w_1 + w_2 - w_3 + w_4)^\gamma - (w_1 - w_2 - w_3 + w_4)^\gamma - (-w_1 - w_2 - w_3 + w_4)^\gamma],$$

$$w_3 = h(w)[(w_1 + w_2 + w_3 + w_4)^\gamma + (-w_1 + w_2 + w_3 + w_4)^\gamma + (w_1 - w_2 + w_3 + w_4)^\gamma +$$



$$(-w_1 - w_2 + w_3 + w_4)^\gamma - (w_1 + w_2 - w_3 + w_4)^\gamma - (-w_1 + w_2 - w_3 + w_4)^\gamma - (w_1 - w_2 - w_3 + w_4)^\gamma - (-w_1 - w_2 - w_3 + w_4)^\gamma],$$

$$w_4 = h(\mathbf{w})[(w_1 + w_2 + w_3 + w_4)^\gamma + (-w_1 + w_2 + w_3 + w_4)^\gamma + (w_1 - w_2 + w_3 + w_4)^\gamma + (-w_1 - w_2 + w_3 + w_4)^\gamma + (w_1 + w_2 - w_3 + w_4)^\gamma + (-w_1 + w_2 - w_3 + w_4)^\gamma + (w_1 - w_2 - w_3 + w_4)^\gamma + (-w_1 - w_2 - w_3 + w_4)^\gamma].$$

**Claim A1.** The system  $\mathbf{w} = h(\mathbf{w})\mathbf{A}\mathbf{t}^T$  has the following properties:

1. If  $\mathbf{w}$  is a solution then so is  $\lambda\mathbf{w}$  for any  $\lambda \neq 0$ .
2. If  $\mathbf{w}$  is a solution then  $\mathbf{v}$  is a solution, where  $\mathbf{v}$  is obtained from  $\mathbf{w}$  by permuting coordinates.

**Proof.** For (1) we have  $h(\lambda\mathbf{w})\mathbf{t}^T(\lambda\mathbf{w}) = \frac{\lambda^2\lambda^{1/2k-1}}{\lambda^{2k/2k-1}}h(\mathbf{w})\mathbf{t}^T = \lambda\mathbf{w}$ . For (2) we have

$$w_1 = h(w_4, w_2, w_3, w_1)r_4\mathbf{t}^T(w_4, w_2, w_3, w_1)$$

and

$$w_4 = h(w_4, w_2, w_3, w_1)r_1\mathbf{t}^T(w_4, w_2, w_3, w_1),$$

and similarly for  $w_2, w_3$ .  $\square$

**Claim A2.** Assume  $w_4 \neq 0$ . Let  $i, j < 4$ . Then

$$|w_i| = |w_j| \text{ or } w_i w_j = 0.$$

**Proof.** We may assume  $w_4 > 0$ . For  $N$  sufficiently large we have  $\|\mathbf{w} - \mathbf{a}\|_2 < \|\mathbf{a}\|_2$  where  $\mathbf{a} = (0, 0, 0, N)$ . Thus, the Taylor series expansion of  $r_i\mathbf{t}^T$  at the point  $\mathbf{a}$  converges at  $\mathbf{w}$ . We have

$$w_1 = 8h(\mathbf{w}) \sum_{\substack{\alpha_1 \in O \\ \alpha_2, \alpha_3 \in E \\ \alpha_4 \in \mathbb{N}}} \frac{(\mathbf{w} - \mathbf{a})^\alpha}{\alpha!} C\left(\sum_{i=1}^4 \alpha_i\right),$$

$$w_2 = 8h(\mathbf{w}) \sum_{\substack{\alpha_2 \in O \\ \alpha_1, \alpha_3 \in E \\ \alpha_4 \in \mathbb{N}}} \frac{(\mathbf{w} - \mathbf{a})^\alpha}{\alpha!} C\left(\sum_{i=1}^4 \alpha_i\right),$$

$$w_3 = 8h(\mathbf{w}) \sum_{\substack{\alpha_3 \in O \\ \alpha_1, \alpha_2 \in E \\ \alpha_4 \in \mathbb{N}}} \frac{(\mathbf{w} - \mathbf{a})^\alpha}{\alpha!} C\left(\sum_{i=1}^4 \alpha_i\right),$$

