

Essay

Not peer-reviewed version

The Code Underneath

[Julio Rives](#) *

Posted Date: 8 November 2024

doi: [10.20944/preprints202411.0594.v1](https://doi.org/10.20944/preprints202411.0594.v1)

Keywords: Keywords Probability Mass Function (PMF), Inverse-Square Law (ISL), Newcomb-Benford Law (NBL), positional Notation (PN), radix economy, harmonic series, trigamma function, likelihood, curious series,



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

The Code Underneath

Julio Rives

Dept. of Artificial Intelligence, Univ. Nacional de Educación a Distancia (UNED), Paseo Senda del Rey, 7, Madrid, 28040, Spain; jrives@dia.uned.es

Abstract: An inverse-square probability mass function (PMF) is at Newcomb-Benford Law's (NBL) root and ultimately at the origin of positional notation and conformality. $\Pr(Z) = (2Z)^{-2}$, where $Z \in \mathbb{Z}^+$. Under its tail, we find information as harmonic likelihood $\mathcal{L}([s, t)) = H_{t-1} - H_{s-1}$, where H_n is the n th harmonic number. The global \mathbb{Q} -NBL is $\Pr(b, q) = \mathcal{L}([q, q+1)) / \mathcal{L}([1, b)) = (qH_{b-1})^{-1}$, where b is the base, and q is a quantum ($1 \leq q < b$). Under its tail, we find information as logarithmic likelihood $\ell([i, j)) = \ln j/i$. The fiducial \mathbb{R} -NBL is $\Pr(r, d) = \ell([d, d+1)) / \ell([1, r)) = \log_r(1 + 1/d)$, where $r < b$ is the radix of a local complex system. In the framework of bijective numeration, we prove that the set of Kempner's series conforms to the global NBL and that the local NBL is length- and position-invariant. The global Bayesian rule multiplies the correlation between two numbers, s and t , by a likelihood ratio that is the NBL probability of bucket $[s, t)$ relative to b 's support. The local Bayesian rule is $\tilde{o}(j : i|r) = i/j \log_r j/i$. To encode the odds of quantum j against i locally, we multiply the prior odds $\Pr(b, j) / \Pr(b, i)$ by a likelihood ratio, which is the NBL probability of bin $[i, j)$ relative to r 's support. This two-factor structure is recurrent under arithmetic operations. The Bayesian rule to recode local data is $\tilde{o}(j : i|r') = \tilde{o}(j : i|r) \ln r / \ln r'$. A particular case of Bayesian data produces the algebraic field of "referential ratios", $\frac{A-B}{A-C}$. The cross-ratio, the central tool in conformal geometry, is a ratio of referential ratios. A one-dimensional coding source reflects the harmonic external world, the annulus $\{x \in \mathbb{Q} \mid 1 \leq |x| < b\}$, into its logarithmic coding space, the ball $\{x \in \mathbb{Q} \mid |x| < 1 - 1/b\}$. The source's conformal encoding function is $y = \log_r(2x - 1)$, where x is the observed Euclidean distance to an object's position. The conformal decoding function is $x = \frac{1}{2}(1 + r^y)$. Both functions, unique under basic requirements, enable information- and granularity-invariant recursion to model the multiscale reality.

Keywords: Probability Mass Function (PMF); Inverse-Square Law (ISL); Newcomb-Benford Law (NBL); positional Notation (PN); radix economy; harmonic series; trigamma function; harmt (harmonic unit of information); likelihood; curious series; bijective numeration; logarithmic scale; global-local duality; coding source; Bayesian Law; secretary problem; referential ratio; cross-ratio; relationalism; conformability; granularity; constructability; divisibility; randomness

MSC: 03D20; 11A63; 11A67; 12F99; 30C35; 30F45; 33B15; 60-08; 60A99; 62C10; 65E10; 68P30; 83-10; 93A13; 94A17

1. Introduction

After Simon Newcomb's public note [1] and Benford's statement [2] that "small things are more numerous than large things, and there is a tendency for the step between sizes to be equal to a fixed fraction of the last preceding phenomenon or event", many scientists [3] tried to explicate the strange high frequency of the micro in nature, the rarity of the macro, and the ebbing progression of the gaps in between.

Nature pivots on exponential powers. Benford underlined that "the geometric series has long been recognized as a common phenomenon in factual literature and in the ordinary affairs of life". Nevertheless, human functions are often arithmetic-centric. Will there be a natural coding system to convert these realms into one another, the observable into our inner world's models, and vice versa? In other words, does nature count on a conformal transformation mechanism [4]?

In modern terms, Newcomb-Benford law (NBL) states that the first digits of randomly chosen original data typically outline a logarithmic curve in an impressive diversity of fields regardless of their physical units. Equivalently, the law remarks that raw natural data usually belong to nearly scale-invariant geometric series. Among its manifestations, it is fascinating that linear coefficients represented by mathematical and physical constants [5] (e.g., proportionality parameters or scalar potentials) adhere to the law.

Although this scenario suggests that NBL might account for an elementary principle, we have yet to clarify its origin, realize a theoretical basis, or encounter a convincing reason [6]. Berger [7] laments that "There is no known back-of-the-envelope argument, not even a heuristic one, that explains the appearance of Benford's law across the board in data that is pure or mixed, deterministic or stochastic, discrete or continuous-time, real-valued or multidimensional."

We claim a primordial probability inverse-square law (ISL) is at NBL's root. This "canonical" probability mass function (PMF) has a double fundamental effect, namely the NBL for the discrete (global and harmonic) and continuous (local and logarithmic) domains. We prefer to anticipate these three laws' properties and affiliated terminology in Table 1, indicating their scope and character, baseline set, physical incarnation, scale, formula, information function, cardinality, and how we will denominate the corresponding item, an item list, and an item range.

Table 1. Nature and terminology of the three foundational contexts, where Z is a nonzero integer, b is the global base, q is a quantum ($1 \leq q < b$), H_n is the n th harmonic number, r is the local radix, and d is a digit ($1 \leq d < r$).

Property ↓ / Law →	Canonical PMF	First NBL	Second NBL
Scope	Mathematical	Global	Local
Character	Discrete	Discrete	Continuous
Baseline set	Natural, Integer	Rational	Real
Physics	Field	Potential	Entropy
Entity at origin	Indeterminate	Observer	Coding source
Scale	Linear	Harmonic	Logarithmic
Probability law	$(2Z)^{-2}$	$(qH_{b-1})^{-1}$	$\log_r(1 + 1/d)$
Information function		Digamma	Logarithm
Cardinality	Infinite	Base	Radix
Item	Number	Quantum	Digit
Item list $\langle \alpha, \beta, \dots \rangle$	String	Chain	Numerical
Item range $[\alpha, \beta)$	Interval	Bucket	Bin

Are these laws naturally predetermined probability distributions? We champion the view that the canonical PMF is a brute fact and, consequently, the global and local versions of NBL are inescapable. For one thing, their mode is one. This number is the base case for almost all proofs by mathematical induction, statistically the most probable cardinal of a natural set (e.g., one cosmos, one black hole at the center of a galaxy, one star ruling an orbital planetary system, one heart pumping a body's blood, one nucleus regulating the cellular activity, et cetera), and seed in the majority of recursive computational processes. We read in [8] that numbers close to the multiplicative unit are not preferably rooted in mathematics, but a simple glance at the Table of Constants in [9] points in the opposite direction. Small leading digits and, in general, small significands (mantissae) of coefficients and magnitudes are the most common in sciences, albeit, of course, we can find cardinalities of all sizes.

That the universe is prone to favor slightness is particularly blatant in physics and chemistry. For instance, following the standard cosmological model [10], the abundance of hydrogen and helium

is roughly 75 % and 23 % of all baryonic matter, respectively [11]. Higher atomic numbers than 26 (iron) are progressively more and more infrequent. Nevertheless, the universe's heaviest elements can comparatively produce the most remarkable galactic phenomena despite their shortage [12] (e.g., the necessary metals to form the Milky Way represent only about 2 % of the galaxy's disk mass). Why do accessibility and reactivity maintain a hyperbolic relationship?

Notwithstanding that NBL assumes standard positional notation (PN) in its fiducial form, our logic also permits obtaining the formulae for non-standard place-value numeral systems. In particular, every NBL's PMF for standard PN has a bijective numeration [13] peer. For example, the standard and bijective decimal system global and local laws are similar but different. These results show that the precision of NBL is nonessential, while the support positional scale is what matters.

This article's field of study is mathematical and computational physics, delving into philosophy, theoretical physics, information theory, probability theory, and number theory. We have organized it as follows. We first examine the challenges researchers historically faced in deducing NBL and the state of the art in this field. Afterward, we present a one-parameter inverse-square PMF for the natural numbers with positive probabilities summing to one, extensible to the integers, and diverging mean (no bias). Next, we deduce the fiducial NBL passing through the global NBL; this two-phase derivation clarifies why the tendency for the minor numbers revealed by the natural sciences can be regular only if we assume that an all-encompassing base exists. To support this view, we substantiate that the set of Kempner's "curious" series conforms to the global NBL for bijective numeration. Further, we surmise a PN resolution, i.e., the prospect of a natural position threshold ascribed to a place-value number system.

Information theory [14] comes into play when we discover that information is prior to probability in the context of NBL. Likewise, a unit fraction is the harmonic likelihood of an elemental quantum gap, and a digit of a numeral written in PN is a bin that covers a proportion of the available logarithmic likelihood.

The odds between two events is a correlation measure whose entropic contribution to a positional scale ushers in Bayes' rule [15], namely the product of two factors, a rational prior and a rational likelihood, precisely the NBL probability of the numeric range involved. This structure is recurrent under arithmetic operations and gives place to the algebraic field of "referential ratios", the ground for Lorentz covariance, and the cross-ratio, a central instrument of conformality. Then, we determine the conformal metric and iterative coding functions that preserve the local Bayesian information and are compatible with a multiscale complex system [16]. Finally, we resolve the canonical PMF's parameter, the proportionality constant that ensures the divisibility of the probability mass for naturals and integers. In the epilogue, we comment on the results and conjecture some ideas that open the door to future research.

The primary motivation of this work is seeking a reason for NBL rather than describing how it works [17] or elucidating its pervasiveness. Although Newcomb was an astronomer and Benford was an electrical engineer and physicist, basic research on NBL has usually been the territory of mathematicians; physics must reconsider NBL. Finding a rational version of the law was also a goal of our investigation, given that real numbers are physically unfeasible, mere mathematical abstractions. \mathbb{Q} fits in a relational world ruled by proportions and approximations, contrasting with the continuum's absolute density and the ultra-accuracy of \mathbb{R} . Another motivation is disclosing how a coding source manipulates information in PN. NBL says nothing about the coding process that leads to a digit's probability of occurrence.

What falls outside our purview? Applications of NBL (e.g., financial) that are irrelevant to computation, information theory, or physics. Neither are we interested in particular virtues of NBL, e.g., the exactness of the law (uncanny, to tell the truth [18]), because they deviate our attention from the critical topics to tackle, to wit, what makes the minor numbers mostly probable, the link to Bayes' rule, and the efficacy and universality of the conformal coding spaces (see Figure 1). Despite the title, this essay is not about cryptographic protocols or codes enabling source compression and decompression

or error detection and correction for data storage or transmission across noisy channels; it is about a source's system of rules for converting global information into local information.

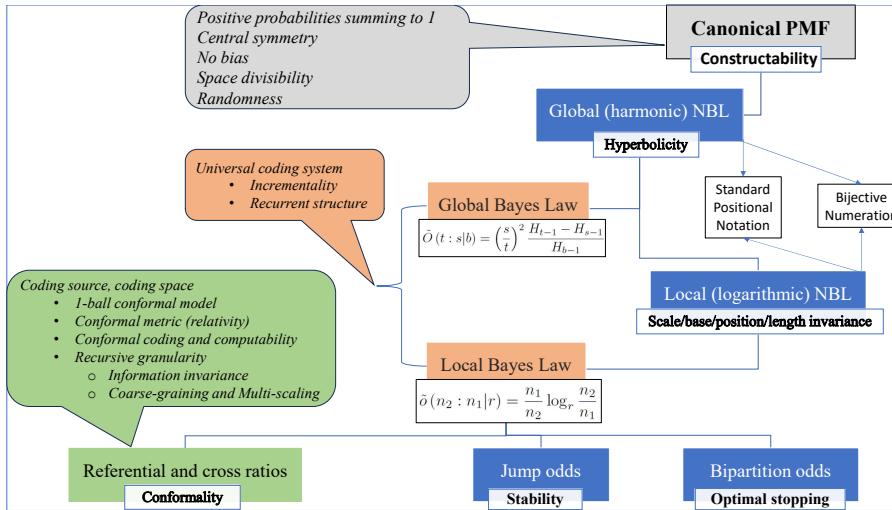


Figure 1. The road to conformal coding.

How did this research develop? Our original rationale was acknowledging that a connection between an NBL and an ISL exists. The rate of change of a significand's probability drops quadratically, i.e.,

$$[P(n)]' \propto \left[\ln \left(1 + \frac{1}{n} \right) \right]' = -\frac{1}{n(n+1)}$$

According to this expression, a numeral's occurrence differential is inversely proportional to the square of its distance (plus its distance) from the coding source. Therefore, we could expect this spatial arrangement around the origin based on an ISL for the natural numbers. If a genuine inverse-square PMF exists, we should arrive at it from just a few essentials. We confirmed that three preconditions, namely positive probabilities summing to one, no bias, and central symmetry, unambiguously define a PMF, except for a proportionality constant. Moreover, requiring probability mass compartmentalization fixes such a constant and completely specifies the canonical PMF for the natural and integer numbers. Because the resulting probability for counting numbers is a unit fraction, a rational version of NBL should accompany the logarithmic counterpart. We ultimately gleaned how to calculate the probability of a quantum in a given base as a value in \mathbb{Q} .

We have encountered that information has a relational character primally conveyed by the likelihood concept, either harmonic ($\mathcal{L}([s, t]) = H_{t-1} - H_{s-1}$ harmt, i.e., "harmonic units" of information) or logarithmic ($\ell([i, j]) = \ln j/i$ nat). Likelihood is not the information obtained by picking an item from a range but the space allocated to encode an item between the range's ends. An NBL probability is a proportion of the information total (likelihood density), and an NBL entropy is the weighted mean of the information total (average likelihood). Moreso, odds, referential ratios, and cross-ratios measure likelihood correlations. Because algebra grows on these rational data, geometry embodies algebraic structures, and physics reflects geometrical rules, information turns out to be physical.

Another high-level achievement was finding a hidden connection between NBL and Bayes' law. This rudimentary rule codes the strength of the relationship between a pair of items normalized in a particular base b or radix r . The global Bayes' rule, in odds form and b -ary harmonic information

units, is the product of a prior, the ratio between the probability of two numbers t and s according to the canonical PMF, by a likelihood factor, the global NBL probability of the bucket $[s, t)$ in base b . The local Bayes' rule, in odds form and r -ary logarithmic information units, is the product of the prior, a ratio between the global NBL probability of two quanta j and i on b 's harmonic scale, by a likelihood factor, the local NBL probability of the bin $[i, j)$ in radix r . Further, Bayesian data conformally encoded constitute normalized likelihood information. Bayes' rule also recodes information after a change of base or radix, a foundation for incremental computation. Lastly, we learned how a source recursively encodes the observable as Bayesian data and decodes these back into the information of the external world. This Bayesian outlook unifies the frequentist, subjective, likelihoodist, and information-theory interpretations.

We have verified that likelihood, probability masses, entropy, and odds are measurable information, the common factor for the universality of the harmonic and logarithmic patterns appearing in real-life raw numerical series. We have even inferred that information divergence is impossible. In the first place, the entropy of the canonical PMF for the natural and integer numbers converges. Likewise, we have defined global and local Bayesian data supported by confined harmonic and logarithmic scales. The jump odds between consecutive quanta or numerals are also delimited. Physically, the entropic cost of crossing entirely the universe or its local copy agrees with the Bekenstein bound [19].

Effectively, information occupies finite space. This essay introduces various examples of how a law, PMF, concept, or formula supports our theory that the cosmos is a hyperbolic, thrifty, and relational information system at a fundamental level. The notion of conformality implemented into a source's coding space subsumes these hallmarks. It employs the NBL invariance of scale, base, length, and position in the Bayes' rule to calculate the entropic contribution of a range of items. This synergy reinforces the thesis that mathematics begets physics and that information is a form of energy. The universe is a natural positional system that rules how a body's local quantum-mechanical degrees of freedom carve the information of its consubstantial properties, backing the Computable Universe Hypothesis [20].

2. Results

We enumerate the research's concrete results and answer what this study adds to human knowledge.

2.1. Specific Achievements

We have found a roundabout but intuitive argument to explain the appearance of NBL in the vast array of contexts in which its effect manifests; NBL issues from an ISL of probability.

When choosing a natural number at random, nature follows a particular PMF where zero is possible and interpretable as "indeterminate", e.g., not-a-number or inaction. We require zero's probability to be $1 - \epsilon S$, where ϵ is a proportionality constant, and ϵS is the probability of picking a counting number, i.e., $\{1, 2, 3, \dots\}$. We also need this one-parameter PMF to have no bias so that no number is prominent (up to its probability), i.e., any number can appear. Moreover, the mass of a counting number N is necessarily ϵ/N^2 if we want the probability function extensible to integer numbers, i.e., a number with the same probability regardless of the sign. Thus, the universe weighs the cost of choosing $\pm N$ as growing quadratically with N .

We have obtained the "global" and "local" NBL from this predetermined PMF. Under its tail, the probability that a natural number exceeds N is proportional to the trigamma function at N . Likewise, the probability of a natural variable's second-order cumulative function falling into $[s, t)$ is a harmonic likelihood ratio that cancels out the constant ϵ , namely the bucket's width $\psi(t) - \psi(s) = H_{t-1} - H_{s-1}$ relative to the base's support width $\psi(b) - \psi(1) = H_{b-1}$, where $1 \leq s < t < b$ and $\{s, t, b\} \in \mathbb{N}^+$. The base b is a global referent that changes the status of a number to a computable elemental entity we call a "quantum". When the bucket is $[q, q + 1)$, we obtain the global NBL of a generic quantum

$q, \Pr(b, q) = \frac{1}{qH_{b-1}}$, an exact and separable function where $q, b \in \mathbb{N}$, $1 \leq q < b$, and H_n is the n th harmonic number. The global NBL represents in information theory the likelihood q encloses concerning the likelihood total, geometrically a share of the surface area swept by q , and physically a scalar potential harmonically diminishing as q moves away from the origin. The odds-version of this PMF (21), also exact and separable, defines the stability of a quantum jump.

We can handle quanta as real variable values when the global base b is giant. Because a coding source does not know the value of b , it must establish a local referent $r \ll b$ to normalize its information separated from the surrounding environment, changing the status of a quantum to a locally computable elemental entity we call a "digit". This scenario involves the canonical PMF's third-order cumulative distribution; the probability of a quantum falling into $[i, j]$ is a logarithmic likelihood ratio that cancels out H_{b-1} , precisely the bin's width $\ln j/i$ relative to the radix support's width $\ln r$. When the bin is $[d, d+1)$, we arrive at the fiducial NBL, i.e., $\log_r(1 + 1/d)$, where $d \in \mathbb{N}$ is a digit such that $1 \leq d < r$. This PMF represents in information theory the likelihood d encloses regarding the likelihood r embraces. It is geometrically a hyperbolic sector equivalent to the surface area swept by d relative to that swept by r and physically a scalar potential r -logarithmically diminishing as d moves away from the origin.

In general, NBL probabilities consider the cutoffs PN imposes as a proportion of the total information. The global and local versions of NBL for standard PN give probability masses similar to a degree. For comparison purposes, 1 in standard ternary occupies $2/3 \approx 66.7\%$ (\mathbb{Q}) and $\ln 2/\ln 3 \approx 63.1\%$ (\mathbb{R}), while 2 occupies $1/3 \approx 33.3\%$ and $\ln 1.5/\ln 3 \approx 36.9\%$, respectively. Likewise, 1 in standard decimal occupies 35.3% (\mathbb{Q}) and 30.1% (\mathbb{R}), while 9 occupies 3.9% and 4.6%, respectively.

Furthermore, we provide NBL for bijective numeration to reinforce the thesis that this law is comprehensively universal. All the formulas of standard PN are translatable to bijective numeration. The NBL with standard radix $r + 1$ corresponds to the NBL with bijective radix r , which is length- and position-invariant in addition to other well-known invariances. Regardless of the numeral system, we must conceive of positional scales as hyperbolic spaces in a broad sense, harmonic in the first place, and logarithmic in the second place.

The sums of Kempner's curious harmonic series [21] echo the bijective harmonic scale traced by the global NBL. This outcome is absolute because every Kempner series is infinite, and the calculations consider every possible numerical chain; extended numerals are increasingly unimportant. For example, in decimal, while removing the terms including less than 10% of 5's in the denominator makes a harmonic series converge, missing the terms including $\geq 10\%$ of 5's does not impede the divergence of the depleted harmonic series. We also figure that the natural span of a positional system in base b is b^b , a measure of the physical quantity of numerals PN can inherently manage. Beyond this computational resolution, quanta or digits could be haphazard for practical purposes.

NBL, a synonym of PN, a subsidiary of the canonical PMF, describes an information field where probability correlates with accessibility, whence, with concentration and durability. Smaller significands occupy more room and enclose less information than greater significands. In other words, the space is denser and more stable near the coding source, while numerals dilute the space and become more reactive as we move away from the origin.

The analysis of NBL from the odds angle drives us to a rudimentary Bayesian framework. The Bayesian view of objectivistic or subjectivistic probability allegedly requires a reasoner to admit ignorance and imperfection expressed by a prior and its likelihood, respectively. The reasoner also accounts for counterhypotheses by considering the product of the prior and its likelihood in the posterior calculation. Natural Bayesianism works similarly but merely involves a coding source supporting PN.

Bayesian encoding, recoding, and decoding are elemental computing routines that handle odds. The Bayesian encoding of the relation between two numbers is the entropic allocation of their correlation for a harmonic scale, i.e., their ratio squared multiplied by the probability of the associated interval in the chosen base. The Bayesian encoding of the relation between two quanta is the entropic contribution of their correlation for a logarithmic scale, i.e., their ratio multiplied by the

probability of the associated bucket in the chosen radix. Therefore, we can interpret a Bayesian rule as the formula to encode the rational point n/d or the corresponding range $[n, d)$ of integers; this duality principle asserting that points and lines are interchangeable is endemic to the cosmos.

The global Bayes' law bridges numbers with information. We measure global Bayesian data in harmonic units of information that depend on the base. The natural harmonic scale uses bucket $[1, 2)$ as a reference. We measure local global Bayesian data in logarithmic units of information that depend on the radix. The natural logarithmic scale uses bin $[1, e)$ as a reference, where e is Euler's number. However, the arithmetic of Bayesian data generally does not refer to the global base or the local radix; it works on natural scales.

The global Bayesian rule allows for calculating a quantum jump probability, with masses decaying similarly to the global NBL as we move away from the source. Likewise, the local version of Bayesian coding drives us to the PMF of a domain's bipartition, an information function applicable to stopping problems. Specifically, we deduce the information gained from splitting a radix's digit set. If we take these digits as generic elements to be processed sequentially, our bipartite odds formula reaches a pair of information maxima involving e . The square root of the radix gives a minimum between the two maxima. We fix ideas by focussing on a variation of the "secretary problem" pursuing "a good" instead of "the optimal" solution. This problem's representativeness joins the overwhelming evidence supporting the overarching character of the NBL.

A kooky discovery is that the structure of Bayesian data whose prior factor is the unit is recurrent under arithmetic operations, giving rise to the algebraic field of "referential ratios" $\frac{A-B}{A-C}$. Moreover, a ratio of referential ratios is a cross-ratio, and the logarithm of a cross-ratio locally provides us with the metric of a conformal space reflecting the observable world and consolidating the universal proclivity towards littleness, lightness, brevity, or shortness.

The coding source calculates the conformal distance from the origin as $2\text{artanh}(Q)/\ln r$, where $\text{artanh}(Q) = \sum_{i=1}^{\infty} \frac{Q^{2i-1}}{2i-1}$, $Q = \text{sgn}(P) - 1/P$, $\text{sgn}()$ is the sign function, and P is the observed Euclidean distance to the point where an external object is. The coding space is the ball $\{Q \in \mathbb{Q} \mid |Q| < 1 - 1/b\}$, with constant curvature of $-\ln r$, where r is the radix used to normalize the information; the harmonic (outside) and logarithmic (inside) scales have a common origin and are separated by the boundary ± 1 when $b \rightarrow \infty$. The "conformal encoding function" using the logarithm is $C = \text{sgn}(P) \log_r(2 \text{sgn}(P) P - 1)$, with inverse "conformal decoding function" $\frac{1}{2} \text{sgn}(C) (1 + r^{\text{sgn}(C)} C)$. Since the metric ranges between $-\infty$ and ∞ , the source can repeat the encoding process inwards until the external object's hyperbolic distance falls within the local coding space, halting the recursion. Likewise, every 1-ball with a radius given by the iterated decoding of $C = 0$ outwards corresponds to a granularity level.

The results of this research stem all from the canonical PMF for the integer numbers, whose characteristics are fundamental and generative, imaging the essence of the cosmos. Physically, positive probabilities summing to one translates into unitarity, central reflection symmetry into parity invariance, fair mean and variance into uncertainty, holistic rationality into discreteness and relationalism, and utmost randomness in picking the number one into the principle of maximum entropy. Likewise, the global NBL (hence Zipf's law [22] with exponent 1), as well as the local NBL (supported by the logarithmic scale), are arguably physical. More generally, our descriptions and derivations introduce diverse instances of how mathematical functions, rules, or algebraic structures emerge as observable dynamics.

2.2. Hyperbolic World

Hyperbolic geometry is non-Euclidean in that it accepts the first four axioms of Euclidean geometry but not the fifth postulate. The n -dimensional hyperbolic space H^n is the unique, simply-connected, and complete Riemannian manifold of constant sectional curvature (equal to -1 [23]). For instance, saddle surfaces resemble the hyperbolic plane H^2 in a neighborhood of a (saddle) point. These are typical ways to introduce the notion of hyperbolicity.

Instead, we prefer to identify a hyperbolic space with a domain whose geometry pivots on the hyperbola, contrasting with flat and elliptic spaces, which are parabola-based and circle-based, respectively. Harmonic scales are part of this world because a logarithmic scale results from summing over a harmonic series with vanishing steps between the values of a rational variable. The computational implementation of this hyperbolic world is PN, i.e., representing numeric entities on a positional scale, either harmonic or logarithmic.

