On the Relativity of Quantumness as Implied by Relativity of Arithmetic and Probability

Marek Czachor
Instytut Fizyki i Informatyki Stosowanej, Politechnika Gdańska, 80-233 Gdańsk, Poland

A hierarchical structure of isomorphic arithmetics is defined by a bijection $g_{\mathbb{R}} : \mathbb{R} \to \mathbb{R}$. It entails a hierarchy of probabilistic models, with probabilities $p_k = g^k(p)$, where g is the restriction of $g_{\mathbb{R}}$ to the interval $[0, 1]$, g^k is the kth iterate of g, and k is an arbitrary integer (positive, negative, or zero; $g^0(x) = x$). The relation between p and $g^k(p)$, $k > 0$, is analogous to the one between probability and neural activation function. For $k \ll -1$, $g^k(p)$ is essentially white noise (all processes are equally probable). The choice of $k = 0$ is physically as arbitrary as the choice of origin of a line in space, hence what we regard as experimental binary probabilities, $p_{\rm exp}$, can be given by any k, $p_{\rm exp} = g^k(p)$. Quantum binary probabilities are defined by $g(p) = \sin^2 \frac{\pi}{2} p$. With this concrete form of g, one finds that any two neighboring levels of the hierarchy are related to each other in a quantum–subquantum relation. In this sense, any model in the hierarchy is probabilistically quantum in appropriate arithmetic and calculus. And the other way around: any model is subquantum in appropriate arithmetic and calculus. Probabilities involving more than two events are constructed by means of trees of binary conditional probabilities. We discuss from this perspective singlet-state probabilities and Bell inequalities. We find that singlet state probabilities involve simultaneously three levels of the hierarchy: quantum, hidden, and macroscopic. As a by-product of the analysis, we discover a new (arithmetic) interpretation of the Fubini–Study geodesic distance.

I. INTRODUCTION

In brief, the quantum measurement problem consists of finding a rule that correlates states of a quantum system with those of a macroscopic observer. When phrased in probabilistic terms, the problem is to find a consistent rule of replacing joint probabilities, $p(a, b)$, by conditional probabilities, $p(a|b)$, where a and b represent states (or properties) of the system and the observer, respectively. In standard quantum mechanics the rule can be inferred from Bayes law by the following sequence of equivalences:

$$p(a|b) = \frac{p(a,b)}{p(b)} = \frac{\mathrm{Tr}\,(\rho P_b P_a P_b)}{\mathrm{Tr}\,(\rho P_b)} = \mathrm{Tr}\,\left(\frac{P_b \rho P_b}{\mathrm{Tr}\,(P_b \rho P_b)} P_a\right) = \mathrm{Tr}\,(\rho_b P_a). \tag{1}$$

Thus, the process of conditioning by the event "b has occurred" can be represented by the "state vector reduction",

$$\rho \mapsto \rho_b = \frac{P_b \rho P_b}{\mathrm{Tr}\,(P_b \rho P_b)}. \tag{2}$$

However, do we really need (2)? From an operational point of view, it is enough if we know the joint probability,

$$p(a, b) = \mathrm{Tr}\,(\rho P_b P_a P_b), \tag{3}$$

and the probability of the condition,

$$p(b) = \mathrm{Tr}\,(\rho P_b). \tag{4}$$

Both numbers are directly related to experimental data, so (2) is redundant.

If we try to generalize the above procedure beyond quantum mechanics, various possibilities arise. In nonlinear quantum mechanics, for example, once we obtain $p(a, b)$ and $p(b)$, we can *deduce* the mathematical form of an effective state vector reduction, but it will not coincide with (2), because the sequence of transformations (1) will no longer be true (cf. [1] for the details). A naive combination of (2) with nonlinear evolution of states implies the inconsistency known as faster-than-light communication [2–5]. Of course, one *can* work with the projection postulate even in nonlinear quantum mechanics (eliminating the faster-than-light effect), but the form of state vector reduction must be first derived in a consistent way from Bayes law [1]. Here, consistency is the keyword.

The Bayes law, when written as $p(a, b) = p(a|b)p(b)$, is known as the *product rule*. Jaynes [6] (following the ideas of Aczél [7] and Cox [8]) derives the product rule from some very general desiderata of consistent and plausible reasoning but, interestingly, what one finds turns out be more general,

$$p(a, b) = g^{-1}\Big(g\big(p(a|b)\big)g\big(p(b)\big)\Big), \tag{5}$$

where g is some monotone non-negative function (cf. Equation (2.27) in [6]). Still, for Jaynes, $p(\dots)$ is not yet a probability. His intuition tells him that the probability (or, rather, a measure of plausibility) is given by $g(p(\dots))$, so that the product rule is reconstructed in the standard form,

$$g\big(p(a,b)\big) = g\big(p(a|b)\big)g\big(p(b)\big). \tag{6}$$

What we will discuss later on in this paper employs a possibility that was not taken into account by Jaynes. Namely, we will treat formulas such as (5) as a definition of a new product, $\odot$, so that

$$p(a,b) = g^{-1}\Big(g\big(p(a|b)\big)g\big(p(b)\big)\Big) = p(a|b) \odot p(b). \tag{7}$$

We will also see that $g(p)$ and its higher iterates have intriguing similarities to neural activation functions, whereas higher iterates of $g^{-1}(p)$ resemble a white noise.

A new product is an element of a new arithmetic, leading us ultimately to a whole hierarchical structure of such generalized models. As one of the conclusions, we will find that both p and $g(p)$ may be treated as genuine probabilities, provided g is restricted to the class discussed in detail in Section II. One of the possibilities, directly related to the measurement problem, is that p are probabilities at a hidden-variable level, whereas $g(p)$ are the quantum ones. We will see that any two neighboring levels of the hierarchy are related to each other in a way that may be regarded as a form of a quantum–subquantum relationship. This will lead to the idea of *relativity of quantumness*.

In any such generalized and fundamental theory one is necessarily confronted with the chicken-or-egg dilemma: What was first, $p(a,b)$ and $p(b)$, or $p(a|b)$ and $p(b)$? The Bayes law that defines the conditional probability in terms of the joint probability, or the product rule that defines the joint probability in terms of the conditional probability?

An alternative form of the dilemma can be expressed in terms of the projection postulate: Do we first define conditional probabilities in terms of some given form of state vector reduction, or we begin with joint probabilities and then infer the form of state vector reduction? In nonlinear quantum mechanics, the latter strategy is superior to the former one. However, in the Bayesian approach to probability, one updates probabilities on the basis of prior information, so the conditional probabilities are superior to the joint ones.

The formalism of arithmetic hierarchies discussed in the present paper clearly prefers the Bayesian approach. The reason is in the three fundamental lemmas, which we will discuss in Section II, which are true only for binary probabilities. There is priority in the binary coding, as we have to construct probabilities involving more than two events in terms of binary trees of conditional probabilities. Binary coding becomes as fundamental for probability theory as the two-spinors are fundamental for relativistic physics [9].

We begin in Section II by recalling the three fundamental lemmas about the functional equation $g(p) + g(1 - p) = 1$. In Section III, we construct a hierarchy of isomorphic arithmetics associated with $g(p)$. The hierarchy of arithmetics leads to a hierarchy of probabilities introduced in Section IV. A hierarchical ordering relation, briefly discussed in Section V, will allow us to unambiguously employ symbols such as $<$ and $>$. A family of product rules, discussed in Section VI, is employed in the problem of hidden-variables representation of singlet-state probabilities in Section VII. We explain, in particular, that one encounters here three types of arithmetic levels in a single formula for joint probabilities: quantum, macroscopic, and hidden. Section VIII introduces some elements of hierarchical calculi, with special emphasis on non-Newtonian integration. We make here a digression on Rényi's entropy which is implicitly based on a generalized arithmetic, but does not take advantage of the possibilities inherent in generalized calculus. Section IX is devoted to local hidden-variable models of singlet-state probabilities constructed in terms of the generalized calculus. This seems to be the most controversial aspect of the formalism, as it clearly contradicts common wisdom about Bell's theorem. Section X brings us to the intriguing role played in quantum mechanics by the geodesic distance in the projective space of quantum states. A typical discussion of the Fubini–Study metric is restricted in the literature to its geometric interpretation. Here, we reveal its unknown aspect: Its role for the arithmetic structure of quantum states. It seems that $g(p) = \sin^2 \frac{\pi}{2} p$ is a fundamental bijection that determines the arithmetic of the subquantum world. In Section XI, we give a simple argument explaining why the effective number of distinguishable probabilistic levels of the hierarchy is finite. We also point out a possible interpretation of the hierarchy of probabilities in terms of neural activation functions. At such a formal level, the only means of relating formal probabilities to experiment is via the laws of large numbers, discussed in Section XII. In Section XIII, we return to the problem of Bell's inequalities. We depart here a little from the formalism we developed in a series of earlier papers where the same arithmetic was used at the hidden and the macroscopic levels. Our current understanding of the problem is that it is better to employ the freedom of combining different arithmetics simultaneously. We end the paper with remarks on open problems, Section XIV, and certain personal perspective is given in Section XV. The Appendix A is devoted to certain technicalities which cannot be found in the literature.

II. THREE FUNDAMENTAL LEMMAS

The hierarchical structure of (binary) probabilities is a consequence of the following three lemmas. They do not have a sufficiently nontrivial generalization beyond the binary case (cf. the discussion in [10]), hence the non-binary case has to be treated in terms of trees of conditional probabilities constructed in analogy to binary Huffman codes [11].

Lemma 1. *$g : [0,1] \to [0,1]$ is a solution of the functional equation $g(p) + g(1-p) = 1$ if and only if*

$$g(p) = \frac{1}{2} + h\left(p - \frac{1}{2}\right), \tag{8}$$

where $h(-x) = -h(x)$, $h : [-1/2, 1/2] \to [-1/2, 1/2]$, i.e., h is an arbitrary odd mapping of the closed interval into itself. Any such g has a fixed point at $p = 1/2$.

Lemma 2. *Consider two functions $g_j : [0,1] \to [0,1]$, $j = 1, 2$, that satisfy assumptions of Lemma 1,*

$$g_j(p) = \frac{1}{2} + h_j\left(p - \frac{1}{2}\right), \tag{9}$$

where $h_j(-x) = -h_j(x)$. Then $g_{12} = g_1 \circ g_2$ also satisfies Lemma 1 with $h_{12} = h_1 \circ h_2$,

$$g_{12}(p) = \frac{1}{2} + h_{12}\left(p - \frac{1}{2}\right). \tag{10}$$

Accordingly,

$$g_{12}(p) + g_{12}(1-p) = 1 \tag{11}$$

for any $p \in [0,1]$.

Lemma 3. *Let $g^k = g \circ \cdots \circ g$, $g^{-k} = g^{-1} \circ \cdots \circ g^{-1}$ (k times), $g^0(x) = x$. If g satisfies Lemma 1,*

$$g(p) = \frac{1}{2} + h\left(p - \frac{1}{2}\right), \tag{12}$$

then the kth iterate g^k also satisfies Lemma 1 for any $k \in \mathbb{Z}$,

$$g^k(p) = \frac{1}{2} + h^k\left(p - \frac{1}{2}\right), \tag{13}$$

where h^k is the kth iterate of h. Accordingly,

$$g^k(p) + g^k(1-p) = 1 \tag{14}$$

for any $p \in [0,1]$, and any integer k. In particular

$$g^{-1}(p) + g^{-1}(1-p) = 1. \tag{15}$$

The proofs can be found in [12, 13].

Armed with the lemmas we can construct a hierarchy of arithmetics, entailing a hierarchy of probabilities.

III. HIERARCHY OF ISOMORPHIC ARITHMETICS

Assume that $g : [0,1] \to [0,1]$ occurring in the above three lemmas is a restriction of a bijection $g_{\mathbb{R}} : \mathbb{R} \to \mathbb{R}$, i.e., $g(x) = g_{\mathbb{R}}(x)$ for $x \in [0,1]$. It does not matter what the properties of $g_{\mathbb{R}}(x)$ are if $x \notin [0,1]$, except for the bijectivity of $g_{\mathbb{R}}$. Put differently, g belongs to the equivalence class $[g_{\mathbb{R}}]$ of bijections whose restrictions to $[0,1]$ are

identical. Following the notation of Lemma 3, we denote $g^k = g_{\mathbb{R}} \circ \cdots \circ g_{\mathbb{R}}$, $g^{-k} = g_{\mathbb{R}}^{-1} \circ \cdots \circ g_{\mathbb{R}}^{-1}$, $g^0(x) = x$. Now, let $x, y \in \mathbb{R}$. Define,

$$x \oplus_k y = g^k\Big(g^{-k}(x) + g^{-k}(y)\Big), \tag{16}$$

$$x \ominus_k y = g^k\Big(g^{-k}(x) - g^{-k}(y)\Big), \tag{17}$$

$$x \odot_k y = g^k\Big(g^{-k}(x) \cdot g^{-k}(y)\Big), \tag{18}$$

$$x \oslash_k y = g^k\Big(g^{-k}(x) / g^{-k}(y)\Big). \tag{19}$$

The arithmetic $\mathbb{R}_k$ is the set $\mathbb{R}$ equipped with the above four operations, i.e., $\mathbb{R}_k = \{\mathbb{R}, \oplus_k, \ominus_k, \odot_k, \oslash_k\}$. The ordering relation is independent of k if g is increasing, which we therefore assume, hence $g^k(x) < g^k(y)$ if and only if $x < y$. The neutral elements of addition, $0_k = g^k(0)$, and multiplication, $1_k = g^k(1)$,

$$x \oplus_k 0_k = x \odot_k 1_k = x, \quad \text{for any } x, \tag{20}$$

can be regarded as bits, in principle applicable to some form of binary coding. Greater natural numbers are obtained by the n-times repeated addition of 1_k,

$$n_k = \underbrace{1_k \oplus_k \cdots \oplus_k 1_k}_{n \text{ times}} = g^k(n), \tag{21}$$

$$n_k \oplus_k m_k = g^k(n+m) = (n+m)_k, \tag{22}$$

$$n_k \odot_k m_k = g^k(nm) = (nm)_k. \tag{23}$$

An nth power of x,

$$x^{n_k} = \underbrace{x \odot_k \cdots \odot_k x}_{n \text{ times}}, \tag{24}$$

satisfies

$$x^{n_k} \odot_k x^{m_k} = x^{(n+m)_k} = x^{n_k \oplus_k m_k}. \tag{25}$$

Rational numbers are those of the form

$$n_k \oslash_k m_k = g^k(n/m) = (n/m)_k, \quad n, m \in \mathbb{Z}. \tag{26}$$

The notion of rationality is arithmetic-dependent. Indeed, let n/m be a rational number in the arithmetic $\mathbb{R}_0 = \{\mathbb{R}, +, -, \cdot, /\}$. Then, typically, $g^k(n/m)$, $k \neq 0$, is not a rational number in $\mathbb{R}_0$. Still, it is a rational number in the arithmetic $\mathbb{R}_k = \{\mathbb{R}, \oplus_k, \ominus_k, \odot_k, \oslash_k\}$ in consequence of (26).

For any $k, l \in \mathbb{Z}$, the four arithmetic operations are related by

$$x \odot_{k+l} y = g^l\Big(g^{-l}(x) \odot_k g^{-l}(y)\Big) = g^k\Big(g^{-k}(x) \odot_l g^{-k}(y)\Big), \tag{27}$$

$$x \oslash_{k+l} y = g^l\Big(g^{-l}(x) \oslash_k g^{-l}(y)\Big) = g^k\Big(g^{-k}(x) \oslash_l g^{-k}(y)\Big), \tag{28}$$

$$x \oplus_{k+l} y = g^l\Big(g^{-l}(x) \oplus_k g^{-l}(y)\Big) = g^k\Big(g^{-k}(x) \oplus_l g^{-k}(y)\Big), \tag{29}$$

$$x \ominus_{k+l} y = g^l\Big(g^{-l}(x) \ominus_k g^{-l}(y)\Big) = g^k\Big(g^{-k}(x) \ominus_l g^{-k}(y)\Big). \tag{30}$$

The bijection $f^k = g^{-k}$ is an isomorphism of $\mathbb{R}_{k+l}$ and $\mathbb{R}_l$, for any $k, l \in \mathbb{Z}$,

$$f^k(x \odot_{k+l} y) = f^k(x) \odot_l f^k(y), \tag{31}$$

$$f^k(x \oslash_{k+l} y) = f^k(x) \oslash_l f^k(y), \tag{32}$$

$$f^k(x \oplus_{k+l} y) = f^k(x) \oplus_l f^k(y), \tag{33}$$

$$f^k(x \ominus_{k+l} y) = f^k(x) \ominus_l f^k(y). \tag{34}$$

The value $l = 0$ is not privileged. The role of a 0th level can be played by any l. The notation where

$$\mathbb{R}_l = \{\mathbb{R}, \oplus_l, \ominus_l, \odot_l, \oslash_l\} = \{\mathbb{R}, +, -, \cdot, /\}, \tag{35}$$

is perfectly acceptable, hence *any* $\mathbb{R}_l$ can be regarded as "the" ordinary arithmetic we are taught at school. The latter statement is the content of the "arithmetic Copernican principle", introduced in [13] and discussed further in [14]. In the present paper we nevertheless simplify notation and assume $\mathbb{R}_0 = \{\mathbb{R}, +, -, \cdot, /\}$. This is analogous to the usual habit of imposing initial conditions in Newtonian dynamics "at $t = 0$" instead of a general $t = t_0$.