$$w_4 = 8h(\mathbf{w}) \sum_{\substack{\alpha_1, \alpha_2, \alpha_3 \in E \\ \alpha_4 \in \mathbb{N}}} \frac{(\mathbf{w} - \mathbf{a})^\alpha}{\alpha!} C\left(\sum_{i=1}^4 \alpha_i\right),$$

where  $\alpha \in \mathbb{N}^4$  is a multi-index,  $\mathbb{N} = E \cup O = \{0, 2, 4, \dots\} \cup \{1, 3, 5, \dots\}$ ,

$$\alpha! = \alpha_1! \alpha_2! \alpha_3! \alpha_4!,$$

$$(\mathbf{w} - \mathbf{a})^\alpha = w_1^{\alpha_1} w_2^{\alpha_2} w_3^{\alpha_3} (w_4 - N)^{\alpha_4},$$

and  $C$  is defined by

$$C(0) = 1, C(1) = \frac{1}{(2k-1)N^{\frac{2k}{2k-1}}}, \text{ and}$$

$$C(n) = \frac{(-1)^{n-1} \prod_{j=1}^{n-1} (j(2k-1) - 1)}{(2k-1)^n N^{\frac{n(2k-1)-1}{2k-1}}} \text{ for } n > 1.$$

Note that  $C(\sum_{i=1}^4 \alpha_i) > 0$  if and only if  $\sum_{i=1}^4 \alpha_i \in O$ . Assume  $w_1, w_2 \neq 0$ . We have

$$w_1 \sum_{\substack{\alpha_2 \in O \\ \alpha_1, \alpha_3 \in E \\ \alpha_4 \in \mathbb{N}}} \frac{(w-a)^\alpha}{\alpha!} C(\sum_{i=1}^4 \alpha_i) = w_2 \sum_{\substack{\alpha_1 \in O \\ \alpha_2, \alpha_3 \in E \\ \alpha_4 \in \mathbb{N}}} \frac{(w-a)^\alpha}{\alpha!} C(\sum_{i=1}^4 \alpha_i),$$

equivalently

$$\sum_{\substack{\alpha_2 \in O \\ \alpha_1, \alpha_3 \in E \\ \alpha_4 \in \mathbb{N}}} \frac{(w-a)^{\alpha+(1,0,0,0)}}{\alpha!} C(\sum_{i=1}^4 \alpha_i) = \sum_{\substack{\alpha_1 \in O \\ \alpha_2, \alpha_3 \in E \\ \alpha_4 \in \mathbb{N}}} \frac{(w-a)^{\alpha+(0,1,0,0)}}{\alpha!} C(\sum_{i=1}^4 \alpha_i).$$

Re-indexing we have

$$\sum_{\substack{\alpha_1, \alpha_2 \in O \\ \alpha_3 \in E \\ \alpha_4 \in \mathbb{N}}} \alpha_1 \frac{(w-a)^\alpha}{\alpha!} C(\sum_{i=1}^4 \alpha_i - 1) = \sum_{\substack{\alpha_1, \alpha_2 \in O \\ \alpha_3 \in E \\ \alpha_4 \in \mathbb{N}}} \alpha_2 \frac{(w-a)^\alpha}{\alpha!} C(\sum_{i=1}^4 \alpha_i - 1).$$

Collecting terms we have

$$\sum_{\substack{\alpha_1 < \alpha_2 \\ \alpha_1, \alpha_2 \in O \\ \alpha_3 \in E \\ \alpha_4 \in \mathbb{N}}} (\alpha_1 - \alpha_2) (w_1^{\alpha_1} w_2^{\alpha_2} - w_1^{\alpha_2} w_2^{\alpha_1}) w_3^{\alpha_3} (w_4 - N)^{\alpha_4} \frac{C(\sum_{i=1}^4 \alpha_i - 1)}{\alpha!} = 0,$$

equivalently

$$\sum_{\substack{\alpha_1 < \alpha_2 \\ \alpha_1, \alpha_2 \in O \\ \alpha_3 \in E \\ \alpha_4 \in \mathbb{N}}} (\alpha_1 - \alpha_2) w_1^{\alpha_1} w_2^{\alpha_2} (1 - (\frac{w_1}{w_2})^{\alpha_2 - \alpha_1}) w_3^{\alpha_3} (w_4 - N)^{\alpha_4} \frac{C(\sum_{i=1}^4 \alpha_i - 1)}{\alpha!} = 0.$$