Various combinations of exponential forms define the hyperbolic functions, so logarithms characterize the corresponding inverse (or area) hyperbolic functions. In geometry, the extent of the hyperbolic angle about the origin between the rays to $(1, 1)$ and $(x, 1/x)$, where $x > 1$, is the sector $\ln x$. The natural logarithmic scale, factually H^1 , rules the cosmos to a great degree, developing systems whose properties echo a scale-invariant and base-invariant frequency.

Physics ties an ISL with "a geometric dilution corresponding to point-source radiation into three-dimensional space" [24]. Math shapes an ISL within a two-dimensional setting [25]. Nonetheless, our brute ISL of probability drives us to various versions of NBL all in one dimension, from which nature can expand the logarithmic scale upon hyperbolic spaces of all ranks to avoid the curse of dimensionality [26]. Remarkably, forming a hyperbolic triangle is more than four times as probable as a non-hyperbolic one. We daresay that hyperbolic geometry beats at the universe's core.

In information theory, we consider that a hyperbolic space is a coding space within which "likelihood" epitomize the physicality of the positional number system. A global NBL probability is a harmonic likelihood ratio, and a local NBL probability is a logarithmic likelihood ratio. A ratio of NBL probabilities determines the relative odds between two buckets of quanta globally or between two bins of digits locally. Typically, a coding source calculates the odds between two numeric events considering the information of the range they embrace regarding the entire informational support provided by the global base or the local radix. These normalized odds are likelihood ratios.

Decoded (prior) odds between two events are correlations that a coding source translates to a positional scale multiplied by a likelihood ratio. This product is Bayes' rule to encode and transform the information. The shock is that first, addition, subtraction, multiplication, and division reproduce this coding pattern, and second, under certain conditions, it collapses into the algebraic field of referential ratios $\frac{A-B}{A-C}$. A quotient of referential ratios is a cross-ratio, the linear fractional transformation's invariant over rings via the action of the modular group upon the real projective plane [27]. Restricted to one dimension, the cross-ratio's logarithm in radix r determines the coding space's metric with curvature $-\ln r$. The canonical encoding function $y = \log_r(2x - 1)$ and the canonical decoding function $x = \frac{1}{2}(1 + r^y)$ are the unique conformal transformations (i.e., preserving orientation and angles) that, if applied iteratively, map $x > 0$ to the coding space's positive side in accord with the minimal information principle. For the same reason, the hyperbolic distance $d_r(A, B) = 2/\ln r (\operatorname{artanh}(B) - \operatorname{artanh}(A))$ between points A and B inside the local coding space is also unique. We conclude that Poincar invariance ultimately stems from the algebraic field of referential ratios.

2.3. Thrifty World

To improve tractability, one can feel tempted to cut the unit uniformly into equal parts. A constant probability distribution assigns the same expected frequency to all the domain values. However, whereas the uniform distribution of probabilities is, in principle, fair and provides maximum entropy, it does not fit well into an open (infinite) outcome space.

Contrariwise, it is noteworthy that [28] "the frequency with which objects occur in 'nature' is an inverse function of their size", indicating that oddity and magnitude usually correlate and conform to a Benford distribution. NBL says the cosmos displays a progressive aversion to larger and larger numbers, somewhat implementing the "parsimonae lex" [29], a principle of frugality [30] that stimulates economy and effectiveness as universal prime movers, drivers of nascent physics, particularly the spacetime geometry.

The canonical PMF exhibits nature's bet on the shortest numbers, but NBL provides further precision, pointing to a conservative policy of significands. For instance, the law favors 123.4 against 12.345 because 12345 is less probable than 1234. For the same reason, the law favors 0.234 against those. The last digits might provide negligible, even arbitrary, information [31]. This innate tendency amounts to restricting the resolution of the representational system to preclude unnecessary precision. Carrying long tails of digits from operation to operation is neither intelligent nor evolutionary. Information is gold, much like energy.

Interestingly, the probability that a randomly chosen natural number between 1 and N is prime is inversely proportional to N 's number of digits, whence to the length of its significand [32], i.e., to its logarithm or equivalently its likelihood. Therefore, primality and information are nearly interrelated. Why is finding a big prime so tricky [33]? Because it demands logarithmically growing energy.

NBL denotes productivity. Radix economy $\mathring{E}(N, r) \approx r \log_r N$ measures the "price" of a numeral N using radix r as a parameter. Cost-saving number systems will employ an efficient coding radix; the optimal radix economy corresponds to Euler's number e , another sign of the preeminence of small numbers. The wider the gap between the economy of consecutive numbers relative to the radix, the higher the expected frequency. Thrifty numbers making a difference are winning, meaning that the probability of a number coded with radix r showing up is the rate of change, or derivative, of its economy concerning the radix, specifically

$$P(N, r) = \frac{\mathring{E}(N+1, r) - \mathring{E}(N, r)}{r} \approx \log_r(N+1) - \log_r N = \log_r \left(1 + \frac{1}{N}\right)$$

This expression indicates the occurrence probability of the numeral N , not necessarily a digit, with radix r . For example, $\log_{10}(1 + 1/22)$ is the probability of running into a decimal number starting with 22, such as 2.29 or 2237. The logarithmic scale knits the linear space toward the coding source; the closer, the higher the spatial density. A large numeral is less likely due to its representational magnitude, so its space is less contracted than that occupied by a numeral with more probability mass. NBL reflects how PN encodes numbers in agreement with this economic criterium.

Therefore, the radix economy establishes a scalar field where the gap between the "potential energies" [34] of two objects only depends on their position as perceived from the source. Thus, the canonical PMF and NBL subsidiaries are fundamentally efficient, balancing probability mass against notation size. Minor numbers are accessible at a lower cost, while spatial dilution and the prospect of likelihood increase, although deceleratingly, as we climb to infinity.

NBL maps (the minor numbers of) the "linear frequency" onto (the least costly digits) of the "logarithmic frequency" through the "harmonic frequency". How does a harmonic scale exhibit its austere nature? The study of constrained harmonic series mainly teaches us that the specific digits involved in the restraining chain do not matter, whereas its length does. Long chains or high densities of quanta are "rare" and deliver slender harmonic terms that hardly occupy space. In contrast, short chains or low quantum densities are regular and cheap, producing heavy harmonic terms that occupy much space, leading to convergence of the series if eliminated. In other words, only usual and economic constraints can impede the divergence of a harmonic series. More generally, increasingly bigger numbers on a linear scale require hyperbolically less and less attention in accord with the room they take up. Nature builds physics upon proximity because "almost all" large numbers are expensive and indiscernible [35].

Our theory also associates efficiency with entropy. We can interpret NBL probabilities as degrees of stability or coherence. The lowest digits maintain distinctness from the surroundings thanks to their solid entropic support. The more significant digits are vulnerable and give rise to more transitions, physically translating into higher reactivity or less resistance to integration with the environment.

Parsimonious management of computational resources is crucial, as optimal stopping problems reveal. In "the secretary problem", selecting "the best" applicant is pragmatically less sensible than simply "a good" one, which requires maximizing the bipartite entropy. The "past partition" emphasizes

the information gathered, while the "future partition" deals with the information we can obtain from forthcoming aspirants. As the number of examined applicants grows, the past information increases, but the future information decreases. In contrast, the probability of taking advantage of both types of information decreases and increases, respectively. The best applicant implies exclusively focusing on the future partition, but a balanced decision also implies contemplating the past.

Information economy enables cosmological evolution. That the universe optimizes computability follows from NBL embracing several invariances. Base invariance ensures even interaction with the environment because changes in the radix value will imply only incremental updates (reencoding), keeping the internal metric up to the curvature. Scale invariance provides the means to recursively perform geometric calculations on nested levels of domain granularity, like a fractal. Rescaling implies only obtaining the powers of any radix using straightforward Moessner's construction [36]. Length and position invariance ensure fault tolerance. Ultimately, PN is effective because it makes the most expected data readily accessible for iterative coding functions.

Because a thrifty world refuses the continuum, computing hyperbolic spaces requires "rationality" to be feasible.

2.4. Relational World

Real numbers are unattainable mathematical objects [37], artificial, mere abstractions; hence, \mathbb{R} -oriented physical laws and principles are suspicious. In contrast, relative odds, i.e., proportions between two numbers, quanta, or digits, are tractable. Rational numbers are the fitting choice in an inaccurate and defective [38] world, where relations are as important as individual entities [39] and comparative quantities predominate over absolute values. A universe built upon the rational setting facilitates divisibility, discreteness, and operability. Calculus of rational information relies on a harmonic scale and uses harmonic numbers. Regardless, we need rational models of reality to prove that the \mathbb{Q} underpins the universe's computational machinery.

Presuming the minimal information principle, we require a fundamental PMF with positive probabilities summing to one, no bias, and central symmetry. To ensure divisibility of the probability space, which enables the operability of the information, the mass distribution we obtain for the natural numbers must be $\frac{1}{2}N^{-2}$ if N is nonzero and $1 - \frac{1}{2}\zeta(2)$ otherwise. Next, we calculate from this PMF the probability of a natural number being odd or even and prime or composite. We also calculate the probability of getting an elliptic, parabolic, or hyperbolic two-dimensional tiling by examining the "triangle group". Similarly, the occurrence probability for a nonzero integer Z is $1/(2Z)^2$. Despite being excluded from the scope of this essay, we can even extend the canonical PMF to rational and algebraic numbers, the computable version of complex numbers. All these laws are rational and inverse square, fulfilling identical requirements.

We underline that the probability mass of a nonzero integer is a unit fraction. Real numbers only appear (in terms of the Riemann zeta function at 2) when the probability of occurrence involves zero or infinity, a sign that these limiting values are virtual. From the canonical PMF, we derive a discrete (global) counterpart of the continuous (local) NBL, where the probability of a significand in a given base is rational. The continuous (local) NBL emanates precisely from the rational (global) NBL by compartmentalizing a one-dimensional hyperbolic space of colossal extent. More generally, whereas the universe originates globally from \mathbb{Q} , it is perceived locally as \mathbb{R} .

The concept of information is fundamentally rational. Harmonic likelihood is global information defined as $[\mathcal{L}(q)]_s^t \equiv [\psi(q)]_s^t / [\psi(q)]_1^2$ harmt, whereas logarithmic likelihood is local information defined as $[\ell(d)]_i^j \equiv [\ln d]_i^j / [\ln d]_1^e$ nat. A "harmt" is the global (harmonic) unit of information, peering the local (logarithmic) unit of information, the "nat". Likewise, NBL PMFs represent normalized information regarding the global base b , $\Pr(b, [s, t)) \equiv [\mathcal{L}(q)]_s^t / [\mathcal{L}(q)]_1^b$, or the local radix locally, $\Pr(r, [i, j)) \equiv [\ell(d)]_i^j / [\ell(d)]_1^e$. Likelihood is space on a harmonic (global) or logarithmic (local) scale; for example, if we assume that the "bit" (2^1 possible states) is the minimal (unit) length [40], one "byte" (2^8 possible states) has length eight. If our world is positional, likelihood and entropy would have

metric units of length for all practical purposes, meaning that information is a physical and manageable resource.

Rationality in its purest form appears as the NBL probability of a quantum $\frac{1}{H_{b-1}} \frac{1}{q}$ or a jump $\frac{1}{H_{b-1} - H_{b-1,2}} \frac{q}{(q+1)^2}$, with masses separable as the product of a function of b and a function of q . However, the relational character of the universal rational setting pops up in all its splendor when we address probability ratios. The odds value between between a pair of numbers, quanta, or numerals is the quotient of their picking probabilities, quantifying the strength of their association. Assuming $a < b$, $\Pr(a)/\Pr(b) > 1$ estimates how uncorrelated a and b are; if $\Pr(a)/\Pr(b) \approx 1$, both events are mutually dependent.

Then, a source encodes, recodes, and decodes odds using Bayes' law, reminding us that ratios are the atoms of a coding process. The global Bayes' rule says that the odds of quantum s against t in base b are the odds of the number s versus t times the probability of the bucket $[s, t)$ in base b . The local Bayes' rule says that the odds of digit i against j with radix r are the global odds of the quantum i versus j times the probability of the bin $[i, j)$ with radix r . Both represent the entropic contribution of the items in a range to a positional scale, confirming that information is relational.

Exceptional cases of Bayesian data are the cross-ratio, a conformality invariant, and the "referential ratios" $\frac{A-B}{A-C}$, the basis for relativity. Despite the conformal coding functions using the logarithm and the exponential function, power (infinite) series by definition, the coding source adds or multiplies incrementally a finite series of referential ratio powers to throw a rational result at any time, bettering the approximation with the number of iterates. Rationality is intricately intertwined with decidability in polynomial time and interruptible algorithms in evolving scenarios [41].

Numeric values do not contain information per se, while a common property makes two entities commensurable, with the global base and the local radix as main referents. We can take global Bayesian data as rational quanta, computable numbers, and local Bayesian data as observable correlations of numerals. In the end, mathematics is \mathbb{Q} -based, and physics is relational.

3. The Whole Story of NBL

We comment on the aspects of the academic story of the fiducial NBL most relevant to our essay and then traverse the deductive road to it. We have discovered many findings on the run related to the nature of the information at a fundamental level.

First, we introduce an inverse-square law as the origin of NBL. This PMF subsumes a probability law of rational masses, giving place to a normalized universal PN system to manage a hyperbolic, thrifty, and relational world. This harmonic scale system employs a global base as a fundamental referent. When the global base is immense, the scale's rational setting approaches a domain of real variables and functions ruled by small radices in local settings. In other words, we prove that the local NBL, as everybody knows it, assumes that a prior all-encompassing base exists. Eventually, the interplay between the global base and the local radix will enable us to determine the canonical metric ascribed to a coding source's conformal space containing an image of the world.

3.1. The Tortuous Road to NBL

The first digits of the numerals found in data series of the most varied sources of natural phenomena [42] do not display a uniform distribution but rather exhibit that the minor ones are the more likely (see [43] for a detailed bibliography and [44,45] for a general overview). Specifically, this "law of anomalous numbers" claims that the universe obeys an exponential distribution to a greater or lesser extent.

Newcomb's insight was, "The law of probability of the occurrence of numbers is such that all mantissae of their logarithms are equally probable." (What Newcomb refers to as "mantissa" is what we will call "significand".) More than half a century later, Benford defined the exact formula of every random variable satisfying the first-digit (and other digits) law [2]. He could not derive it formally, although seeded a line of research asserting that "The basic operation $F = \int \frac{dx}{x}$ or $F = \sum \frac{x}{X}$ in

converting from the linear frequency of the natural numbers to the logarithmic frequency of natural phenomena and human events can be interpreted as meaning that, on the average, these things proceed on a logarithmic or geometric scale."

However, this transition from \mathbb{N} to $\log \mathbb{N}$, when the baseline set is unlimited, implies tackling the problem of "picking an integer at random" [46], and then mathematical difficulties arise. To commence, numerals beginning with a specific digit do not have a natural density. The decimal sequence $\{1, 11, 12, 13, \dots, 100, 101, 102, 103 \dots\}$ that groups the first digits does not converge (e.g., oscillates). Moreover, suppose each natural occurs with equal probability. In that case, the whole space must have probability 0 or ∞ , violating countable additivity (by which the measure of a set must be nonzero, finite, and equal to the sum of the measures of the disjoint subsets); hence, we cannot construct a viable discrete probability distribution. The attempt to choose $P(N) = 1/N$ fails because it diverges in the limit; it is not countably additive. Furthermore, a universal law such as NBL is supposed to be scale-invariant. However, there are no scale-invariant probability distributions on the Borel (measurable) subsets of the positive reals because the probability of the sets $[0, 1]$ and $[0, s]$ would be equal for every scale $0 \leq s \leq 1$, disobeying once more countable additivity [47].

Hill [48] resumed Newcomb's idea; logarithm's significands of sequences conformant to NBL trace a uniform distribution. He identified an appropriate domain for the "natural probability space" and, based on the decimal mantissa σ -algebra (where countable unions and intersections of subsets can be assigned a gauge), formally deduced the law for the first digit and joint distribution of the leading digits. He also provided a new statistical log-limit central-limit-like significant-digit law theorem that stated the scale-invariance, base-invariance, sum-invariance, and uniqueness of NBL. The cumulative distribution function is $\Pr(r, d \leq m) = \log_r(1 + m)$, where $d, m \in [1..r]$ and r is the radix.

Since Hill's publication in 1995, more derivations have come to light, one of the subtlest appearing in [49] (section 14.2). Nonetheless, they all ignore foundational causes.

3.2. Properties of the Distribution

A vehicle of NBL is how different measurement records spread and repositories aggregate data. For one thing, the significant-digit frequencies of random samples from random distributions converge to conform to NBL, even though some of the individual distributions selected may not [50]. Besides, many real-world examples of NBL arise from multiplicative fluctuations [51]. What happens is that the absorptive property, exclusive of the fiducial NBL, kicks in [52]; "if X obeys Benford's law, and Y is any positive statistic independent of X , then the product XY also obeys Benford's law – even if Y did not obey this law". To boot, "variable multipliers (or variable growth rates) not only preserve Benford's law but stabilize it by averaging out the errors".

Which standard probability distributions obey NBL? Rarely does a distribution of distributions disagree with NBL [53]. The ratio distribution of two uniform, two exponential, and two half-normal distributions approximately stick to NBL. The Pareto distribution enjoys the scale-invariance property as long as we move from discrete to continuous variables, and Zipf's law ($\propto 1/z^\alpha$ with $\alpha \approx 1$) satisfies the abovementioned absorptive property if one stays over the median number of digits [52]. More generally, right-tailed distributions putting most mass on small values of the random variable (i.e., survival or monotonically decreasing like the log-logistic distribution) are just about compliant with NBL [28] (e.g., the tail of the Yule–Simon distribution [54]). The Log-normal distribution fits NBL, and the Weibull and Inverse Gamma distributions are close to NBL under certain conditions [55]. In short, NBL embraces an ample range of statistical models and mixtures of probability distributions.

Empirical testing of random numerals generated according to the exponential and the generalized normal distributions reveals adherence to NBL [56]. More precisely, almost every exponentially increasing positive sequence is Benford (e.g., sequences of power a^n , where $a > 1$), and every super-exponentially increasing or decreasing positive sequence (e.g., the factorial) is Benford for almost every starting point [57]. Further, an NBL-compliant data series is inherently sturdy because of its invariance to changes concerning sign, base, and scale [58]; for instance, mining data about the

lifetime of mesons or antimesons in microseconds in decimal or seconds in binary results in strict observance of the law.

All these mathematical circumstances we have summarized about NBL explain why it is so widespread but not its reason. Failure to comprehend this distinction has generated confusion and is a typical scientific misunderstanding [59]. In other cases, authors have deemed specific remarks about NBL its cause when they are indeed consequences [60].

We will explain why discrete distributions decaying as $\propto 1/z^\alpha$ with $\alpha \in \mathbb{Q}^{(1,2]}$ are indirectly NBL-compliant. The common factor of all the quasi-NBL distributions is that proportional data intervals approximately fit their heavy tail (the fatter, the better). Notably, this work does not deal with the NBL invariances as presumed properties but derives them from basic requirements demanded from the canonical PMF, producing a subsidiary global NBL and, thereon, the fiducial NBL. The appearance of NBL in power sequences indeed concerns how PN codes probability ratios (odds), where the logarithm and the exponential constitute a fundamental functional duality. The intricate and critical linkage of the law with the rational numbers jumps out.

3.3. A Fundamental Probability Law

We seek a well-defined PMF, i.e., positive probabilities summing to 1. Not all Zipfian distributions [61] can do the job, for $\Pr(N) \propto N^{-\alpha}$ eludes divergence only if $\alpha > 1$. In particular, linear forms for the denominator of a natural's probability cannot fulfill countable additivity.

We assume that \mathbb{N} is an inductively constructible set from which all physical phenomena can crop up from the source outward, a basis of reductionism and weak emergency [62]. By including "nil", we also ponder "infinity" as its reciprocal. However, both projective concepts are only potential and limiting numbers in the offing; employing the successor and predecessor as symmetric constructors, we must be able to choose any number strictly between 0 and ∞ so that no counting number is extraordinary. Again, many Zipfian distributions cannot do the job, for $\Pr(N) \propto N^{-\alpha}$ has a diverging mean only if $\alpha \leq 2$. For instance, cubic or higher polynomials lead to convergent expected values.

Additionally, we require a sound and dependable extension to the integers. Zipfian distributions where α is an even natural do the job, but in the range $\alpha \in (1, 2]$ defined by the two previous requirements, $\alpha = 2$ is the fitting choice, the only value assuring central reflection symmetry. To cap it all, $\Pr(N) \propto N^{-2}$ agrees with the minimal information principle [63]; considering other quadratic polynomials for the denominator of a natural's probability does not yield a better law because it would introduce unwarranted assumptions in vain. For instance, the Zipf–Mandelbrot law [64] $\Pr(N) = 1/(N^2 - 5N + 7)$ deals with unexplained coefficients and is not centrally symmetric.

Therefore, the PMF of a random variable X taking natural numbers is

$$\Pr(X = N \in \mathbb{N}) = \begin{cases} N \in \mathbb{N} - \{0\} : & \frac{\epsilon}{N^2} \\ \text{else :} & 1 - \epsilon \zeta(2) \end{cases} \quad (1)$$

We will suppose the proportionality parameter $\epsilon \in \mathbb{Q}^+$ to comply again with the minimal information principle. $\zeta(2) = \pi^2/6$ is the value of the Riemann zeta function at 2, brewing gently as a factor of endless aggregation of occurrence probabilities. Because the "else" (null) case is possible, this PMF is not a pure zeta distribution [65].

Countable additivity holds; the probabilities sum to 1 owing to

$$\sum_{N=1}^{\infty} \frac{\epsilon}{N^2} = \epsilon \zeta(2)$$

The picking event X is fair owing to the indeterminacy of the expected value of a natural number, i.e.,

$$\widehat{E}(X) = 0(1 - \epsilon\zeta(s)) + \sum_{N=1}^{\infty} N \frac{\epsilon}{N^2} = \epsilon \sum_{N=1}^{\infty} \frac{1}{N} = \infty$$

Indeed, the n th-order moment diverges for all nonzero $n \in \mathbb{N}$.

This PMF does not assume the law of large numbers or the law of rare events. On the contrary, it works under the statistical assumption of independence of occurrences and no bias. Outcomes of the picking event are unpredictable, even considering an indefinite trail of repetitions. No predetermined constant mean exists in space or time, nor is there an absolute measure of "rarity"; the relative frequency between two events solely depends on their probability mass. We can regard it as a brute law.

Let us leave the rational ϵ unfixed for the time being, given that it is unimportant for the derivation of NBL. Remember that $\epsilon \in (0, 6/\pi^2)$ holds the constraint $\Pr(N \in \mathbb{N}) > 0$ (i.e., $\epsilon/N^2 > 0$ and $\epsilon\zeta(2) < 1$), and we will return to it in subsection 7.1.

3.4. The Rational (Global) Version of NBL

In analytic number theory, the mesmerizing Euler-Mascheroni constant γ ([66], section 1.5) is the limiting difference between the harmonic series and the logarithm, i.e.,

$$\lim_{N \rightarrow \infty} H_N - \ln N \sim \gamma$$

where $H_N \equiv \sum_{k=1}^N \frac{1}{k}$ is the N th harmonic number. If our universe is as harmonic as logarithmic [49], the discrete version of the NBL must exist connected to but separated from the continuous (fiducial) one.

The cumulative distribution function of a random variable X obeying (1) is

$$\Pr(1 \leq X < N) = \epsilon \sum_{k=1}^{N-1} \frac{1}{k^2}$$

which tells us how often the random variable X is below N . We call its complementary function "natural exceedance probability", quantifying how often X is on level N or above. This dwindling distribution function is

$$\Pr(X \geq N) = \Pr(X \geq 1) - \Pr(1 \leq X < N) = \epsilon(\zeta(2) - H_{N-1,2})$$

where $H_{N,2} \equiv \sum_{k=1}^N 1/k^2$ is the generalized N th harmonic number in power 2.

We can express this probability in terms of the second derivative of the gamma function $\Gamma(x)$'s logarithm, i.e., the digamma function's first normal derivative, defined as

$$\psi'(x) = \left[\frac{d}{dx} \ln(\Gamma(x)) \right]' = \frac{d}{dx} \frac{\Gamma'(x)}{\Gamma(x)}$$

Since $\zeta(2) = \psi'(1)$ and $\psi'(N+1) = \psi'(N) - 1/N^2 = \psi'(1) - \sum_{k=1}^N 1/k^2$, the natural exceedance of N is

$$\Pr(X \geq N) = \epsilon\psi'(N)$$

Numbers lack physicality. If numbers were frequencies, the trigamma function would represent a probability fractal signal such that the occurrence probability density (i.e., per frequency range) decays proportionally with the signal's frequency.

Regardless of the scale, let us divide the natural line into concatenated strings of numbers of the same length, which we name "quanta". Then, the second-order cumulative function arrives on the scene for global computability. The plot of $\epsilon\psi(q) + \text{constant}$, the natural exceedance's antiderivative, has an informational flavor. A significant value of the quantum q is more unpredictable and influential

than a minor one; this harmonic surprise needs a medium to reify the event occurrence, and the extent of the resulting log note $\psi(q)$ is its only measure.

So, how likely is the event $X = q$ to fall into bucket $[s, t)$, assuming a harmonic scale underneath? The "natural harmonic likelihood" \mathcal{L} depends on the bucket's extent, namely

$$[\mathcal{L}_X(q)]_s^t \equiv \frac{[\epsilon\psi(q) + \text{constant}]_s^t}{[\epsilon\psi(q) + \text{constant}]_1^2} = \frac{[\psi(q)]_s^t}{[\psi(q)]_1^2} = H_{t-1} - H_{s-1} = \mathcal{L}([s, t)) \in \mathbb{Q} \quad (2)$$

in natural harmonic units of (global) information, where we have considered the generalized recurrence relation $\psi(t+1) - \psi(s+1) = H_t - H_s$. Note that (2) is a proportion, canceling the constant ϵ .

The natural harmonic likelihood is neither the probability of a quantum falling into $[s, t)$ nor the probability that $[s, t)$ is the truth given the observation $X = q$. It is the information obtained by picking a quantum from the bucket $[s, t)$ or the information that $X = q$ gives when $s \leq q < t$, i.e., the space allocated to encode a quantum between the bucket's ends, which is why it does not refer to q .

The harmonic number function (interpolated to cope with rational arguments) parallels the continuous world's logarithmic function in information theory, like in analytic number theory. $\mathcal{L}([1, 2)) \equiv H_1 - H_0 = 1$ represents the "harmt" (a portmanteau of "harmonic unit"), just as the natural local information unit, the "nat", corresponds to $[1, e)$ by $[\ln x]_1^e = 1$. Thus, natural harmonic and logarithmic likelihoods are analogous, as we will explain in Section 3.6. In particular, $\mathcal{L}[q, q+1) = \psi(q+1) - \psi(q) = H_q - H_{q-1} = 1/q$ implies that q 's reciprocal denotes information, precisely the natural likelihood of an elemental quantum gap.