The hierarchy of arithmetics leads to the hierarchy of probabilities.

IV. HIERARCHY OF PROBABILITIES

Let $g(1) = 1$, so that $1_k = g^k(1) = 1$ and $0_k = g^k(0) = 0$, for any k. Now, let p, q, $p + q = 1$, be probabilities. Assuming that g satisfies the assumptions of Lemma 1, we find (in consequence of Lemmas 2 and 3, and $g^k(1) = 1$ for any $k \in \mathbb{Z}$)

$$p + q = 1, \tag{36}$$
$$g^k(p) + g^k(q) = 1, \tag{37}$$
$$p \oplus_{-k} q = g^{-k}\left(g^k(p) + g^k(q)\right) = 1, \tag{38}$$

for any $k \in \mathbb{Z}$. The Copernican aspect is visible at the level of probabilities as well, if we define $P = g^k(p)$, $Q = g^k(q)$, so that

$$g^{-k}(P) + g^{-k}(Q) = 1, \tag{39}$$
$$P + Q = 1, \tag{40}$$
$$P \oplus_k Q = g^k\left(g^{-k}(P) + g^{-k}(Q)\right) = 1, \tag{41}$$

for any $k \in \mathbb{Z}$. Indeed, how to distinguish between (36)–(38) and (39)–(41), if we bear in mind that k can be positive, negative, or zero, and the formulas are true for all k? How to distinguish between the two levels if in both cases we find $p+q = 1$ and $P+Q = 1$? Which of the probabilities, p or P, is the one we measure in experiment? Which iterate, k, 0, or $-k$, is the one that defines *our* probabilities we experimentally define in terms of frequencies of successes? Which natural numbers n_k, $n = n_0$, or n_{-k}, are the ones we use to define numbers of trials and successes?

Formula (38) shows that probabilities p and q sum to 1 in infinitely many ways, corresponding to infinitely many values of k in $\oplus_{-k}$. Formula (37) shows that probabilities p and q generate infinitely many probabilities $p_k = g^k(p)$ and $q_k = g^k(q)$ that sum to 1 by means of the same addition $+ = \oplus_0$. The Arithmetic Copernican Principle is a relativity principle which states that any value of k can correspond to the arithmetic and probability that we regard as "the human and experimental one".

Still, this is not the end of the story. Replacing in (37) k by $k - l$,

$$g^{k-l}(p) + g^{k-l}(q) = 1, \tag{42}$$

and acting on both sides with g^l, we find

$$g^l\left(g^{k-l}(p) + g^{k-l}(q)\right) = g^k(p) \oplus_l g^k(q) = 1, \tag{43}$$

for any $k, l \in \mathbb{Z}$. The resulting wealth of available probability models implied by a single bijection g is truly overwhelming, yet ignored by those who study quantum probabilities and the hidden variables problem.

Let us now consider the concrete case of the equivalence class of a function $g_{\mathbb{R}}$ whose restriction to $[0, 1]$ is given by $g(x) = \sin^2 \frac{\pi}{2} x$. Then,

$$h(x) = g\left(x + \frac{1}{2}\right) - \frac{1}{2} = \frac{1}{2}\sin \pi x, \quad -\frac{1}{2} \leqslant x \leqslant \frac{1}{2}, \tag{44}$$
$$g(p) = \frac{1}{2} + h\left(p - \frac{1}{2}\right) = \frac{1}{2} + \frac{1}{2}\sin \pi\left(p - \frac{1}{2}\right), \quad 0 \leqslant p \leqslant 1, \tag{45}$$

Let $p = (\pi - \theta)/\pi$ be the probability of finding a point belonging to the overlap of two half-circles rotated by $\theta \in [0, \pi]$. Then, for $k = 1$, $q = \theta/\pi$,

$$P = g(p) = g^k(p) = \sin^2 \frac{\pi}{2} \frac{\pi - \theta}{\pi} = \cos^2 \frac{\theta}{2}, \tag{46}$$
$$Q = g(q) = g^k(q) = \sin^2 \frac{\pi}{2} \frac{\theta}{\pi} = \sin^2 \frac{\theta}{2}, \tag{47}$$

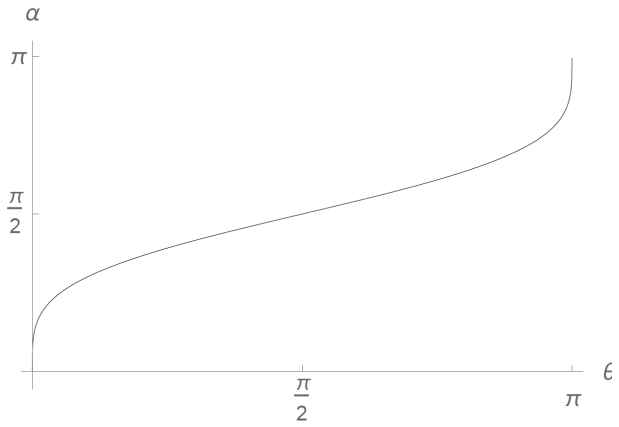

Figure 1. The relation between α and θ as given by (51). There are three fixed points: $\alpha(0) = 0$, $\alpha(\pi/2) = \pi/2$, $\alpha(\pi) = \pi$. Here, α is the geometric angle between the two Stern–Gerlach devices, whereas θ is a hidden parameter.

in which we recognize the conditional probabilities for two successive measurements of spin-1/2 in two Stern–Gerlach devices placed one after another, with relative angle θ.

By Lemma 3, we have in fact much more, because $k = 1$ can be replaced by any integer. For example, the second iterate

$$P = g^2(p) = g\big(g(p)\big) = \sin^2\frac{\pi}{2}\left(\cos^2\frac{\theta}{2}\right), \tag{48}$$

satisfies $g^2(p)+g^2(q) = 1$, of course, as can be proved by a straightforward but instructive calculation [14]. The minus-first iterate,

$$P = g^{-1}(p) = \frac{2}{\pi}\arc\sin\sqrt{p} = \frac{2}{\pi}\arc\sin\sqrt{\frac{\pi-\theta}{\pi}}, \tag{49}$$

satisfies $g^{-1}(p) + g^{-1}(q) = 1$, and so on and so forth.

Clearly, we have absolutely no criterion that could indicate which level of the hierarchy is the one we regard as our human one, a fact that justifies the adjective "Copernican". For example, rewriting (49) as

$$P \;=\; g^{1-2}(p) = g^1\left(g^{-2}(p)\right) = g^1\left(1-g^{-2}(q)\right) = g^1\left(1-\frac{\alpha}{\pi}\right) = \cos^2\frac{\alpha}{2}, \tag{50}$$

we find the relation between the two parameters, α and θ, corresponding to the two levels of the hierarchy (see Figure 1),

$$\alpha(\theta) = \pi g^{-2}(q) = 2\arc\sin\sqrt{\frac{2}{\pi}\arc\sin\sqrt{\frac{\theta}{\pi}}}. \tag{51}$$

The usual tests of classicality and quantumness are based on inequalities. However, in order to discuss an inequality we have to control ordering relations such as $\leqslant$ and $\geqslant$. Fortunately, with our assumptions about g the problem is trivial.

V. HIERARCHICAL ORDERING RELATION

We assume that the bijection g is strictly increasing, i.e., $x < y$ if and only if $g(x) < g(y)$. A composition of two strictly increasing functions is strictly increasing, hence $x < y$ implies $g^k(x) \ominus_l g^k(y) < 0_l = 0$ for any $k, l \in \mathbb{Z}$. The latter leads to a unique ordering relation at the level of the entire hierarchy of arithmetics. This is why it is safe to use the symbols $<$, $>$, $\leqslant$, $\geqslant$ at any level of the hierarchy.

So far, we have restricted our analysis to binary events. An extension to higher dimensional problems needs the notion of a product rule.

VI. HIERARCHICAL PRODUCT RULES

The standard product rule states that probability of a sequence of two events, first a_1 then a_2, is given by the product of the prior $p(a_1)$ (a probability of the condition) with the posterior $p(a_2|a_1)$ (a conditional probability of a_2 under the condition that a_1 has happened). The sums of binary probabilities,

$$g^{k_1}\big(p(0)\big) \oplus_l g^{k_1}\big(p(1)\big) \;=\; 1, \quad \text{for any } k_1, l \in \mathbb{Z}, \tag{52}$$

$$g^{k_2}\big(p(0|a_1)\big) \oplus_l g^{k_2}\big(p(1|a_1)\big) \;=\; 1, \quad \text{for any } k_2, l \in \mathbb{Z}, \tag{53}$$

as implied by the lemmas, are naturally related to

$$g^{k_2}\big(p(a_2|a_1)\big) \odot_l g^{k_1}\big(p(a_1)\big), \quad \text{for any } k_1, k_2, l \in \mathbb{Z}, \tag{54}$$

because

$$\bigoplus_{a_1,a_2 l} g^{k_2}\big(p(a_2|a_1)\big) \odot_l g^{k_1}\big(p(a_1)\big) = 1, \quad \text{for any } k_1, k_2, l \in \mathbb{Z}. \tag{55}$$

A sequence of results, $a_n, a_{n-1}, \ldots, a_1$, implies their joint probability

$$g^{k_n}\big(p(a_n|a_{n-1}\ldots a_1)\big) \odot_l \cdots \odot_l g^{k_2}\big(p(a_2|a_1)\big) \odot_l g^{k_1}\big(p(a_1)\big) \tag{56}$$

normalized by

$$\bigoplus_{a_1\ldots a_n l} g^{k_n}\big(p(a_n|a_{n-1}\ldots a_1)\big) \odot_l \cdots \odot_l g^{k_2}\big(p(a_2|a_1)\big) \odot_l g^{k_1}\big(p(a_1)\big) = g^l(1) = 1. \tag{57}$$

In particular, for $l = 0$,

$$g^{k_1}\big(p(0)\big) + g^{k_1}\big(p(1)\big) \;=\; 1, \quad \text{for any } k_1 \in \mathbb{Z}, \tag{58}$$

$$g^{k_2}\big(p(0|a_1)\big) + g^{k_2}\big(p(1|a_1)\big) \;=\; 1, \quad \text{for any } k_2 \in \mathbb{Z}, \tag{59}$$

and

$$\sum_{a_1,a_2} g^{k_2}\big(p(a_2|a_1)\big) g^{k_1}\big(p(a_1)\big) = 1, \quad \text{for any } k_1, k_2 \in \mathbb{Z}. \tag{60}$$

At the other extreme is the case of $l = k_1 = k_2 = k$,

$$g^k\big(p(a_2|a_1)\big) \odot_k g^k\big(p(a_1)\big) = g^k\big(p(a_2|a_1)p(a_1)\big), \tag{61}$$

with normalization

$$\bigoplus_{a_1,a_2 k} g^k\big(p(a_2|a_1)p(a_1)\big) = g^k\left(\sum_{a_1,a_2} p(a_2|a_1)p(a_1)\right) = 1, \quad \text{for any } k \in \mathbb{Z}. \tag{62}$$

It is striking that in formulas such as (56) each of the k-indices can be in principle different. In effect, (56) may be regarded as a component of a tensor.

A truly nontrivial application of generalized product rules occurs in the problem of singlet-state probabilities, quantum entangled states, and Bell's theorem.

VII. SINGLET-STATE PROBABILITIES

Singlet-state probabilities occur in experiments where two parties ("Alice" and "Bob") are macroscopically separated, but the measurements they perform are the quantum ones. Such probabilities naturally occur in the context of the hierarchical product rule. Indeed, consider the following probabilities,

$$p(0) \;=\; p(1) = \frac{1}{2}, \tag{63}$$

$$p(0|0) \;=\; p(1|1) = \frac{\theta}{\pi}, \tag{64}$$

$$p(1|0) \;=\; p(0|1) = \frac{\pi - \theta}{\pi}, \tag{65}$$

whose geometric interpretation is evident. As the bijection take the one occurring in (45)–(47). Then,

$$g\big(p(0|0)\big)g\big(p(0)\big) \;=\; g\big(p(1|1)\big)g\big(p(1)\big) = \frac{1}{2}\sin^2\frac{\theta}{2}, \tag{66}$$

$$g\big(p(1|0)\big)g\big(p(0)\big) \;=\; g\big(p(0|1)\big)g\big(p(1)\big) = \frac{1}{2}\cos^2\frac{\theta}{2}, \tag{67}$$

are the probabilities typical of the singlet state. Let us note that we have employed the product rule,

$$g^{k_2}\big(p(a|b)\big) \odot_l g^{k_1}\big(p(b)\big) = g^1\big(p(a|b)\big) \odot_0 g^{k_1}\big(p(b)\big), \tag{68}$$

with $k_2 \neq l$. k_1 can be arbitrary because $g(1/2) = 1/2 = g^{k_1}(1/2)$ for any g that satisfies Lemma 1. For simplicity, we set $k_1 = 1$. Now, the joint probability can be interpreted as follows:

$$P(a,b) = \underbrace{g\big(\overbrace{p(a|b)}^{\text{hidden}}\big)}_{\text{quantum}} \underbrace{\odot_0}_{\text{macroscopic}} \underbrace{g\big(\overbrace{p(b)}^{\text{hidden}}\big)}_{\text{quantum}}. \tag{69}$$

Let us further note that we could have started with the following:

$$g^k\big(p(0)\big) \;=\; g^k\big(p(1)\big) = \frac{1}{2}, \tag{70}$$

$$g^k\big(p(0|0)\big) \;=\; g^k\big(p(1|1)\big) = \frac{\theta}{\pi}, \tag{71}$$

$$g^k\big(p(0|1)\big) \;=\; g^k\big(p(1|0)\big) = \frac{\pi-\theta}{\pi}. \tag{72}$$

Then, $g^{k+1}\big(p(a_2|a_1)\big)g^{k+1}\big(p(a_1)\big)$ would be the singlet-state probabilities.

One concludes that the notion of a quantum level is a relative one. In fact, any level is quantum, and any level is hidden; moreover, any $\odot_l$ can play the role of the macroscopic arithmetic. What counts is the neighboring location in the hierarchy. The so-called violation of Bell's inequality is an inconsistency that occurs if we apply the arithmetic of a hidden level to calculations performed at the neighboring quantum one. An analogous inconsistency that occurs between non-neighboring levels leads to violations beyond the Tsirelson bound [13, 15].

In order to perform calculations at different levels of the hierarchy, we have to understand what the consequences are of the hierarchical structure of arithmetics for the resulting hierarchy of calculi.

VIII. HIERARCHY OF CALCULI

A hierarchy of arithmetics leads to a hierarchy of "non-Newtonian" calculi [16–21]. Here, functions such as $A : \mathbb{R} \to \mathbb{R}$ have to be treated as mappings between arithmetics and not between sets, hence it is more appropriate to write

$$A_{lk} : \mathbb{R}_k \to \mathbb{R}_l, \tag{73}$$

with some $k, l \in \mathbb{Z}$. Otherwise the notions of derivative and integral are ambiguous. The derivative of A_{lk} is

$$\frac{\mathrm{D}_l A_{lk}(x)}{\mathrm{D}_k x} \;=\; \lim_{\delta\to 0}\Big(A_{lk}(x \oplus_k \delta_k) \ominus_l A_{lk}(x)\Big) \oslash_l \delta_l. \tag{74}$$

As before, $\delta_k = g^k(\delta)$, $\delta_l = g^l(\delta)$. The derivative is $\mathbb{R}_l$-linear and satisfies an appropriate Leibniz rule,

$$\frac{\mathrm{D}_l\big(A_{lk}(x) \oplus_l B_{lk}(x)\big)}{\mathrm{D}_k x} \;=\; \frac{\mathrm{D}_l A_{lk}(x)}{\mathrm{D}_k x} \oplus_l \frac{\mathrm{D}_l B_{lk}(x)}{\mathrm{D}_k x}, \tag{75}$$

$$\frac{\mathrm{D}_l\big(A_{lk}(x) \odot_l B_{lk}(x)\big)}{\mathrm{D}_k x} \;=\; \left(\frac{\mathrm{D}_l A_{lk}(x)}{\mathrm{D}_k x} \odot_l B_{lk}(x)\right) \oplus_l \left(A_{lk}(x) \odot_l \frac{\mathrm{D}_l B_{lk}(x)}{\mathrm{D}_k x}\right). \tag{76}$$

Integration of $A_{lk} : \mathbb{R}_k \to \mathbb{R}_l$ is defined in a way that guarantees the two fundamental theorems of calculus (under standard assumptions about differentiability and continuity):

$$\int_a^b \frac{\mathrm{D}_l A_{lk}(x)}{\mathrm{D}_k x}\mathrm{D}_k x \;=\; A_{lk}(b) \ominus_l A_{lk}(a), \tag{77}$$

$$\frac{\mathrm{D}_l}{\mathrm{D}_k x}\int_a^x A_{lk}(y)\mathrm{D}_k y \;=\; A_{lk}(x). \tag{78}$$

The formulas become less abstract if one considers the following commutative diagram ($f = g^{-1}$)

$$\begin{array}{ccc} \mathbb{R}_k & \xrightarrow{A_{lk}} & \mathbb{R}_l \\ {\scriptstyle f^k}\big\downarrow & & \big\uparrow{\scriptstyle g^l} \\ \mathbb{R}_0 & \xrightarrow{A_{00}} & \mathbb{R}_0 \\ {\scriptstyle g^n}\big\downarrow & & \big\uparrow{\scriptstyle f^m} \\ \mathbb{R}_n & \xrightarrow{A_{mn}} & \mathbb{R}_m \end{array} , \tag{79}$$

leading to a very simple and useful form of the derivative (74),

$$\frac{\mathrm{D}_l A_{lk}(x)}{\mathrm{D}_k x} \;=\; g^l\left(\frac{\mathrm{d}A_{00}\big(f^k(x)\big)}{\mathrm{d}f^k(x)}\right), \tag{80}$$

while the integral reads,

$$\int_a^b A_{lk}(x)\mathrm{D}_k x \;=\; g^l\left(\int_{f^k(a)}^{f^k(b)} A_{00}(r)\mathrm{d}r\right). \tag{81}$$

Here, dr denotes the usual (Riemann, Lebesgue, etc.) integral in $\mathbb{R}_0$. Formula (80) is derived under the assumption that $g : \mathbb{R} \to \mathbb{R}$ is continuous (in the usual meaning of the term employed in ordinary "Newtonian" real analysis), which is however, automatically guaranteed by the fact that g is a bijection. What is important, neither g nor its inverse f have to be differentiable in the standard Newtonian sense. The latter makes an important difference with respect to the ordinary differential geometry where functions such as $g(x) = x^{1/3}$ would be excluded as non-differentiable at $x = 0$. In the non-Newtonian formalism, any bijection g, as well as its inverse f, are automatically smooth with respect to the non-Newtonian differentiation defined by the same g. Various explicit examples can be found in [22, 23].