The key point is that

$$(\alpha_1 - \alpha_2) w_1^{\alpha_1} w_2^{\alpha_2} w_3^{\alpha_3} (w_4 - N)^{\alpha_4} \frac{C(\sum_{i=1}^4 \alpha_i - 1)}{\alpha!}$$

always has the same sign as  $w_1 w_2$ . Thus, since  $\alpha_2 - \alpha_1 \in E$  we conclude that  $|w_1| = |w_2|$ . □

**Claim A3.** If  $w_i, w_j \neq 0$  then  $|w_i| = |w_j|$ .

**Proof.** By Claim A1 and Claim A2 we may assume that  $w_1, w_4 \neq 0$  and  $w_2, w_3 = 0$ . We may further assume that  $w_4 = 1$ . Thus, the equation for  $w_1$  becomes

$$w_1 = \frac{(w_1 + 1)^{1/2k-1} - (-w_1 + 1)^{1/2k-1}}{(w_1 + 1)^{1/2k-1} + (-w_1 + 1)^{1/2k-1}}$$

so

$$(w_1 + 1)(-w_1 + 1)^{1/2k-1} = (-w_1 + 1)(w_1 + 1)^{1/2k-1},$$

from which we conclude that  $w_1 \in \{-1, 1\}$  as desired.  $\square$

We have thus shown that for every  $i, j \leq 4$  either  $|w_i| = |w_j|$  or  $w_i w_j = 0$  so we need only consider critical points with

$$w_i \in \{0, 1, -1\}$$

for each  $i = 1, \dots, 4$ . It is easy to check that the maximum of the original function  $f$  occurs when exactly two of the weights are 0 and this maximum value is  $(\frac{1}{2})^{\frac{k-1}{k}}$ , completing the proof of Theorem 1.