A global base b marks the boundary between the mathematical and physical world. We define the probability mass of bucket $[s, t)$ regarding b 's support as the harmonic likelihood ratio

$$\Pr(b, [s, t)) = \frac{\mathcal{L}([s, t))}{\mathcal{L}([1, b))} = \frac{H_{t-1} - H_{s-1}}{H_{b-1}} \in \mathbb{Q} \quad (3)$$

where $1 \leq s < t < b$ and $s, t, b \in \mathbb{N}$. This probability is separable as a product of $[s, t)$'s and b 's functions, expressing a part of the information total that is the b -normalized rational quantum t/s 's length or bucket $[s, t)$'s width.

The reader can object that the concept of likelihood is unnecessary to define (3) since we can directly define the probability of a bucket as $[\psi(q)]_s^t / [\psi(q)]_1^b$. However, we aim to stress that we get information regardless of the base, only relative to the natural harmonic bucket $[\psi(q)]_1^2$. Because $[1, b)$ gives the maximum likelihood estimate, $\Pr(b, [s, t))$ is the relative likelihood function [67] of the bucket $[s, t)$ given $1 \leq s < t < b$.

When $s = q$ and $t = q+1$, we obtain

$$\Pr(b, q) = \frac{\mathcal{L}([q, q+1))}{\mathcal{L}([1, b))} = \frac{H_q - H_{q-1}}{H_{b-1}} = \left(\frac{1}{H_{b-1}} \right) \left(\frac{1}{q} \right) \in \mathbb{Q} \quad (1 \leq q < b) \quad (4)$$

We measure this PMF in b -ary harmonic information units. It is the simplest case of Zipf's law, geometrically an embryonic form of progressive one-dimensional circle inversion. Further, if q represented a frequency, we could understand the probability of a quantum with a given base as a (physical) potential diminishing hyperbolically with the distance from the source, i.e., a flicker [68] or pink [69] noise.

We have described how the harmonic series bridges equations (1) and (4). Both laws point to minor numbers as the most frequent significands, amassing more probability around the source to increase accessibility. However, we find three main differences between them:

1. (1)'s probability masses are rational numbers. Instead, a quantum's probability represents an area ratio measured through the digamma function; hence, a quantum's probability is a quota of information.

2. The global NBL outlines a hyperbola instead of an ISL. Thus, while the probability of a number is inversely proportional to its norm (the number's square), the probability assigned to a quantum is inversely proportional to its modulus (the quantum's absolute value).
3. (4) gives us the thing-as-it-appears (perceived potential) stemming from the thing-in-itself (field per se) [70] expressed by (1), two sides of the same property or object, the dual essence of the world.

3.5. Analysis of the Global NBL

The global (\mathbb{Q} -based) NBL's average probability for the decimal system is $1/9 \sum_{q=1}^{q=9} (qH_{10-1})^{-1} = 1/9 = .\bar{1} \approx 11\%$, which is equal to the local (\mathbb{R} -based) NBL's average probability (due to $\sum_{d=1}^{d=9} \log(1 + 1/d) = 1$). The mean value for the quanta 1 to 9 following the global NBL is $q \approx 3.18$ (from $\Pr(10, q) = 1/9$), whereas it is $d \approx 3.43$ (from $\log(1 + 1/d) = 1/9$) for the local NBL. The harmonic mean value for the quanta 1 to 9 following the global NBL is $q = (9+1)/2 = 5$.

As expected, Equation (4) brings $\Pr(2, 1) = 1$, i.e., 1 occupies 100% of the space in binary. In base 3, the appearance probabilities of 1 and 2 as the first quantum are $\Pr(3, 1) = 2/3$ and $\Pr(3, 2) = 1/3$, respectively, a 2/1 sharing out. We deem this Pareto rule so rudimentary that it might be fundamental in physics. The corresponding Pareto rule is $63/37$ if we utilize the local NBL. Quantum 1 in decimal occupies $\Pr(10, 1) = 2520/7129 \approx 35.3\%$, while it is 30.1% using the local NBL. Figure 2 compares the probability of a decimal datum's first position value between the global, discrete, rational, countable, harmonic NBL and the local, continuous, real, uncountable, logarithmic one. Regardless of the cardinality, the former is always steeper.

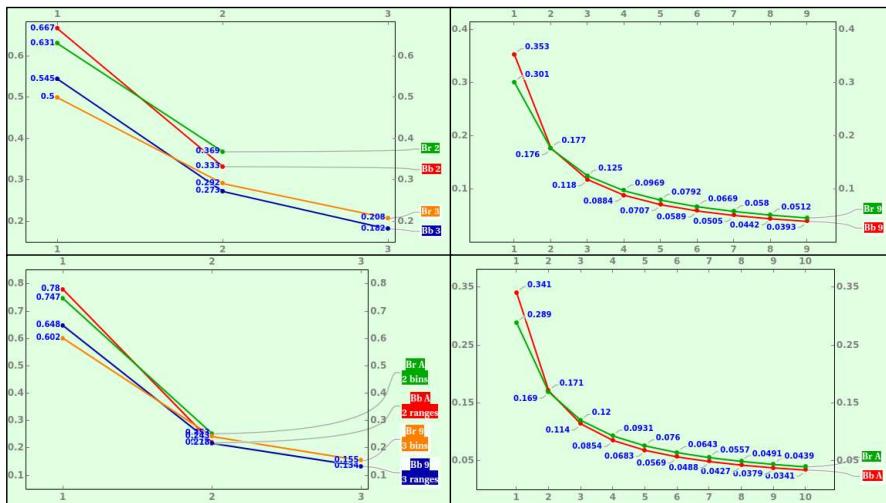


Figure 2. A comparison of the global with the local (fiducial) NBL, where "Bb" stands for "Bijective (global) base" and "Br" for "Bijective (local) radix". Vertical axes represent the occurrence probability of the horizontal axes' quanta or digits. The plot on the top left shows the PMFs of the global and local standard ternary (bijective binary) numeral system along with the PMFs of the global and local standard quaternary (bijective ternary) numeral system. The plot on the top right shows the PMFs of the global and local standard decimal (bijective nonary) numeral system. The plot on the bottom right shows the PMFs of the global and local standard undecimal (bijective decimal) numeral system. The plot on the bottom left shows the PMFs of the global and local standard undecimal numeral system divided into [1..6] and [6..11] and the PMFs of standard decimal divided into [1..4], [4..7], and [7..10].

The unit bucket a quantum represents can be of any size, so we can recursively perform the integration and normalization process that gave rise to (4) "within" every quantum attributed to base b ,

obtaining a chain of nested quanta. The probability of getting the leading chain c of quanta with any length in b -ary is simply

$$\Pr(b, c) = \frac{H_c}{H_{b-1}} - \frac{H_{c-1}}{H_{b-1}} = \frac{1}{cH_{b-1}} \in \mathbb{Q}$$

It represents c 's likelihood in b -ary harmonic units and becomes (4) when c is a base's quantum. For example, the probability masses that a decimal chain starts with "10" (e.g., 0.1071) and "99" (e.g., 992) are $\Pr(10, 10) = (10H_9)^{-1} \approx 0.03535$ and $\Pr(10, 99) = (99H_9)^{-1} \approx 0.00357$.

3.6. The Fiducial (Local) NBL

The global NBL furnishes the frame for constructing a sheer logarithmic system that conserves base and scale. To achieve such a pursuit, we must turn to the local context of a coding source and analyze how it represents a numeral in PN.

We call a "bin" of "digits" to a bucket of quanta in the source's proximity. The third-order cumulative function of (1) arrives on the scene to facilitate local computability. When the base b is enormous, we can handle digits like real values to calculate the antiderivative of (4), $\ln d / H_{b-1} + \text{constant}$, which outlines how unexpected and momentous digit d is. Large values locally transmit more information than small ones; for whom? Logarithmic surprise needs an observer to reify the event occurrence. The harmonic information perceived by a receiving system, a coding source, becomes local information with extension $\ln d$. Consequently, broad bins are more likely than narrow ones as supporting evidence.

Assuming a logarithmic scale underneath, we define the "natural logarithmic likelihood" ℓ_Y of the event $Y = d < b$ to fall into bin $[i, j)$ as the ratio

$$[\ell_Y(d)]_i^j \equiv \frac{\left[\frac{\ln d}{H_{b-1}} + \text{constant} \right]_i^j}{\left[\frac{\ln d}{H_{b-1}} + \text{constant} \right]_1^e} = \frac{[\ln d]_i^j}{[\ln d]_1^e} = \ln \frac{j}{i} = \ell([i, j)) \in \mathbb{R} \quad (5)$$

Note that this proportion no longer refers to base b ; a coding source is unaware of the global setting for calculation purposes.

The natural logarithmic likelihood is neither the probability of a digit falling into $[i, j)$ nor the probability that $[i, j)$ is the truth given the observation $Y = d$. It is the information obtained by picking a digit from the bin $[i, j)$ or the information that $Y = d$ gives when $i \leq d < j$, i.e., the space allocated to encode a digit between the bin's ends, which is why it does not refer to d . However, it has nothing to do with surprisal [71]; ℓ denotes informative space rather than information content. Indeed, we can take it as the natural positional length of j/i or the natural width of $[i, j)$. We can also take (5) as the differential entropy of the uniform probability density function $\Pr(x) = i/j \forall x < j/i$.

We measure the natural logarithmic likelihood in natural units ("nats") because of $[\ln d]_1^e = 1$. It is manifestly scale-invariant; since the area of a hyperbolic sector (in standard position) from $(1, 1)$ to $(x, 1/x)$ is $\ln x$, another way to define invariance of scale is that a squeeze (geometrical) mapping boosts the logarithmic likelihood up or down arithmetically (see [49] chapter I).

The domain of a digit d spans from the unit to $r - 1$, where $r \ll b$ is the cardinality of the local coding space, precisely the source's "radix". We define the r -ary probability mass of bin $[d_1, d_2)$ relative to the radix's support as the logarithmic likelihood ratio

$$\Pr(r, [d_1, d_2)) = \frac{\ell([d_1, d_2))}{\ell([1, r))} = \frac{\ln \frac{d_2}{d_1}}{\ln r} = \log_r \frac{d_2}{d_1} \in \mathbb{R} \quad (1 \leq d_1 < d_2 < r) \quad (6)$$

with $d_1, d_2, r \in \mathbb{N}$. We can take it as the representation length of d_2/d_1 or the width of $[d_1, d_2)$ in r -ary logarithmic information units, in correspondence with equation (3), reckoning the probability

of a bucket as a normalized harmonic likelihood. Therefore, in PN, the probability is a quota of the available space, a view we will develop in subsections 5.1 and 5.2.

Geometrically, the probability of event $d_1 \leq d < d_2$ conditioned to r is the ratio between the areas under the hyperbola delimited by bins $[d_1, d_2)$ and $[1, r)$, equivalent to the area enclosed by the rays $1/d_1$ and $1/d_2$ relative to the span of the hyperbolic angle r . Because the hyperbola preserves scale changes, the logarithm uniformly distributes the significant digits of a geometrical sequence, as Newcomb underlined in his note; $k \ln x = \ln x^k$ implies that, for example, x must drop to $\sqrt[3]{x}$ to divide the natural likelihood by three ($k = 1/3$).

By setting in (6) $d_1 = d$ and $d_2 = d + 1$, we fit the Y 's occurrences into the digits of a standard PN system with radix r , obtaining

$$\Pr(r, d) = \log_r \left(1 + \frac{1}{d} \right) \in \mathbb{R} \quad (7)$$

The original natural random variable $Y \in \mathbb{N}$ and the underlying global base b are absent. This expression is the local (fiducial) NBL, which tells us the PMF of a r -ary numeral's first digit.

A coding system (observer or source) that uses standard PN handles the unit range as a concatenation of the sub-bins $[\log_r 1, \log_r 2) = [0, \log_r 2), [\log_r 2, \log_r 3), \dots [\log_r (r-1), \log_r r) = [\log_r (r-1), 1)$, covering intervals of $\log_r 2/1, \log_r 3/2, \dots \log_r r/(r-1)$ units of space, and corresponding to the symbols 1, 2, ... and $r-1$, respectively; the addition of these areas is the unit.

More fundamentally, common digits are near the coding source, i.e., the probability of a digit correlates with its accessibility and declines logarithmically. If we liken probability mass to space, smaller digits induce more density than significant digits. In other words, accessibility concentrated around the origin progressively dilutes as we move away, contrasting with the linear scale that distributes the space evenly.

We can generalize (6) to cope with bins outside the radix. The resulting expression is not generally a probability anymore, given that we can have bins of any size, but it is again an r -normalized likelihood that retains the geometric interpretation. In other words,

$$\ell([n_1, n_2) | r) = \frac{\ell([n_1, n_2))}{\ell([1, r))} = \log_r \frac{n_2}{n_1} \in \mathbb{R} \quad (1 \leq n_1 < n_2) \quad (8)$$

is the r -normalized n_2/n_1 's length or $[n_1, n_2)$'s width. We can regard it as a fractal dimension where r is the scaling factor, n_2 is the number of measurement units, and n_1 is the number of fractal copies. For instance, (8) might explain the Weber-Fechner law [72] in psychophysics, where $\ell([n_1, n_2) | r)$ is the intensity of human sensation, $1/\ln r$ is a perception- and stimulus-dependent proportionality constant, n_2 is the strength of the stimulus, and n_1 is the zeroing strength threshold.

When $n_1 = n$ and $n_2 = n + 1$, we can again interpret this likelihood as the probability of getting a leading r -ary numeral $n \in \mathbb{N}^+$ of any length, i.e.,

$$\Pr(r, n) = \ell([n, n + 1) | r) = \log_r (n + 1) - \log_r n = \log_r \left(1 + \frac{1}{n} \right) \in \mathbb{R}$$

The efficiency of a r -ary numeral system worsens as $r \rightarrow 1^+$ or $r \rightarrow b \rightarrow \infty$ [73] because r diverges from the optimal radix economy, namely Euler's number e , destroying the information. In the former case, we encounter the unary system, which boils down to a linear frequency. In the latter case, the numerals $n < r$ that only use the first position increase limitlessly. Both are no-coding cases.

4. A Curious Effect

We prove that the Kempner distribution reflects the rational version of NBL for bijective numeration, allows figuring a natural resolution in PN, and confirms a global tendency towards smallness.

Watch the notation; we display the base and the radix underlined to denote bijective numeration rather than standard notation.

4.1. NBL for Bijective Numeration

Suspicion about the authenticity of the number cero [74] suggests that bijective PN is likely more natural than standard PN, the number system we use daily. Various curious series we will analyze in the following subsection, specifically the Kempner distribution, append additional evidence that NBL for bijective numeration [75] is foundational and universal.

Every formula about the NBL for standard PN has a bijective peer. Following the same plot thread we developed in Section 3.4, a sample of chains encoded using bijective \underline{b} -ary satisfies the global NBL if the leading quantum falls in bucket $[s, t)$ relative to the area swept by base \underline{b} with probability

$$\Pr(\underline{b}, [s, t)) = \frac{H_{t-1} - H_{s-1}}{H_{\underline{b}}} \in \mathbb{Q}$$

where $1 \leq s < t \leq \underline{b}$ and $s, t, \underline{b} \in \mathbb{N}$. When $s = q$ and $t = q + 1$ we obtain the probability with base \underline{b} of leading quantum q ,

$$\Pr(\underline{b}, q) = \frac{1}{qH_{\underline{b}}} \in \mathbb{Q} \quad (1 \leq q \leq \underline{b}, \{q, \underline{b}\} \in \mathbb{N}) \quad (9)$$

Thus, NBL for the standard PN in base $b + 1$ corresponds to NBL for bijective \underline{b} -ary numeration. For example, we obtain $\Pr(\underline{1}, 1) = 100\%$, $\Pr(\underline{3}, 1) = 6/11 \approx 54.5\%$, $\Pr(\underline{3}, 2) = 3/11 \approx 27.3\%$, $\Pr(\underline{3}, 3) = 2/11 \approx 18.2\%$, $\Pr(\underline{A}, 1) = 0.34142$, and $\Pr(\underline{A}, A) = 0.03414$, where " \underline{A} " symbolizes the bijective decimal base. Owing to $\Pr(\underline{2}, 1) = 2/3$ and $\Pr(\underline{2}, 2) = 1/3$, the odds $o(2 : 1|\underline{2}) = \Pr(\underline{2}, 2)/\Pr(\underline{2}, 1) = 1/2$ constitute an essential sharing out.

The entropy of PMF (9), $\tilde{E}(\underline{b})$, is the expected value (weighted arithmetic mean) of the harmonic likelihood function $(\psi(x) - \psi(1)) = H_{x-1}$ evaluated at the probability mass reciprocal, i.e.,

$$\tilde{E}(\underline{b}) = \hat{E}(\Pr(\underline{b}, q)) = \sum_{q=1}^{\underline{b}} \frac{H_{qH_{\underline{b}}-1}}{qH_{\underline{b}}} \text{harmt}$$

For example, $\tilde{E}(\underline{1}) = 0$, $\tilde{E}(\underline{2}) = 0.90914$, $\tilde{E}(\underline{3}) = 1.35432$, $\tilde{E}(\underline{10}) = 2.47676$, and $\tilde{E}(\underline{100}) = 4.2269$. When \underline{b} acquires a gargantuan value, we can take the summation as an integral and the harmonic number function as the natural logarithm, so that the "differential entropy" [76] of the global NBL approximately tends to

$$\int_1^{\underline{b}} \frac{\ln(q \ln \underline{b})}{q \ln \underline{b}} dq = \frac{1}{2} \ln \underline{b} + \ln(\ln \underline{b})$$

Thus, the global entropy is finite, which agrees with the Bekenstein bound in physics.

The probability of picking a chain of any length starting with c is the likelihood gap it induces on the \underline{b} -ary harmonic scale, i.e.,

$$\Pr(\underline{b}, c) = \frac{H_c}{H_{\underline{b}}} - \frac{H_{c-1}}{H_{\underline{b}}} = \frac{1}{cH_{\underline{b}}} \in \mathbb{Q}$$

which becomes (9) when c is a base's quantum. For example, the probability that a bijective decimal chain starts with 11 (e.g., .111) and AA (e.g., AAAA) is $\Pr(\underline{A}, 11) = 1/(11H_{10}) \approx 0.03104$ and $\Pr(\underline{A}, AA) = 1/(110H_{10}) \approx 0.003104$, respectively.

This result allows us to derive the probability of picking a length- l bijective \underline{b} -ary chain starting with the quantum q ,

$$\Pr(\underline{b}, l, q) = \frac{\sum_{k=\underline{b}^{l-1} + \frac{\underline{b}^{l-1}-1}{\underline{b}-1}-1}^{k=(q+1)\underline{b}^{l-1} + \frac{\underline{b}^{l-1}-1}{\underline{b}-1}-1} \frac{1}{k}}{H_{\frac{\underline{b}^{l+1}-1}{\underline{b}-1}-1} - H_{\frac{\underline{b}^l-1}{\underline{b}-1}-1}} \in \mathbb{Q}$$

where $1 \leq q \leq \underline{b}$, $\underline{b} > 1$, and $\{\underline{b}, l, q\} \in \mathbb{N}^+$. For instance, the probability of running into 1 to 3 as the first quantum of a bijective ternary chain with length 5 is $\{0.46565, 0.30602, 0.22833\}$, and the chances of choosing 1 to A as the first quantum of a bijective decimal chain with length 2 is $\{0.2842, 0.1688, 0.1205, 0.09377, 0.07677, 0.065, 0.05637, 0.04976, 0.04454, 0.04031\}$. Watch that this equation boils down to (9) if $l = 1$.

Too, the probability that we run into q as the p -th quantum of a bijective \underline{b} -ary chain is

$$\Pr(\underline{b}, q, p) = \frac{\sum_{k=\frac{\underline{b}^{p-1}-1}{\underline{b}-1}}^{k=\frac{\underline{b}^p-1}{\underline{b}-1}-1} \frac{1}{\underline{b}k+q}}{H_{\frac{\underline{b}^{p+1}-1}{\underline{b}-1}-1} - H_{\frac{\underline{b}^p-1}{\underline{b}-1}-1}} \in \mathbb{Q} \quad (10)$$

where $1 \leq q \leq \underline{b}$, $\underline{b} > 1$, and $\{\underline{b}, q, p\} \in \mathbb{N}^+$. For instance, the probability of getting 1 to 3 as the fifth quantum of a bijective ternary chain is $\{0.335011, 0.333327, 0.331662\}$, and the chances of encountering 1 to A as the second quantum of a bijective decimal chain is $\{0.1183, 0.113, 0.1083, 0.1041, 0.1004, 0.09694, 0.09381, 0.09094, 0.08829, 0.08583\}$. Watch that this equation reduces to (9) if $p = 1$. Figure 3 shows the PMF of various bijective bases for consecutive positions and the hyperbolic progression of the bijective ternary digits as the position increases.

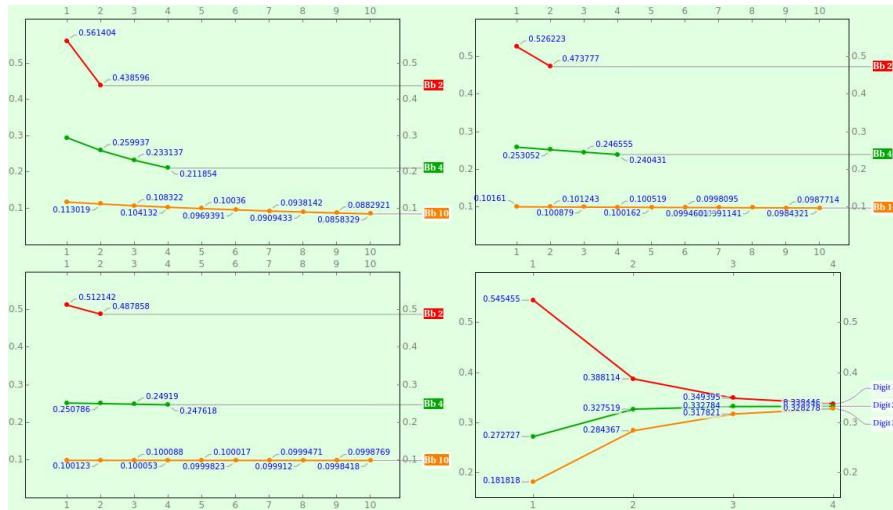


Figure 3. Leading quantum's PMF for bijective bases 2, 4, and 10 at positions 2 (top-left), 3 (top-right), and 4 (bottom-left), which quickly tend to the uniform distribution. On the bottom right, we show the hyperbolic plot of the bijective ternary digits as a function of their position; only the first few quanta make a coding difference.

Following the plot thread we developed in Section 3.6, the ratio between the area under the hyperbola delimited by the bin $[d_1, d_2]$ and the radix support $[1, r \ll \underline{b}]$ is

$$\Pr(\underline{r} [d_1, d_2]) = \log_{\underline{r}+1} \frac{d_2}{d_1} \in \mathbb{R} \quad (1 \leq d_1 < d_2 \leq \underline{r}, \{d_1, d_2, \underline{r}\} \in \mathbb{N}^+)$$

We arrive at the NBL for bijective notation by putting $d_1 = d$ and $d_2 = d + 1$. A sample of numerals expressed in bijective r -ary PN satisfies the local NBL if the leading digit d occurs with probability

$$\Pr(r, d) = \log_{r+1} \left(1 + \frac{1}{d} \right) \in \mathbb{R} \quad (1 \leq d \leq r) \quad (11)$$

The NBL with radix $r + 1$ corresponds to the bijective r -ary numeration's NBL; for example, the standard ternary system assigns to 1 and 2 the probabilities 63 % and 37 %, which is the PMF of bijective binary numeration. In the usual case where the radix is $r = 10$, the standard decimal system assigns to digits 1 and 9 probabilities of 30.1 % and 4.6 %. In contrast, the bijective decimal numeration assigns to digits 1 and A $\equiv 10$ probabilities of 28.9 % and 4.0 %. Likewise, the local bijective ternary numeration assigns to 1, 2, and 3 the probabilities 50 %, 29 %, and 21 %, contrasting with the percentages 54.5 %, 27.3 %, and 18.2 % the global bijective ternary numeration assigns.

The entropy of PMF (11) for radix r , $\tilde{e}(r)$, is the expected value (weighted arithmetic mean) of the likelihood function ($\ln(x)$) evaluated at the probability mass reciprocal, i.e.,

$$\tilde{e}(r) = \hat{E}(\Pr(r, d)) = \sum_{d=1}^{d=r} \log_{r+1} \left(1 + \frac{1}{d} \right) \ln \left(\frac{1}{\log_{r+1} \left(1 + \frac{1}{d} \right)} \right) \text{ nat}$$

For example, $\tilde{e}(1) = 0$, $\tilde{e}(2) = 0.65846$, $\tilde{e}(3) = 1.03247$, $\tilde{e}(10) = 2.08134$, and $\tilde{e}(100) = 3.84099$. Because $r < b$ and we assume that b is a positive natural number, the local entropy is finite, in agreement with the Bekenstein bound.

Note that (11) is also valid for the unary system ($r = 1$), unlike (7) in standard PN; bijective unary assigns the probability of 100 % to 1. A system "encoding" data in bijective unary has no curvature and keeps a linear scale. In bijective numeration, (re)coding from unary into r -ary means summing the number of ones and executing an iterative procedure based on Euclidean division. Figure 4 describes the encoding algorithm; e.g., it converts the representation of 1567 into 1233231_3 .

		Position	Remainder	Quotient	Digit	Term weigh	
Radix	Datum	0	1567	522	1	1	
		1	522	173	3	9	$Digit \times Radix^{Position}$
		2	173	57	2	18	
		3	57	18	3	81	
		4	18	5	3	243	
		5	5	1	2	486	
		6	1	0	1	729	
			Total			1567	
							Procedure halts
							$1567 \equiv 1233231_3$

Figure 4. Data encoding in bijective numeration. Note that we use the ceiling function; specifically, we use $\lceil \cdot \rceil - 1$ instead of the floor function $\lfloor \cdot \rfloor$, which standard PN uses.

We can generalize the PMF given by (11) to the probability of getting a leading r -ary numeral $n \in \mathbb{N}^+$ of any length. It is the likelihood gap it induces on the logarithmic scale, i.e.,

$$\Pr(\underline{r}, n) = \log_{\underline{r}+1} (n+1) - \log_{\underline{r}+1} n = \log_{\underline{r}+1} \left(1 + \frac{1}{n}\right) \in \mathbb{R} \quad (12)$$

For example, the probability that a bijective decimal numeral starts with "2A1", say $2.A1\underline{10}$ or $2A17\underline{10}$, is $\log_{11} \left(1 + \frac{1}{301}\right) = 0.13832\%$.