Linearity of the integral must be understood in the sense of $\mathbb{R}_l$,

$$\int_a^b A_{lk}(x) \oplus_l B_{lk}(x)\mathrm{D}_k x \;=\; \int_a^b A_{lk}(x)\mathrm{D}_k x \oplus_l \int_a^b B_{lk}(x)\mathrm{D}_k x, \tag{82}$$

$$\int_a^b A_l \odot_l B_{lk}(x)\mathrm{D}_k x \;=\; A_l \odot_l \int_a^b B_{lk}(x)\mathrm{D}_k x, \quad \text{for a constant } A_l \in \mathbb{R}_l, \tag{83}$$

a property of fundamental importance for Bell-type inequalities [13]. An analogous form of generalized linearity of integrals occurs in fuzzy calculus [24–28].

Diagram (79) implies

$$A_{lk} = g^l \circ A_{00} \circ f^k = g^{l-m} \circ g^m \circ A_{00} \circ f^n \circ f^{k-n} = g^{l-m} \circ A_{mn} \circ f^{k-n}, \tag{84}$$

which leads to a new type of a chain rule, relating derivatives and integrals at different levels of the hierarchy,

$$\frac{\mathrm{D}_l A_{lk}(x)}{\mathrm{D}_k x} \;=\; g^{l-m}\left(\frac{\mathrm{D}_m A_{mn}\left(f^{k-n}(x)\right)}{\mathrm{D}_n f^{k-n}(x)}\right), \tag{85}$$

$$\int_a^b A_{lk}(x)\mathrm{D}_k x \;=\; g^{l-m}\left(\int_{f^{k-n}(a)}^{f^{k-n}(b)} A_{mn}(x)\mathrm{D}_n x\right). \tag{86}$$

Formulas (85) and (86) do not seem to appear in the literature, so we prove them in Appendix A.

Digression: Logarithm and Rényi Entropies

Exponential function is defined by the differential equation,

$$\frac{\mathrm{D}_l \exp_{lk}(x)}{\mathrm{D}_k x} \;=\; g^l\left(\frac{\mathrm{d}\exp_{00}\left(f^k(x)\right)}{\mathrm{d}f^k(x)}\right) = \exp_{lk}(x) = g^l\left(\exp_{00}\left(f^k(x)\right)\right), \tag{87}$$

$$\exp_{lk}(0_k) \;=\; 1_l. \tag{88}$$

The solution is given by $\exp_{00}(x) = e^x$ and satisfies

$$\exp_{lk}(x \oplus_k y) = \exp_{lk}(x) \odot_l \exp_{lk}(y). \tag{89}$$

The inverse is given by

$$\ln_{kl}(x) = g^k \left(\ln_{00} \left(f^l(x) \right) \right), \tag{90}$$

where $\ln_{00}(x) = \ln x$, and

$$\ln_{kl}(x \odot_l y) = \ln_{kl}(x) \oplus_k \ln_{kl}(y). \tag{91}$$

Now, consider $\phi_\alpha(x) = e^{(1-\alpha)x}$, $\phi_\alpha^{-1}(x) = \frac{1}{1-\alpha} \ln x$. Rényi introduced his α-entropy as a Kolmogorov–Nagumo average [29–36] of the Shannon amount of information [37] (we prefer the natural logarithm to the original $\log_2$ from [33], but this is just a choice of units of information),

$$S_\alpha \;=\; \phi_\alpha^{-1} \left(\sum_p p \phi_\alpha(-\ln p) \right) = \frac{1}{1-\alpha} \ln \left(\sum_p p^\alpha \right). \tag{92}$$

It is clear that (92) can be expressed in several different ways by means of generalized arithmetics. For example,

$$\ominus_1 \ln_{1,0}(x) = g^1 \left(-\ln \left(f^0(x) \right) \right) = g \left(-\ln x \right), \tag{93}$$

has the same functional form as $\phi_\alpha\big(-\ln p\big)$. Alternatively, defining

$$x \oplus y \;=\; \phi_\alpha^{-1} \left(\phi_\alpha(x) + \phi_\alpha(y) \right), \tag{94}$$

$$x \odot y \;=\; \phi_\alpha^{-1} \left(\phi_\alpha(x) \phi_\alpha(y) \right), \tag{95}$$

and $\phi_\alpha^{-1}(p) = P$, we find

$$S_\alpha = \phi_\alpha^{-1} \left(\sum_P \phi_\alpha(P) \phi_\alpha\big(-\ln \phi_\alpha(P)\big) \right) = \bigoplus_P P \odot \ln \big(1/\phi_\alpha(P)\big). \tag{96}$$

Rényi's choice of $\phi_\alpha(x) = e^{(1-\alpha)x}$ was dictated by the assumed additivity of entropy for independent (i.e., uncorrelated) systems. Our general formalism suggests various hierarchical generalizations of the notion of entropy, automatically inheriting the additivity properties from the arithmetics involved. Some examples can be found in [10].

IX. APPLICATION: LOCAL HIDDEN-VARIABLE MODELS BASED ON NON-NEWTONIAN INTEGRATION

Consider an integral representation of the standard $\mathbb{R}_0$-valued probability, with probability densities ρ_{00} and characteristic functions

$$\chi_{\varphi,00}(\lambda) = \begin{cases} 1 & \text{if } \lambda \in [\varphi - \pi/2, \varphi + \pi/2] \\ 0 & \text{if } \lambda \notin [\varphi - \pi/2, \varphi + \pi/2] \end{cases} \tag{97}$$

treated as mappings $\mathbb{R}_0 \to \mathbb{R}_0$. For example, setting $\theta = \alpha - \beta$ in (63) and (64) one can express the probabilities in integral forms,

$$\frac{1}{2} \;=\; \int \chi_{\alpha,00}(\lambda) \rho_{00}(\lambda) \, \mathrm{d}\lambda = \frac{1}{2\pi} \int_{\alpha-\pi/2}^{\alpha+\pi/2} \mathrm{d}\lambda, \tag{98}$$

$$\frac{1}{2} \frac{\alpha - \beta}{\pi} \;=\; \int \chi_{\alpha,00}(\lambda) \chi_{\beta+\pi,00}(\lambda) \rho_{00}(\lambda) \, \mathrm{d}\lambda = \frac{1}{2\pi} \int_{\beta+\pi/2}^{\alpha+\pi/2} \mathrm{d}\lambda, \quad \beta \leqslant \alpha. \tag{99}$$

$\chi_{\varphi,00}(\lambda)$ is the characteristic function of the half-circle located symmetrically with respect to the angle φ; $\rho_{00}(\lambda) = 1/(2\pi)$ is the uniform probability density on the circle. Formula (99) is *local* in the sense of Bell [38] and Clauser and Horne [39], because of the *product structure* of the term

$$\chi_{\alpha,00}(\lambda) \chi_{\beta+\pi,00}(\lambda) = \chi_{\alpha,00}(\lambda) \odot_0 \chi_{\beta+\pi,00}(\lambda). \tag{100}$$

The case $k = l = 0$ of Bayes law discussed in Section VI is (with $\theta = \alpha - \beta$)

$$\frac{\alpha-\beta}{\pi} = \frac{\int \chi_{\beta+\pi,00}(\lambda)\chi_{\alpha,00}(\lambda)\rho_{00}(\lambda)\,\mathrm{d}\lambda}{\int \chi_{\alpha,00}(\lambda')\rho_{00}(\lambda')\,\mathrm{d}\lambda'} = \frac{p(0_2,0_1)}{p(0_1)} = \frac{p(1_2,1_1)}{p(1_1)} \tag{101}$$

$$= \int \chi_{\beta+\pi,00}(\lambda)\frac{\chi_{\alpha,00}(\lambda)\rho_{00}(\lambda)}{\int \chi_{\alpha,00}(\lambda')\rho_{00}(\lambda')\,\mathrm{d}\lambda'}\mathrm{d}\lambda, \tag{102}$$

which is equivalent to the assumption that the first measurement reduces the probability density according to

$$\rho_{00}(\lambda) \mapsto \frac{\chi_{\alpha,00}(\lambda)\rho_{00}(\lambda)}{\int \chi_{\alpha,00}(\lambda')\rho_{00}(\lambda')\,\mathrm{d}\lambda'}. \tag{103}$$

Equation (103) is an example of a classical projection postulate in theories based on $\mathbb{R}_0$ arithmetic.

Returning to the singlet case, corresponding to $k = 1$, $l = 0$, we can write it in analogy to (98) and (99),

$$g\big(p(a_2|a_1)\big)g\big(p(a_1)\big) = g\left(\frac{\int \chi_{a_1,00}(\lambda)\chi_{a_2,00}(\lambda)\rho_{00}(\lambda)\,\mathrm{d}\lambda}{\int \chi_{a_1,00}(\lambda)\rho_{00}(\lambda)\,\mathrm{d}\lambda}\right) g\left(\int \chi_{a_1,00}(\lambda)\rho_{00}(\lambda)\,\mathrm{d}\lambda\right) \tag{104}$$

$$= \frac{1}{2}g\left(2\int \chi_{a_1,00}(\lambda)\chi_{a_2,00}(\lambda)\rho_{00}(\lambda)\,\mathrm{d}\lambda\right) \tag{105}$$

$$= G\left(\int \chi_{a_1,00}(\lambda)\chi_{a_2,00}(\lambda)\rho_{00}(\lambda)\,\mathrm{d}\lambda\right) \tag{106}$$

$$= G\left(\int \chi_{a_1\wedge a_2,00}(\lambda)\rho_{00}(\lambda)\,\mathrm{d}\lambda\right), \tag{107}$$

where $G(x) = \frac{1}{2}g(2x)$, and

$$\chi_{a_1\wedge a_2,00}(\lambda) = \chi_{a_1,00}(\lambda)\chi_{a_2,00}(\lambda) \tag{108}$$

is the characteristic function representing the conjunction "a_1 and a_2". Notice that (106) is a non-Newtonian integral

$$G\left(\int \chi_{a_1,00}(\lambda)\chi_{a_2,00}(\lambda)\rho_{00}(\lambda)\,\mathrm{d}\lambda\right) = \int \chi_{a_1,11}(\lambda)\odot_1\chi_{a_2,11}(\lambda)\odot_1\rho_{11}(\lambda)\,\mathrm{D}_1\lambda, \tag{109}$$

of the function

$$\chi_{a_1,11}\odot_1\chi_{a_2,11}\odot_1\rho_{11} : \mathbb{R}_1 \to \mathbb{R}_1, \tag{110}$$

where

$$\begin{array}{ccc} \mathbb{R}_1 & \xrightarrow{\chi_{a_1,11}} & \mathbb{R}_1 \\ {\scriptstyle G^{-1}}\downarrow & & \uparrow{\scriptstyle G} \\ \mathbb{R}_0 & \xrightarrow{\chi_{a_1,00}} & \mathbb{R}_0 \end{array}, \qquad \begin{array}{ccc} \mathbb{R}_1 & \xrightarrow{\rho_{11}} & \mathbb{R}_1 \\ {\scriptstyle G^{-1}}\downarrow & & \uparrow{\scriptstyle G} \\ \mathbb{R}_0 & \xrightarrow{\rho_{00}} & \mathbb{R}_0 \end{array}, \tag{111}$$

and the multiplication is given by

$$x\odot_1 y = G\big(G^{-1}(x)\odot_0 G^{-1}(y)\big) = G\big(G^{-1}(x)G^{-1}(y)\big). \tag{112}$$

The right-hand side of (109) has again the Bell–Clauser–Horne product form, the only difference being that instead of $\odot_0$ one employs $\odot_1$. This is why (109) can be regarded as a local hidden-variable representation of singlet-state probabilities, hence a counterexample to Bell's theorem. This is the main idea of the approach to singlet-state correlations introduced in [12] and further discussed in [10, 13, 14].

A formal basis of the construction from [10, 12–14] is given by the following:

Lemma 4. *Consider four joint probabilities $p_{0_1 0_2}$, $p_{1_1 1_2}$, $p_{0_1 1_2}$, $p_{1_1 0_2}$, satisfying*

$$\sum_{ab} p_{ab} = 1, \tag{113}$$

$$\sum_a p_{aa_2} = \sum_a p_{a_1 a} = \frac{1}{2}. \tag{114}$$

A sufficient condition for

$$\sum_{ab} G(p_{ab}) \;=\; 1, \tag{115}$$

is given by $G(p) = \frac{1}{2}g(2p)$*, where* g *satisfies Lemma 1. Any such* G *has a fixed point at* $p = 1/4$*.*

A disadvantage of the construction based on Lemma 4 is its restriction to "rotationally symmetric" probabilities, i.e., those fulfilling (114). Moreover, being in itself sufficient as a counterexample to Bell's theorem, it lacks the generality typical of arbitrary $k, l \in \mathbb{Z}$.

The fundamental structure of the quantum probability model seems to be best described by Formula (69).

So far, the angles occurring in singlet-state probabilities were interpretable as experimental parameters (angles between polarizers or Stern–Gerlach devices). But what about arbitrary quantum states, even those described by infinite-dimensional Hilbert spaces? It turns out that the parameter in question can be interpreted in geometric terms, independently of the physical nature of the problem.

X. FUBINI–STUDY GEODESIC DISTANCE AS A HIDDEN VARIABLE

The scalar product $\langle a|b\rangle$ of two vectors belonging to some Hilbert space defines their Fubini–Study geodesic distance $\theta(a,b)$ [40–45],

$$|\langle a|b\rangle|^2 = \langle a|a\rangle\langle b|b\rangle \cos^2\theta(a,b). \tag{116}$$

Let P_b be a projector, $|b\rangle = P_b|a\rangle$, and $\langle a|a\rangle = 1$, so that $\langle b|b\rangle = \langle a|b\rangle = \langle a|P_b|a\rangle = P(b|a)$ is a conditional quantum probability. The geodesic distance between $|a\rangle$ and $|b\rangle$ satisfies

$$|\langle a|b\rangle|^2 = \langle a|P_b|a\rangle^2 = \langle a|P_b|a\rangle \cos^2\theta(a,b), \tag{117}$$

and thus,

$$P(b|a) = \cos^2\theta(a,b). \tag{118}$$

The formal angle $\theta(a,b)$ between the two vectors in the Hilbert space acquires a direct physical interpretation if a and b represent linear polarizations of photons: $\theta(a,b)$ becomes the angle between two polarizers. In the analogous case of the electrons, $\theta(a,b)$ would represent one half of the angle between two Stern–Gerlach devices.

Next, let us rewrite (118) as

$$P(b|a) = \cos^2\theta(a,b) = \sin^2\frac{\pi}{2}p(b|a) = g\big(p(b|a)\big) = \cos^2\frac{\pi}{2}\big(1-p(b|a)\big), \tag{119}$$

where $g : [0,1] \to [0,1]$ is the bijection we have introduced in the context of the singlet state. Probabilities $p(b|a)$ and $P(b|a) = g(p(b|a))$ represent, respectively, the hidden and the quantum neighboring levels of the hierarchy of (conditional) probabilities. The hidden probability is thus directly related to the Fubini–Study geodesic distance,

$$\theta(a,b) \;=\; \frac{\pi}{2}\big(1-p(b|a)\big), \tag{120}$$

$$p(b|a) \;=\; 1-\frac{\theta(a,b)}{\pi/2}, \tag{121}$$

$$q(b|a) \;=\; 1-p(b|a) = \frac{\theta(a,b)}{\pi/2}, \tag{122}$$

where $q(b|a)$ is the probability that two randomly chosen and intersecting straight lines intersect at an angle not exceeding $\theta(a,b) \in [0,\pi/2]$.