## References

- Jaynes, E. Information Theory and Statistical Mechanics. *Phys. Rev.* **1957**, *106*, 620–630.
- Birkhoff, G.; von Neumann, J. The Logic of Quantum Mechanics. *Ann. Math.* **1934**, *37*, 823–843.
- Hardy, L. Quantum Theory From Five Reasonable Axioms. 2001. Available online: <https://arxiv.org/abs/quant-ph/0101012> (accessed on 29 September 2022).
- Van Dam, W. Implausible Consequences of Superstrong Nonlocality. 2005. Available online: <https://arxiv.org/abs/quant-ph/0501159> (accessed on 29 September 2022).
- Pawlowski, M.; Paterek, T.; Kaszlikowski, D.; Scarani, V.; Winter, A.; Zukowski, M. Information Causality as a Physical Principle. *Nature* **2009**, *461*, 1101–1104.
- Dakić, B.; Brukner, Č. Quantum Theory and Beyond: Is Entanglement Special? In *Deep Beauty: Understanding the Quantum World through Mathematical Innovation*; Halvorson, H., Ed.; Cambridge University Press: Cambridge, UK, 2011; pp. 365–392.
- Chiribella, G.; D’Ariano, G.; Perinotti, P. Informational Derivation of Quantum Theory. *Phys. Rev. A* **2011**, *84*, 012311.
- Wootters, W. A Wigner-Function Formulation of Finite-State Quantum Mechanics. *Ann. Phys.* **1987**, *176*, 1–21.
- Gibbons, K.; Hoffman, M.; Wootters, W. Discrete Phase Space Based on Finite Fields. *Phys. Rev. A* **2004**, *70*, 062101.
- Bell, J. On the Einstein–Podolsky–Rosen Paradox. *Physics* **1964**, *1*, 195–200.
- Wigner, E. On the Quantum Correction For Thermodynamic Equilibrium. *Phys. Rev.* **1932**, *40*, 749–759.
- Dirac, P. The Physical Interpretation of Quantum Mechanics. *Proc. R. Soc. A* **1942**, *180*, 1–40.
- Feynman R. Negative Probability. In *Quantum Implications: Essays in Honour of David Bohm*; Hiley, B., Peat, F., Eds.; Routledge: London, UK, 1987; pp. 235–248.
- Abramsky, S.; Brandenburger, A. The Sheaf-Theoretic Structure of Non-Locality and Contextuality. *New J. Phys.* **2011**, *13*, 113036.
- Popescu, S.; Rohrlich, D. Quantum Nonlocality as an Axiom. *Found. Phys.* **1994**, *24*, 379–385.
- Von Neumann, J. *Mathematische Grundlagen der Quantenmechanik*; Springer: Berlin, Germany, 1932.
- Shannon, C. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423, 623–656.
- Rényi, A.. On Measures of Information and Entropy. In *the 4th Berkeley Symposium on Mathematical Statistics and Probability*, Neymann, J., Ed.; University of California Press: Berkeley, CA, USA, 1961; pp. 547–561.
- Daróczy, Z. Über die gemeinsame Charakterisierung der zu den nicht vollständigen Verteilungen gehörigen Entropien von Shannon und von Rényi. *Z. Wahrscheinlichkeitstheorie Verwandte Geb.* **1963**, *1*, 381–388.
- Wehner, S.; Winter, A. Entropic Uncertainty Relations—A Survey. *New J. Phys.* **2010**, *12*, 025009.
- Bialynicki-Birula, I.; Rudnicki, Ł. Entropic Uncertainty Relations in Quantum Physics. In *Statistical Complexity: Applications in Electronic Structure*; Sen, K., Ed.; Springer: Berlin, Germany, 2011; pp. 1–34.
- Coles, P.; Beta, M.; Tomamichel, M.; Wehner, S. Entropic Uncertainty Relations and Their Applications. *Rev. Mod. Phys.* **2017**, *89*, 015002.
- Everett, H. *On the Foundations of Quantum Mechanics*. Ph. D. Dissertation, Princeton University, Princeton, NJ, USA, 1957.
- Hirschman, I. A Note on Entropy. *Am. J. Math.* **1957**, *79*, 152–156.
- Beckner, W. Inequalities in Fourier Analysis. *Ann. Math.* **1975**, *102*, 159–182.
- Bialynicki-Birula, I.; Mycielski, J. Uncertainty Relations for Information Entropy in Wave Mechanics. *Commun. Math. Phys.* **1975**, *44*, 129–132.
- Wootters, W.; Sussman, D. Discrete Phase Space and Minimum-Uncertainty States. 2007. Available online: <https://arxiv.org/abs/0704.1277> (accessed on 29 September 2022).
- Fuchs, C. Quantum Mechanics as Quantum Information (and Only a Little More). In *Quantum Theory: Reconsideration of Foundations*; Khrennikov, A., Ed.; Växjö University Press: Växjö, Sweden, 2002; pp. 463–543.
- Onggadinata, K.; Kurzynski, P.; Kaszlikowski, D. Qubit from the Classical Collision Entropy. 2022. Available online: <https://arxiv.org/abs/2205.00773> (accessed 4 September 2022).
- Heisenberg, W. Über den anschulichen Inhalt der quantentheoretischen Kinematik und Mechanik. *Z. Phys.* **1927**, *43*, 172–198.
- Sakurai, J.; Napolitano, J. *Modern Quantum Mechanics*, 2nd ed.; Addison-Wesley: Boston, MA, USA, 2011.

- 
32. Brandenburger, A.; La Mura, P. Axioms for Rényi Entropy with Signed Measures. 2019. Available online at [www.adambrandenburger.com](http://www.adambrandenburger.com) (accessed on 29 September 2022).
  33. Boyd, S.; Vandenberghe, L. *Convex Optimization*; Cambridge University Press: Cambridge, UK, 2004.