This result allows us to derive the probability of picking a bijective \underline{r} -ary numeral with length l starting with the digit d ,

$$\Pr(\underline{r}, l, d) = \sum_{k=d\underline{r}^{l-1} + \frac{r^{l-1}-1}{r-1}-1}^{k=(d+1)\underline{r}^{l-1} + \frac{r^{l-1}-1}{r-1}-1} \log_{\underline{r}^{l-1}-1} \left(1 + \frac{1}{k}\right) \in \mathbb{R}$$

where $1 \leq d \leq \underline{r}$, $\underline{r} > 1$, and $\{\underline{r}, l, d\} \in \mathbb{N}^+$. For instance, the probability of picking 1 to 3 as the first digit of a bijective ternary numeral with length 5 is $\{0.465312, 0.306147, 0.228541\}$, and the probability of choosing 1 to A as the first digit of a bijective decimal numeral with length 2 is $\{0.2797, 0.1685, 0.1209, 0.09442, 0.07746, 0.06567, 0.057, 0.05036, 0.0451, 0.04084\}$. Owing to this equation boils down to (11) if $l = 1$, the local NBL is length-invariant! Figure 5 shows the PMF of various bijective radices for consecutive lengths and the hyperbolic progression of the bijective ternary digits as the numeral's length expands.

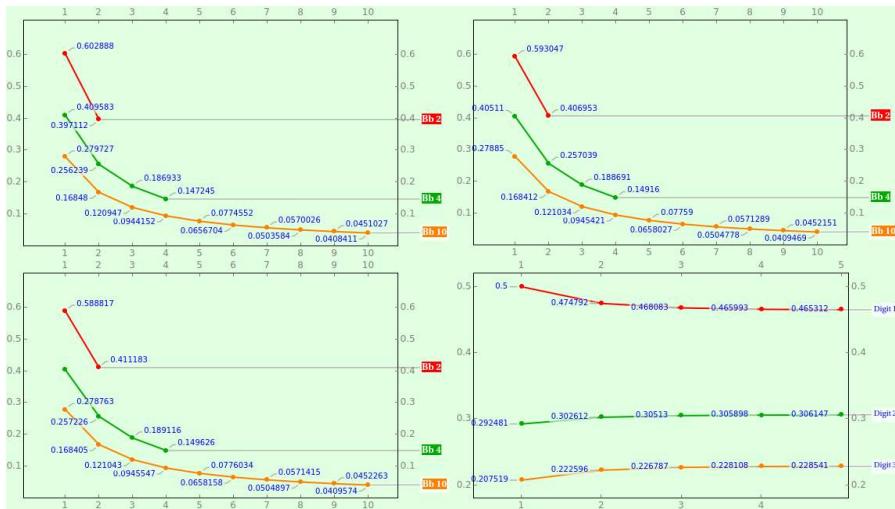


Figure 5. Leading digit's PMF for numerals in bijective radices 2, 4, and 10 with lengths 2 (top-left), 3 (top-right), and 4 (bottom-left). On the bottom right, we show the probability plot of the bijective ternary digits as a function of the numeral's length; the probability gap between consecutive digits tends to stabilize.

Likewise, Equation (12) allows us to derive the law for digits beyond the first; the probability of getting a \underline{r} -ary digit d at position p is

$$\Pr(\underline{r}, d, p) = \sum_{k=\frac{r^{p-1}-1}{r-1}-1}^{k=\frac{r^p-1}{r-1}-1} \log_{\frac{r^p-1}{r-1}-1} \left(1 + \frac{1}{rk+d}\right) \in \mathbb{R}$$

where $1 \leq d \leq \underline{r}$, $\underline{r} > 1$, and $\{\underline{r}, d, p\} \in \mathbb{N}^+$. Because this equation reduces to (11) if $p = 1$, the local NBL is position-invariant! For instance, the chance of picking 1 to 3 as the fifth digit of a bijective ternary numeral is $\{0.335006, 0.333327, 0.331667\}$, and

the probability of choosing 1 to A as the second digit of a bijective decimal numeral is $\{0.1177, 0.1126, 0.1081, 0.1041, 0.1004, 0.09707, 0.09402, 0.09121, 0.08862, 0.0862\}$.

4.2. Depleted and Constrained Harmonic Series

The global NBL for bijective numeration suddenly appears in the set of Kempner's curious series. We say a series is curious when the infinite summation of a harmonic series, divergent, is depleted by constraining its terms to satisfy specific convergence conditions. For example, consider the harmonic series missing the terms where "66" appears in their denominator. Most researchers in this fieldwork use decimal representation, but we can generalize the results to any base. Although their terminology refers to the items of a unit fraction's denominator as digits, for us, these are quanta of a chain because we are handling terms of a harmonic series.

The point is that most depletions result in an absolute mass because a harmonic series is on the verge of divergence. In particular, a harmonic series becomes convergent by omitting a single quantum. For example, the shrunk harmonic series without the terms in which "4" appears anywhere in the decimal representation of the denominator is K_4 of the Kempner series. Offhand, convergence comes up because we withdraw most of the terms; $1/10$ of the terms contain a "4" if the random variable ranges from 0 to 9, 20% have at least one "4" if the random variable ranges from 0 to 99, and eventually, most of the terms of any random chain with 100 quanta will contain at least one "4" and will not sum. However, this explanation needs to be corrected.

A K_N series converges slowly [77]. We will reason that this property is due to large numerical chains' relative and geometrically short contribution to the total. Table 2 summarizes the outcomes of approximated calculations from 1 (K_1) to $A \equiv 10$ (K_A). Nonetheless, the most stunning feature of the Kempner summations (third column) is that they outline a curve that decreases harmonically.

Every quantum eliminates the same number of terms. $K_1 < K_2 < \dots < K_A$ means not that "1" is in more terms than "2" or "3" but a heavier mass attributed to the terms with the minor quanta; if we take out $1/1$, the resulting summation is smaller than when we take out $1/2$ or $1/3$, and "A" is the quantum that contributes less to the total. (Although "A" is taken as "0" for calculation purposes, the value of K_A proves that bijective numeration is underneath.) Considering that a Kempner series is infinite and the set of Kempner series embraces all quanta q represented in bijective decimal, how could we find a better proof that a default probability potential outlines a hyperbolically decreasing function of q ?

Since a curious series converges by default of unit fraction terms, the mass share of a quantum globally depends on the reciprocals of the Kempner summations; the third column of the table includes K_q 's reciprocals normalized to 100 % (e.g., K_1 's relative mass is $M_1 = \left(K_1 \sum_{q=1}^A 1/K_q\right)^{-1} \approx 13\%$). We must underline the relevance of these summations and percentages, reflecting the mass of every quantum irrespective of where it is, in contrast with the global NBL, which indicates the probability mass of a quantum at a given position in a given base.

We introduce two caveats to analyze the NBL weights (fourth column). First, the Kempner distribution conforms with NBL via the average of NBL distributions for different positions, which is NBL, too. For instance, W_1 is, in principle, the average of quantum 1's probabilities at first (34.14%), second (11.89%), third (10.18%), fourth (10.01%), et cetera position according to (10). Second, because the distribution of the n th quantum quickly tends to be uniform (10% for each of the ten quanta from the fifth position), we must suspect that there exists a threshold position above which the contributions to the quantum's weight do not count; otherwise, the resulting mean distribution will end up reaching uniformity despite the differences that the Benford distribution makes at the first positions. Consequently, the last column calculates W_q as the NBL frequency averaged only over the first nine positions. Averaging ten positions also gives an excellent approximation (with a mean error of .091 %) to the distribution of Kempner masses, but nine positions deliver the minimal total mean error of .024 %.

Table 2. These are the absolute and relative masses of the Kempner series compared to NBL averaged over the first nine positions.

Decimal quantum (q)	Kempner K_q summations	Kempner M_q mass	NBL average weight W_q (9 positions)
1	16.1770	13.00	12.91
2	19.2573	10.92	10.95
3	20.5699	10.22	10.26
4	21.3275	9.86	9.89
5	21.8346	9.63	9.65
6	22.2056	9.47	9.48
7	22.4935	9.35	9.36
8	22.7264	9.25	9.25
9	22.9207	9.18	9.16
A	23.1034	9.10	9.09
Total	212.6158	100	100

Can we extrapolate this result in $\underline{b} = \underline{A}$ to any value of \underline{b} ? If affirmative, PN would ignore a natural significand's quanta from the \underline{b} th place, agreeing with claims often made by mathematicians [78], physicists [79,80], and engineers [81] about the illogicality of a PN system carrying excessive digits in calculations of any type, regardless of the discipline.

We surmise that a bijective b -ary chain c that fulfills $\log_b c > \underline{b}$ is physically elusive. The universe in base \underline{b} would cope with at most \underline{b} nesting levels, each distinguishing between b possible quanta. The "physical resolution"

$$\mathring{R}(\underline{b}) = \underline{b}^{\underline{b}}$$

would estimate the scope of quanta a computational system like the cosmos can naturally operate, much as a native resolution describes the number of pixels a screen can display.

In [82], the author contrives an efficient algorithm for summing a series of harmonic numbers whose denominator contains no occurrences of a particular numerical chain. As a result of the calculations, a harmonic series in base b omitting a chain of length n (regardless of its specific quanta) might converge approximately to

$$b^n \ln b.$$

This conjecture means that the contribution of linearly more extended chains to an endless series is geometrically lesser. For instance, the harmonic series where we impede the occurrence of the decimal numeral "314159" is about 2302582.334, whereas the same sum omitting "only" "3" is 22.921, 10^5 times as low. Thus, large numerical chains would be exponentially inconsequential.

More general constraints allow several occurrences of a given quantum to calculate summations positively. Let $S(n, q, b)$ be the sums of the b -base reciprocals of naturals that have precisely n instances of the quantum q . For example, omitting the terms whose denominator in decimal representation contains one or more 6 is the particular case $S(0, 6, 10)$. The sequence of values S decreases and tends to

$$\lim_{n \rightarrow \infty} S(n, q, b) = b \ln b \approx \ln \mathring{R}(b)$$

regardless of q [83].

Except for the gap from $n = 0$ to $n = 1$, where the total increases, the summation falls as we raise the constraining quantity of quanta. What is the reason? It is not that we get more terms with n qs

than terms containing $n + 1$ qs , but that the longer the chain, the lighter the contribution. Furthermore, when $n \gg 1$, $S(n, q, b) \gtrsim S(n + 1, q, b)$, whereas if $n \gtrsim 1$, $S(n, q, b) \gg S(n + 1, q, b)$, i.e., increments of n near the origin produce significant drops and vice versa, increments of n far from the origin produce negligible drops. Although we have not statistically tested "the number of quanta" for compliance with NBL, we can again conclude that while small is a synonym for solid and discernible, huge numerical chains are fragile and hardly convey differences.

Instead of imposing absolute constraints, we can allow in a term arbitrarily many quanta q irrespective of the position and number so long as the proportion of qs remains below a fixed parameter $\lambda \in [0..1]$. In [84], the authors prove that the series converges if and only if $\lambda < 1/b$. In decimal, while Kempner's original series implies $\lambda = 0$, where no term containing a given quantum contributes to the summation, the complete harmonic series means $\lambda = 1$, where any density is allowed, i.e., we keep all the reciprocals.

For instance, if we consider the constraint "allow a rate of $\lambda = 5\%$ of 7s at most", the term $1/98765432109876543210$ disappears (10 % of 7s), but neither $1/98654321098$ (no 7s) nor $1/98865432109876543210$ (5 % of 7s) does. While the series converges in $\lambda \in [0..1/10]$, it no longer converges above the threshold $\lambda = 1/10$. Note that the archetype of the Pareto law appears naturally; on average, 90% of the unit fractions, those with the highest quantum density, offset the remaining 10%. Moreover, this result engages with our surmise concerning the physical resolution $\mathring{R}(b)$ of a universal computational system. Again, densities of b quanta or more are intractable. A PN system must restrict itself to chains with less than b quanta to guarantee the operability of coded data and avoid overflow conditions.

5. Odds

Although odds typically appear in gambling and statistics, this section illustrates how they are central to the computational processes of a coding source, including an application to physics and another to decision theory.

We usually define the odds of an outcome as the ratio of the number of events that generate that particular result to those that do not. In this sense, odds constitute another measure of the chance of a result. Likewise, the ratio between the probabilities of two events determines their relative odds; the higher the odds of an outcome compared with another, the more informative the latter's occurrence is.

Indeed, odds highlight the rational character of a probability. For instance, we can interpret the one-parameter PMF (1) in terms of odds. Since the odds O of picking a nonzero natural N against piking N^2 are precisely N^2 , we can establish $\Pr(N) = \epsilon O(N^2 : N)$.

The encoded odds between a pair of events are the product of their probability ratio and likelihood factor. The coding rule agrees with Bayes' law. Odds between propensities or degrees of belief become information correlations representing entropic contributions in Bayesian coding. Thus, we attribute a metric sense to this theorem, embracing the objectivistic [85] and subjectivistic [86] interpretations.

Our description will exclusively focus on standard PN, omitting the corresponding bijective numeration's derivations and formulas for conciseness.

5.1. Global Bayesian Coding

The probability ratio between two events diverges from the unit as their correlation weakens. A PN system must multiply this value by a coding factor to fit into the base's harmonic scale. This operation is rigorously Bayes' theorem. Specifically, global Bayesian coding employs the formula

$$\tilde{O}(t : s | X = b) = \tilde{O}(t : s) \Lambda_{\tilde{O}}(t : s | X = b) \in \mathbb{Q} \quad (1 \leq s < t < b) \quad (13)$$

to encode the odds between two numbers.

- $\tilde{O}(t : s | X = b)$ represents the global (encoded or posterior) odds of getting quantum t against s in base b . We can consider it the rational quantum s/t on a b -ary harmonic scale.
- $\tilde{O}(t : s)$ is the ratio between the probabilities of the two events according to (1), namely

$$\tilde{O}(t:s) = \tilde{O}(t:s|1) = \frac{\Pr(t)}{\Pr(s)} = \left(\frac{s}{t}\right)^2$$

straightforwardly measuring the (decoded or prior) odds of picking the number t against s on a linear scale. If we fix the center of the range, the narrower the interval, the higher the odds, whereas if we fix the interval width, the minor s (or t), the lower the odds. Note that the odds of two concatenated intervals calculated separately are the product of the interval's odds,

$$\tilde{O}(z:x|1) = \left(\frac{x}{y}\right)^2 \left(\frac{y}{z}\right)^2 = \left(\frac{x}{z}\right)^2$$

- $\Lambda_{\tilde{O}}(t:s|X=b)$ is the global coding (Bayes) factor, which measures the degree to which the outcome b of the random variable X supports "hypothesis" t against s , assuming both are independent numbers. Because interval $[s, t)$ is not yet encoded, the coding law establishes a likelihood difference instead of a likelihood ratio, namely

$$\Lambda_{\tilde{O}}(t:s|X=b) = \mathcal{L}(t|X=b) - \mathcal{L}(s|X=b) = \frac{H_{t-1} - H_{s-1}}{H_{b-1}} = \Pr(b, [s, t))$$

where

$$\mathcal{L}(q|X=b) = \frac{H_{q-1}}{H_{b-1}} \quad (14)$$

is the likelihood function of q with b fixed; since $\mathcal{L}(1|X=b)$ vanishes and $\mathcal{L}(b|X=b) = 1$, we can understand this function as a measure of the nearness between q and $b > 1$ normalized to one. The coding factor is precisely Equation (3), measured in b -ary harmonic information units.

Compiling, PN calculates (13) as

$$\tilde{O}(t:s|b) = \left(\frac{s}{t}\right)^2 \frac{H_{t-1} - H_{s-1}}{H_{b-1}} \in \mathbb{Q} \quad (1 \leq s < t < b) \quad (15)$$

It is the cost of computing the bucket's harmonic width, i.e., the entropic contribution of bucket $[s, t)$ to b 's harmonic scale. Because $\tilde{O}(2:1|b)$ is maximally informative irrespective of the base, the global information unit corresponds to the natural harmonic bucket $[\psi(q)]_1^2$ we use in (2). The global odds of a quantum against itself vanish, having no representation on a harmonic scale. The reciprocal

$$\tilde{O}(s:t|b) = \frac{1}{\tilde{O}(t:s|b)} = \left(\frac{t}{s}\right)^2 \frac{H_{b-1}}{H_{t-1} - H_{s-1}} \in \mathbb{Q} \quad (1 \leq s < t < b)$$

measures the odds of quantum s against t , with a maximum approaching b^2 as b climbs to infinity. A PN system must employ (13)'s variation

$$\tilde{O}(t:s|b') = \tilde{O}(t:s|b) \Lambda_{\tilde{O}}(b':b|q \in [s, t)) \in \mathbb{Q}$$

to recode globally, where $1 \leq s \leq q < t < b$ and $t < b'$. The coding (Bayes) factor is a likelihood ratio when it deals with previously encoded data, as usual in statistics; using (14),

$$\Lambda_{\tilde{O}}(b':b|q \in [s, t)) = \frac{\mathcal{L}(b|q)}{\mathcal{L}(b'|q)} = \frac{H_{b-1}/H_{q-1}}{H_{b'-1}/H_{q-1}} = \frac{H_{b-1}}{H_{b'-1}}$$

measures the degree to which a given quantum supports "hypothesis" b' against b , assuming both are independent quanta. Hence, the PN system can change to base b' utilizing the rule

$$\tilde{O}(t:s|b') = \tilde{O}(t:s|b) \frac{H_{b-1}}{H_{b'-1}} \in \mathbb{Q} \quad (16)$$

which coincides with the odds of $[s, t)$ in base b' for the first time because of

$$\tilde{O}(t : s | b') = \left(\left(\frac{s}{t} \right)^2 \left(\frac{H_{t-1} - H_{s-1}}{H_{b'-1}} \right) \right) \frac{H_{b-1}}{H_{b'-1}} = \left(\frac{s}{t} \right)^2 \Pr(b', [s, t])$$

Equation (16)' Bayes factor $H_{b-1}/H_{b'-1}$ is the classical Bayes factor replacing probabilities by global likelihoods. Thus, the transformation $b \rightarrow b'$ constitutes a primal memory (incremental) process that decreases the global odds if $b < b'$, and vice versa, increases the global odds if $b > b'$. For instance, an asymmetry such that b grows every tic of a global clock would mean an unstoppable progressive information loss for a fixed universe region; this connection between time and entropy is crucial to theoretical physics and cosmology [87].

For example, the PN system encodes bucket [4, 13) to base $b = 100$ as

$$\tilde{O}(13 : 4 | 100) = \left(\frac{4}{13} \right)^2 \frac{H_{13-1} - H_{4-1}}{H_{100-1}} \approx 0.0232$$

This value is the entropic contribution of bucket [4, 13) to 100's harmonic scale. When the PN system changes the base to $b' = 110$, using (16), it delivers

$$\tilde{O}(13 : 4 | 110) = \tilde{O}(13 : 4 | 100) \frac{H_{100-1}}{H_{110-1}} \approx 0.0228$$

meaning that the bucket's entropic contribution decreases. Then, changing to base $b' = 90$ yields

$$\tilde{O}(13 : 4 | 90) = \tilde{O}(13 : 4 | 110) \frac{H_{110-1}}{H_{90-1}} \approx 0.0237$$

i.e., the bucket's entropic contribution increases. Finally, the PN system decodes the odds with base 90 by solving the prior from (15), i.e.,

$$\tilde{O}(13 : 4) = \frac{\tilde{O}(13 : 4 | 90)}{\Pr(90, [4, 13])} = \tilde{O}(13 : 4 | 90) \frac{H_{90-1}}{H_{13-1} - H_{4-1}} = \left(\frac{4}{13} \right)^2$$

5.2. Local Bayesian Coding

Local Bayesian coding assumes that (13), the global Bayesian law, governs the universe's information. We express the informational correlation between two numerals by multiplying their harmonic correlation by a coding factor, obtaining a point on a logarithmic scale. This operation is rigorously the Bayes' theorem that settles down the basis of a conformal metric space. Specifically, local Bayesian coding employs the formula

$$\tilde{o}(n_2 : n_1 | Y = r) = \tilde{o}(n_2 : n_1 | b) \Lambda_{\tilde{o}}(n_2 : n_1 | Y = r) \in \mathbb{R} \quad (17)$$

to encode the probability ratio between n_1 and n_2 , where $(1 \leq n_1 < n_2 < b) \wedge (r < b)$.

- $\tilde{o}(n_2 : n_1 | Y = r)$ represents the local (encoded or posterior) odds of getting n_2 against n_1 with radix r .
- $\tilde{o}(n_2 : n_1 | b)$ is the (prior) probability ratio between the two events only assuming that a global base exists; using (4),

$$\tilde{o}(n_2 : n_1 | b) = \frac{\Pr(b, n_2)}{\Pr(b, n_1)} = \frac{\frac{1}{n_2 H_{b-1}}}{\frac{1}{n_1 H_{b-1}}} = \frac{n_1}{n_2}$$

It measures the strength of the association between n_1 and n_2 on the harmonic scale provided by b .

- $\Lambda_{\tilde{o}}(n_2 : n_1 | Y = r)$ is the local coding (Bayes) factor, which measures the degree to which the outcome r of the random variable Y supports "hypothesis" n_2 against n_1 , assuming both are

independent. Because the bucket $[n_1, n_2)$ is not locally encoded yet, the coding law establishes a likelihood difference instead of a likelihood ratio, namely

$$\Lambda_{\delta}(n_2 : n_1 | Y = r) = \ell(n_2 | Y = r) - \ell(n_1 | Y = r)$$

where $\ell(n | Y = r) = \log_r n$ quantifies the likelihood of n when the actual value r occurs. In short, the local coding factor is the log-odds of n_2 relative to n_1 , equivalent to the support of $[n_1, n_2)$ with radix r according to (8), i.e.,

$$\Lambda_{\delta}(n_2 : n_1 | Y = r) = \log_r n_2 - \log_r n_1 = \log_r \frac{n_2}{n_1} = \ell([n_1, n_2) | r) \quad (18)$$

Compiling, the PN system calculates (17) as

$$\tilde{o}(n_2 : n_1 | r) = \frac{n_1}{n_2} \log_r \frac{n_2}{n_1} \quad (1 \leq n_1 < n_2) \quad (19)$$

Note that the local odds of a numeral against itself vanish, having no representation on a logarithmic scale. Consequently, the local Bayes' rule measures the entropic contribution of bin $[n_1, n_2)$ on r 's logarithmic scale, with a minimum approaching $\tilde{o}(r : 1 | r) = 1/r$ as r climbs to infinity. Euler's number has an extraordinary meaning in this setting; a Bayesian datum in this form is maximally informative irrespective of the radix when $n_2/n_1 = e = 2.718\ldots$, an ideal proportion that induces the local information unit associated with the natural logarithmic bin $[\ln n]_1^e$ we use in (5).

A PN system must employ (17)'s variation

$$\tilde{o}(n_2 : n_1 | r') = \tilde{o}(n_2 : n_1 | r) \Lambda_{\delta}(r' : r | n \in [n_1, n_2)) \in \mathbb{R}$$

$$\forall \{n_1, n, n_2, r, r'\} \in \mathbb{N} \mid (1 \leq n_1 \leq n \leq n_2)$$

to recode locally. When it deals with previously encoded data, the local Bayes factor is a likelihood ratio, as usual in statistics,

$$\Lambda_{\delta}(r' : r | n \in [n_1, n_2)) = \frac{\ell(r | n)}{\ell(r' | n)} = \frac{\log_n r}{\log_n r'} = \frac{\ln r}{\ln r'}$$

Thus, the degree to which the outcome $n \in [n_1, n_2)$ of the random variable Y supports "hypothesis" r' against r is independent of n . Then, a coding source can change to standard radix r' utilizing

$$\tilde{o}(n_2 : n_1 | r') = \tilde{o}(n_2 : n_1 | r) \frac{\ln r}{\ln r'} \quad (20)$$

which coincides with the odds of n_2 against n_1 with radix r' for the first time

$$\tilde{o}(n_2 : n_1 | r) \frac{\ln r}{\ln r'} = \left(\frac{n_1}{n_2} \log_r \frac{n_2}{n_1} \right) \frac{\ln r}{\ln r'} = \frac{n_1}{n_2} \log_{r'} \frac{n_2}{n_1}$$

Equation (20)' Bayes factor $\ln r / \ln r'$ is the classical Bayes factor replacing probabilities by local likelihoods. Note that the transformation $r \rightarrow r'$ increases the odds if $r > r'$, and vice versa, decreases the odds if $r < r'$.

For example, a coding source locally encodes the bin $[4, 13)$ using radix $r = 100$ as

$$\tilde{o}(13 : 4 | 100) = \frac{4}{13} \log_{100} \frac{13}{4} \approx 0.07875$$

This value measures the entropic contribution of bin $[4, 13)$ to standard radix 100. When the coding source changes the radix to $r' = 110$, it delivers using (20)

$$\tilde{o}(13 : 4 | 110) = \tilde{o}(13 : 4 | 100) \frac{\ln 100}{\ln 110} \approx 0.07715$$

Then, changing the radix to $r' = 90$ yields

$$\tilde{o}(13:4|90) = \tilde{o}(13:4|110) \frac{\ln 110}{\ln 90} \approx 0.0806$$

Finally, the coding source decodes the odds with radix 90 by solving the prior from (19), i.e.,

$$\tilde{o}(13:4) = \frac{\tilde{o}(13:4|90)}{\Pr(90, [4, 13])} = \frac{0.0806}{\log_{90} \frac{13}{4}} \approx 0.30769 \approx \frac{4}{13}$$

Remember that local Bayesian coding copes not only with ratios of digits but with ratios of numerals in general. For example, the rational $95/971$ (bin $[95, 971]$) encoded with radix 4 is

$$\tilde{o}(971:95|4) = \frac{1133_4}{33023_4} \log_4 \frac{33023_4}{1133_4} \approx 0.02213333_4$$

If environmental conditions cast a change to radix 3, the coding source would decode the datum

$$\tilde{o}(971:95|3) = 0.02213333_4 \frac{\ln 4}{\ln 3} \approx 0.01212022_3$$

as

$$\tilde{o}(971:95) = \frac{\tilde{o}(971:95|3)}{\ell([95, 971]|3)} \approx \frac{0.01212022_3}{\log_3 \frac{1022222_3}{10112_3}} \approx \frac{95}{971}$$

5.3. Elemental Jumps

Using odds instead of probabilities is especially powerful when we measure the gap between successive quanta or digits.