The Fubini–Study geodesic distance has been turned into a classical measure of a subset of a quarter-circle. It defines the whole hierarchy of probabilities, $g^k\big(p(b|a)\big)$, where $k = 1$ is the quantum one. Note that $g(p) = \sin^2\frac{\pi}{2}p$ has been elevated to the role of a universal bijection, defining an arithmetic applicable to all the possible (pure) quantum states.

Explicitly, we find

$$\vdots$$

$$g^{-1}\big(p(b|a)\big) = \frac{1}{\pi/2}\arcsin\sqrt{1-\frac{\arccos\sqrt{P(b|a)}}{\pi/2}}, \tag{123}$$

$$g^{0}\big(p(b|a)\big) = 1-\frac{\arccos\sqrt{P(b|a)}}{\pi/2}, \tag{124}$$

$$g^{1}\big(p(b|a)\big) = \sin^2\frac{\pi}{2}\left(1-\frac{\arccos\sqrt{P(b|a)}}{\pi/2}\right) = P(b|a), \tag{125}$$

$$g^{2}\big(p(b|a)\big) = \sin^2\left(\frac{\pi}{2}P(b|a)\right), \tag{126}$$

$$g^{3}\big(p(b|a)\big) = \sin^2\left[\frac{\pi}{2}\sin^2\left(\frac{\pi}{2}P(b|a)\right)\right], \tag{127}$$

$$\vdots$$

Since $\langle a|P_b|a\rangle = P(b|a)$ is real, it can be written as a real quadratic form,

$$\langle a|P_b|a\rangle = \sum_{rs}\Re(a_r)A_{rs}\Re(a_s) + \sum_{rs}\Im(a_r)B_{rs}\Im(a_s) + \sum_{rs}\Re(a_r)C_{rs}\Im(a_s). \tag{128}$$

Hence,

$$g^{2}\big(p(b|a)\big) = g^{1}\big(P(b|a)\big) \tag{129}$$

$$= g^{1}\left(\sum_{rs}\Re(a_r)A_{rs}\Re(a_s) + \sum_{rs}\Im(a_r)B_{rs}\Im(a_s) + \sum_{rs}\Re(a_r)C_{rs}\Im(a_s)\right) \tag{130}$$

$$= \bigoplus_{rs} g\big(\Re(a_r)\big)\odot g\big(A_{rs}\big)\odot g\big(\Re(a_s)\big)\bigoplus_{rs} g\big(\Im(a_r)\big)\odot g\big(B_{rs}\big)\odot g\big(\Im(a_s)\big)$$
$$\bigoplus_{rs} g\big(\Re(a_r)\big)\odot g\big(C_{rs}\big)\odot g\big(\Im(a_s)\big) \tag{131}$$

$$= \langle g(a)|\odot g(P_b)\odot|g(a)\rangle = \langle a_1|\odot_1 P_{b,1}\odot_1|a_1\rangle, \tag{132}$$

where $\langle g(a)|\odot g(P_b)\odot|g(a)\rangle$ in (132) is defined in a way that parallels the form of

$$\langle a|P_b|a\rangle = \langle a_0|\odot_0 P_{b,0}\odot_0|a_0\rangle \tag{133}$$

in (128), but with all the "standard" sums $+ = \oplus_0$ and products $\cdot = \odot_0$ replaced by $\oplus_1$ and $\odot_1$, and all the coefficients transformed by g. In effect, the difference between (128) and (132) is purely notational, as one can write the whole hierarchy of probabilities in a "quantum" form as well,

$$\vdots$$

$$g^{0}\big(p(b|a)\big) = \langle a_{-1}|\odot_{-1} P_{b,-1}\odot_{-1}|a_{-1}\rangle, \tag{134}$$

$$g^{1}\big(p(b|a)\big) = \langle a_0|\odot_0 P_{b,0}\odot_0|a_0\rangle, \tag{135}$$

$$g^{2}\big(p(b|a)\big) = \langle a_1|\odot_1 P_{b,1}\odot_1|a_1\rangle, \tag{136}$$

$$g^{3}\big(p(b|a)\big) = \langle a_2|\odot_2 P_{b,2}\odot_2|a_2\rangle \tag{137}$$

$$\vdots$$

This is the Copernican principle in action. The choice of the "quantum" level of the hierarchy is just a matter of convention. In fact, any formula from (123)–(127) can represent quantum mechanics known from textbooks.

It is perhaps more striking that any of these levels can be regarded as a hidden-variable level, where the hidden variable is given by an appropriate geodesic distance.

The concrete example of $g(p) = \sin^2\frac{\pi}{2}p$ can help us to understand the structure of the whole hierarchy. We will see that, in spite of the infinite dimension of the hierarchy, one effectively deals with a finite dimensional structure.

XI. EFFECTIVE TRUNCTION OF THE INFINITE HIERARCHY OF PROBABILITIES

Figure 2 explains why in spite of the infinite number of levels, those that statistically differ between one another may be limited to a finite "band" in the hierarchy. What it practically means is that if our level of the hierarchy is given by some l (say, $l = 0$) then, depending on the available precision of our experiments, we may restrict the analysis to a finite collection of probabilities. In the example depicted in Figure 2, we can restrict the analysis to 31 levels,

$$\{g^{-15}(p), \ldots, g^{-1}(p), p, g(p), \ldots, g^{15}(p)\}, \tag{138}$$

because the full infinite hierarchy is indistinguishable from

$$\{\ldots, g^{-15}(p), \ldots, g^{-15}(p), \ldots, g^{-1}(p), p, g(p), \ldots, g^{15}(p), \ldots, g^{15}(p), \ldots\}, \tag{139}$$

When increasing k in g^k, we effectively obtain a theory that may look discrete, because $g^k(p)$, $k > k_{\max}$, are indistinguishable from the red step function in Figure 2. For $g^k(p)$, $k < k_{\min}$, we obtain an analogous behavior of the inverse functions.

Let us stress that the above argument for indistinguishability has been formulated only for probabilities, $p \in [0, 1]$, hence for $g(p)$, and not for $g_{\mathbb{R}}(x)$, $x \notin [0, 1]$. In principle, for $x \notin [0, 1]$, all the levels of the hierarchy may be distinguishable.

Notice that for this concrete $g(p) = \sin^2 \frac{\pi}{2} p$, one finds $g^{15}(p) \approx 0$ if $p < 1/2$, $g^{15}(1/2) = 1/2$, and $g^{15}(p) \approx 1$ if $p > 1/2$. Thus, the higher-level probabilities possess several obvious analogies to neural activation functions [46], making links between the hierarchical structure and the measurement problem even more intriguing. An observer who measures $g^{15}(p)$ probabilities ignores practically all the events whose probability is smaller than $1/2$, and treats all $p > 1/2$ as certain.

This type of behavior is the essence of learning algorithms. An intriguing possibility occurs that $g(p)$ is a probability related to the act of learning that events with probability p are true. Hence, the natural question: Is the stabilization of large $k > 0$ iterates on effectively the step function a formal counterpart of stabilization of self-observation, a creation of self-awareness?

For the negative iterates, instead of a threshold function we tend toward a "white noise": $g^{-15}(0) = 0$, $g^{-15}(1/2) = 1/2$, $g^{-15}(1) = 1$, and $g^{-15}(p) \approx 1/2$, for $0 < p < 1$. The lower levels of the hierarchy become less and less diverse from the point of view of a higher-level observer. Here, the analogy is with observations of micro-scale events is quite evident. The relativity of probability becomes analogous to the "relativity of smallness"—what is small to us, may be large for a bacteria or an atom.

It is worth recalling that $g^{-15}(p)$ and $g^{15}(p)$ only *look* discrete due to our limited resolution—in reality, both maps are continuous bijections of $[0, 1]$ into itself.

Now, what about experiment and laws of large numbers? Can they somehow discriminate between all these probabilities?

XII. HIERARCHICAL LAWS OF LARGE NUMBERS

Laws of large numbers formalize the relations between probabilities (real numbers), (natural) numbers of trials and successes, and (rational) numbers of their relative frequencies. However, as we already know, all these notions are arithmetic dependent: a natural number $n_k = g^k(n) \in \mathbb{R}_k$ may not be a natural number from the point of view of some other $\mathbb{R}_l$, a rational number $(n/m)_k = g^k(n/m) \in \mathbb{R}_k$ may not be a rational number from the point of view of $\mathbb{R}_l$, and so on. The most general law of large numbers should involve all the levels of the hierarchy simultaneously. Dealing with binary events, we need an appropriate generalization of the Bernoulli law of large numbers.

To begin with, let us imagine we "live" in a world where all the possible computations are performed in terms of the arithmetic $\mathbb{R}_l$. If we toss a coin, say, one hundred times, and observe heads forty times, the arithmetic formulation of the experiment involves $n_l = 40_l$ heads in $N_l = 100_l$ trials. The experimental ratio is $n_l \oslash_l N_l = 40_l \oslash_l 100_l$. This is a rational number in $\mathbb{R}_l$.

If the same experiment is described by an observer who employs arithmetic $\mathbb{R}_j$, $j \neq l$, the experimental ratio is given by $n_j \oslash_j N_j = 40_j \oslash_j 100_j$. In terms of g^l and g^j we can write $40_l \oslash_l 100_l = g^l(40/100)$ and $40_j \oslash_j 100_j = g^j(40/100)$. Yet, if we demanded $g^l(40/100) = g^j(40/100)$, it would imply that $g^{l-j}(40/100) = 40/100$, i.e., $40/100$ is a fixed point of g^{l-j}. Since the same argument can be applied to any rational number, one arrives at the conclusion that the trivial case $g(x) = g^0(x)$ is the only solution.

One concludes that a nontrivial g generically implies $n_j \oslash_j N_j \neq n_l \oslash_l N_l$ for $l \neq j$. In other words, the same experiment can be described by different probabilities, $p_l = g^l(p) \neq p_j = g^j(p)$, although from the frequentist perspective both descriptions involve forty successes in one hundred trials. We inevitably arrive at the whole hierarchy.

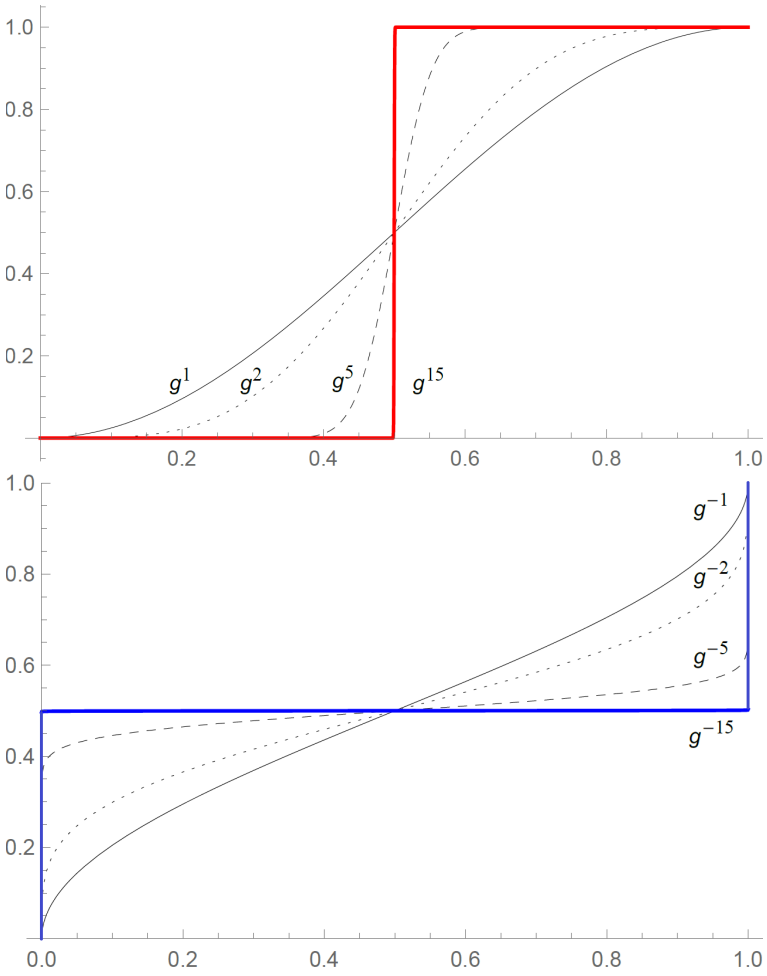

Figure 2. A total of 1, 2, 5 and 15 iterations of $g(p) = \sin^2 \frac{\pi}{2} p$ (upper plots). All the curves cross at $p = 1/2$ and are of the sigmoidal form, analogously to activation functions occurring in learning algorithms. Is it just a coincidence, or are there deeper connections to the problem of measurement, learning, or consciousness? Iterates g^k with $k > 15$ are practically indistinguishable within the precision of the plot: They all look like the red step function. An analogous phenomenon occurs for the negative iterates: $k = -1, -2, -5, -15$, but here almost all events described by $g^{-15}(p)$ are equally probable, hence indistinguishable for level-0 observers (lower plots). Effectively, even though the number of levels is infinite, the distinguishable ones are restricted to a finite "band" $k_{\min} \leqslant k \leqslant k_{\max}$. Of course, the Copernican aspect of the hierarchy means that the same happens in a neighborhood of any l, and not only $l = 0$ depicted here.

This is my tentative interpretation of the hierarchical structure. However, the links with neural activation functions deserve a separate study.

In order to formulate a generalized Bernoulli law of large numbers, we have to estimate the probability that

$$\left|g^k(p) \ominus_l n_l \oslash_l N_l\right| = \left|g^l\left(g^{k-l}(p) - n/N\right)\right| = g^l\left(\left|g^{k-l}(p) - n/N\right|\right) \geqslant \varepsilon_l = g^l(\varepsilon). \tag{140}$$

The modulus is defined in $\mathbb{R}_l$ in the standard way,

$$|x| = \begin{cases} x & \text{if } x \geqslant 0_l \\ \ominus_l x & \text{if } x < 0_l \end{cases}, \tag{141}$$

where we keep in mind that, by assumption, $0_l = g^l(0) = 0$ and the ordering relation is unaffected by a strictly increasing g. Inequality (140) effectively boils down to

$$\left|g^{k-l}(p) - n/N\right| \geqslant \varepsilon. \tag{142}$$

Next, we note that probabilities depicted in the lower part of Figure 3 are normalized in consequence of the identity

$$\left(g^k(p) \oplus_l g^k(q)\right)^{N_l} = \bigoplus_{n=0}^{N}{}_l \binom{N_l}{n_l} \odot_l g^k(q)^{(N-n)_l} \odot_l g^k(p)^{n_l} = 1_l = 1, \tag{143}$$

$$\binom{N_l}{n_l} = g^l\left[\binom{N}{n}\right]. \tag{144}$$

The probability

$$p(n_l, N_l)_k = \binom{N_l}{n_l} \odot_l g^k(q)^{(N-n)_l} \odot_l g^k(p)^{n_l} \tag{145}$$

$$= g^l\left[\binom{N}{n} g^{k-l}(q)^{N-n} g^{k-l}(p)^n\right] \tag{146}$$

corresponds to n_l sucessess in N_l trials. The expected number of successes and the corresponding variance read,

$$\langle n_l \rangle_k = \bigoplus_{n=0}^{N}{}_l n_l \odot_l \binom{N_l}{n_l} \odot_l g^k(q)^{(N-n)_l} \odot_l g^k(p)^{n_l} \tag{147}$$

$$= g^l\left[N g^{k-l}(p)\right] = N_l \odot_l g^k(p), \tag{148}$$

$$\langle n_l^{2_l} \rangle_k \ominus_l \langle n_l \rangle_k^{2_l} = g^l\left(N g^{k-l}(p) g^{k-l}(q)\right) = N_l \odot_l g^k(p) \odot_l g^k(q) \tag{149}$$

$$= N_l^{2_l} \odot_l \bigoplus_{n=0}^{N}{}_l \left[n_l \oslash_l N_l \ominus_l g^k(p)\right]^{2_l} \odot_l p(n_l, N_l)_k. \tag{150}$$

Applying g^{-l} to (149) and (150), we find

$$g^{k-l}(p) g^{k-l}(q)/N = \sum_{n=0}^{N} \left[n/N - g^{k-l}(p)\right]^2 \binom{N}{n} g^{k-l}(q)^{N-n} g^{k-l}(p)^n. \tag{151}$$

Now, let $n \in \mathbb{N}_{\varepsilon, g^{k-l}(p), N}$ if $\left|n/N - g^{k-l}(p)\right| \geqslant \varepsilon$. Then,

$$g^{k-l}(p) g^{k-l}(q)/N \geqslant \sum_{n \in \mathbb{N}_{\varepsilon, g^{k-l}(p), N}} \left[n/N - g^{k-l}(p)\right]^2 \binom{N}{n} g^{k-l}(q)^{N-n} g^{k-l}(p)^n \tag{152}$$

$$\geqslant \varepsilon^2 \sum_{n \in \mathbb{N}_{\varepsilon, g^{k-l}(p), N}} \binom{N}{n} g^{k-l}(q)^{N-n} g^{k-l}(p)^n \tag{153}$$

$$= \varepsilon^2 p\left(n \in \mathbb{N}_{\varepsilon, g^{k-l}(p), N}\right), \tag{154}$$

where $p\left(n \in \mathbb{N}_{\varepsilon, g^{k-l}(p), N}\right) \in \mathbb{R}_0$ is the 0th-level probability that $\left|n/N - g^{k-l}(p)\right| \geqslant \varepsilon$. In this way we have arrived at the standard Bernoulli law of large numbers in $\mathbb{R}_0$,

$$p\left(n \in \mathbb{N}_{\varepsilon, g^{k-l}(p), N}\right) \leqslant \frac{g^{k-l}(p) g^{k-l}(q)}{N\varepsilon^2}. \tag{155}$$

Of course, the left-hand side of (155) cannot be greater than 1, so the number of trials N must be chosen so that

$$\frac{g^{k-l}(p) g^{k-l}(q)}{\varepsilon^2} \leqslant N. \tag{156}$$

For $p_l = g^l(p)$ we find, denoting $\varepsilon_l = g^l(\varepsilon)$, $N_l = g^l(N)$,

$$p_l\left(n \in \mathbb{N}_{\varepsilon, g^{k-l}(p), N}\right) \leqslant g^l\left(\frac{g^{k-l}(p) g^{k-l}(q)}{N\varepsilon^2}\right) = g^l\left(\frac{g^{k-l}(p) g^{k-l}(q)}{g^{-l}(N_l) g^{-l}(\varepsilon_l)^2}\right)$$

$$= g^k(p) \odot_l g^k(q) \oslash_l \left(N_l \odot_l \varepsilon_l^{2_l}\right), \tag{157}$$

for any $k \in \mathbb{Z}$.