The odds (15) between consecutive quanta

$$\tilde{O}(q+1:q|b) = \left(\frac{q}{q+1} \right)^2 \Pr(b, [q, q+1]) = \left(\frac{q}{q+1} \right)^2 \frac{1}{qH_{b-1}} = \frac{q}{(q+1)^2 H_{b-1}}$$

measure the associated harmonic likelihood gap in a given base b , where we have used equations (3) and (4). b -normalized quantum jumps define the PMF

$$\begin{aligned} \zeta_b &= \frac{1}{H_{b-1} - H_{b-1,2}} \\ \Pr_{\tilde{O}}(b, q) &= \zeta_b \frac{q}{(q+1)^2} \in \mathbb{Q} \end{aligned} \tag{21}$$

an exact and multiplicatively separable function where $H_{N,2} \equiv \sum_{k=1}^N 1/k^2$ is the generalized N th harmonic number in power two and $1 \leq q < b - 1$. Note that the summation only goes until the penultimate quantum $q = b - 2$ because $q = b - 1$ cannot jump to b .

PMF (21) is well-defined because of

$$\sum_{q=1}^{b-2} \Pr_{\tilde{O}}(b, q) = 1$$

so we can take as the odds version of PMF (4). For example, $\zeta_4 = 36/17$, $\Pr_{\tilde{O}}(4, 1) = \zeta_4 1/4 = 9/17$, and $\Pr_{\tilde{O}}(4, 2) = \zeta_4 2/9 = 8/17$. With $b = 7$, we get $\Pr_{\tilde{O}}(7, 1) = 900/3451 \approx 0.261$, $\Pr_{\tilde{O}}(7, 2) = 800/3451 \approx 0.232$, $\Pr_{\tilde{O}}(7, 3) = 675/3451 \approx 0.196$, $\Pr_{\tilde{O}}(7, 4) = 576/3451 \approx 0.167$, and $\Pr_{\tilde{O}}(7, 5) = 500/3451 \approx 0.145$. Fig. 6 outlines in red the PMF corresponding to standard undecimal in a global setting. The information gap decays harmonically from the second quantum so that transiting from the greatest quanta is easier than from the minor ones. Indeed, only the first few quanta remain stable.

Developing similar reasoning in a local setting, using equations (6), (7), and (19) ($d_1 = d$ and $d_2 = d + 1$), the odds between consecutive digits

$$\tilde{o}(d+1 : d|r) = \frac{d}{d+1} \log_r \left(1 + \frac{1}{d} \right) \quad (22)$$

measure the associated likelihood gap in radix r . Then, we can calculate the PMF that normalizes these digit gaps in a given radix; the larger the digit, the lesser the information differential. For example, the PMF corresponding to standard quaternary is $\{0.561814, 0.438186\}$. With radix $r = 7$, we get $\text{Pr}_{\tilde{o}}(7, 1) \approx 29.8\%$, $\text{Pr}_{\tilde{o}}(7, 2) \approx 23.24\%$, $\text{Pr}_{\tilde{o}}(7, 3) \approx 18.55\%$, $\text{Pr}_{\tilde{o}}(7, 4) \approx 15.35\%$, and $\text{Pr}_{\tilde{o}}(7, 5) \approx 13.06\%$.

Figure 6 outlines in green the logarithmic PMF of standard undecimal, measuring the improbability of a random local jump through its contribution to the coding source's entropy. The lowest digits maintain discernibility from the environment, while the decreasing entropic support of the more significant digits makes them more vulnerable.

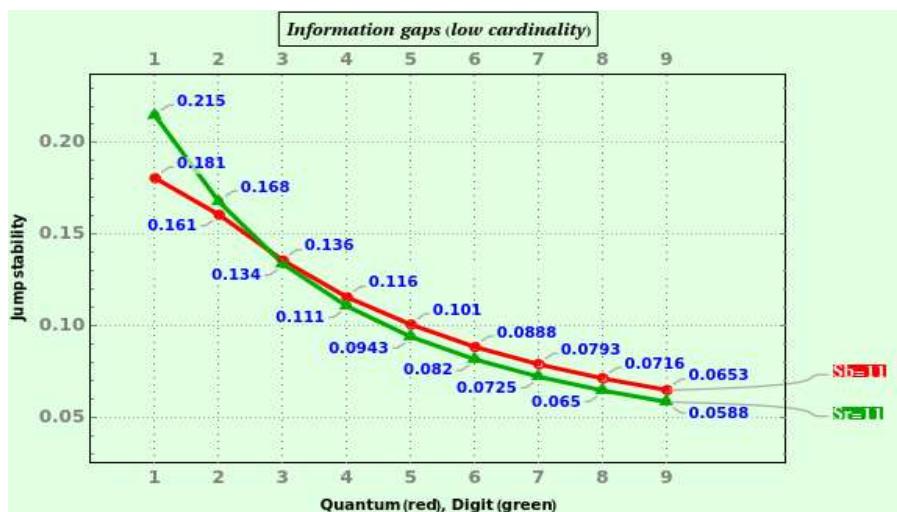


Figure 6. These are the information gaps the undecimal numeral system induces. Note that the fiducial NBL for decimal numerals is steeper than the digit plot (in green), and the digit plot is steeper than the quantum (in red).

Although the fiducial NBL is steeper than the corresponding $\text{Pr}_{\tilde{o}}$ regardless of the radix, and this is steeper than $\text{Pr}_{\tilde{o}}$ irrespective of the base, these three plots are hardly distinguishable for large cardinalities (see Figure 7), meaning that an NBL probability is synonym with stability. A transition from the greatest quanta or digits is generally much more frequent than a transition from the minor ones. This condition resembles the reactivity of the chemical elements periodic table concerning the electron shell (i.e., principal quantum number). More generally, ascending order (of numbers, quanta, digits, or shells) correlates with unsteadiness, which explains why closeness prevails over farness.

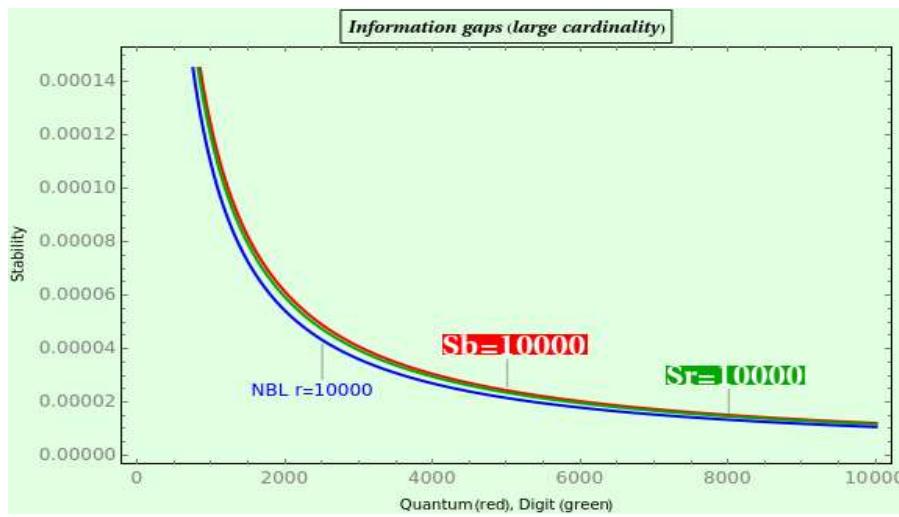


Figure 7. These are information gaps induced by the quanta of global base 10000 (in red) and the digits of local radix10000 (in green), compared with the fiducial NBL.

5.4. Optimal Stopping

A PN system assigns an information value to the concepts of likelihood, probability, and odds. In subsections 5.1 and 5.2, we argue that Bayes' rule is the entropic contribution of a bucket to a harmonic scale or a bin to a logarithmic scale. In particular, eq. (19) allows us to calculate the information we can extract from a bipartition by nailing the first and last domain digits. Assuming $1 \ll r < b$, the local odds of getting digit x against 1 and r against x estimate the information aggregate of the two parts. Inherent to the X 's dichotomy $\{[1, x), [x, r)\}$,

$$\begin{aligned}\ddot{o}_r(x) &= \ddot{o}(\{[1, x), [x, r)\}) \\ &= \ddot{o}(x : 1|r) + \ddot{o}(r : x|r) \\ &= \frac{1}{x} \log_r x + \frac{x}{r} \log_r \frac{r}{x}\end{aligned}$$

gives the bipartite odds in logarithmic r -ary units of information, where $1 < x < r$.

We obtain additive countability by making $\kappa \int_1^r \ddot{o}_r(x) dx = 1$. The entropy (local likelihood) distribution function

$$\kappa = \frac{4}{r^2 + 4(r \ln r - 1) \ln \sqrt{r} - 1}$$

$$\overleftarrow{o}_r(x) = \kappa \frac{r}{x} \ln x$$

$$\overrightarrow{o}_r(x) = \kappa x \ln \frac{r}{x}$$

$$\overleftrightarrow{o}_r(x) = \overleftarrow{o}_r(x) + \overrightarrow{o}_r(x) \quad (23)$$

gives the normalized bipartite odds so that $\overleftrightarrow{o}_r(x) = \kappa \ddot{o}_r(x)$ acquires a value between 0 and 1.

For $\overleftrightarrow{o}_r(1) = \overleftrightarrow{o}_r(r) = \kappa \ln r$, both $\overleftrightarrow{o}_r(1)$ and $\overleftrightarrow{o}_r(r)$ tend to vanish in the limit $r \rightarrow b \rightarrow \infty$. Where does (23) become stationary? When $r \geq 55$, the normalized bipartite odds produce two maxima corresponding to $\overleftarrow{o}_r(x)$ and $\overrightarrow{o}_r(x)$; as $r \rightarrow b \rightarrow \infty$, the first maximum tends to $x = e$ and the second to $x = r/e$. These maxima optimize the total information transmission of the system. We find at $x = \sqrt{r}$,

between the two maxima, a digit that minimizes the distinguishability between the two partitions, which is the analog of the middle point of a segment on the linear scale.

For example, with radix $r = 10000$, the bipartitions are maximally entropic about $\{[1, 2.7329], [2.7329, 10000]\}$ and $\{[1, 3659.1], [3659.1, 10000]\}$, and $\overleftrightarrow{o}_r(100)$ is the minimum. Figure 8 repeats this exercise and shows the results with $r = 100$ applicants.

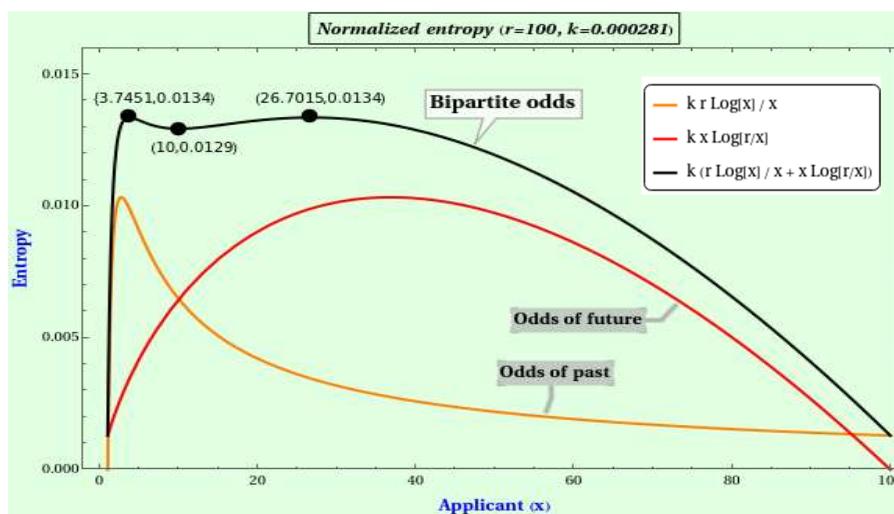


Figure 8. Plots of the odds that yield the past, future, and bipartite entropy (local likelihood distribution function) with radix 100 (99 applicants). The three points correspond to the maxima and the minimum, whose abscissas give place to the bipartitions with utmost information and tiniest bipartite distinguishability, respectively.

Supposing that $1/x$ and x/r are probabilities, Equation (23) is the addition of the corresponding entropies. Both maxima separate a stage of "retention" from a "decision" stage. Retention implies input processing, which raises entropy, whereas decision involves output processing, which lowers entropy. Maximum entropy indicates the best resource efficiency between the ascent and descent sections. Overall, the plot of $\overleftrightarrow{o}_r(x)$ reflects a natural entropic imbalance toward the small values; $\overleftrightarrow{o}_r(x)$ dominates in the short term, whereas $\overrightarrow{o}_r(x)$ dominates in the middle and long term. Computationally, it induces the bulk of processing far before reaching $x = r/2$, while physically, it implies a bias of space or time.

The bipartite odds function can have interesting consequences in computational physics, especially in sequential decision-making to solve optimal stopping (or planning) problems with solutions such as the odds algorithm [88]. Specifically, the secretary problem [89] is a mathematical trope to grasp how computation closely ties with incremental (Bayesian) inference, hence with the asymmetric management of fundamental resources. Shortly, one of $r - 1$ sequentially interviewed applicants must be nominated, with the proviso that they will be either chosen or rejected just after being examined; past the first $\lfloor x \rfloor$ applicants (typically a secretary, but also a lead actor or actress or a car), the judges select the next one that is better than any of the previous ones. Well, $x \approx r/e$ maximizes $\overrightarrow{o}_r(x)$, i.e., the probability of success in choosing "the best" applicant.

Instead, $\overleftrightarrow{o}_r(x)$ answers a different question. What is the optimal size $\lceil x \rceil$ of examined applicants to maximize the odds of choosing "a good" one? This nuance implies a crucial difference in approaching a solution; in this case, we must consider both terms of (23). We define "a good" prospect as a "candidate" in terms of the classic secretary problem, i.e., a seeker (or contender, or claimant) better than the previously examined applicants.

Considering that x is the current applicant, the bipartition separates the past from the future because $\overleftarrow{o}_r(x)$ and $\overrightarrow{o}_r(x)$ focus on the expected benefit before and after x and the practicality of preceding against succeeding data.

Regarding the "past" term, $\ln x$ is the amount of information ascribed to examined applicants, and $1/x$ is the probability of using such information. As $x \rightarrow 1$, we take advantage of less and less gathered information, albeit more likely, whereas if $x \rightarrow r$, we can leverage more and more references, albeit less likely. There is a compromise between choosing the first applicant (i.e., utterly uninformed decision-making) and selecting the last applicant (i.e., assuring to miss all the acquired information). We obtain the maximum of $\overleftarrow{o}_r(x)$ at $x \rightarrow e$ as $r \rightarrow b \rightarrow \infty$.

Regarding the "future" term, $\ln r/x$ is the information we can obtain from forthcoming applicants, and x/r is the probability of using such information. As $x \rightarrow 1$, we will surely miss the most suitable prospects; if $x \rightarrow r$, we will hardly find a suitable applicant. There is a compromise between choosing the first applicant (i.e., ignoring the information the remaining applicants can provide) and selecting the last applicant (i.e., information exhausted). We obtain the maximum of $\overrightarrow{o}_r(x)$ at $x \rightarrow r/e$ as $r \rightarrow b \rightarrow \infty$.

Summing both terms implies balancing the partition behind against the partition ahead. If x is too low ($x \gtrsim 1$), you have the most information ahead for an acceptable selection, and if x is too high ($x \lesssim r$), you have many references for a good choice. Unfortunately, if x is too low ($x \gtrsim 1$), you have less probability of making an acceptable selection, and if x is too high ($x \lesssim r$), you have probably missed the finest choices. While bipartition $\{[1, r/e], [r/e, r)\}$ implies a probability of $1/e$ % of skipping and selecting the best alternative, bipartition $\{[1, e], [e, r)\}$ reduces this percentage significantly. Thus, $\overleftarrow{o}_r(x)$ enables promptness and $\overrightarrow{o}_r(x)$ quality.

The entropy distribution function of a bipartition rises to the first maximum, falls and rises again to reach the second maximum, and decays until it almost vanishes. The right holistic strategy is to wait for the information to stop rising so that $\partial^{\overleftarrow{\overrightarrow{o}}_r(x)}/\partial x$ vanishes and $\partial^2 \overleftarrow{\overrightarrow{o}}_r(x)/\partial x^2$ decreases, i.e., in agreement with the maximum entropy principle for isolated systems (and the minimum energy principle for closed systems) in thermodynamics.

Exclusively concentrating on the past term also makes sense. The idea is to assess the general level after examining only a few applicants. Assuming that ours behaves as a linear time-invariant system, deviations decay as e^{-x} , so the probability that the mean of the three first interviewed applicants is close to the pool mean is $1 - e^{-3} \approx 95\%$. Since a threesome reasonably represents the whole set of applicants, we can confidently pick a forthcoming candidate. $\overleftarrow{o}_r(x)$ considers the cost of the processing; it is a precursor of human intuition and opens the door to computational methods of solution refinement. For instance, assuming that we can retain a (preliminary) solution, we can progressively renew candidates between the two maxima. If the selection process continues after the second maximum, we are in the same scenario as the classic secretary problem.

Deciding near $x = \sqrt{r}$, between the maxima, is questionable because having already spent substantial resources on getting information, the probability of picking the best applicant still needs to reach the optimum. Nonetheless, it is a separator of the two partitions that a living being, for instance, can seek on purpose to maximize internal order or coherence.

5.5. Bayesian Recurrence

PN coding is fundamentally relational. It compares pairs of buckets by figuring out probability double ratios. Using (15), we define

$$\tilde{O}(j : i | h : g) = \frac{\tilde{O}(j : i | b)}{\tilde{O}(h : g | b)} = \frac{\left(\frac{i}{j}\right)^2 \Pr(b, [i, j])}{\left(\frac{g}{h}\right)^2 \Pr(b, [g, h])} = \left(\frac{hi}{gj}\right)^2 \left(\frac{H_{j-1} - H_{i-1}}{H_{h-1} - H_{g-1}} \right) \in \mathbb{Q}$$

where $(1 \leq g < h) \wedge (1 \leq i < j)$. The first factor is the prior (decoded) ratio $\tilde{O}(hi : gj|r = 1)$, and the second is the likelihood ratio measuring the strength of the correlation between the information that the buckets $[i, j)$ and $[g, h)$ keep. Note that the formula does not refer to the global base anymore.

This structure is repetitive; a generic probability quadruple ratio is the rational

$$\begin{aligned} \frac{\tilde{O}(j : i|h : g)}{\tilde{O}(n : m|l : k)} &= \tilde{O}(j : i|h : g|n : m|l : k) = \\ &= \left(\frac{\frac{hi}{gj}}{\frac{lm}{kn}} \right)^2 \left(\frac{(H_{l-1} - H_{k-1})(H_{j-1} - H_{i-1})}{(H_{n-1} - H_{m-1})(H_{h-1} - H_{g-1})} \right) = \\ &= \left(\frac{hikn}{gjlm} \right)^2 \left(\frac{(H_{j-1}H_{l-1} + H_{i-1}H_{k-1}) - (H_{j-1}H_{k-1} + H_{i-1}H_{l-1})}{(H_{g-1}H_{m-1} + H_{h-1}H_{n-1}) - (H_{g-1}H_{n-1} + H_{h-1}H_{m-1})} \right) \end{aligned}$$

Thus, we can formulate a probability ratio of order 2^n as coded odds, i.e., the product of a rational squared and a likelihood ratio between two rational differences. Moreover, the product of probability ratios also fits Bayes' coding pattern because of

$$\begin{aligned} \tilde{O}(j : i|h : g)\tilde{O}(n : m|l : k) &= \tilde{O}(j : i|h : g|l : k|n : m) = \\ &= \left(\frac{hilm}{gjkn} \right)^2 \left(\frac{(H_{j-1}H_{n-1} + H_{i-1}H_{m-1}) - (H_{j-1}H_{m-1} + H_{i-1}H_{n-1})}{(H_{g-1}H_{k-1} + H_{h-1}H_{l-1}) - (H_{g-1}H_{l-1} + H_{h-1}H_{k-1})} \right) \end{aligned}$$

It is paramount to highlight that PN coding also copes with probability double ratios in a local setting, with no extra apparatus. A coding source uses the rule

$$\tilde{o}(j : i|h : g) = \frac{\tilde{o}(j : i|r)}{\tilde{o}(h : g|r)} = \frac{\frac{i}{j} \log_r \frac{i}{j}}{\frac{g}{h} \log_r \frac{h}{g}} = \left(\frac{hi}{gj} \right) \left(\frac{\ln j - \ln i}{\ln h - \ln g} \right) \in \mathbb{Q}$$

where $(1 \leq g < h) \wedge (1 \leq i < j)$, to compare a pair of odds. The first factor is the prior (locally decoded) ratio $\tilde{o}(hi : gj|r = 1)$, and the second is the likelihood ratio (Bayes factor), which measures the strength of the informational correlation between bins $[g, h)$ and $[i, j)$ and obliterates the local radix.

In the particular case where the bins have a joint event, either initial or final, we obtain

$$\tilde{o}(j : g|h : g) = \left(\frac{h}{j} \right) \left(\frac{\ln j - \ln g}{\ln h - \ln g} \right) \in \mathbb{Q}$$

If the joint event is the unit, we obtain the "subjective ratio"

$$(j; h) = \tilde{o}(j : 1|h : 1) = \left(\frac{h}{j} \right) \left(\frac{\ln j}{\ln h} \right) = \frac{\ln j^h}{\ln h^j} \in \mathbb{Q} \quad (24)$$

where $1 < h < j$; e.g., $(10^2; 10) = 1/5$, $(10^{10}; 10^9) = 1/9$, $(10^3; 10) = 3/100$, $(10^4; 10) = 1/250$, $(10^5; 10) = 1/2000$, and $(10^8; 10) = 1/1250000$. It is the relative mutual likelihood between a pair of numerals regarded from the source, satisfying $(j; h) = 1/(h; j)$; for example, $(10^5; 10^4) = 1/8$ and $(10^4; 10^5) = 8$.

The structure of a Bayesian datum $(A/D)(B-C/E-F)$ is locally repetitive, like in global coding. The original posterior odds $\tilde{o}(j : i|r) = \frac{i}{j} \log_r \frac{i}{j}$ imply $A/D = i/j$, $B = \ln j$, $C = \ln i$, $E = \ln r$, and $F = \ln 1$, the original prior odds $\frac{i}{j}$ imply $A/D = i/j$ and $B - C = E - F$, and the odds arithmetic always yields the same format; an essential operation "*" invariably results in the product of a probability ratio (the prior factor) and a likelihood ratio of differences (the Bayes factor). Let

$$\left(\frac{G}{K} \right) \left(\frac{H-J}{L-M} \right) * \left(\frac{N}{R} \right) \left(\frac{P-Q}{S-T} \right) = \left(\frac{A}{D} \right) \left(\frac{B-C}{E-F} \right)$$

then,

$$* \equiv " + " \Rightarrow \begin{cases} A/D = 1/KR \\ B = GHRS + GJRT + KLN + KMNQ \\ C = GHRT + GJRS + KLNQ + KMNP \\ E = LS + MT \\ F = MS + LT \end{cases}$$

$$* \equiv " - " \Rightarrow \begin{cases} A/D = 1/KR \\ B = GHRS + GJRT + KMNP + KLNQ \\ C = GJRS + GHRT + KLN + KMNQ \\ E = LS + MT \\ F = MS + LT \end{cases}$$

$$* \equiv " \times " \Rightarrow \begin{cases} A/D = GN/KR \\ B = HP + JQ \\ C = JP + HQ \\ E = LS + MT \\ F = MS + LT \end{cases}$$

$$* \equiv " \div " \Rightarrow \begin{cases} A/D = GR/KN \\ B = HS + JT \\ C = JS + HT \\ E = LP + MQ \\ F = MP + LQ \end{cases}$$

For instance, the arithmetic of a probability quadruple ratio

$$\tilde{o}(n : p | r : s) * \tilde{o}(v : w | y : z) = \left(\frac{A}{D} \right) \left(\frac{B - C}{E - F} \right)$$

determines

$$* \equiv " + " \Rightarrow \begin{cases} A/D = 1/nsvz \\ B = prvz (\ln n \ln y + \ln p \ln z) + nswy (\ln r \ln v + \ln s \ln w) \\ C = prvz (\ln n \ln z + \ln p \ln y) + nswy (\ln r \ln w + \ln s \ln v) \\ E = \ln r \ln y + \ln s \ln z \\ F = \ln r \ln z + \ln s \ln y \end{cases}$$

$$* \equiv " - " \Rightarrow \begin{cases} A/D = 1/nsvz \\ B = prvz (\ln n \ln y + \ln p \ln z) + nswy (\ln r \ln w + \ln s \ln v) \\ C = prvz (\ln n \ln z + \ln p \ln y) + nswy (\ln r \ln v + \ln s \ln w) \\ E = \ln r \ln y + \ln s \ln z \\ F = \ln r \ln z + \ln s \ln y \end{cases}$$

$$* \equiv " \times " \Rightarrow \begin{cases} A/D = prwy/nsvz \\ B = \ln n \ln v + \ln p \ln w \\ C = \ln n \ln w + \ln p \ln v \\ E = \ln r \ln y + \ln s \ln z \\ F = \ln r \ln z + \ln s \ln y \end{cases}$$

$$* \equiv " \div " \Rightarrow \begin{cases} A/D = pryz/nswy \\ B = \ln n \ln y + \ln p \ln z \\ C = \ln n \ln z + \ln p \ln y \\ E = \ln r \ln v + \ln s \ln w \\ F = \ln r \ln w + \ln s \ln v \end{cases}$$

As an example, we can express the original local odds (19) as an arithmetic combination of quadruple ratios, namely

$$\tilde{o}(j:i|r) = \frac{\tilde{o}(j:i|r:1)}{\tilde{o}(r:1|r^2:1) + \tilde{o}(r:1|r^2:1)} = \frac{\frac{r_i}{j} \log_r \frac{j}{i}}{2 \frac{r^2}{r} \frac{\ln r}{\ln r^2}} = \frac{i}{j} \log_r \frac{j}{i}$$

This property reinforces a PN system's recoding process, which takes advantage of the most recent information. Note that one thing is that Bayesian odds boost incremental computing, and a horse of another color is that the structure of Bayesian data is recurrent under arithmetic operations. By exploiting the same representational pattern for all its calculation methods, a coding source can "accumulate experience".

5.6. Referential Ratio and Cross-Ratio

The probability quadruple ratio

$$\frac{\tilde{O}(j:i|j:g)}{\tilde{O}(n:i|n:g)} = \left(\frac{\frac{ji}{gj}}{\frac{ni}{gn}} \right)^2 \left(\frac{(H_{n-1} - H_{g-1})(H_{j-1} - H_{i-1})}{(H_{n-1} - H_{i-1})(H_{j-1} - H_{g-1})} \right)$$

is of foremost interest. Since the prior is the unit, it is a sheer likelihood ratio, i.e., a genuine proportion of information we can rewrite as

$$\frac{\tilde{O}(j:i|j:g)}{\tilde{O}(n:i|n:g)} = \frac{\frac{H_{j-1} - H_{i-1}}{H_{j-1} - H_{g-1}}}{\frac{H_{n-1} - H_{i-1}}{H_{n-1} - H_{g-1}}} \quad (25)$$

When $n \rightarrow b \rightarrow \infty$, the denominator drops from the formula. Using the change of variables $A = H_{j-1}$, $B = H_{i-1}$, and $C = H_{g-1}$, this singular Bayes factor tends to the "referential ratio"

$$(A; B : C) \equiv \frac{A - B}{A - C} \quad (26)$$

which is a ratio between the likelihood of two buckets with a joint referent.

Locally, the probability quadruple ratio

$$\frac{\tilde{o}(r^A : r^C | r^A : r^B)}{\tilde{o}(r^D : r^C | r^D : r^B)} = \frac{\tilde{o}(e^A : e^C | e^A : e^B)}{\tilde{o}(e^D : e^C | e^D : e^B)}$$

also leads to (26) when $r^D \rightarrow b \rightarrow \infty$, e.g.,

$$\frac{\ln e^3/e^5}{\ln e^3/e^7} = (3; 5 : 7) = \frac{1}{2}$$

and

$$\frac{\ln 2.5163/4.5091}{\ln 2.5163/6.5065} \approx (0.9228; 1.5061 : 1.8728) \approx \frac{35}{57} = \frac{H_2 - H_4}{H_2 - H_6}$$

Note that the product, quotient, sum, and difference of two referential ratios are referential ratios. In other words, the set of referential ratios, which we will represent by $\tilde{\mathbb{Q}}$, is an ordered algebraic field (of characteristic zero) where

$$(F; G : H) + (J; K : L) = \frac{(FL + HJ - FJ) - (FJ + GL + HK - FK - GJ)}{(FL + HJ - FJ) - (HL)}$$

$$(F; G : H) - (J; K : L) = \frac{(FL) - (FK + GL + HJ - GJ - HK)}{(FL) - (FJ + HL - HJ)}$$

$$(F; G : H) \times (J; K : L) = \frac{(FJ) - (FK + GJ - GK)}{(FJ) - (FL + HJ - HL)}$$

$$(F; G : H) \div (J; K : L) = \frac{(FJ) - (FL + GJ - GL)}{(FJ) - (FK + HJ - HK)}$$

We can represent a referential ratio $(A; B : C)$ as the point $(A - B, A - C, A)$ on a three-dimensional grid where the x component is the numerator, the y component is the denominator, and the z component is the reference. Figure 9 displays the referential ratios $(F; G : H) = (2; 3 : 1)$ and $(J; K : L) = (4; 1 : 5)$ and the result of the basic operations between them, i.e., addition, subtraction, multiplication, and division.

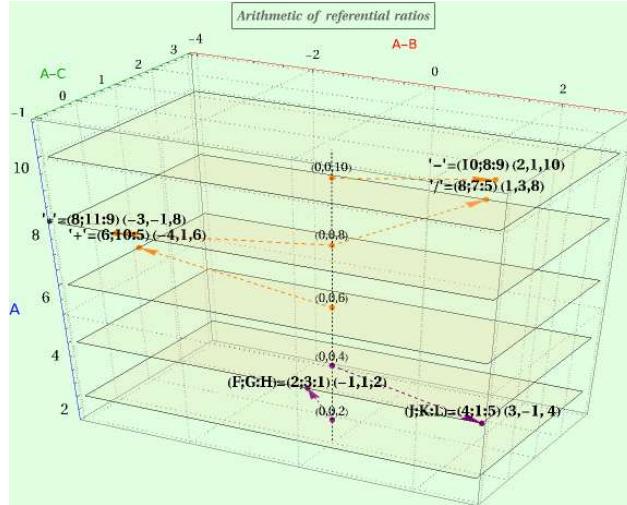


Figure 9. Example of how to represent a referential ratio and its arithmetic.

We can formally define referential ratios as equivalence classes (symbol " \sim ") of integer triplets where $(A; B : C) \sim (A; D : E) \iff BE = CD$. Mind that $\mathbb{Q} \subset \tilde{\mathbb{Q}}$ due to

$$\frac{N}{D} = (; N : D)$$

This view lets us comprehend a referential ratio as a dislocated or transported rational number; (26) is B/C observed from A . So, $\tilde{\mathbb{Q}}$ is an extension field of \mathbb{Q} such that the latter's operations are those of the former by restricting the referent to the origin. Specifically, the multiplicative units of the field are $(\pm 1 : 1)$. Besides, $(A; B : B) = (1; :) = 1$, $(A; B : A - 1) = A - B$, $(A; B : C) = (-A; -B : -C)$, and $(A; B : C) = (A; C : B)^{-1}$.

Since a detailed algebraic analysis of this field, its meaning, representation, and potential applications would need a specific article, we will focus herein only on a couple of manifestations.

In physics, we must generically understand the concept of correlation as a ratio between magnitudes of the same physical unit. The most straightforward embodiment of the referential ratio gives the Doppler effect's relationship between the frequency perceived by the receiver f_r and emitted frequency f_s , i.e.,

$$\frac{f_r}{f_s} = (s_w; s_r : -s_s)$$

where s_w is the propagation speed of waves in the medium, $s_r < s_w$ is the speed of the receiver relative to the medium, and $s_s < s_w$ is the speed of the source relative to the medium, assuming that they are getting away from each other [90]. Likewise, the formula of the relativistic Doppler effect of the source's frequency relative to the receiver's frequency moving away at speed v is [91]

$$\left(\frac{f_s}{f_r}\right)^2 = (1; (-v : c) : (v : c)) = (1; -v/c : v/c) = \frac{1 + v/c}{1 - v/c} \quad (27)$$

A referential ratio also appears subsumed into the "cross-ratio" of four distinct points [92]

$$(A, B; C, D) = \frac{\frac{A-C}{A-D}}{\frac{B-C}{B-D}} = \frac{(A; C : D)}{(B; C : D)} = (AB + CD; AD + BC : AC + BD) \quad (28)$$

where the alphabetical order indicates that A , B , C , and D are consecutive on the rational projective line, and $A - B$ and $C - D$ have the same sign. This likelihood ratio is the central tool that characterizes the projective line's geometry. The cross-ratio calculates how much the quadruple's crossing symmetries deviate from the ideal proportion 1, precisely the extent to which the ratio of how C divides $[A, B]$ is proportional to how D divides $[A, B]$. For example, the substitutions $\{A = H_{j-1}, B = H_{n-1}, C = H_{i-1}, D = H_{g-1}\}$ produce the cross-ratio defined by (25).

6. Conformality

Departing from an inverse-square PMF for the naturals, we gleaned the global and local NBL, implying that a double scale is necessary to support a universal place-value system. A global base specifies the harmonic scale, while a local radix fixes the logarithmic scale that a coding source uses to represent numerals in PN.

In the previous section, we managed odds as probability ratios, i.e., information double ratios. Bayes-compliant information 2^n -ratios are closed under division. An exceptional case of these is the referential ratio. A ratio of referential ratios is precisely a cross-ratio whose logarithm determines the conformal metric of a local coding space.

Conformal maps preserve angles, hence the shape of the figures, which also implies scale invariance. These properties are critical to translating the elements of a global harmonic space into a local logarithmic subspace, the latter reflecting the state of the former. Conformality is a requirement for coding information that drives complexity; "there is a shared very particular characteristic of all complex systems. And that is they internally encode the world in which they live" [93].

6.1. the Conformal 1-Annulus Model

The cross-ratio paves the way to conformality because it is invariant under linear fractional transformations over rings [94]. Noteworthy, the group of linear fractional transformations $(Az+B)/(Cz+D)$, where $\{A, B, C, D\} \in \mathbb{Z}$, called the "modular group" (a subset of the M bius group), acts transitively on the points of the grid \mathbb{Z}^2 visible from the origin, i.e., the irreducible fractions [95], so preserving the form of polygonal shapes through the cross-ratio. A modular map also conserves the referential ratio given by (28).

Because the harmonic and logarithmic scales handle the concept of cross-ratio, we can find a modular transformation between four specific points in a global space S and four points in a given S 's subspace, the coding space where the source makes a local model.

The most powerful application of the cross-ratio is the Poincar disk (The Non-Euclidean World in [96]), a conformal model of hyperbolic geometry that projects the whole \mathbb{H}^2 in the unit disk. Circle-preserving M bius transformations are the isometries of the complex plane. Assuming that the disk center is at the plane's origin, points z_2 and z_3 within the disk connected by the arc of a geodesic circle perpendicularly intersecting the disk's boundary at z_1 and z_4 are at a hyperbolic distance of $\ln(z_1, z_2; z_3, z_4)$ [97]. This measure is invariant under the subset of M bius maps acting transitively on the unit disk, the space of the coding source.

In one dimension, the complex plane augmented by the point at infinity can be considered the real projective line [98], and the disk becomes the unit 1-ball. More specifically, the set of irreducible fractions augmented by the point at infinity is the rational projective line; hence the unit 1-ball becomes the rational open unit interval.

While $(-1, 1)$ is the mathematical domain where the modular group acts, we are interested in the global computational space where Bayesian processes and transformative calculation methods occur. We assume that global Bayesian data, i.e., rational quanta, populate a cosmos of information a source perceives and codes to create a continuous world model. Outside a coding source, the information resides on a harmonic scale, whereas inside, a logarithmic scale lodges local Bayesian data.

Suppose that an object is at position P outside $(-1, 1)$. We are ignorant of the actual computation of P , but we know that it is a rational number resulting from applying the rule (15). Be that as it may, we can use Equation (19) to locally figure the odds of $P - 1$ against $P + 1$ in radix r , whose Bayes factor is the logarithm in r -ary units of a cross-ratio where $z_1 = -1$, $z_2 = 1$, $z_3 = P$, and $z_4 = \infty$, i.e.,

$$\log_r(-1, 1; P, \infty) \equiv \log_r(P; -1 : 1) = \log_r \frac{P+1}{P-1}$$

In information theory, this expression is the representational length in radix r of the rational number $P+1/P-1$ and, according to NBL (8), the r -normalized width of bin $[P - 1, P + 1]$. We can unite these outlooks by interpreting this Bayes factor as the hyperbolic distance from P to $b \rightarrow \infty$, i.e.,

$$d_r(P, \infty) \equiv \log_r(-1, 1; P, \infty) = \frac{2}{\ln r} \operatorname{arcoth}(P)$$

where arcoth is the inverse function of the hyperbolic cotangent.

A neat inversion conformally maps the outside of the coding source to its inside,

$$(-1, 1; P, \infty) \xrightarrow{z \mapsto 1/z} (-1, 1; D, 0) \quad (29)$$

conserving the cross-ratio. (Other inversions $z \mapsto \alpha^{-1}/z$ also serve but violate the minimal information principle.) For example, if $P = 2$, $(-1, 1; P, \infty) = \frac{-1-2}{1-2} = 3 = \frac{1+\frac{1}{2}}{1-\frac{1}{2}} = (-1, 1; D, 0)$. Therefore, the r -normalized hyperbolic distance between the origin and $|D| = |1/P| < 1$ is

$$d_r(0, D) \equiv \log_r(-1, 1; D, 0) \equiv \log_r(0, D; 1, -1) \equiv$$

$$\equiv \log_r \frac{1+D}{1-D} \equiv \log_r (1; -D : D) \equiv \frac{2}{\ln r} \operatorname{artanh}(D)$$

where artanh is the inverse function of the hyperbolic tangent, and $\ln r$ is the (constant) curvature's absolute value of the source's coding space.

The inverse function of $d_r(0, 1/P)$ is $P = \coth((\frac{1}{2} \ln r) d_r(0, 1/P))$. For example, an object at a Euclidean distance $P = 10^6$ from the origin is at a natural hyperbolic distance of $d_e(P, \infty) = 2 \operatorname{arcoth}(10^6) \approx 2 \times 10^{-6}$ from $b \rightarrow \infty$. The coding source positions the object at $D = 1/P = 10^{-6}$, at a natural hyperbolic distance of $d_e(0, D) = 2 \operatorname{artanh}(10^{-6}) = 2 \operatorname{arcoth}(10^6)$ from the origin, and decodes it as $P = \coth \operatorname{artanh}(10^{-6}) = 10^6$.

Mind that the local coding space is the 1-annulus

$$\mathring{\mathbb{a}}^- \equiv \{D \in \mathbb{Q} \mid -1 < D < -1/b\}$$

$$\mathring{\mathbb{a}}^+ \equiv \{D \in \mathbb{Q} \mid 1/b < D < 1\}$$

$$\mathring{\mathbb{a}} \equiv \mathring{\mathbb{a}}^- \cup \mathring{\mathbb{a}}^+ \equiv \{D \in \mathbb{Q} \mid 1/b < |D| < 1\}$$

reflecting what the source observes in the 1-annulus

$$\mathring{\mathbb{A}}^- \equiv \{P \in \mathbb{Q} \mid -b < P \leq -1\}$$

$$\mathring{\mathbb{A}}^+ \equiv \{P \in \mathbb{Q} \mid 1 \leq P < b\}$$

$$\mathring{\mathbb{A}} \equiv \mathring{\mathbb{A}}^- \cup \mathring{\mathbb{A}}^+ \equiv \{P \in \mathbb{Q} \mid 1 \leq |P| < b\} \quad (30)$$

For instance, if $D \in \mathring{\mathbb{a}}^+$, $d_r(0, D)$ vanishes if $P \rightarrow b \rightarrow \infty$, is > 1 if P is at a Euclidean distance closer than $\coth \ln r / 2$ from the origin, and diverges if $P \rightarrow 1^+$.

6.2. The Conformal 1-Ball Model

The r -normalized hyperbolic distance between two points A and B in $\mathring{\mathbb{a}}$ is $d_r(0, Q_B) - d_r(0, Q_A)$, i.e.,

$$d_r(Q_A, Q_B) = \frac{2}{\ln r} (\operatorname{artanh}(Q_B) - \operatorname{artanh}(Q_A))$$

where Q_A and Q_B result from the conformal transformation (29), i.e., $P \mapsto 1/Q$, which mirrors the external world concerning $\mathring{\mathbb{a}}$'s outward boundary, to wit ± 1 . Thus,

$$d_r(Q) = d_r(0, Q) = \frac{2}{\ln r} \operatorname{artanh}(Q) \quad (31)$$

reflects how far an object at Q is from infinity, situated at the origin.

Nonetheless, we want the origins of the coding source and $\mathring{\mathbb{A}}$ to coincide and ± 1 to be the infinite points of the local model. This requirement implies calculating Q 's complement to one, a logical negation that varies on the left and the right. Recall that all negations are derivations of the canonical one [99], so we will use $Q \rightarrow -1 - Q$ on the left and $Q \rightarrow 1 - Q$ on the right to satisfy the minimal information principle. The coding space is now the open 1-ball

$$\mathring{\mathbb{C}} \equiv \{Q \in \mathbb{Q} \mid |Q| < 1 - 1/b\} \quad (32)$$

reflecting what the source observes in the 1-annulus $\mathring{\mathbb{A}}$.

An object at $P \in \mathbb{A}^+$ will be in \mathbb{B}^+ at a Euclidean distance of

$$Q = (P; 1 :) = 1 - \frac{1}{P}$$

from the origin and hence at a hyperbolic distance of

$$d_r^+(Q) = \frac{2}{\ln r} \operatorname{artanh}(Q) = \log_r(2P - 1) = \log_r(P; 1 - P : P - 1) \quad (33)$$

with inverse (decoding) function $P = \frac{1}{2} \left(1 + r^{d_r^+(Q)} \right)$.

Similarly, an object at $P \in \mathbb{A}^-$ will be in \mathbb{B}^- at a Euclidean distance of

$$Q = (; P + 1 : -P) = -1 - \frac{1}{P}$$

from the origin and hence at a hyperbolic distance of

$$d_r^-(Q) = \frac{2}{\ln r} \operatorname{artanh}(Q) = -\log_r(-(2P + 1)) = \log_r(P; P + 1 : -P - 1) \quad (34)$$

with inverse (decoding) function $P = -\frac{1}{2} \left(1 + r^{-d_r^-(Q)} \right)$.

For example, a coding source places an object observed at a Euclidean distance of $P = \pm 10^6$ at $Q = \pm 0.999999$ in \mathbb{C} at a natural hyperbolic distance $d_e(Q) = 2 \operatorname{artanh}(\pm 0.999999) = \pm \ln(\pm 2 \times 10^6 - 1) = \pm 14.50866$ from the origin, and $P = \pm \frac{1}{2} (1 + e^{14.50866}) = \pm 10^6$. On the positive side, the odds are $\delta(2P - 1 : 1|e) = \frac{\ln(2P-1)}{2P-1} = \frac{\ln 1999999}{1999999} \approx 7.254332 \times 10^{-6}$. Suppose that, later, the coding source calculates this value as 7.254265×10^{-6} , meaning that either the object has moved to $P \approx \pm 1000010$ (because $\delta(2P - 1 : 1|e) = \frac{\ln 2000019}{2000019} \approx 7.254265 \times 10^{-6}$) or the radix has changed to $r \approx 2.71831$ (because $\delta(2 \times 10^6 - 1 : 1|r) = \frac{\ln 1999999}{1999999 \ln 2.718307} \approx 7.254265 \times 10^{-6}$). Even a combination of these two cases could produce the same odds value.

Note that functions in the form $|\theta_1| \exp(|\theta_2 Q|) \operatorname{artanh}^{\theta_3}(|\theta_4| Q)$, where $\{\theta_1, \theta_2, \theta_4 \in \mathbb{Q}\}$ and θ_3 is an odd power, e.g., $2 \exp(3Q) \operatorname{artanh}^5(7Q)$, also give rise to an odd hyperbolic distance that complies with boundary conditions $d_r(\pm 1) = \pm \infty$ and a vanishing distance when Q vanishes ($|P| \rightarrow 1$). However, they would introduce new factors and parameters we cannot explain; $\operatorname{artanh}(Q)$ is the only conformal function that retains the origin and conforms with the minimal information principle. Besides, it agrees with the canonical PMF (the first power is the most probable) and satisfies the additional condition of having a non-vanishing derivative at the origin, i.e., the origin is not stationary so that the function can keep its increasing tendency from left to right.

The effect of the radix on (31) is to adjust the point of maximum curvature. If $r \rightarrow 1^+$ (physically, at very low energies), such a point's curvature vanishes as $Q \rightarrow \pm \frac{1}{2} \pm$; if $r \rightarrow b \rightarrow \infty$ (physically, at very high energies), it diverges as $Q \rightarrow \pm 1^\mp$. As we mentioned in Section 3.6, these extreme values convey no-coding cases. When $r = e = 2.718 \dots$, i.e., when \mathbb{C} 's curvature radius is $1/\ln e = 1$, the coding source is maximally efficient, and the maximum curvature approximately corresponds to a hyperbolic distance of $d_e(0.72543) = 2 \operatorname{artanh}(0.72543) = 1.83803$.

6.3. Conformal Relativity

We must take the hyperbolic distance (31) as an abstract concept that does not have to be a physical length.

Imagine that the global base b physically represents the speed of light. On the right hand, $P = (b; : v) = b/(b-v) \equiv 1/(1-v/c)$ produces $Q = 1 - 1/P = v/b \equiv v/c$; if an object's speed is $v \rightarrow b \equiv c$, then $P \rightarrow \infty^+$ and $Q \rightarrow 1^-$, and if $v \rightarrow 0^+$, we get $P \rightarrow 1^+$ and $Q \rightarrow 0^+$. On the left hand, $P = (; b : - (b+v)) = -b/(b+v) \equiv -1/(1+v/c)$ produces $Q = -1 - 1/P = v/b \equiv v/c$; if an object's speed is $v \rightarrow -b \equiv -c$, then $P \rightarrow \infty^-$ and $Q \rightarrow -1^+$, and if $v \rightarrow 0^-$, we get $P \rightarrow -1^-$ and $Q \rightarrow 0^-$. In either case, $Q = v/b$, and we can write the relativistic Doppler effect (27) in the form

$$\frac{f_s}{f_r} = e^{\text{artanh}(v/b)} = e^{\text{artanh}(Q)}$$

Since the rapidity corresponding to velocity v is, by definition,

$$v \equiv \text{artanh}(v/b) \quad (35)$$

the Special Relativity's Lorentz factor is $\gamma \equiv \cosh v$ [100]. Thus, the coding source resolves the composition of Doppler shifts as the exponential of the addition of hyperbolic distances, i.e.,

$$\frac{f_s}{f_m} \frac{f_m}{f_r} = e^{\nu_{ms} + \nu_{rm}}$$

where f_m is the frequency perceived by the first receiver, ν_{ms} the rapidity of the first receiver relative to the source, and ν_{rm} the rapidity of the second receiver relative to the first one.

The special relativity theory is only conformal in terms of rapidity. Visualize two inertial frames, A and B , cruising at relativistic speed ratios of $P_A = 3$ and $P_B = 18$ about the origin of the coding source. These correspond in \mathbb{C} at ratios $Q_A = 1 - 1/3 = 2/3$ and $Q_B = 1 - 1/18 = 17/18$ of the speed v to $b \equiv c$, defining rapidities $\nu_A = \text{artanh}(2/3)$ and $\nu_B = \text{artanh}(17/18)$, and encoded in ternary as hyperbolic (relativistic) speeds $d_3(Q_A) = 2^{\nu_A}/\ln 3 = 1.465$ and $d_3(Q_B) = 2^{\nu_B}/\ln 3 = 3.2362$. The difference in hyperbolic speeds is linear in \mathbb{B} , i.e., $d_3(Q_A, Q_B) = d_3(Q_B) - d_3(Q_A) = 1.7712$. Within \mathbb{A} (30), the difference in (Euclidean) velocities is $18 - 3 = 15$, but the difference in hyperbolic speeds is

$$\frac{1}{2} \left(1 + 3^{\frac{2}{\ln 3}(\nu_B - \nu_A)} \right) = \frac{1}{2} \left(1 + 3^{1.7712} \right) = 4$$

Rapidity arithmetic is more straightforward than calculating Einstein's subtraction formula of (Euclidean) velocities, which calculates $\nu_B - \nu_A$ as $\text{artanh} \left(\frac{1 - (17/18)(2/3)}{(17/18) - (2/3)} \right)$.

Another way to obtain the same result is directly using the cross-ratio, i.e.,

$$\frac{1}{2} \left(1 + 3^{\log_3 \frac{\frac{2/3-1}{2/3+1}}{\frac{17/18-1}{17/18+1}}} \right) = \frac{1}{2} \left(1 + 3^{\log_3 7} \right) = 4$$

These results mean the weave of Lorentz invariance, and more generally, Poincar invariance, is the algebraic field of referential ratios. Lorentz symmetry [101] locally preserves central reflections and boosts, the latter maintaining constant the speed of light (the global base) when transforming to a reference frame with a different velocity. Poincar symmetry, the entire symmetry group of any relativistic field theory, additionally preserves the laws of physics for inertial coding sources situated at different quantum positions.

6.4. Conformal Coding and Computability

artanh has a protagonist role in a conformal space not only due to its manifestations in physics, mainly the metric (31), but also because its Taylor series allows calculating iteratively the natural logarithm itself based on the odd powers of the referential ratio $\frac{x-1}{x+1}$ ([102], 4.1.27), i.e.,

$$\ln x = 2 \sum_{i=1}^{\infty} \frac{(x; 1 : -1)^{2i-1}}{2i-1} \quad (36)$$

It is valid for any $x \in \mathbb{R}^+$, especially when $x \approx 1$. For example, let us calculate the ternary logarithm of $P = 10^6$. Since $10^6 = 1212210202001_3$, a numeral with 13 digits, its logarithm's characteristic is 12. Then, the coding source calculates the logarithm's mantissa from (36), where $x = P - 3^{12} = 1.88168$; after five iterations, the mantissa's error is less than one millionth. So, we calculate $\log_3 P = 12 + \frac{2}{\ln 3} \sum_{i=1}^5 \frac{(1.88168; 1 : -1)^{2i-1}}{2i-1} = 12.57541925$ against the real value of 12.57541965.

Because the coding source can calculate the logarithm in its coding space (32), \mathbb{B} 's curvature $-\ln r$ is a built-in value. Likewise, \mathbb{B} 's Euler characteristic $\chi = \int_0^1 \ln x \, dx = -1$, a topological invariant [103] corresponding to no vertices, one edge, and no faces, is a built-in value. Moreover, the coding source is "aware" of the PMF (1) through the digamma function $\psi(x) = \Gamma'(x)/\Gamma(x)$ (see Section 3.4) because the gamma function results from integrating the powers of \mathbb{B} 's curvature over the unit segment, namely $\Gamma(n+1) = \int_0^1 (-\ln x)^n \, dx$.

Let us denote the r -normalize hyperbolic distance (equations 33 and 34) in logarithmic terms the "conformal encoding function" of $P \in \mathbb{A}$, namely

$$\vec{C}_r(P) \equiv \operatorname{sgn}(P) \log_r(2 \operatorname{sgn}(P) P - 1) \quad (37)$$

with inverse "conformal decoding function"

$$\overleftarrow{C}_r(C) \equiv \frac{1}{2} \operatorname{sgn}(C) \left(1 + r^{\operatorname{sgn}(C)} \vec{C}_r(P) \right) \quad (38)$$

where $\operatorname{sgn}()$ is the signum function (see Figure 10).

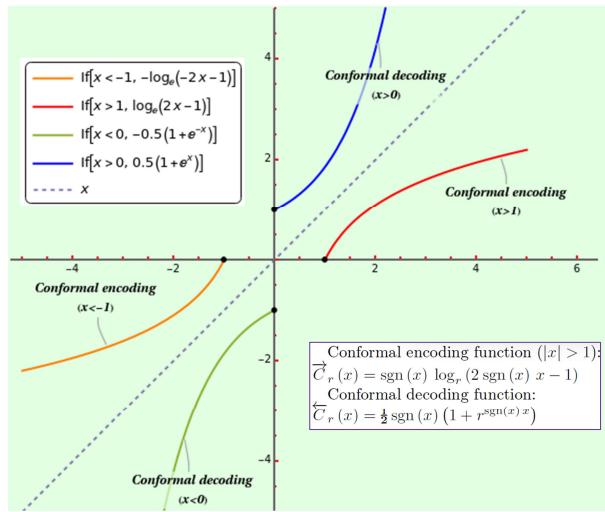


Figure 10. The coding functions of the 1-ball conformal model.

Because the source places an object observed at a Euclidean distance $P \in \mathbb{A}$ at a Euclidean distance

$$Q = \operatorname{sgn}(P) - \frac{1}{P} = \operatorname{sgn}(P) (|P|; 1 :)$$

from the origin, we can calculate the conformal encoding function using (36) as the infinite summation

$$\vec{C}_r(Q) = \frac{2}{\ln r} \sum_{i=1}^{\infty} \frac{Q^{2i-1}}{2i-1}$$

Consequently, the coding source can calculate the conformal decoding function as the infinite product

$$\overleftarrow{C}_r(Q) = \frac{1}{2} \operatorname{sgn}(Q) \left(1 + r^{\operatorname{sgn}(Q)} \vec{C}_r(Q) \right) = \frac{1}{2} \operatorname{sgn}(Q) \left(1 + \prod_{i=1}^{\infty} e^{2\operatorname{sgn}(Q) \frac{Q^{2i-1}}{2i-1}} \right)$$

which does not depend on r .