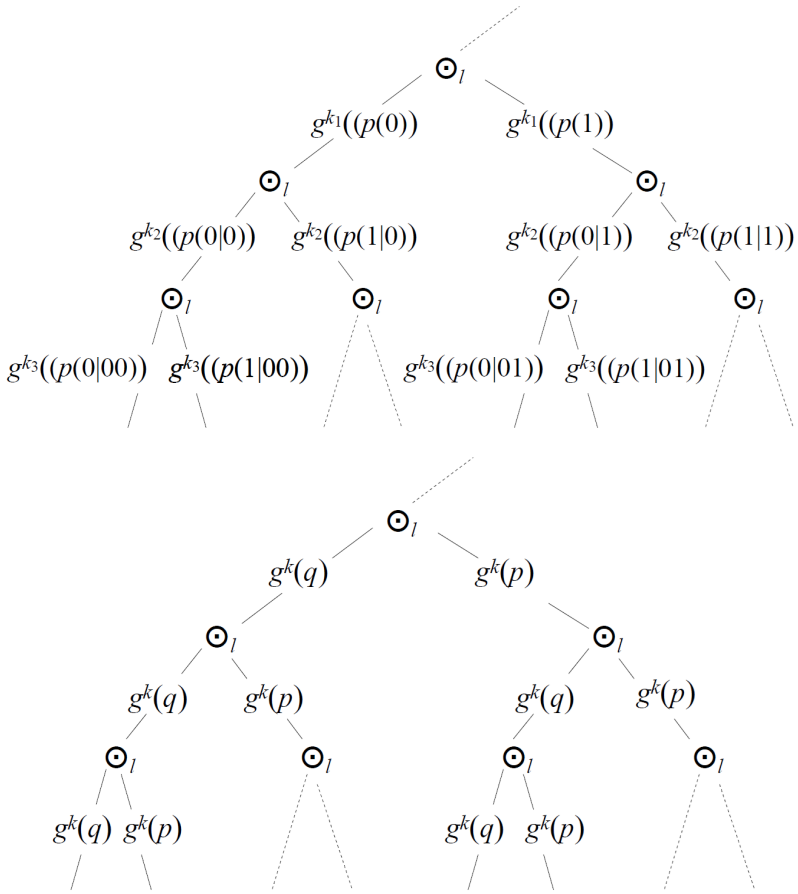

Figure 3. The upper diagram: An $\mathbb{R}_l$-valued branch of a binary tree of conditional probabilities. This is how one can include events with more results than just two. Assuming independent events and the same value of all k_j (the lower diagram), we can derive a hierarchical analog of the Bernoulli law of large numbers. Laws of large numbers are the places where theory and experiment meet.

In order to have a feel of the influence of $l \in \mathbb{Z}$ on the rate of convergence of experimental ratios to probabilities, consider the simple case of a symmetric coin, $p = q = 1/2$, and the universal quantum bijection $g(x) = \sin^2 \frac{\pi}{2} x$. Since $g^{k-l}(1/2) = 1/2$ for any k, l, we have to estimate

$$p_l \left(n \in \mathbb{N}_{\varepsilon, g^{k-l}(p), N} \right) \leqslant g^l \left(\frac{1}{4N\varepsilon^2} \right), \tag{158}$$

$$\frac{1}{4\varepsilon^2} \leqslant N. \tag{159}$$

Figure 4 illustrates the right-hand side of (158) for $\varepsilon = 0.1$ and $25 \leqslant N \leqslant 75$, for the first four iterates of g, from $g^1(x) = \sin^2 \frac{\pi}{2} x$ to

$$g^4(x) = \sin^2 \frac{\pi}{2} \left(\sin^2 \frac{\pi}{2} \left(\sin^2 \frac{\pi}{2} \left(\sin^2 \frac{\pi}{2} x \right) \right) \right). \tag{160}$$

The graphs are intriguing. Their interpretation is additionally obscured by the fact that Wolfram Mathematica operates in the arithmetic $\mathbb{R}_0$, which is not used by any of the four observers. The problem requires further studies.

XIII. HIERARCHICAL APPROACH TO BELL'S THEOREM—REVISITED

If we are able to reconstruct singlet-state probabilities in a hidden-variable way, it means that Bell's inequality (in any form) cannot be proved for the model. In the hierarchical context the obstacle for proving the inequality

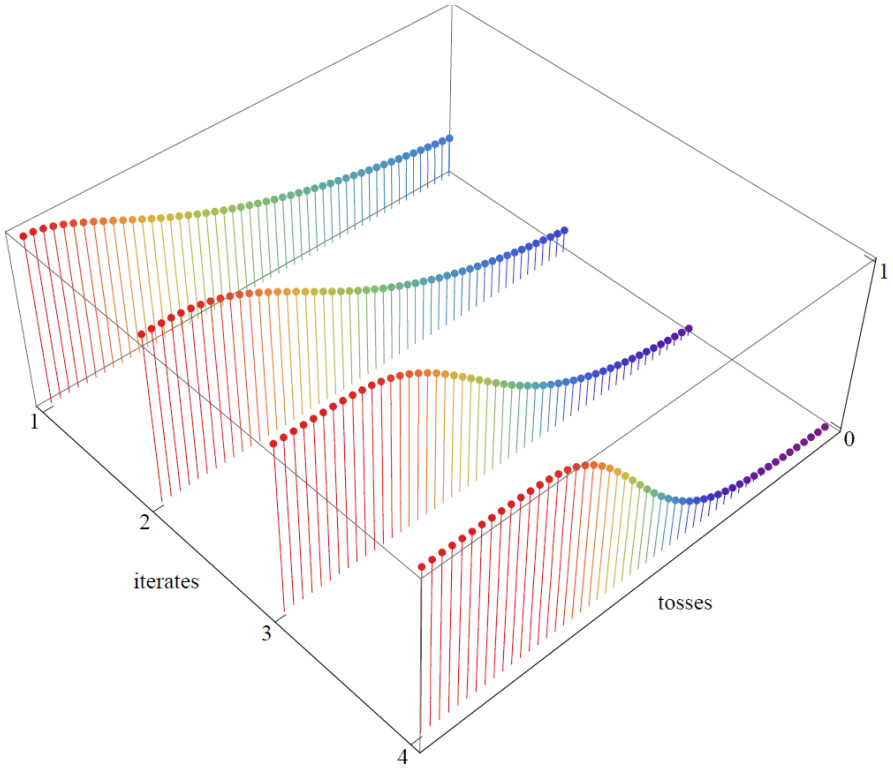

Figure 4. Hierarchical law of large numbers in action. Upper bound on probability of disagreement between theory and experiment in N tosses of a symmetric coin for four different arithmetics $\mathbb{R}_l$ of the observer. Plot of the right-hand side of (158) with $\varepsilon = 0.1$, for the four iterates g^l, $l = 1, 2, 3, 4$, of $g(x) = \sin^2 \frac{\pi}{2} x$. The number of coin tosses $25 \leqslant N \leqslant 75$. Plots are made in the arithmetic $\mathbb{R}_0$, implicitly assumed in Wolfram Mathematica 14.

lies in the lack of the k-level additivity of the l-level integrals, if $l \neq k$. The usual derivation, when seen from the hierarchical perspective, assumes $\oplus_0$-additivity of $\mathrm{D}_1\lambda$ integrals, which is untrue for a nontrivial g, and $g(p) = \sin^2 \frac{\pi}{2}$ in particular, hence the inequality derived at level zero does not apply to level 1: Level-0 formulas are "violated" by level-1 probabilities (and the other way around).

Let us see how it works. Consider the joint probabilities

$$\begin{aligned} P(a_1, a_2) &= P(\text{first } a_1 \text{ then } a_2) = g\big(p(a_2|a_1)\big)g\big(p(a_1)\big) = g\big(p(a_1|a_2)\big)g\big(p(a_2)\big) \\ &= P(\text{first } a_2 \text{ then } a_1) = P(a_2, a_1), \end{aligned} \tag{161}$$

where we assume the independence of the order in which the measurements are performed. This is typical of the scenarios involving "observer 1 measuring a_1" ("Alice") and "observer 2 measuring a_2" ("Bob") who are space-like separated and thus the order is undefined.

Now, we will derive an analog of the Clauser–Horne inequality [39]. We will work with probabilities (161). Let us stress that an analogous derivation was presented in [13], but was based on the form occurring in (109), that is by means of the bijection G. The derivation we will discuss now is based on $g(x)$, and not on $G(x) = \frac{1}{2} g(2x)$. Why? Because we want a proof that is easy to generalize to any $k, l \in \mathbb{Z}$.

We assume a local-hidden variable form of the probabilities that occur at the hidden level (level zero), hence

$$p(a_2|a_1) = \frac{\int \chi_{a_1,00}(\lambda)\chi_{a_2,00}(\lambda)\rho_{00}(\lambda)\,\mathrm{d}\lambda}{\int \chi_{a_1,00}(\lambda)\rho_{00}(\lambda)\,\mathrm{d}\lambda}, \tag{162}$$

$$p(a_1) = \int \chi_{a_1,00}(\lambda)\rho_{00}(\lambda)\,\mathrm{d}\lambda. \tag{163}$$

Level-one conditional probabilities

$$g\big(p(a_2|a_1)\big) \;=\; g\left(\frac{\int \chi_{a_1,00}(\lambda)\chi_{a_2,00}(\lambda)\rho_{00}(\lambda)\,\mathrm{d}\lambda}{\int \chi_{a_1,00}(\lambda)\rho_{00}(\lambda)\,\mathrm{d}\lambda}\right) \tag{164}$$

$$=\; \int \chi_{a_1,11}\odot_1\chi_{a_2,11}\odot_1\rho_{11}(\lambda)\,\mathrm{D}_1\lambda \oslash_1 \int \chi_{a_1,11}\odot_1\rho_{11}(\lambda)\,\mathrm{D}_1\lambda, \tag{165}$$

can be rewritten in several useful forms. First of all, introducing the reduced (conditional) probability density we obtain the "projection postulate",

$$\rho_{11}(\lambda) \;\mapsto\; \rho_{a_1,11}(\lambda) = \chi_{a_1,11}\odot_1\rho_{11}(\lambda)\oslash_1\int \chi_{a_1,11}\odot_1\rho_{11}(\lambda')\mathrm{D}_1\lambda', \tag{166}$$

$$g\big(p(a_2|a_1)\big) \;=\; \int \chi_{a_2,11}\odot_1\rho_{a_1,11}(\lambda)\,\mathrm{D}_1\lambda. \tag{167}$$

Secondly, we can explicitly express the conditional probability in a local Clauser–Horne form (in the arithmetic $\mathbb{R}_1$),

$$g\big(p(a_2|a_1)\big) \;=\; \int x_{a_1,11}(\lambda)\odot_1 y_{a_2,11}(\lambda)\odot_1\rho_{11}(\lambda)\,\mathrm{D}_1\lambda, \tag{168}$$

where

$$x_{a_1,11}(\lambda) \;=\; \chi_{a_1,11}(\lambda)\oslash_1\int \chi_{a_1,11}(\lambda')\odot_1\rho_{11}(\lambda')\mathrm{D}_1\lambda' \tag{169}$$

$$=\; g_{\mathbb{R}}\left(\frac{\chi_{a_1,00}(\lambda)}{\int \chi_{a_1,00}(\lambda')\rho_{00}(\lambda')\mathrm{d}\lambda'}\right), \tag{170}$$

$$y_{a_2,11}(\lambda) \;=\; \chi_{a_2,11}(\lambda) = g\big(\chi_{a_2,00}(\lambda)\big) = \chi_{a_2,00}(\lambda), \tag{171}$$

(because $g(0)=0$, $g(1)=1$).

Repeating step by step the derivation of the Clauser–Horne inequality [39], but here in the arithmetic $\mathbb{R}_1$, we can derive an analogous inequality which *must* be satisfied at the quantum level of the hierarchy. Such an inequality *cannot be violated* by quantum probabilities. For simplicity let us reduce the analysis to singlet-state probabilities and $g_{\mathbb{R}}(n)=n$ for any $n\in\mathbb{Z}$. Then,

$$\int x_{a_1,11}(\lambda)\odot_1\rho_{11}(\lambda)\mathrm{D}_1\lambda \;=\; 1, \tag{172}$$

$$\int y_{a_2,11}(\lambda)\odot_1\rho_{11}(\lambda)\mathrm{D}_1\lambda \;=\; 1/2, \tag{173}$$

$$0\leqslant x_{a_1,11}(\lambda)\leqslant X = g_{\mathbb{R}}\left(\frac{1}{\int \chi_{a_1,00}(\lambda')\rho_{00}(\lambda')\mathrm{d}\lambda'}\right) = 2, \tag{174}$$

$$0\leqslant y_{a_2,11}(\lambda)\leqslant Y = 1, \tag{175}$$

and

$$g\big(p(a_2|a_1)\big) = g\big(p(a_1|a_2)\big). \tag{176}$$

Next, we consider the Clauser–Horne linear combination

$$\begin{aligned} CH(\lambda) \;=\;& x_{a_1,11}\odot_1 y_{b_2,11}(\lambda)\ominus_1 x_{a_1,11}\odot_1 y_{b'_2,11}(\lambda)\oplus_1 x_{a'_1,11}\odot_1 y_{b_2,11}(\lambda)\oplus_1 x_{a'_1,11}\odot_1 y_{b'_2,11}(\lambda)\\ &\ominus_1 x_{a'_1,11}\odot_1 Y\ominus_1 X\odot_1 y_{b_2,11}(\lambda). \end{aligned} \tag{177}$$

Repeating in $\mathbb{R}_1$ the reasoning from [39], we obtain

$$-2\leqslant CH(\lambda)\leqslant 0. \tag{178}$$

$\mathbb{R}_1$-multiplying the latter by $\rho_{11}(\lambda)$, integrating with $\mathrm{D}_1\lambda$, and taking into account the $\mathbb{R}_1$-linearity of the $\mathrm{D}_1\lambda$ integral, we find

$$0\leqslant g\big(p(a_1|b_2)\big)\ominus_1 g\big(p(a_1|b'_2)\big)\oplus_1 g\big(p(a'_1|b_2)\big)\oplus_1 g\big(p(a'_1|b'_2)\big)\leqslant 2. \tag{179}$$

Notice that inequality (179) involves conditional probabilities, as opposed to the original Clauser–Horne one which was based on joint probabilities. The inequalities derived in the arithmetic induced by $G(x)$ and discussed in [12, 13] were also based on joint probabilities. However, joint probabilities involve the "macroscopic" level-0 multiplication of $1/2$ by $\cos^2(\alpha/2)$, whereas the conditional probabilities involve only the arithmetic of the "microscopic" level-1 probability $\cos^2(\alpha/2)$.

When investigating the violation of inequalities such as (179) one should keep in mind the difference between $g(x) = \sin^2 \frac{\pi}{2}x$, for $x \in [0,1]$, and its extension $g_{\mathbb{R}}(\mathrm{x})$ beyond the interval $[0,1]$. Here, (179) is derived under the assumption that $g_{\mathbb{R}}(n) = n$, for any integer n. Readers interested in explicit examples of $g_{\mathbb{R}}$ may consult [12–14].

The inequality that can indeed be violated is

$$0 \leqslant g\big(p(a_1|b_2)\big) - g\big(p(a_1|b_2')\big) + g\big(p(a_1'|b_2)\big) + g\big(p(a_1'|b_2')\big) \leqslant 2, \tag{180}$$

but it *cannot be proved* for the model, so is simply untrue. The technical difficulty in proving (180) is the lack of $\mathbb{R}_0$-linearity of the $\mathrm{D}_1\lambda$ integral.

The notion of "violation" of a formula is, in my opinion, very confusing. In the same sense one could say that the real-number inequality $x^2 \geqslant 0$ is violated by complex numbers. Instead of saying that $i^2 = -1$ violates $x^2 \geqslant 0$ one rather says that $x^2 \geqslant 0$ cannot be proved for all $x \in \mathbb{C}$. The same happens with the Bell inequality, derived in $\mathbb{R}_0$ but not valid in $\mathbb{R}_1$. On the other hand, the inequalities that *can* be derived in $\mathbb{R}_1$ are never "violated" in $\mathbb{R}_1$, but certainly will be untrue in some other $\mathbb{R}_k$.

XIV. INTERFERENCE, PROPAGATORS, DYNAMICS...

Formulas (134)–(137) show that the conditional probabilities can be written in scalar-product forms,

$$\begin{aligned}
&\vdots \\
g^0\big(p(b|a)\big) &= \langle b_{-1}| \odot_{-1} |a_{-1}\rangle = \langle b|a\rangle_{-1}, &(181)\\
g^1\big(p(b|a)\big) &= \langle b_0| \odot_0 |a_0\rangle = \langle b|a\rangle_0, &(182)\\
g^2\big(p(b|a)\big) &= \langle b_1| \odot_1 |a_1\rangle = \langle b|a\rangle_1, &(183)\\
g^3\big(p(b|a)\big) &= \langle b_2| \odot_2 |a_2\rangle = \langle b|a\rangle_2, &(184)\\
&\vdots
\end{aligned}$$

where we have introduced the compact notation,

$$\langle b_k| \odot_k |a_k\rangle = \langle b|a\rangle_k = g\left(\langle b|a\rangle_{k-1}\right). \tag{185}$$

These concrete scalar products are real. However, a complex scalar product can be always treated as a pair of reals with, in principle, different arithmetics for real and imaginary parts (see Appendix A). This type of generalized complex numbers was applied to non-Newtonian Fourier analysis on fractals [47], and proved very useful in circumventing certain impossibility theorems about Fourier transforms on the triadic Cantor set. Scalar products (181)–(184) when generalized to complex numbers (see Appendix A) can be used to generalize Feynman's path integral formalism to its hierarchical form, ultimately leading to propagators and time evolution.

We leave it for a future paper.

XV. AN OPEN ENDING

Standard modern physics involves a three-level hierarchy: quantum, classical and cosmological. As human observers, we are positioned at the center of this hierarchy, but the connections with the remaining two levels remain unclear. We do not understand how is it that we observe quantum properties (the measurement problem). Similarly, we do not understand our relation with the large-scale universe (the dark energy problem). In both cases the arithmetic freedom is probably essential [12, 48] but generally overlooked by our scientific community.

Bell's theorem is generally believed to eliminate levels lower than the quantum one, but the hierarchical picture questions this viewpoint: Quantum and classical probabilities typical of the singlet state belong to neighboring levels in the hierarchy—*any* two neighboring levels. Elimination of any of the levels, thus, would destroy the whole hierarchical structure, all quantum levels included [14].

To the best of my knowledge, the first systematic study of generalized arithmetics in physics was initiated by my paper [49], in which the relativity of arithmetic was interpreted in terms of a fundamental symmetry. However, I merely rediscovered a structure that had previously been introduced to calculus by Grossman and Katz (non-Newtonian calculus) [16–18], Maslov (idempotent analysis) [50] and Pap (g-calculus) [19]. The origins of the idea of generalized arithmetic and calculus can be traced back to the works of Volterra on the product integral [51], Kolmogorov [34], and Nagumo [35] on generalized means, and Rényi on generalized entropies [33]. Studies of a nonstandard number theory were initiated by Rashevsky [52] and, in a concrete form of non-Diophantine arithmetic, developed by Burgin [53–56]. Generalized forms of arithmetic can be found in Bennioff's attempts to formulate a coherent theory of physics and mathematics [57–61]. Mathematical constructions such as Lad's impediment functions [62], cepstral signal analysis [63, 64], fractal F^{α}-calculus [65–70], or nonextensive statistics [71–75], involve certain formal elements analogous to non-Newtonian integration or differentiation. The first application of non-Newtonian calculus to probability of which I am aware was provided by Meginniss in his analysis of the objectivity of p versus the subjectivity of $g(p)$, with applications to gambling theory [76]. Another field in which generalized arithmetic and non-Newtonian calculus are starting to attract attention is mathematical finance [77, 78]. From my personal perspective, the most important achievements of the new formalism include circumventing the limitations of Bell's theorem and Tsirelson bounds in quantum mechanics [12, 13]; the arithmetic of time, which appears to eliminate dark energy from cosmology in the same way that the arithmetic of velocities eliminated the luminiferous aether from special relativity [48]; formulating wave propagation along fractal coastlines [23]; and overcoming the limitations of Fourier analysis on Cantor sets [47].

The two most important observations of the present study seem to be the interpretation of the singlet-state probabilities in terms of several different arithmetic levels occurring in a single Formula (69),

$$P(a,b) = \underbrace{g\big(\overbrace{p(a|b)}^{\text{hidden}}\big)}_{\text{quantum}} \underbrace{\odot_0}_{\text{macroscopic}} \underbrace{g\big(\overbrace{p(b)}^{\text{hidden}}\big)}_{\text{quantum}}. \tag{186}$$

and the possible links with neural network learning algorithms.

The hierarchical structure is clearly "there". What we have understood so far is just the tip of the iceberg.

ACKNOWLEDGMENTS

Calculations in Mathematica were carried out at the Academic Computer Center in Gdańsk (CI TASK project pt01234).

Appendix A

1. Proof of (85):

$$
\begin{aligned}
\frac{\mathrm{D}_l A_{lk}(x)}{\mathrm{D}_k x} &= \lim_{\delta\to 0}\Big(A_{lk}(x\oplus_k \delta_k)\ominus_l A_{lk}(x)\Big)\oslash_l \delta_l \\
&= \lim_{\delta\to 0}\Big(g^{l-m}\circ A_{mn}\circ f^{k-n}(x\oplus_k \delta_k)\ominus_l g^{l-m}\circ A_{mn}\circ f^{k-n}(x)\Big)\oslash_l \delta_l \\
&= \lim_{\delta\to 0} g^l\Big(g^{-l+l-m}\circ A_{mn}\circ f^{k-n}\circ g^k\big(f^k(x)+f^k(\delta_k)\big)-g^{-l+l-m}\circ A_{mn}\circ f^{k-n}(x)\Big)\oslash_l \delta_l \\
&= \lim_{\delta\to 0} g^l\Big(g^{-m}\circ A_{mn}\circ f^{-n}\big(f^k(x)+f^k(\delta_k)\big)-g^{-m}\circ A_{mn}\circ f^{k-n}(x)\Big)\oslash_l \delta_l \\
&= \lim_{\delta\to 0} g^l\Big[\Big(g^{-m}\circ A_{mn}\circ f^{-n}\big(f^k(x)+\delta\big)-g^{-m}\circ A_{mn}\circ f^{k-n}(x)\Big)/\delta\Big] \\
&= \lim_{\delta\to 0} g^{l-m}\circ g^m\Big[f^m\circ g^m\Big(g^{-m}\circ A_{mn}\circ f^{-n}\big(f^k(x)+\delta\big)-g^{-m}\circ A_{mn}\circ f^{k-n}(x)\Big)/f^m(\delta_m)\Big] \\
&= \lim_{\delta\to 0} g^{l-m}\Big[g^m\Big(g^{-m}\circ A_{mn}\circ f^{-n}\big(f^k(x)+\delta\big)-g^{-m}\circ A_{mn}\circ f^{k-n}(x)\Big)\oslash_m \delta_m\Big] \\
&= g^{l-m}\lim_{\delta\to 0}\Big[\Big(A_{mn}\circ g^n\big(f^k(x)+\delta\big)\ominus_m A_{mn}\circ f^{k-n}(x)\Big)\oslash_m \delta_m\Big] \\
&= g^{l-m}\lim_{\delta\to 0}\Big[\Big(A_{mn}\circ g^n\big(f^n\circ f^{k-n}(x)+f^n(\delta_n)\big)\ominus_m A_{mn}\circ f^{k-n}(x)\Big)\oslash_m \delta_m\Big] \\
&= g^{l-m}\lim_{\delta\to 0}\Big[\Big(A_{mn}\big(f^{k-n}(x)\oplus_n \delta_n\big)\ominus_m A_{mn}\big(f^{k-n}(x)\big)\Big)\oslash_m \delta_m\Big] \\
&= g^{l-m}\left(\frac{\mathrm{D}_m A_{mn}\left(f^{k-n}(x)\right)}{\mathrm{D}_n f^{k-n}(x)}.\right)
\end{aligned}
$$

2. Proof of (86):

Define

$$B_{lk}(x) = \int_a^x A_{lk}(y)\mathrm{D}_k y = g^l\left(\int_{f^k(a)}^{f^k(x)} A_{00}(r)\mathrm{d}r\right) = g^l\left(B_{00}\big(f^k(x)\big)\right), \tag{A1}$$

$$C_{lk}(x) = g^{l-m}\left(\int_{f^{k-n}(a)}^{f^{k-n}(x)} A_{mn}(y)\mathrm{D}_n y\right) = g^{l-m}\left(C_{mn}\big(f^{k-n}(x)\big)\right) \tag{A2}$$

Now compute the derivatives:

$$
\begin{aligned}
\frac{\mathrm{D}_l}{\mathrm{D}_k x}B_{lk}(x) &= \frac{\mathrm{D}_l}{\mathrm{D}_k x}\int_a^x A_{lk}(y)\mathrm{D}_k y && \text{(A3)}\\
&= \frac{\mathrm{D}_l}{\mathrm{D}_k x}g^l\left(B_{00}\big(f^k(x)\big)\right) && \text{(A4)}\\
&= g^l\left(\frac{\mathrm{d}}{\mathrm{d}f^k(x)}B_{00}\big(f^k(x)\big)\right) && \text{(A5)}\\
&= g^l\left(\frac{\mathrm{d}}{\mathrm{d}f^k(x)}\int_{f^k(a)}^{f^k(x)} A_{00}(r)\mathrm{d}r\right) && \text{(A6)}\\
&= g^l\left(A_{00}\big(f^k(x)\big)\right) = A_{lk}(x) && \text{(A7)}\\
\frac{\mathrm{D}_l C_{lk}(x)}{\mathrm{D}_k x} &= g^{l-m}\left(\frac{\mathrm{D}_m C_{mn}\left(f^{k-n}(x)\right)}{\mathrm{D}_n f^{k-n}(x)}\right) && \text{(A8)}\\
&= g^{l-m}\left(\frac{\mathrm{D}_m}{\mathrm{D}_n f^{k-n}(x)}\int_{f^{k-n}(a)}^{f^{k-n}(x)} A_{mn}(y)\mathrm{D}_n y\right) && \text{(A9)}\\
&= g^{l-m}\left(A_{mn}\big(f^{k-n}(x)\big)\right) = A_{lk}(x) && \text{(A10)}
\end{aligned}
$$

The derivatives are identical,

$$\frac{\mathrm{D}_l}{\mathrm{D}_k x} g^{l-m}\left(\int_{f^{k-n}(a)}^{f^{k-n}(x)} A_{mn}(y)\mathrm{D}_n y\right) \;=\; \frac{\mathrm{D}_l}{\mathrm{D}_k x}\int_a^x A_{lk}(y)\mathrm{D}_k y, \tag{A11}$$

which implies

$$g^{l-m}\left(\int_{f^{k-n}(a)}^{f^{k-n}(x)} A_{mn}(y)\mathrm{D}_n y\right) = \int_a^x A_{lk}(y)\mathrm{D}_k y \oplus_l \text{constant} \tag{A12}$$

Setting $x = a$ we find

$$0_l = 0_l \oplus_l \text{constant} = \text{constant} \tag{A13}$$

hence

$$g^{l-m}\left(\int_{f^{k-n}(a)}^{f^{k-n}(x)} A_{mn}(y)\mathrm{D}_n y\right) = \int_a^x A_{lk}(y)\mathrm{D}_k y \tag{A14}$$

3. Powers (Repeated Multiplications)

In order to introduce generalized arithmetics of complex numbers we need a useful concept of a "first power" [22]. To this end, consider two sets X and Y and a map $A : X \to Y$ which can be described by a convergent power series. If X and Y are equipped with different arithmetics we first have to clarify the meaning of "power". This will be performed as follows. Consider two bijections $f_X : X \to \mathbb{R}$, $f_Y : Y \to \mathbb{R}$, and their composition $f = f_Y^{-1} \circ f_X$. The map

$$X \ni x \mapsto x^{1_{XY}} = f(x) \in Y \tag{A15}$$

defines a first Y-valued power of $x \in X$.

Lemma 5.

$$(x^{1_{XY}})^{1_{YZ}} \;=\; x^{1_{XZ}}, \tag{A16}$$
$$x^{1_{XY}1_{YX}} \;=\; x = x^{1_{XX}}, \tag{A17}$$
$$(x \odot_X y)^{1_{XY}} \;=\; x^{1_{XY}} \odot_Y y^{1_{XY}}, \tag{A18}$$
$$(x \oplus_X y)^{1_{XY}} \;=\; x^{1_{XY}} \oplus_Y y^{1_{XY}}. \tag{A19}$$

Proof.

$$\begin{aligned}
(x^{1_{XY}})^{1_{YZ}} &= x^{1_{XY}1_{YZ}} = f_Z^{-1}\circ f_Y \circ f_Y^{-1}\circ f_X(x)\\
&= f_Z^{-1}\circ f_X(x) = x^{1_{XZ}},\\
(x \odot_X y)^{1_{XY}} &= f_Y^{-1}\circ f_X(x \odot_X y) = f_Y^{-1}\big(f_X(x) f_X(y)\big)\\
&= f_Y^{-1}\Big(f_Y\big(x^{1_{XY}}\big) f_Y\big(y^{1_{XY}}\big)\Big)\\
&= x^{1_{XY}} \odot_Y y^{1_{XY}},\\
(x \oplus_X y)^{1_{XY}} &= f_Y^{-1}\circ f_X(x \oplus_X y) = f_Y^{-1}\big(f_X(x) + f_X(y)\big)\\
&= f_Y^{-1}\Big(f_Y\big(x^{1_{XY}}\big) + f_Y\big(y^{1_{XY}}\big)\Big) = x^{1_{XY}} \oplus_Y y^{1_{XY}}.
\end{aligned}$$

The remaining properties are obvious. □

Let $n_Y = f_Y^{-1}(n)$, with $n \in \mathbb{N}$ being a natural number satisfying the arithmetic defined by $f_{\mathbb{R}}(x) = x$. Then, first of all,

$$\begin{aligned}
n_Y &= f_Y^{-1}\circ f_{\mathbb{R}}(n) = n^{1_{\mathbb{R}Y}},\\
n &= n_{\mathbb{R}}.
\end{aligned}$$

More generally,

$$(n_Y)^{1_{YZ}} \;=\; f_Z^{-1}\circ f_Y\circ f_Y^{-1}(n) = f_Z^{-1}(n) = n_Z,$$

and, in particular,

$$\begin{aligned}(0_Y)^{1_{YZ}} &= f_Z^{-1}(0) = 0_Z,\\ (1_Y)^{1_{YZ}} &= f_Z^{-1}(1) = 1_Z,\end{aligned}$$

are the relations between neutral elements in Y and Z.
An nth Y-valued power reads

$$\begin{aligned}x^{n_{XY}} &= x^{1_{XY}}\odot_Y\cdots\odot_Y x^{1_{XY}} \quad (n \text{ times}),\\ &= f_Y^{-1}\Big(f_Y\big(x^{1_{XY}}\big)\ldots f_Y\big(x^{1_{XY}}\big)\Big)\\ &= f_Y^{-1}\Big(f_X(x)\ldots f_X(x)\Big)\\ &= f_Y^{-1}\Big(f_X\big(x\odot_X\cdots\odot_X x\big)\Big)\\ &= \big(x\odot_X\cdots\odot_X x\big)^{1_{XY}},\\ x^{n_{XY}}\odot_Y x^{m_{XY}} &= x^{(n+m)_{XY}}.\end{aligned}$$