Therefore, Euclidean distances measured in \mathbb{B} (32) are the only inputs necessary to compute the coding functions.

Furthermore, a coding source can calculate its first digit probability as

$$\Pr(r, 1) = \log_r \left(1 + \frac{1}{1} \right) = \frac{\int_0^1 \operatorname{artanh}(Q) \, dQ}{\ln(r)} = \frac{1}{2} \int_0^1 \vec{C}_r(Q) \, dQ$$

6.5. Local Bayesian Entropy

The conformal encoding function (37) represents the likelihood of the local Bayesian odds, namely

$$\tilde{o}(2 \operatorname{sgn}(P) P - 1 : 1 | r) = \frac{\log_r(2 \operatorname{sgn}(P) P - 1)}{2 \operatorname{sgn}(P) P - 1} = \frac{\log_r(2|P| - 1)}{2|P| - 1}$$

which expresses the entropic contribution of bin $[1, 2|P| - 1]$, hence of P , to \mathbb{A} 's information total, where (30) defines \mathbb{A} . Because a cross-ratio is invariant under a conformal transformation, so is the Bayesian information defined by the local odds. The transformed Bayesian datum is

$$\tilde{o}(1 + \operatorname{sgn}(Q) Q : 1 - \operatorname{sgn}(Q) Q | r) = \frac{1 - \operatorname{sgn}(Q) Q}{1 + \operatorname{sgn}(Q) Q} \log_r \left(\frac{1 + \operatorname{sgn}(Q) Q}{1 - \operatorname{sgn}(Q) Q} \right)$$

which expresses the entropic contribution of bin $[1 - |Q|, 1 + |Q|]$, hence of Q , to \mathbb{C} 's information total, where (32) defines \mathbb{C} . Moreover, because the mapping is bijective, \mathbb{A} and \mathbb{C} contain the same absolute likelihood information; therefore, reflects exactly the Bayesian entropy of \mathbb{A} .

The limiting function of the rationals in \mathbb{A} to approximate a piece of "real" average information would require an analysis analogous to [104] (chapter 4b), which pivots on the differential entropy [76]. Assuming $b \rightarrow \infty$, such a "differential Bayesian entropy" measures the continuous weighted likelihood from the coding source boundary to a point P ; it is precisely the integral

$$\tilde{e}_{\mathbb{A}^+}(P) = \int_1^P \frac{\log_r(2x - 1)}{2x - 1} \, dx = \frac{1}{\ln r} \left(\frac{\ln(2P - 1)}{2} \right)^2$$

on the right and

$$\tilde{e}_{\mathbb{A}^-}(P) = \int_P^{-1} \frac{\log_r(-2x - 1)}{(-2x - 1)} \, dx = \frac{1}{\ln r} \left(\frac{\ln(-2P - 1)}{2} \right)^2$$

on the left.

Then, artanh comes up again to estimate the coding source's entropy $\tilde{e}_{\mathbb{C}}$. Using (33) and (34),

$$\tilde{e}_{\mathbb{B}}(Q) = \tilde{e}_{\mathbb{A}}(P) = \frac{\operatorname{artanh}^2(Q)}{\ln r} \text{ nat} \quad (39)$$

would dominate the coding source's Bayesian entropy from the origin to infinite points ($Q \rightarrow 1$). In the special theory of relativity, this result means that the entropy grows quadratically with the rapidity (35) when $v \rightarrow b \equiv c$. Using (31),

$$\tilde{e}_{\mathbb{B}}(Q) = \ln r (d_r(Q)/2)^2 \text{ nat}$$

in terms of distance; note that this expression peers Bekenstein-Hawking's formula of black hole entropy in quantum gravity [105]. Since entropy measures confusion, this result means that objects in remarkably curved coding spaces or at huge distances are indiscernible.

6.6. Conformal Iterated Coding

Because of the hyperbolic distance (equations 31 or 37) and the rapidity range between $-\infty$ and ∞ , we can presume the coding source's logarithmic scale is a (new) whole external world, defined by Euclidean distances, and repeat the encoding process.

If we apply the conformal encoding function on the right recursively

$$\log_r (2 \log_r (2 \log_r (2 \dots - 1) - 1) - 1)$$

the source encodes an external object's position sooner or later in ζ^+ , and the recursion halts. For example, if $r = 3$, we can map an object observed at a Euclidean distance of googol from the origin, after five nested conformal transformations, onto a point in ζ^+ at an approximated hyperbolic distance of 0.096773 from the origin.

Repeatedly applying the encoding function (37) or the decoding function (38) is information-preserving iterated coding. We will use the notation $r \circ n$ to express the n th iterate ($n \geq 1$) of the encoding function $\vec{C}_r(P)$ so that $\vec{C}_{r \circ 1}(P) \equiv \vec{C}_r(P)$ and

$$\vec{C}_{r \circ (n+1)}(P) \equiv \vec{C}_r(P) \circ \vec{C}_{r \circ n}(P)$$

where " \circ " denotes function composition holding the properties

$$\vec{C}_{r \circ (m+n)}(P) \equiv \vec{C}_{r \circ n}(P) \circ \vec{C}_{r \circ m}(P)$$

and

$$\vec{C}_{r \circ mn}(P) \equiv \vec{C}_{r \circ n}(\vec{C}_{r \circ m}(P))$$

Note that the limits of the coding space remain unaltered irrespective of the iteration because $\vec{C}_{r \circ n}(\pm 1)$ vanishes for all n .

The iterated logarithm of $N \in \mathbb{N}^+$, written $\log_* N$, is the number of times the natural logarithm function must be recursively applied before the result is less than or equal to the unit. Similarly,

$$\vec{C}_{r_*}(P)$$

is the number of times we must iteratively apply (37) until the absolute value of the result is less than one.

We call the sequence of values $\vec{C}_{r \circ n}(P)$, where $1 \leq n \leq \vec{C}_{r_*}(P)$, the "conformal orbit" of P , which outlines a tetrational plot [106]. For example, the orbit with radix $r = 3$ of the quantum minus googol is $\{-10^{100}, -210.221, -5.49687, -2.09533, -1.05609, -0.096773\}$. No value of the orbit can be "identically 1" because ± 1 represents $\pm\infty$, while the global base b is our universe's maximum. Indeed, $\vec{C}_{e_*}(b)$ gives us the universe's maximum natural depth.

We recover the original point by applying (38) iterated the same number of times, i.e.,

$$\overleftarrow{C}_{r \circ \vec{C}_{r_*}(P)}(\vec{C}_{r \circ \vec{C}_{r_*}(P)}(P))$$

where $\overleftarrow{C}_{r \circ 1}(H) \equiv \overleftarrow{C}_r(H)$ and $\overleftarrow{C}_{r \circ (n+1)}(H) \equiv \overleftarrow{C}_r(H) \circ \overleftarrow{C}_{r \circ n}(H)$.

Every 1-ball of radius $\overleftarrow{C}_{r \circ n}(0)$ might correspond to a granularity level [107], a local setting belonging to the nested information of a (global) complex system such as the universe. Considering that $\vec{C}_{r_*}(P)$ grows with P exceptionally slowly, the natural granularity levels are likely few; the ternary granularity depth for the currently estimated universe size in Planck units would be $\vec{C}_{3_*}(10^{61}) = 5$, and the binary granularity depth for the currently estimated number of atoms in the known universe would be $\vec{C}_{2_*}(10^{88}) = 8$.

Granularity and primality might be related. A PN system encodes a number as a polynomial in a single (integer) variable [108], the base or radix, where the coefficients are the possible outcomes of the variable, either quanta or digits. Imagine now that $b = r^k$, where r is a prime, and $k = \vec{C}_{r*}(b)$ is a nonzero natural that indicates the r -ary depth of the universe. Then, the universe could be a finite (or Galois) field [109] of order b and characteristic r (addition of r copies of any quantum vanishes) where the operations of multiplication, addition, subtraction, and division are well-defined, and equation $(q_1 + q_2)^r = q_1^r + q_2^r$ holds. A granularity tier would constitute a prime field of order r represented by its digits $1 \leq d < r$ (roots of the polynomial $X^{r-1} - 1$), and b 's quanta would correspond to the factors of $X^{b-1} - 1$ over r 's field.

Because every iteration conserves the local likelihood information, a granularity realm has an identical copy of the Bayesian data that matches its range of distances in the external world. Nevertheless, it is autonomous in creating new information elements, such as those resulting from clustering points or lumping together states of similar behavior, defining emerging organizational layers. We can even take this combination of iterated coding with coarse-grained modeling [110] as a principle of multiscale modeling [111].

From a computational point of view, conformal coding might use a representation similar to the level-index number system [112]. The quantum $P \in \mathbb{A}^+$ encoded as the (true normalized form of the) significand $0 < s < 1$ after $n = \overrightarrow{C}_{r*}(P) \geq 1$ iterations would be represented as s_{ron} so that

$$P \approx \overleftarrow{C}_{ron}(s) = \frac{1}{2} \left(1 + r \left(\frac{\frac{1}{2} \left(1 + r^{\frac{1}{2} \left(1 + r^{\dots \frac{1}{2} \left(1 + r^s \right)} \right)} \right)}{1 + r} \right) \right)$$

where the order (height) of the power tower is n .

For example, a conformal representation of the number googol is $.0021211221\dots_{305}$ (see Figure 11) owing to $.0021211221\dots_3 \equiv .096773\dots$ and

We similarly obtain that $10^{88} = \overleftarrow{C}_{2 \circ 8} (.805897 \dots) \equiv .1100111001 \dots_{2 \circ 8}$.

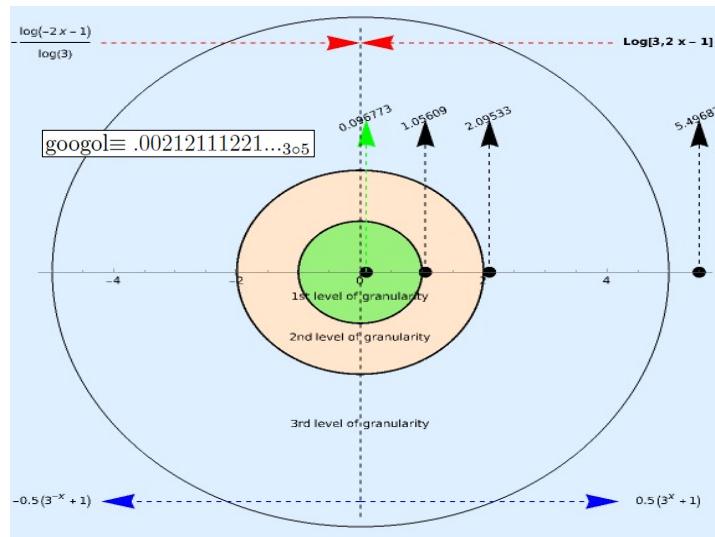


Figure 11. The most profound three levels of granularity $\{\pm \overleftarrow{C}_r(0) = \pm 1, \pm \overleftarrow{C}_{r02}(0) = \overleftarrow{C}_r(\pm 1) = \pm 2, \pm \overleftarrow{C}_{r03}(0) = \overleftarrow{C}_r(\pm 2) = \pm 5\}$ (out of 5) a coding source generates to encode the number googol with radix $r = 3$, the minor 4 points of the conformal orbit, and the conformal encoding (in red) and decoding (in blue) functions.

7. Primordial Distributions

We construct the one-parameter canonical PMF from binomial generators. Then, we fix the parameter of the canonical PMF for the natural and integer numbers by introducing a requirement for the divisibility of the probability mass. This provision is equivalent to making the event "picking the number one" a Bernoulli process.

The resulting PMF for the natural numbers allows calculating the probability and entropy of dichotomies like odd-even and prime-composite and trichotomies such as negative-zero-positive and elliptic-Euclidean-hyperbolic. The canonical PMF for the integer numbers could be the germ of a fundamentally unitary, parity-invariant, uncertain, discrete, and maximally entropic universal field.

7.1. Ensuring Constructability, Rationality, and Randomness

We can not irrefutably prove that the probability ISL (1) is a foremost PMF beating at the core of the cosmos, but NBL emanating from it is at least evidence supporting that possibility. Because NBL is pervasive and reflects the properties of PN, efficiency must be a rudimentary feature of nature. Following this trend of thought, we will describe how to grow this one-parameter PMF from elemental geometric distributions, guaranteeing universal constructability. We will also tune the PMF's parameter to achieve the expected divisibility of \mathbb{N} 's probability mass and cast probability values as unit fractions, which backs the idea that nature has a prominent rational character.

Euler product formula for the zeta function (equations 1.6 and 1.13 in [113]) allows us to define PMF (1)'s primordial random variable $X \in \mathbb{N}^+$ as the infinite product

$$X = \prod_{p \in \mathbb{P}} p^{X(1/p^2)}$$

of independent identically distributed random variables of parameter $1/p^2$ with a geometric sequence of probabilities $\Pr_{p \in \mathbb{P}} (X(1/p^2) = k) = (1/p^2)^k (1 - 1/p^2)$ [65], corresponding to the PMF of k failures before the first success, each binomial trial (see below) with failure probability $1/p^2$. This

construction means that the canonical PMF (up to ϵ) consists of geometric distributions that, in turn, are "memoryless" [114] and "infinitely divisible" [115]. Of course, this infinitude is theoretical because there must be a maximum prime.

Once we have constructed the one-parameter canonical PMF from elemental generators, which constitutes a rudimentary notion of emergence, we can improve its computability by digging into rationality and making the probability mass of the natural numbers divisible on average to ensure its compartmentalization.

At the end of Section 3.3, we determined that $\epsilon \in (0, 6/\pi^2)$ by merely demanding $\Pr(N \in \mathbb{N}) > 0$, i.e., $(1 - \epsilon\zeta(2)) = (1 - \epsilon\pi^2/6) > 0$. Given that the fraction of square grid \mathbb{Z}^2 's points visible from the origin is precisely $6/\pi^2$ [116], our requirement is equivalent to setting ϵ as a rational equal to the size of a subset of them.

We now introduce a new vital constraint on the number of divisors of a natural number intended to narrow the range of possible values of ϵ . The rationale is that divisibility, as a dual concept of primality, is paramount to understanding our universe.

Remember that a nonzero natural's number of divisors d includes one and the number itself; for instance, $d(1^2) = 1$, $d(2^2) = 3$ (1, 2, and 4), $d(3^2) = 3$ (1, 3, and 9), and $d(4^2) = 5$ (1, 2, 4, 8, and 16). Notwithstanding that the distribution $\Pr(d(N^2))$ is unknown, we can calculate the mean of divisors of counting numbers squared employing the general law of the expected values (or "unconscious statistician") [117]. We require such an expected value to be at least two to guarantee that the divisibility of the entire probability room defined by (1) takes place non-trivial and naturally, for if N^2 splits into $d(N^2) \geq 2$ parts, so can $\Pr(N)$, i.e.,

$$2 \leq \hat{E}(d(N^2)) = \epsilon \sum_{N=1}^{\infty} \frac{d(N^2)}{N^2} = \epsilon \frac{\zeta^3(2)}{\zeta(4)} = \epsilon \frac{(\pi^2/6)^3}{\pi^4/90} = \epsilon \frac{5}{12} \pi^2$$

where the summation agrees the equation 3.41 of [118]. Considering the high probability of picking the unit, this constraint is more rigid than it might seem at first sight.

Let us recap. We are imposing only a pair of constraints to provide ϵ with an accurate value; first, the probability of a natural to be nonzero, and second, the expected probability mass of the set of natural numbers to be splittable. Thus,

$$(0 < \epsilon < 6/\pi^2) \wedge \left(\epsilon \frac{5}{12} \pi^2 \geq 2 \right)$$

constricts the possible values of ϵ to the narrow range

$$\epsilon \in \frac{1}{\pi^2} \left[\frac{24}{5}, 6 \right) \approx (.4863417, .6079271)$$

We aim to define NBL in a strict rational setting to increase operability through multiplication and division (see subsections 3.4 and 5.1). Although rationals such as $49/100$ or $6/10$ satisfy this constraint, the most probable numerator and denominator in agreement with (1) are precisely 1 and 2, so

$$\epsilon = \frac{1}{2} \tag{40}$$

In passing, this value assures that the probability mass of a nonzero natural number is splittable. Moreover, it is a unit fraction.

What does (40) mean from the information theory perspective? It means converting the event "picking the unit" into a Bernoulli (binomial) experiment equivalent to flipping a coin. A sequence of these independent identically distributed picks is a Bernoulli process, unique and universal in that it is the single most random non-mixing process possible [119].

More generally, the corresponding variable that considers obtaining exactly k ones in an experiment with N trials, each with a probability of success ϵ , follows the binomial distribution [120]

$$\binom{N}{k} = \frac{N!}{k!(N-k)!}$$

$$\Pr(k, N, \epsilon) = \binom{N}{k} \epsilon^k (1-\epsilon)^{(N-k)} \quad (41)$$

isomorphic to "exactly k tosses out of $N \geq k$ tosses resulting in a head". The logarithmic scale awesomely comes up when we calculate the entropy of this distribution, which tends to $\frac{1}{2} \ln(2\pi e N \epsilon (1-\epsilon))$ as N approaches infinity.

Well, the binary entropy function of "picking the number one" (and the complementary event "picking a number with splittable probability") attains its maximum when (40) holds, like a fair coin, where the odds for and against are 1 (heads or tails) by definition. For the same reason, $\frac{1}{2}$ maximizes the entropy of the binomial distribution (41), too.

In summary, (40) ensures five conditions, to wit, positive rational probabilities congruent with the PMF itself, divisibility of N 's probability mass, maximum number of naturals with splittable occurrence probability, counting numbers with unit fraction probability masses, and maximum universal randomness.

7.2. Canonical PMF for the Natural and Integer Numbers

Assuming (40), we can establish that

$$\Pr(X = N) = \begin{cases} N \in \mathbb{N} - \{0\} : & \frac{1}{2N^2} \\ \text{else} : & 1 - \frac{1}{2}\zeta(2) \end{cases} \quad (42)$$

is the canonical PMF for a random variable X that takes natural values. Its mode is one, with undefined mean and variance, and entropy requiring 3 bits because of

$$-(1 - \frac{1}{2}\zeta(2)) \log_2(1 - \frac{1}{2}\zeta(2)) - \frac{1}{2} \sum_{N=1}^{\infty} \frac{\log_2(\frac{1}{2}N^{-2})}{N^2} = 2.6178 \text{ bit}$$

As required, the expected divisibility value of the global probability mass surpasses two, namely

$$\hat{E}(\delta(N^2)) = \frac{1}{2} \frac{\zeta^3(2)}{\zeta(4)} \approx 2.05617$$

Indeterminacy has chances every tick of the clock because of $1 - \frac{1}{2}\zeta(2) \approx 17.75\%$. Therefore, the probability of a counting number coming out is

$$\frac{1}{2} \sum_{N=1}^{\infty} \frac{1}{N^2} = \frac{\zeta(2)}{2} = \frac{\pi^2}{12} \approx 82.25\%$$

The probability of a string of numbers is the product of the individual probabilities; for example, the probability of picking s and t in a row is $\Pr(< s, t >) = \Pr(< t, s >) = \Pr(s) \Pr(t)$. The probability of a choice between a set of numbers is the sum of the individual probabilities; in particular, the probability of picking a number in the interval $[s..t]$ is the sum of $\Pr(s) + \Pr(s+1) + \dots + \Pr(t-1)$.

The probabilities of a natural number being odd and even are $(3/4)(\zeta(2)/2) = (\pi/4)^2 \approx 5/8$ and $(1/4)(\zeta(2)/2) = \pi^2/48 \approx 5/24$, respectively (see equation 1.12 in [121]). So, getting an odd natural number is $\frac{3/4}{1/4} = 3$ times as probable as picking an even, in sharp contrast with the "intuitive" $\frac{1}{2}$ size in [48].

The probability of the event "picking a natural number greater than 1", i.e., a number with splittable probability, is the unit minus the probabilities of indeterminate and one, namely

$$\Pr(X \geq 2) = 1 - (1 - \frac{1}{2}\zeta(2) + \frac{1}{2}) = \frac{1}{2}(\zeta(2) - 1) \approx 0.32247$$

The probability of the event "picking a prime" is half the sum of the reciprocals of prime numbers \mathbb{P} squared, i.e.,

$$\Pr(X = N \in \mathbb{P}) = \sum_{N \in \mathbb{P}} \frac{1}{2N^2} = \frac{1}{2}P_\zeta(2) \approx 0.22612$$

where $P_\zeta(\cdot)$ is the prime zeta function. Thus, picking a composite number has a probability of

$$\Pr(X \geq 2 \wedge X \notin \mathbb{P}) = \frac{1}{2}(\zeta(2) - 1 - P_\zeta(2)) \approx 0.09634$$

and the probability that $N \in \mathbb{P}$, conditioned to be greater than 1, exceeds 70% due to

$$\Pr(X = N \in \mathbb{P} | X \geq 2) = \frac{\Pr(X = N \in \mathbb{P})}{\Pr(X \geq 2)} = \frac{P_\zeta(2)}{\zeta(2) - 1} \approx \frac{0.22612}{0.32247} \approx 0.70123$$

Therefore, observing primes in nature is expected; regarding number theory, the odds of prime versus composite are 70/30.

On the other hand, the canonical PMF (42) straightforwardly explains the proclaimed supremacy of the hyperbolic configurations concerning the two-dimensional tilings algebraically associated with the finite reflection groups [122].

The probability of a natural number greater than one occurring thrice is

$$(\Pr(X \geq 2))^3 = (\frac{1}{2}(\zeta(2) - 1))^3 \approx 0.033532$$

In other words, this value is the probability of producing three naturals to form a triangle.

The probability of picking three naturals forming a Euclidean triangle (i.e., $1/l + 1/m + 1/n = 1$) is the probability of picking $\{2, 3, 6\}$, $\{2, 4, 4\}$, or $\{3, 3, 3\}$, namely

$$(\frac{1}{2})^3 \left(\frac{1}{2^2 3^2 6^2} + \frac{1}{2^2 4^2 4^2} + \frac{1}{3^2 3^2 3^2} \right) \approx 0.000390$$

The probability of picking three naturals forming a spherical triangle (i.e., $1/l + 1/m + 1/n > 1$) is

$$(\frac{1}{2})^3 \left(\frac{\zeta(2) - 1}{2^2 2^2} + \frac{1}{2^2 3^2 3^2} + \frac{1}{2^2 3^2 4^2} + \frac{1}{2^2 3^2 5^2} \right) \approx 0.005780$$

where the first term is the probability of a triplet of two 2s and any natural greater than one.

The probability of picking three naturals forming a hyperbolic triangle ($1/l + 1/m + 1/n < 1$) equals the occurrence probability of a triangle that is neither Euclidean nor spherical,

$$0.033532 - 0.000390 - 0.005780 \approx 0.0273615$$

So, the odds of hyperbolic cases against non-hyperbolic cases point to the former's predominance, specifically

$$\frac{0.0273615}{0.000390 + 0.005780} \approx 4.434429$$

In summary, two-dimensional tilings are usually hyperbolic. Remember that a star of the triangle group [123] is the set of all rotational triangle subgroups $D(2, 3, n)$ ($1/2 + 1/3 + 1/n < 1$), isomorphic to the modular group we broached at the end of Section 5.6. This group is a subset of the almighty M bius group, which permeates through many fields of geometry, number theory, and significant areas of physics.

In particular, $D(2, 3, 7)$ is the triangle subgroup most probable in the hyperbolic class. It is algebraically the quotient of the group of unit quaternions by its center ± 1 and topologically the cover for all Hurwitz surfaces. The Hurwitz group of a Hurwitz surface maximizes the order of the automorphism group for a given genus through orientation, scale, and angle-preserving transformations, i.e., conformal mappings, with grand importance in Quantum Field Theory, gravitation, and cosmology. The Hurwitz surface of the minor genus (three) is the Klein quartic [124,125] with automorphism group the self-dual Fano plane, closely related to the octonions [126], both appearing in game theory and fundamental physics. Moreover, the mysterious Monster group can grow from $D(2, 3, 7)$ [127].

These facts indicate again that inversive and conformal geometry might be at the world's heart, built from the simplest simplices (generalizations of a triangle to any dimension).

Finally, we can extend (42) to the set of integer numbers to establish

$$\Pr(Z) = \begin{cases} Z \in \mathbb{Z} - \{0\} : & \frac{1}{(2Z)^2} \\ \text{else :} & 1 - \frac{1}{2}\zeta(2) \end{cases} \quad (43)$$

The canonical PMF for the integers, illustrated in Figure 12, has mode ± 1 , with entropy

$$-(1 - \frac{1}{2}\zeta(2)) \log_2 (1 - \frac{1}{2}\zeta(2)) - 2 \sum_{Z=1}^{\infty} \frac{\log_2 ((2Z)^{-2})}{(2Z)^2} = 3.44027 \text{ bit}$$

For nonzero integers x and y , it is the solution to the functional equation

$$\Pr(x)\Pr(y) = (xy\Pr(xy))^2$$

satisfying the condition $\Pr(\pm 1) = \frac{1}{2}$.

This PMF, which accounts for the weight of the ISL in physics [128], cannot be more buildable; take every second integer from the coding source point of view, i.e., $\{\dots, -8, -6, -4, -2, 2, 4, 6, 8, \dots\}$, then their reciprocals squared, obtaining $\{\dots, 1/64, 1/36, 1/16, 1/4, 1/4, 1/16, 1/36, 1/64, \dots\}$, which coincides with (43). It satisfies positive probability masses summing to one, central reflection symmetry, fair (i.e., undefined) mean and variance, holistic rationality, and randomness. Unitarity, parity invariance, indeterminism, discreteness, and the principle of maximum entropy are the physical reflections of these properties, which are all fundamental. Besides, the square root of the probability mass (physically, the probability amplitude) of a nonzero integer is precisely half its reciprocal.

Again, the resulting NBL is valid irrespective of the proportionality constant ϵ following Section 3.6, namely

$$\Pr(r, d) = \log_r \left(1 + \frac{1}{|d|} \right)$$

where $1 \leq |d| < r$, $r \in \mathbb{N} \geq 2$, and $d \in \mathbb{Z}$.