Note that one naturally arrives at the definition of

$$\begin{aligned}x^{n_X} &= x\odot_X\cdots\odot_X x\\ &= x^{1_{XX}}\odot_X\cdots\odot_X x^{1_{XX}} = x^{n_{XX}},\end{aligned}$$

which coincides with the definition of x^{n_k} discussed earlier. So, n_X, understood as a power, can be identified with $n_X = f_X^{-1}(n)$, since

$$n_X\oplus_X m_X \;=\; f_X^{-1}(n+m) = (n+m)_X,$$

and thus one obtains the expected relation between products and sums,

$$x^{n_X}\odot_X x^{m_X} = x^{(n+m)_X} = x^{n_X\oplus_X m_X}.$$

Finally, let us compute

$$\begin{aligned}x^{n_{XY}m_{YZ}} &= \big(\underbrace{x\odot_X\cdots\odot_X x}_{n}\big)^{1_{XY}1_{YZ}}\odot_Z\cdots\odot_Z\big(\underbrace{x\odot_X\cdots\odot_X x}_{n}\big)^{1_{XY}1_{YZ}}\\ &= \big(x\odot_X\cdots\odot_X x\big)^{1_{XZ}}\odot_Z\cdots\odot_Z\big(x\odot_X\cdots\odot_X x\big)^{1_{XZ}}\\ &= \big(x^{1_{XZ}}\odot_Z\cdots\odot_Z x^{1_{XZ}}\big)\odot_Z\cdots\odot_Z\big(x^{1_{XZ}}\odot_Z\cdots\odot_Z x^{1_{XZ}}\big)\\ &= x^{1_{XZ}}\odot_Z\cdots\odot_Z x^{1_{XZ}} \quad (nm \text{ times})\\ &= x^{(nm)_{XZ}}\end{aligned}$$

4. Generalized Complex Numbers

Let $(x,y)\in X_1\times X_2$. The arithmetic of complex numbers is defined as

$$\begin{aligned}x\oplus y &= (x_1,x_2)\oplus(y_1,y_2) = (x_1\oplus_{X_1} y_1, x_2\oplus_{X_2} y_2),\\ x\odot y &= (x_1,x_2)\odot(y_1,y_2)\\ &= \Big(x_1\odot_{X_1} y_1\ominus_{X_1} x_2^{1_{X_2X_1}}\odot_{X_1} y_2^{1_{X_2X_1}}, x_1^{1_{X_1X_2}}\odot_{X_2} y_2\oplus_{X_2} x_2\odot_{X_2} y_1^{1_{X_1X_2}}\Big)\end{aligned}$$

(we feel free to represent pairs of numbers as either columns or rows). Neutral elements of multiplication and addition are given by

$$1_X \;=\; (1_{X_1},0_{X_2}), \tag{A20}$$

$$0_X \;=\; (0_{X_1},0_{X_2}). \tag{A21}$$

The "imaginary unit" is represented by

$$i_X \;=\; (0_{X_1}, 1_{X_2})\,. \tag{A22}$$

In order to simplify notation i_X will be sometimes denoted by i'. We get the standard "i squared equals minus one" rule,

$$\begin{aligned}
i'\odot(y_1,y_2) \;&=\; \Big(0_{X_1}\odot_{X_1} y_1 \ominus_{X_1} 1_{X_2}^{\,1_{X_2}x_1}\odot_{X_1} y_2^{\,1_{X_2}x_1}, 0_{X_1}^{\,1_{X_1}x_2}\odot_{X_2} y_2 \oplus_{X_2} 1_{X_2}\odot_{X_2} y_1^{\,1_{X_1}x_2}\Big)\\
&=\; \Big(\ominus_{X_1} y_2^{\,1_{X_2}x_1}, y_1^{\,1_{X_1}x_2}\Big) = (z_1,z_2),\\
i'\odot(z_1,z_2) \;&=\; \Big(\ominus_{X_1} z_2^{\,1_{X_2}x_1}, z_1^{\,1_{X_1}x_2}\Big) = \Big(\ominus_{X_1} y_1^{\,1_{X_1}x_2 1_{X_2}x_1}, \Big(\ominus_{X_1} y_2^{\,1_{X_2}x_1}\Big)^{1_{X_1}x_2}\Big)\\
&=\; (\ominus_{X_1} y_1, \ominus_{X_2} y_2) = i'\odot i'\odot(y_1,y_2) = \ominus(y_1,y_2).
\end{aligned}$$

Thinking of the plane as a representation of complex numbers, we identify real and imaginary parts as follows:

$$\begin{aligned}
\Re x \;&=\; \begin{pmatrix} x_1 \\ 0_{X_2}\end{pmatrix},\\
\Im x \;&=\; \begin{pmatrix} x_2^{\,1_{X_2}x_1} \\ 0_{X_2}\end{pmatrix},\\
i'\Im x \;&=\; i'\odot\begin{pmatrix} x_2^{\,1_{X_2}x_1} \\ 0_{X_2}\end{pmatrix} = \begin{pmatrix} 0_{X_1} \\ x_2\end{pmatrix},
\end{aligned}$$

(the imaginary part is also real!). Decomposition of a general complex x into its real and imaginary parts can be expressed in the usual way by means of addition,

$$\begin{aligned}
x \;&=\; \Re x \oplus i'\Im x = \begin{pmatrix} x_1 \\ 0_{X_2}\end{pmatrix} \oplus i'\odot\begin{pmatrix} x_2^{\,1_{X_2}x_1} \\ 0_{X_2}\end{pmatrix} = \begin{pmatrix} x_1 \\ 0_{X_2}\end{pmatrix} \oplus \begin{pmatrix} \ominus_{X_1} 0_{X_2}^{\,1_{X_2}x_1} \\ x_2^{\,1_{X_2}x_1 1_{X_1}x_2}\end{pmatrix}\\
&=\; \begin{pmatrix} x_1 \ominus_{X_1} 0_{X_1} \\ 0_{X_2}\oplus_{X_2} x_2\end{pmatrix} = \begin{pmatrix} x_1 \\ x_2\end{pmatrix}.
\end{aligned}$$

Complex conjugation reads

$$\begin{aligned}
x^* \;&=\; \Re x \ominus i'\Im x = \begin{pmatrix} x_1 \\ 0_{X_2}\end{pmatrix} \ominus i'\odot\begin{pmatrix} x_2^{\,1_{X_2}x_1} \\ 0_{X_2}\end{pmatrix} = \begin{pmatrix} x_1 \\ 0_{X_2}\end{pmatrix} \ominus \begin{pmatrix} \ominus_{X_1} 0_{X_2}^{\,1_{X_2}x_1} \\ x_2^{\,1_{X_2}x_1 1_{X_1}x_2}\end{pmatrix}\\
&=\; \begin{pmatrix} x_1 \oplus_{X_1} 0_{X_1} \\ 0_{X_2}\ominus_{X_2} x_2\end{pmatrix} = \begin{pmatrix} x_1 \\ \ominus_{X_2} x_2\end{pmatrix}.
\end{aligned}$$

Modulus squared is real,

$$\begin{aligned}
x\odot x^* \;&=\; \begin{pmatrix} x_1 \\ x_2\end{pmatrix}\odot\begin{pmatrix} x_1 \\ \ominus_{X_2} x_2\end{pmatrix} = \begin{pmatrix} x_1^{\,2_{X_1}} \oplus_{X_1} x_2^{\,2_{X_2}x_1} \\ \ominus_{X_2} x_1^{\,1_{X_1}x_2}\odot_{X_2} x_2 \oplus_{X_2} x_2\odot_{X_2} x_1^{\,1_{X_1}x_2}\end{pmatrix}\\
&=\; \begin{pmatrix} x_1^{\,2_{X_1}} \oplus_{X_1} x_2^{\,2_{X_2}x_1} \\ 0_{X_2}\end{pmatrix}.
\end{aligned}$$

The definition of addition is obvious, but let us take a closer look at multiplication. Recalling that $x^{1_{X_2}x_1} = f_{X_1}^{-1}\circ f_{X_2}(x)$, $x^{1_{X_1}x_2} = f_{X_2}^{-1}\circ f_{X_1}(x)$, we rephrase real and imaginary parts of the product as

$$\begin{aligned}
(x\odot y)_1 \;&=\; f_{X_1}^{-1}\Big(f_{X_1}(x_1)f_{X_1}(y_1) - f_{X_1}\Big(x_2^{\,1_{X_2}x_1}\Big) f_{X_1}\Big(y_2^{\,1_{X_2}x_1}\Big)\Big)\\
&=\; f_{X_1}^{-1}\Big(f_{X_1}(x_1)f_{X_1}(y_1) - f_{X_2}(x_2)f_{X_2}(y_2)\Big) = f_{X_1}^{-1}\big(\Re(\tilde{x}\tilde{y})\big), \qquad\qquad (A23)\\
(x\odot y)_2 \;&=\; f_{X_2}^{-1}\Big(f_{X_2}\Big(x_1^{\,1_{X_1}x_2}\Big) f_{X_2}(y_2) + f_{X_2}(x_2)f_{X_2}\Big(y_1^{\,1_{X_1}x_2}\Big)\Big)\\
&=\; f_{X_2}^{-1}\Big(f_{X_1}(x_1)f_{X_2}(y_2) + f_{X_2}(x_2)f_{X_1}(y_1)\Big) = f_{X_2}^{-1}\big(\Im(\tilde{x}\tilde{y})\big), \qquad\qquad (A24)
\end{aligned}$$

where $\tilde{x} = f_{X_1}(x_1) + i f_{X_2}(x_2) = \tilde{x}_1 + i\tilde{x}_2$, etc. Analogously,

$$\begin{aligned}(x \oplus y)_1 &= f_{X_1}^{-1}\Big(f_{X_1}(x_1) + f_{X_1}(y_1)\Big) = f_{X_1}^{-1}\big(\Re(\tilde{x}+\tilde{y})\big),\\ (x \oplus y)_2 &= f_{X_2}^{-1}\Big(f_{X_2}(x_2) + f_{X_2}(y_2)\Big) = f_{X_2}^{-1}\big(\Im(\tilde{x}+\tilde{y})\big).\end{aligned}$$

Accordingly,

$$x \odot y = \Big(f_{X_1}^{-1}\big(\Re(\tilde{x}\tilde{y})\big), f_{X_2}^{-1}\big(\Im(\tilde{x}\tilde{y})\big)\Big), \tag{A25}$$

$$x \oslash y = \Big(f_{X_1}^{-1}\big(\Re(\tilde{x}/\tilde{y})\big), f_{X_2}^{-1}\big(\Im(\tilde{x}/\tilde{y})\big)\Big), \tag{A26}$$

$$x \oplus y = \Big(f_{X_1}^{-1}\big(\Re(\tilde{x}+\tilde{y})\big), f_{X_2}^{-1}\big(\Im(\tilde{x}+\tilde{y})\big)\Big), \tag{A27}$$

$$x \ominus y = \Big(f_{X_1}^{-1}\big(\Re(\tilde{x}-\tilde{y})\big), f_{X_2}^{-1}\big(\Im(\tilde{x}-\tilde{y})\big)\Big). \tag{A28}$$

One can still further simplify operations with complex numbers. Note that

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} f_{X_1}^{-1} \circ f_{X_1}(x_1) \\ f_{X_2}^{-1} \circ f_{X_2}(x_2) \end{pmatrix} = \begin{pmatrix} f_{X_1}^{-1}(\Re\tilde{x}) \\ f_{X_2}^{-1}(\Im\tilde{x}) \end{pmatrix}, \tag{A29}$$

so that

$$\begin{pmatrix} f_{X_1}^{-1}(\Re\tilde{x}) \\ f_{X_2}^{-1}(\Im\tilde{x}) \end{pmatrix} \odot \begin{pmatrix} f_{X_1}^{-1}(\Re\tilde{y}) \\ f_{X_2}^{-1}(\Im\tilde{y}) \end{pmatrix} = \begin{pmatrix} f_{X_1}^{-1}\big(\Re(\tilde{x}\tilde{y})\big) \\ f_{X_2}^{-1}\big(\Im(\tilde{x}\tilde{y})\big) \end{pmatrix}, \tag{A30}$$

$$\begin{pmatrix} f_{X_1}^{-1}(\Re\tilde{x}) \\ f_{X_2}^{-1}(\Im\tilde{x}) \end{pmatrix} \oplus \begin{pmatrix} f_{X_1}^{-1}(\Re\tilde{y}) \\ f_{X_2}^{-1}(\Im\tilde{y}) \end{pmatrix} = \begin{pmatrix} f_{X_1}^{-1}\big(\Re(\tilde{x}+\tilde{y})\big) \\ f_{X_2}^{-1}\big(\Im(\tilde{x}+\tilde{y})\big) \end{pmatrix}, \tag{A31}$$

which are, perhaps, the most convenient forms of generalized complex arithmetic. The operations $\Re$ and $\Im$ are denoted by identical symbols no matter which arithmetic is used. This will not lead to inconsistencies and should be clear from a context. The standard Diophantine complex numbers are denoted either as $x = x_1 + ix_2$ or $x = (x_1, x_2)$ and it is understood that the two notations mean the same, i.e., it is allowed to write $x_1 + ix_2 = (x_1, x_2)$.

Lemma 6. *$\oplus$ and $\odot$ are associative and commutative, and $\oplus$ is distributive with respect to $\odot$.*

The proof is an immediate consequence of (A30) and (A31).

5. Complex-Valued Scalar Product via Non-Newtonian Integration

A complex-valued function is defined by the diagram

$$\begin{array}{ccc} X & \xrightarrow{A} & Y = Y_1 \times Y_2 \\ {\scriptstyle f_X}\big\downarrow & & \big\downarrow{\scriptstyle f_Y = f_{Y_1} \times f_{Y_2}} \\ \mathbb{R} & \xrightarrow{\tilde{A}} & \mathbb{R}^2 \end{array}. \tag{A32}$$

Consider two functions $A, B : X \to Y_1 \times Y_2$ associated with diagrams of the form (A32). Let $\tilde{A}(r) = \tilde{A}_1(r) + i\tilde{A}_2(r)$, $\tilde{B}(r) = \tilde{B}_1(r) + i\tilde{B}_2(r)$, and

$$\langle \tilde{A} | \tilde{B} \rangle = \int_{-f_X(T)/2}^{f_X(T)/2} \overline{\tilde{A}(r)} \tilde{B}(r) \mathrm{d}r.$$

$f_X(T)$ can be finite or infinite. The scalar product of two functions $A, B : X \to Y_1 \times Y_2$ is defined as

$$\langle A | B \rangle = \int_{\ominus_X T \oslash_X 2_X}^{T \oslash_X 2_X} A(x)^* \odot B(x) \mathrm{D}x. \tag{A33}$$

The same symbol of the scalar product for both $\langle \tilde{A} | \tilde{B} \rangle$ and $\langle A | B \rangle$ will not lead to ambiguities.

Lemma 7. *(Properties of the scalar product)*

$$\begin{aligned}
\langle A|B\rangle &= \begin{pmatrix} f_{Y_1}^{-1}\big(\Re\langle\tilde{A}|\tilde{B}\rangle\big) \\ f_{Y_2}^{-1}\big(\Im\langle\tilde{A}|\tilde{B}\rangle\big) \end{pmatrix}, && \text{(A34)}\\
\langle A|B\rangle^* &= \langle B|A\rangle, && \text{(A35)}\\
\langle A|B\oplus C\rangle &= \langle A|B\rangle\oplus\langle A|C\rangle, && \text{(A36)}\\
\langle A|\lambda\odot B\rangle &= \lambda\odot\langle A|B\rangle. && \text{(A37)}
\end{aligned}$$

Proof. The integrand

$$\begin{aligned}
A(x)^*\odot B(x) &= \begin{pmatrix} A_1(x) \\ \ominus_{Y_2}A_2(x) \end{pmatrix}\odot\begin{pmatrix} B_1(x) \\ B_2(x) \end{pmatrix}\\
&= \begin{pmatrix} f_{Y_1}^{-1}\Big(f_{Y_1}\big(A_1(x)\big)f_{Y_1}\big(B_1(x)\big) - f_{Y_2}\big(\ominus_{Y_2}A_2(x)\big)f_{Y_2}\big(B_2(x)\big)\Big) \\ f_{Y_2}^{-1}\Big(f_{Y_1}\big(A_1(x)\big)f_{Y_2}\big(B_2(x)\big) + f_{Y_2}\big(\ominus_{Y_2}A_2(x)\big)f_{Y_1}\big(B_1(x)\big)\Big) \end{pmatrix}\\
&= \begin{pmatrix} f_{Y_1}^{-1}\Big(\tilde{A}_1\big(f_X(x)\big)\tilde{B}_1\big(f_X(x)\big) + \tilde{A}_2\big(f_X(x)\big)\tilde{B}_2\big(f_X(x)\big)\Big) \\ f_{Y_2}^{-1}\Big(\tilde{A}_1\big(f_X(x)\big)\tilde{B}_2\big(f_X(x)\big) - \tilde{A}_2\big(f_X(x)\big)\tilde{B}_1\big(f_X(x)\big)\Big) \end{pmatrix}\\
&= \begin{pmatrix} f_{Y_1}^{-1}\Big(\tilde{C}_1\big(f_X(x)\big)\Big) \\ f_{Y_2}^{-1}\Big(\tilde{C}_2\big(f_X(x)\big)\Big) \end{pmatrix} = C(x).
\end{aligned}$$

So,

$$\begin{aligned}
\langle A|B\rangle &= \int_{\ominus_X T\oslash_X 2_X}^{T\oslash_X 2_X} C(x)\mathrm{D}x = \begin{pmatrix} f_{Y_1}^{-1}\left(\int_{f_X(\ominus_X T\oslash_X 2_X)}^{f_X(T\oslash_X 2_X)}\tilde{C}_1(r)\mathrm{d}r\right) \\ f_{Y_1}^{-1}\left(\int_{f_X(\ominus_X T\oslash_X 2_X)}^{f_X(T\oslash_X 2_X)}\tilde{C}_2(r)\mathrm{d}r\right) \end{pmatrix}\\
&= \begin{pmatrix} f_{Y_1}^{-1}\left(\int_{-f_X(T)/2}^{f_X(T)/2}\tilde{C}_1(r)\mathrm{d}r\right) \\ f_{Y_2}^{-1}\left(\int_{-f_X(T)/2}^{f_X(T)/2}\tilde{C}_2(r)\mathrm{d}r\right) \end{pmatrix} = \begin{pmatrix} f_{Y_1}^{-1}\left(\int_{-f_X(T)/2}^{f_X(T)/2}\Re\left(\overline{\tilde{A}(r)}\tilde{B}(r)\right)\mathrm{d}r\right) \\ f_{Y_2}^{-1}\left(\int_{-f_X(T)/2}^{f_X(T)/2}\Im\left(\overline{\tilde{A}(r)}\tilde{B}(r)\right)\mathrm{d}r\right) \end{pmatrix} && \text{(A38)}\\
&= \begin{pmatrix} f_{Y_1}^{-1}\big(\Re\langle\tilde{A}|\tilde{B}\rangle\big) \\ f_{Y_2}^{-1}\big(\Im\langle\tilde{A}|\tilde{B}\rangle\big) \end{pmatrix}.
\end{aligned}$$

By definition of complex conjugation

$$\begin{aligned}
\langle A|B\rangle^* &= \begin{pmatrix} f_{Y_1}^{-1}\big(\Re\langle\tilde{A}|\tilde{B}\rangle\big) \\ \ominus_{Y_2}f_{Y_2}^{-1}\big(\Im\langle\tilde{A}|\tilde{B}\rangle\big) \end{pmatrix} = \begin{pmatrix} f_{Y_1}^{-1}\big(\Re\langle\tilde{A}|\tilde{B}\rangle\big) \\ f_{Y_2}^{-1}\big(-\Im\langle\tilde{A}|\tilde{B}\rangle\big) \end{pmatrix} = \begin{pmatrix} f_{Y_1}^{-1}\big(\Re\langle\tilde{B}|\tilde{A}\rangle\big) \\ f_{Y_2}^{-1}\big(\Im\langle\tilde{B}|\tilde{A}\rangle\big) \end{pmatrix}\\
&= \langle B|A\rangle. && \text{(A39)}
\end{aligned}$$

The remaining two properties follow from associativity and distributivity of the arithmetic operations, supplemented by linearity of the integral. □

6. Continuous Transition Between Two Levels of the Hierarchy

There exists a simple generalization of the hierarchies, allowing for a continuous transition between the levels. It is based on the following

Lemma 8. *Consider a collection (finite or not) of parameters λ_j, $\sum_j\lambda_j = 1$, and a collection of bijections g_j that satisfy Lemma 1. Then $g = \sum_j\lambda_j g_j$ also satisfies Lemma 1.*

Proof. Each g_j can be written as

$$g_j(p) = \frac{1}{2} + h_j\left(p - \frac{1}{2}\right), \qquad \text{(A40)}$$

where $h_j(-x) = -h_j(x)$, so

$$g(p) = \sum_j \lambda_j g_j(p) = \underbrace{\frac{1}{2}\sum_j \lambda_j}_{1/2} + \underbrace{\sum_j \lambda_j h_j\left(p - \frac{1}{2}\right)}_{h(p-1/2)}. \tag{A41}$$

The function $h(x) = \sum_j \lambda_j h_j(x)$ is odd, as a linear combination of odd functions. □

As a side remark, let us note that $g(p) = \sin^2 \frac{\pi}{2} p$ satisfies all the concavity and monotonicity properties required for the existence of non-integer iterations g^r, $r \notin \mathbb{Z}$, (cf. [79], Chapter XV). The problem is intriguing as an alternative possibility of continuously switching between quantum and subquantum levels of the hierarchy (cf. the discussion of zero-multiplier iterative roots in [80], Chapter 11.5).

[1] Czachor, M.; Doebner, H.-D. Correlation experiments in nonlinear quantum mechanics. *Phys. Lett. A* **2002**, *301*, 139–152.
[2] Gisin, N. Stochastic quantum dynamics and relativity. *Helv. Phys. Acta* **1989**, *62*, 363–371.
[3] Gisin, N. Weinberg's non-linear quantum mechanics and supraluminal communications. *Phys. Lett. A* **1990**, *143*, 1–2.
[4] Polchinski, J. Weinberg's nonlinear quantum mechanics and the Einstein-Podolsky-Rosen paradox, *Phys. Rev. Lett.* **1991**, *66*, 397–400.
[5] Jordan, T.F. Reconstructing a nonlinear dynamical framework for testing quantum mechanics. *Ann. Phys.* **1993**, *225*, 83–113.
[6] Jaynes, E.T. *Probability Theory. The Logic of Science*; Bretthorst, G.L., Ed.; Cambridge University Press: Cambridge, UK, 2003.
[7] Aczél, J. *Lectures on Functional Equations and Their Applications*; Academic Press: New York, NY, USA, 1966.
[8] Cox, R.T. *The Algebra of Probable Inference*; Johns Hopkins University Press: Baltimore, MD, USA, 1961.
[9] Penrose, R.; Rindler, W. *Spinors and Space-Time. Volume 1: Two-Spinor Calculus and Relativistic Fields*; Cambridge University Press: Cambridge, UK, 1984.
[10] Czachor, M. Unifying aspects of generalized calculus. *Entropy* **2020**, *22*, 1180.
[11] Cormen, T.H.; Leiserson, ; C.E.; Rivest, R.L.; Stein, C. *Introduction to Algorithms*, 4th ed.; MIT Press and McGraw-Hill: Cambridge, MA, USA, 2022.
[12] Czachor, M. Arithmetic loophole in Bell's Theorem: Overlooked threat to entangled-state quantum cryptography. *Acta Phys. Polon. A* **2021**, *139*, 70–83.
[13] Czachor, M.; Nalikowski, K. Imitating quantum probabilities: Beyond Bell's theorem and Tsirelson bounds. *Found. Sci.* **2024**, *29*, 281–305.
[14] Czachor, M. Contra Bellum: Bell's theorem as a confusion of languages. *Acta Phys. Polon. A* **2023**, *143*, S158–S170.
[15] Cirel'son, B.S. Quantum generalizations of Bell's inequality. *Lett. Math. Phys.* **1980**, *4*, 93–100.
[16] Grossman, M.; Katz, R. *Non-Newtonian Calculus*; Lee Press: Pigeon Cove, MA, USA, 1972.
[17] Grossman, M. *The First Nonlinear System of Differential and Integral Calculus*, Mathco: Rockport, TX, USA, 1979.
[18] Grossman, M. *Bigeometric Calculus: A System with Scale-Free Derivative*; Archimedes Foundation: Rockport, TX, USA, 1983.
[19] Pap, E. g-calculus. *Zb. Rad. Prirod.-Fak. Ser. Mat.* **1993**, *23*, 145–156.
[20] Pap, E. Generalized real analysis and its applications. *Int. J. Approx. Reason.* **2008**, *47*, 368–386.
[21] Grabisch, M.; Marichal, J.-L.; Mesiar, R.; Pap, E. *Aggregation Functions*; Cambridge University Press, Cambridge, UK, 2009.
[22] Burgin, M.; Czachor, M. *Non-Diophantine Arithmetics in Mathematics, Physics, and Psychology*; World Scientific: Singapore, 2020.
[23] Czachor, M. Waves along fractal coastlines: From fractal arithmetic to wave equations. *Acta Phys. Polon. B* **2019**, *50*, 813–831.
[24] Zimmermann, H.-J. *Fuzzy Set Theory—and its Applications*, 3rd ed.; Kluwer: Boston, MA, USA, 1996.
[25] Dubois, D.; Prade, H. Towards fuzzy differential calculus. Part 1: Integration of fuzzy mappings. *Fuzzy Sets Syst.* **1982**, *8*, 1–17.
[26] Dubois, D.; Prade, H. Towards fuzzy differential calculus. Part 2: Integration on fuzzy intervals. *Fuzzy Sets Syst.* **1982**, *8*, 105–116.
[27] Dubois, D.; Prade, H. Towards fuzzy differential calculus. Part 3: Differentiation. *Fuzzy Sets Syst.* **1982**, *8*, 225–233.
[28] Zhang, D.; Mesiar, R.; Pap, E. Pseudo-integral and generalized Choquet integral. *Fuzzy Sets Syst.* **2022**, *446*, 193–221.
[29] Czachor, M.; Naudts, J. Thermostatistics based on Kolmogorov-Nagumo averages: Unifying framework for extensive and nonextensive generalizations. *Phys. Lett. A* **2002**, *298*, 369–374.
[30] Jizba, P.; Arimitsu, T. Observability of Rényi's entropy, *Phys. Rev. E* **2004**, *69*, 026128.
[31] Jizba, P.; Arimitsu, T. The world according to Rényi: Thermodynamics of fractal systems. *AIP Conf. Proc.* **2001**, *597*, 341.

[32] Jizba, P.; Korbel, J. When Shannon and Khinchin meet Shore and Johnson: Equivalence of information theory and statistical inference axiomatics. *Phys. Rev. E* **2020**, *101*, 042126.
[33] Rényi, A. Some fundamental questions of information theory. *MTA III. Oszt. Közl.* **1960**, *10*, 251–282; Reprinted in *Selected Papers of Alfréd Rényi*, Turán, P., Ed.; Akadémiai Kiadó: Budapest, Hungary, 1976.
[34] Kolmogorov, A.N. Sur la notion de la moyenne. *Atti. Acad. Naz. Lincei. Rend.* **1930**, *12*, 388–391; Reprinted in *Selected Works of A. N. Kolmogorov. Vol.1. Mathematics and Mechanics*; Tikhomirov, V. M., Ed.; Kluwer: Dordrecht, The Netherlands, 1991.
[35] Nagumo, M. Über eine Klasse der Mittelwerte, *Jpn. J. Math.* **1930**, *7*, 71–79; Reprinted in *Mitio Nagumo Collected Papers*; Yamaguti, M., Nirenberg, L., Mizohata, S., Sibuya, Y., Eds.; Springer: Tokyo, Japan, 1993.
[36] Naudts, J. *Generalized Thermostatistics*; Springer: London, UK, 2011.
[37] Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423, 623–653.
[38] Bell, J.S. On the Einstein-Podolsky-Rosen paradox. *Physics* **1964**, *1*, 195.
[39] Clauser, J.F.; Horne, M.A. Experimental consequences of objective local theories. *Phys. Rev. D* **1974**, *10*, 526.
[40] Fubini, G. Sulle metriche definite da una forme Hermitiana. *Atti Istit. Veneto* **1904**, *63*, 502–513.
[41] Study, E. Kürzeste Wege im komplexen Gebiet. *Math. Ann.* **1905**, *60*, 321-378.
[42] Field, T.R.; Hughston, L.P. The geometry of coherent states. *J. Math. Phys.* **1999**, *40*, 2568–2583.
[43] Brody, D.C.; Hughston, L.P. Geometric quantum mechanics. *J. Geom. Phys.* **2001**, *38*, 19–53.
[44] Bengtsson, I.; Życzkowski, K. *Geometry of Quantum States. An Introduction to Quantum Entanglement*; Cambridge University Press: Cambridge, UK, 2006.
[45] Chruściński, D. Geometric aspects of quantum mechanics and quantum entanglement. *J. Phys. Conf. Ser.* **2006**, *30*, 9.
[46] Kunc, V.; Kléma, J. Three decades of activations: A comprehensive survey of 400 activation functions for neural networks. *arXiv* **2024**, arXiv:2402.09092.
[47] Aerts, D.; Czachor, M.; Kuna, M. Fourier transforms on Cantor sets: A study in non-Diophantine arithmetic and calculus. *Chaos Solitons Fractals* **2016**, *91*, 461–468.
[48] Czachor, M. Non-Newtonian mathematics instead of non-Newtonian physics: Dark matter and dark energy from a mismatch of arithmetics. *Found. Sci.* **2021**, *26*, 75–95.
[49] Czachor, M. Relativity of arithmetic as a fundamental symmetry of physics. *Quantum Stud. Math. Found.* **2016**, *3*, 123–133.
[50] Maslov, V.P. On a new superposition principle for optimization problems. *Uspekhi Mat. Nauk* **1987**, *42*, 43. (In Russian)
[51] Volterra, V.; Hostinský, B. *Opérations Infinitésimales Linéaires. Applications Aux équations Différentielles et Fonctionnelles*; Gauthier-Villars: Paris, France, 1938.
[52] Rashevsky, P.K. On the dogma of the natural numbers. *Uspekhi Mat. Nauk.* **1973**, *28*, 243–246. (In Russian)
[53] Burgin, M.S. Nonclassical models of the natural numbers. *Uspekhi Mat. Nauk* **1977**, *32*, 209–210. (In Russian)
[54] Burgin, M.S. *Non-Diophantine Arithmetics, or Is It Possible That 2 + 2 Is Not Equal to 4?*; Ukrainian Academy of Information Sciences: Kiev, Ukraine, 1997. (In Russian)
[55] Burgin, M. Introduction to projective arithmetics. *arXiv* **2010**, arXiv:1010.3287.
[56] Burgin, M.; Meissner, G. $1 + 1 = 3$: Synergy arithmetics in economics. *Appl. Math.* **2017,** *8*, 133–134.
[57] Benioff, P. Towards a coherent theory of physics and mathematics. *Found. Phys.* **2002**, *32*, 989–1029.
[58] Benioff, P. Towards a coherent theory of physics and mathematics. The theory-experiment connection. *Found. Phys.* **2005**, *35*, 1825–1856.
[59] Benioff, P. Fiber bundle description of number scaling in gauge theory and geometry. *Quant. Stud.: Math. Found.* **2015**, *2*, 289–313.
[60] Benioff, P. Space and time dependent scaling of numbers in mathematical structures: Effects on physical and geometric quantities. *Quant. Inf. Proc.* **2016**, *15*, 1081–1102.
[61] Benioff, P. Effects of a scalar scaling field on quantum mechanics. *Quant. Inf. Proc.* **2016**, *15*, 3005–3034.
[62] Lad, F. Embedding Bayes' theorem in general learning rules: Connections between idealized behaviour and empirical research on learning. *Brit. J. Math. Stat. Psychol.* **1978**, *31*, 113–125.
[63] Childers, D.G.; Skinner, D.P.; Kemerait, R.C. The cepstrum: A guide to processing. *Proc. IEEE* **1977**, *65*, 1428–1443.
[64] Oppenheim, A.V.; Schafer, R.W. From frequency to quefrency: A history of the cepstrum. *IEEE Signal Process. Mag.* **2004**, *21*, 95.
[65] Parvate, A.; Gangal, A.D. Calculus on fractal subsets of real line. (I) Formulation. *Fractals* **2009**, *17*, 53.
[66] Parvate, A.; Gangal, A.D. Calculus on fractal subsets of real line. (II) Conjugacy with ordinary calculus. *Fractals* **2011**, *19*, 271.
[67] Golmankhaneh, A.K.; Baleanu, D. New derivatives on the fractal subset of real line. *Entropy* **2016**, *18*, 1.
[68] Golmankhaneh, A.K.; Baleanu, D. Non-local integrals and derivatives on fractal sets with applications. *Open Phys.* **2016**, *14*, 542–548.
[69] Golmankhaneh, A.K.; Tunc, C. On the Lipschitz condition in the fractal calculus. *Chaos Soliton Fract.* **2017**, *95*, 140–147.
[70] Golmankhaneh, A.K. *Fractal Calculus and Its Applications: F^{α}-calculus*; World Scientific: Singapore, 2022.
[71] Tsallis, C.; Mendes, R.; Plastino, A. The role of constraints within generalized nonextensive statistics. *Physica A* **1998**, *261*, 543–554.
[72] Touchette, H. When is a quantity additive, and when is it extensive? *Physica A* **2002**, *305*, 84–88.
[73] Nivanen, L.; Le Mehaute, A.; Wang, Q.A. Generalized algebra within a nonextensive statistics. *Rep. Math. Phys.* **2003**, *52*, 437–444.
[74] Kaniadakis, G. Nonlinear kinetics underlying generalized statistics. *Physica A* **2001**, *296*, 405.

[75] Kaniadakis, G. Theoretical foundations and mathematical formalism of the power-law tailed statistical distributions. *Entropy* **2013**, *15*, 3983–4010.
[76] Meginniss, J.R. Non-Newtonian calculus applied to probability, utility, and Bayesian analysis. In *Proceedings of the Business and Economic Statistics Section*; American Statistical Association: Washington, DC, USA, 1980; pp. 405–410.
[77] Carr, P.; Cherubini, U. Option pricing generators. *Front. Math. Financ.* **2023**, *2*, 150–169.
[78] Carr, P.; Cirillo, P. A pseudo-analytic generalization of the memoryless property for continuous random variables and its use in pricing contingent claims. *R. Soc. Open Sci.* **2024**, *11*, 231690.
[79] Kuczma, M. *Functional Equations in a Single Variable*; Polish Scientific Publishers: Warszawa, Poland, 1968.
[80] Kuczma, M.; Choczewski, B.; Ger, R. *Iterative Functional Equations*; Cambridge University Press: Cambridge, UK, 1990.