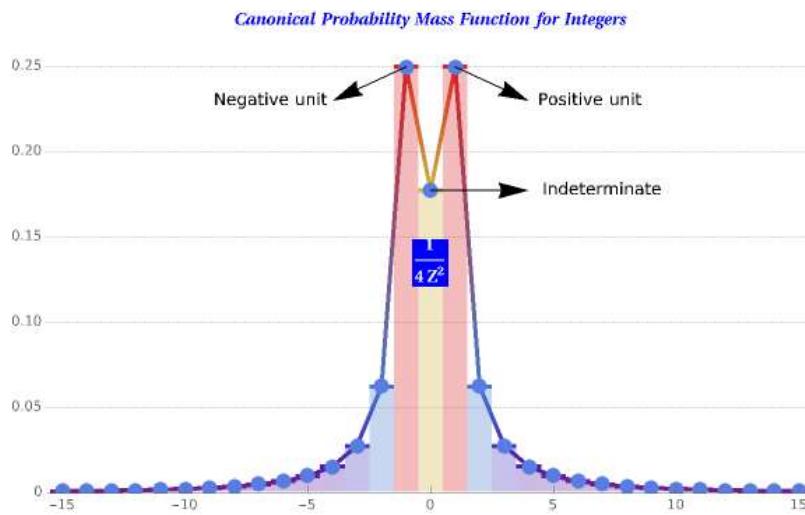


Figure 12. The canonical PMF for the integer numbers could be the germ of a fundamentally unitary, parity-invariant, indeterministic, discrete, and maximally random universal field. The upper points resemble the (inverted) cross-section of a sombrero potential (e.g., the quartic function $-(Z/2)^2 + 2(Z/2)^4$) of a scalar field with an unstable (indeterminate) center and a nonzero "vacuum expectation value"; the multiplicative units ∓ 1 provide this field with the ground (vacuum) state enabling spontaneous symmetry-breaking. Thus, the sombrero potential would be the physical manifestation of a fundamental improbability mass function.

8. Epilogue

Our research shows how discreteness and the continuum interact under the shelter of (1). A complex system and its environment embody the continuous local and discrete global. The NBL probability's derivative of the local takes us to the global, and vice versa; the global's integral situates us in a local setting of likelihood-based probability. A harmonic scale of rational numbers supports the global realm, while a logarithmic scale of "real values" supports the local realm. The harmonic scale's base is a universal reference that establishes the concept of likelihood-based information, and a logarithmic scale's radix is an exponentiation constant that normalizes the system's conformal space of local information coded in PN.

We assume the system is a coding source that observes the outside, operates internally the gathered information, and takes action on the environment. More precisely, a coding source uses the synergy between Benford's and Bayes's laws to reflect (encode) the external world, process (recode or arithmetically transform) the information, and return (decode) the results to its immediate surroundings. These laws connect mathematics with physics.

8.1. Canonical PMF

NBL bets on smallness; little objects are more numerous than extensive ones. Why? The fact that many probability distributions partially adhere to NBL does not reveal its root. Nor can we glean its origin from the fact that merging methods via sampling or multiplication of real-world data series produces adherence to NBL. "Mathematics alone cannot justify a first-digit law", wrote Raimi. Given that the effects of NBL are well-known in physics, we need to be aware of its fundamental character and ultimate cause.

NBL is not so mysterious if we concede that it originates from an ISL of probability. Whenever the canonical PMF governs a system behavior, we can infer its properties are data spaces that record information on a positional scale. This constructible primordial PMF states an absolute hyperbolic

relation between the square of a nonzero number $N \in \mathbb{N}$ and its probability mass $\Pr(N)$, specifically that $N^2 \Pr(N)$ is constant. We stress that the phenomenology of the canonical PMF and NBL run in parallel, to the point that we aver that ours is an inverse-square world for the most part, much as Havil affirms that "It's a Harmonic World" and "It's a Logarithmic World".

If we assume the minimal information principle, PMF (43) is the only way to satisfy that the probabilities are positive summing to 1, guarantee that our random variable's average value is not finite if we repeat the experiment often enough, ensure mass divisibility, and cope with the extension to all the integers, hence to the rational and algebraic numbers. These requirements' logic, sturdiness, and feasibility suggest a full-fledged tenet at the heart of mathematics, physics, and higher integrative levels. Besides, this PMF implements the Axiom of Induction; "at least one inductive set does exist determined uniquely by its members that has a substrative probability distribution".

Although the canonical PMF consists of Bernoulli generators, we introduce (43) as a brute fact deprived of tangible information. Notwithstanding, this probability ISL could be the embryo for crucial experimental laws of physics, such as Newton's Universal Gravitation and Coulomb's Electrostatic Force. Understood as an improbability field, it could even give place to the energy density of space characterized by a sombrero potential with an unstable center and a nonzero vacuum expectation value.

The canonical PMF indicates a manner of arranging availability versus transcendence. Occasional numbers are more startling and influential than abundant ones. While frequent numbers provide resilience, infrequent numbers have the capacity for transformation. Therefore, the universal equilibrium is not enforced via uniformity but achieved by hyperbolically balancing accessibility or stability (position) against magnitude or reactivity (momentum); does not it sound to the Uncertainty Principle?

In particular, our study of depleted harmonic series teaches us that the specific digits involved in a constraining numeral do not matter. In contrast, the length of such a numeral does. Short numerals or low digit densities are accessible and cheap, producing heavy terms that condense the space. In contrast, long numerals or high densities of digits are "rare" and deliver slender harmonic terms. In other words, increasingly bigger numbers on a linear scale have less and less weight on a positional scale. Moreover, "almost all" large numbers have a high cost of accessibility and are indistinguishable except for their order of magnitude.

The canonical PMF naturally copes with indeterminacy by introducing the "indeterminate" value (interpreted as inaction or not-a-number and symbolized by zero) and dodging a finite expected value. Additionally, uncertainty appears with a less metaphysical flavor and a quantum touch. A number's value and probability represent a foundational (position-momentum) inaccuracy,

$$N \cdot \Pr(N) = \frac{1}{2N} \leq \frac{1}{2}$$

Assuming that induction is a rudiment of our cosmos, this inequality is a sort of "certainty principle", i.e., a clue that uncertainty is finite.

The canonical PMF defines a large number by its probabilities, proportional to the chance of fitting its tail and inversely proportional to its opportunity, i.e.,

$$N \sim \frac{\epsilon \psi'(N)}{\frac{\epsilon}{N^2}} = \frac{\Pr(\geq N)}{\Pr(N)}$$

Equality exclusively occurs in the infinite limit when the trigamma function approaches a hyperbola asymptotically (see Section 3.4). Alternatively, for $\Pr'(\geq N) \sim -\Pr(N)$, $\lim_{N \rightarrow \infty} \Pr(\geq N) + N \Pr'(\geq N)$ vanishes. In principle, scale invariance is unreachable because the exact solution $\Pr(\geq N) = \epsilon/N$ exists only in the offing.

Nevertheless, summing from the unit to base b , say the superlative natural number, immediately drives us to the global NBL. b confines the physical framework where numbers become quanta. The

trigamma function measures a number's probability exceedance, i.e., its cumulative distribution function's upper tail. When normalizing the trigamma function relative to the base's support, we obtain the probability of quantum q in base b , which is the product's reciprocal of q and the harmonic number $b - 1$. We can interpret $\Pr(r, q) = 1/q/H_{b-1}$ as measuring cruising speed because $1/q$ and H_{b-1} estimate a quantum's span (space) and a base's scope (time); minor quanta or bases raise promptness, while high quanta or bases yield delay.

Pure conformality is only possible within a finite global scope; b ultimately enables implementing a PN-based coding space where the accessibility potential

$$\Pr(b, q) + q \frac{\partial \Pr(b, q)}{\partial q}$$

vanishes for all quanta q and the logarithm can germinate.

8.2. the Logarithm Measures Local Information

Integrating under the global NBL's hyperbola and normalizing concerning a local radix $r \ll b$ immediately drives us to the fiducial, \mathbb{R} -based, logarithmic form of NBL.

The logarithm and its inverse (the exponential) are fundamental functions because they appear everywhere in mathematics and physics, to wit generalized means, primes, fractals, solutions to many differential equations, power laws, information transmission, von Neumann entropy, et cetera. This plenty demonstrates that hyperbolic spaces proliferate. Indeed, we prove that the canonical PMF leads to such dominance in Section 7.2. Within the scope of this essay, not only does the logarithm measure the local natural likelihood (5) and probability of a bin (equations 6 and 7), but the local Bayes factor (18), the elemental jumps (22), the entropy distribution function of a bipartition (23), the subjective ratio (24), the hyperbolic distance (31), the rapidity (35), the canonical coding (37) and decoding (38) functions, the differential Bayesian entropy (39), and the repetition of the Bernoulli event "picking the unit" (41) also involve it.

In geometry, the logarithm gives the normalized area under the hyperbola. The logarithm bridges information and physics, especially thermodynamics, via the Gibbs entropy formula in statistical mechanics and the Bekenstein-Hawking formula in quantum gravity. In information theory, the logarithm mainly estimates the representational extent of a given numeral written in PN. Further, our essay proves that the logarithm resolves the metric of a conformal space by recasting correlations into distances or rapidities, as Section 6.2 describes. This conversion is critical in iterated coding, especially in coarse-grained and multiscale modeling. The radix's logarithm is precisely the (absolute value of the) coding space's curvature. In particular, radix $r = e$ defines the natural logarithmic scale, i.e., the standard one-dimensional hyperbolic space.

The logarithm is central to comprehending how profoundly NBL connects with recurrence and incrementality. A coding source implements Bayes' rule by multiplying the prior odds between two quanta by a likelihood factor represented by a local NBL probability, precisely the logarithm of their ratio's reciprocal. We say a local datum is Bayesian if it admits this structure, which is recurrent under iterative processes of encoding, recoding, arithmetic operations, and decoding. A local Bayesian datum represents likelihood information, e.g., the encoded odds of an elemental gap or a stopping choice.

Assuming that a digit is an orbit and the radix is the number of orbits, e.g., of an atom, a jump between consecutive orbits introduces significant entropy differences only in the origin's immediacy. Lower orbits have more difficulty achieving a transition, whereas the reactivity associated with the farthest orbits is logarithmically more probable. This behavior resembles the chemical elements' periodic table concerning the electron shells.

Assuming that a digit is an item and the radix r is the number of items in a pool, Bayesian coding solves a version of "the secretary problem" that considers the strategy to select "a good" item rather than "the best" one. It belongs to a class called "last-success problems" with universal scope. Its objective is to determine the last item x on the fly that maximizes the probability of success in accomplishing

the stopping rule, i.e., rejecting the first x items and stopping afterward at the next that is better than the preceding ones. We approach the solution by aggregating the odds $\tilde{o}(x : 1|r) \propto \frac{r}{x} \ln x$ and $\tilde{o}(r : x|r) \propto x \ln \frac{r}{x}$. We feature a pair of characteristic properties of all these good-choice (best-choice included) and optimal-stopping problems; first, they require a phase for incremental information gathering to deliver a preliminary proper output only refined if there is time left, and second, the solution x generates asymmetry by making the past partition (bin $[1, x)$) smaller than the future partition (bin $[x, r)$). Is this procedure not primary management of memory in real-time?

8.3. Conjectures

On the one hand, the canonical PMF and its NBL subsidiaries explain why proximity or slightness provides more stability than distance or heftiness. Occurrence probability attracts information toward a central source; apart from the second thermodynamical law, no other law allows inferring such a fundamental imbalance. However, the entropic leaks from the most outlying digits offset this mass accumulation in the source's proximity. Therefore, data encoded in PN would induce alternating uphill and downhill flows, reflecting a brute fluctuation between the dual elementary concepts of concentration and dispersion.

On the other hand, Newcomb-Benford and Bayesian laws regulate the implementation of a conformal space through tractable hyperbolic functions. An NBL probability is a likelihood, i.e., a Bayes factor. A referential ratio is a particular case of Bayesian datum with a unit prior (i.e., a sheer Bayes factor) and a point at infinity. A cross-ratio is a quotient of referential ratios, and the logarithm of a cross-ratio yields a conformal metric. The canonical coding functions define how a complex system, say a coding source, creates an image of the world, which can render crucial consequences in physics, principally enabling scalability and boosting efficiency as leitmotifs and chief drivers of cosmological development.

The universal hyperbolicity that the inverse-square, Newcomb-Benford, and Bayesian laws steer can illuminate the measurement problem. If the integer line \mathbb{Z} were the position space of a generic object, the canonical PMF for the integers (43) would match with a default wave function with one degree of freedom expressible as a linear combination of the position eigenstates $|Z\rangle$, namely

$$\Psi = \dots + 1/4 | -2 \rangle + 1/2 | -1 \rangle + \sqrt{1 - \frac{1}{2}\zeta(2)} | 0 \rangle + 1/2 | +1 \rangle + 1/4 | +2 \rangle + 1/6 | +3 \rangle + \dots$$

A nonzero integer $Z \in \mathbb{Z}$ represents an actual eigenvalue corresponding to the eigenstate with rational amplitude $(2Z)^{-1}$, and the origin is the idle or vacuum (beable) state with quantum amplitude $\sqrt{1 - \frac{1}{2}\zeta(2)}$. By the Born rule, the quantum-mechanical probability of being at place Z is the square modulus of its rational amplitude, precisely its canonical probability mass. From a complementary point of view, the canonical PMF tells us the probability $\Pr(Z) = (4E|Z\rangle)^{-1}$ of having energy $E|Z\rangle = Z^2$, corresponding to a wave function's rational amplitude $(2\sqrt{E|Z\rangle})^{-1}$ in the momentum space. Thus, the physical existence of the canonical PMF would imply a default wave function and vice versa.

A system object of observation and the measurement entity (e.g., an instrument, an experimenter, or the environment) interact, entangling into a mixture of states represented by their joint wave function. Suppose the observed object is a test particle, and the observing entity is a measurement device whose wave function corresponds to our universally binding canonical PMF. The joint wave function is untestable, but the coding device can measure the particle's position up to inherent uncertainty.

We assume particles materialize as wave packets with arbitrary widths rather than points. During measurement, a particle in the observation field progressively decoheres, losing its pure quantum nature and dynamics while transferring the information into the coding source, which recursively assigns the particle's state to nested bins forming a numeral. For example, suppose that a particle has triangular wave function $\frac{1}{2}| -1 \rangle + 1/\sqrt{2}| 0 \rangle + \frac{1}{2}| +1 \rangle$ with center of masses objectively at position 200 (and triangular momentum compatible with the uncertainty principle). We can locate

the particle's wave function projected onto a bin of digits at any time. The odds of registering "2" to "1" are $\left(\frac{((1/\sqrt{2})^2 + (1/2)^2) \log_{10}(1+1/2)}{((1/2)^2 \log_{10}(1+1/1))} \right) = 1.755$, the odds of registering "20" to "19" are $\left(\frac{((1/\sqrt{2})^2 + (1/2)^2) \log_{10}(1+1/20)}{((1/2)^2 \log_{10}(1+1/19))} \right) = 2.854$, and the odds of registering "200" to "199" are $\left(\frac{((1/\sqrt{2})^2 \log_{10}(1+1/200)}{((1/2)^2 \log_{10}(1+1/199))} \right) = 1.99$. Note that position data are biased towards the source regarding the expected genuine odds, to wit $3 = \frac{((1/\sqrt{2})^2 + (1/2)^2)}{(1/2)^2}$, 3, and $2 = \frac{(1/\sqrt{2})^2}{(1/2)^2}$; objects would be farther than they appear! After measurement, we have no more coherent position state but a position datum; the particle's information now lies in the measurement numeral.

8.4. Some Metaphysics

Literature about NBL overlooks its rational aspect. Ours is primarily a harmonic world, where rational numbers guarantee calculability. PMF (42) and PMF (43) tell us that the probabilities of natural and integer numbers are unit fractions. The global NBL, stemming from (1), means that quantum frequencies are unit fractions, too. Besides, although the fiducial NBL appears to be an absolute law that assigns a probability to every digit separately, a digit acquires its informational meaning only compared with another, i.e., in odds form. Contrary to the Special Relativity premises, we take the global base manifesting as the speed of light as proof of the universe's "rational" rather than "real" essence (see Section 6.2).

More fundamentally, holistic rationality implies relationalism (i.e., reciprocity) and operability (i.e., arithmetical tractability), basic properties of a physical transformation. Rationality is budding relativity. Indeed, the quotidian continuum we perceive from our local outlook approximates the discrete reality; the continuum emerges from the rational instead of vice versa. We conclude that \mathbb{Q} is the universal number system at the heart of computability, contrasting with the contrived real numbers.

Infinity has no place in the algebraic field of rational numbers. Further, we have endorsed the universe's finiteness in many other ways. To begin with, we feature a global base closely related to the maximum natural (or prime) number. Raw data statistics of natural phenomena indicate that no counting system can avoid this universal tendency towards littleness. Perceivable things in the cosmos are typically small but always rationally commensurable from some standpoint; otherwise, they would be incomparable, whence indiscernible. An infinite host universe reduces all its finite guests to zero, an unobservable number. PN ignores numeral positions surpassing a certain threshold. Because a (quantum) measurement obliterates the least significant digits, the models a coding source supports necessarily give rise to non-deterministic mechanics, such as Gisin's proposal, "which is empirically equivalent to classical mechanics but uses only finite-information numbers." A transfinite universe prone to productivity is counterintuitive. The universe is an economic system precisely owing to its limited scope and resources.

This essay generally points to mathematics having a physical status. We have put laws midway between mathematics and physics on the table. The canonical PMF for integer numbers defines a pervasive numeric field of stability that is the germ of a constitutively unitary, parity-invariant, uncertain, discrete, and maximally random universe. The linkage between probability space and physical space, especially within a coding space, is so intricate that we hardly find discrepancies. Further, because the notion of logarithmic likelihood results from comparing two logarithmic sectors (5) and a local NBL probability mass is a ratio of logarithmic likelihoods (6), stating that information is physical means probability is physical.

We have told a hegira from information coding to physics, presuming the Galilean idea that nature is mathematical per se. William K. Clifford (1976, *On the Bending of Space*, https://doi.org/10.1007/978-94-010-1727-5_49) underlined that we might "be treating merely as physical variations effects which are really due to changes in the curvature of our space", although "Whether one associates 'geometric' ideas with a theory is [illegible] a private matter", stated Einstein in a letter written to Reichenbach (Google translation from Doc. AEA 20-117 of the Albert Einstein Archive,

8 April 1926). Our partway philosophical worldview supports the theory that every mathematical concept has a physical peer because physics emerges from algebra via geometry, supported by hyperbolicity, economy, and relationalism. The embodiment of these pillars makes Tegmark's hypothesis that the observable reality is a mathematical structure defined by computable functions plausible. We must add that such a structure consists of conformal spaces and transformations.

Acknowledgments: We herein express our tribute to some personalities and entities influencing this work. The UNED has allowed the development of a scientific instinct and a generalistic approach to problem-solving. Working for Indra enabled a pragmatic approach to figure out any crisis and awareness of the productive character of our universe. We greatly thank Jos Mira, Joan Baez, Lee Smolin, Sean Carroll, Norman J. Wildberger, and Sabine Hossenfelder for their wisdom and explanatory capacity. We apologize in advance for omitting laudable references. We sincerely appreciate those offering comments and critical insight on this version.

Funding: The author (ORCID 0000-0003-3980-5829) asserts that only scientific rigor, significance, and clarity drive the high-level goals of this work. It contains no known minor or significant incongruencies, errors, or inaccuracies. The author did not receive support from any organization for the submitted work and declares that he has no competing financial or non-financial interests directly or indirectly related to the work submitted for publication. No personal relationships have influenced the content of this work. This work is an original one-piece that must be kept unsplit into several parts. It has not been published elsewhere in any form or language, partially or in whole. The author claims to have committed no intentional ethical wrongdoing related to this paper, including self-plagiarism, plagiarism, far-fetched self-citations, conflict of interest, inaccurate authorship declarations, and unacceptable biases concerning the references. This work respects third-party rights such as copyright and moral rights.

References

1. Newcomb, S. Note on the Frequency of Use of the Different Digits in Natural Numbers. *American Journal of Mathematics* **1881**, *4*, 39–40.
2. Benford, F. The Law of Anomalous Numbers. *Proceedings of the American Philosophical Society* **1938**, *78*, 551–572.
3. Hürlimann, W. Benford's law from 1881 to 2006. *Research Gate* **2006**, [math/0607168].
4. Blair, D.E. *Inversion Theory and Conformal Mapping*; Graduate Studies in Mathematics, American Mathematical Society, 2000.
5. Burke, J.; Kincanon, E. Benford's Law and Physical Constants: the Distribution of Initial Digits. *American Journal of Physics* **1991**, *59*, 952–952.
6. Berger, A.; Hill, T. Benford's Law Strikes Back - No Simple Explanation in Sight for Mathematical Gem. *The Mathematical Intelligencer* **2011**, *33*, 85–91. doi:10.1007/s00283-010-9182-3.
7. Berger, A.; Hill, T.P. What is Benford's Law? *Notices of the AMS* **2017**, *64*, 132–134. doi:10.1090/noti1477.
8. Hossenfelder, S. *Lost in Math: How Beauty Leads Physics Astray*; Basic Books, 2018.
9. Finch, S.R. *Mathematical Constants II*; Encyclopedia of Mathematics and its Applications, Cambridge University Press, 2018.
10. Planck Collaboration.; Aghanim, N.; Akrami, Y.; Ashdown, M.; Aumont, J.; Baccigalupi, C.; Ballardini, M.; Banday, A. J.; Barreiro, R. B.; Bartolo, N.; Basak, S.; Battye, R.; Benabed, K.; Bernard, J. P.; Bersanelli, M.; Bielewicz, P.; Bock, J. J.; Bond, J. R.; Borrill, J.; Bouchet, F. R.; Boulanger, F.; Bucher, M.; Burigana, C.; Butler, R. C.; Calabrese, E.; Cardoso, J. F.; Carron, J.; Challinor, A.; Chiang, H. C.; Chluba, J.; Colombo, L. P. L.; Combet, C.; Contreras, D.; Crill, B. P.; Cuttaia, F.; de Bernardis, P.; de Zotti, G.; Delabrouille, J.; Delouis, J. M.; Di Valentino, E.; Diego, J. M.; Doré, O.; Douspis, M.; Ducout, A.; Dupac, X.; Dusini, S.; Efstathiou, G.; Elsner, F.; Enblin, T. A.; Eriksen, H. K.; Fantaye, Y.; Farhang, M.; Fergusson, J.; Fernández-Cobos, R.; Finelli, F.; Forastieri, F.; Frailis, M.; Fraisse, A. A.; Franceschi, E.; Frolov, A.; Galeotta, S.; Galli, S.; Ganga, K.; Génova-Santos, R. T.; Gerbino, M.; Ghosh, T.; González-Nuevo, J.; Górski, K. M.; Gratton, S.; Gruppuso, A.; Gudmundsson, J. E.; Hamann, J.; Handley, W.; Hansen, F. K.; Herranz, D.; Hildebrandt, S. R.; Hivon, E.; Huang, Z.; Jaffe, A. H.; Jones, W. C.; Karakci, A.; Keihänen, E.; Kesktalo, R.; Kiiveri, K.; Kim, J.; Kisner, T. S.; Knox, L.; Krachmalnicoff, N.; Kunz, M.; Kurki-Suonio, H.; Lagache, G.; Lamarre, J. M.; Lasenby, A.; Lattanzi, M.; Lawrence, C. R.; Le Jeune, M.; Lemos, P.; Lesgourgues, J.; Levrier, F.; Lewis, A.; Liguori, M.; Lilje, P. B.; Lilley, M.; Lindholm, V.; López-Caniego, M.; Lubin, P. M.; Ma, Y. Z.; Macías-Pérez, J. F.; Maggio, G.; Maino, D.; Mandolesi, N.; Mangilli, A.; Marcos-Caballero, A.; Maris, M.; Martin, P. G.; Martinelli, M.; Martínez-González, E.; Matarrese, S.; Mauri, N.; McEwen, J. D.; Meinhold, P.

R.; Melchiorri, A.; Mennella, A.; Migliaccio, M.; Millea, M.; Mitra, S.; Miville-Deschénes, M. A.; Molinari, D.; Montier, L.; Morgante, G.; Moss, A.; Natoli, P.; Nørgaard-Nielsen, H. U.; Pagano, L.; Paoletti, D.; Partridge, B.; Patanchon, G.; Peiris, H. V.; Perrotta, F.; Pettorino, V.; Piacentini, F.; Polastri, L.; Polenta, G.; Puget, J. L.; Rachen, J. P.; Reinecke, M.; Remazeilles, M.; Renzi, A.; Rocha, G.; Rosset, C.; Roudier, G.; Rubiño-Martín, J. A.; Ruiz-Granados, B.; Salvati, L.; Sandri, M.; Savelainen, M.; Scott, D.; Shellard, E. P. S.; Sirignano, C.; Sirri, G.; Spencer, L. D.; Sunyaev, R.; Suur-Uski, A. S.; Tauber, J. A.; Tavagnacco, D.; Tenti, M.; Toffolatti, L.; Tomasi, M.; Trombetti, T.; Valenziano, L.; Valiviita, J.; Van Tent, B.; Vibert, L.; Vielva, P.; Villa, F.; Vittorio, N.; Wandelt, B. D.; Wehus, I. K.; White, M.; White, S. D. M.; Zacchei, A.; Zonca, A. Planck 2018 results - VI. Cosmological parameters. *Astronomy & Astrophysics* **2020**, *641*, A6. doi:10.1051/0004-6361/201833910.

11. Helmenstine, A. Composition of the Universe - Element Abundance. *Science Notes* **2022**.
12. Mo, H.; van den Bosch, F.; White, S. *Galaxy Formation and Evolution*; Galaxy Formation and Evolution, Cambridge University Press, 2010.
13. Wikipedia contributors, the Free Encyclopedia. Bijective Numeration, 2020.
14. Witten, E. A Mini-introduction to Information Theory. *La Rivista del Nuovo Cimento* **2020**, *43*, 187–227. doi:10.1007/s40766-020-00004-5.
15. Bayes, T. LII. An Essay Towards Solving a Problem in the Doctrine of Chances. By the late Rev. Mr. Bayes, F. R. S. Communicated by Mr. Price, in a Letter to John Canton, A. M. F. R. S. *Philosophical Transactions of the Royal Society of London* **1763**, *53*, 370–418.
16. Oldershaw, R.L. Nature is Startling - Conformal Symmetry. *Medium* **2017**. doi:10.13140/RG.2.2.12399.74407.
17. Fewster, R.M. A Simple Explanation of Benford's Law. *The American Statistician* **2009**, *63*, 26–32.
18. Cai, Z.; Faust, M.; Hildebrand, A.J.; Li, J.; Zhang, Y. The Surprising Accuracy of Benford's Law in Mathematics. *The American Mathematical Monthly* **2020**, *127*, 217–237. doi:10.1080/00029890.2020.1690387.
19. Bekenstein, J.D. How does the Entropy/Information Bound Work? *Foundations of Physics* **2005**, *35*, 1805–1823.
20. Tegmark, M. The Mathematical Universe. *Foundations of Physics* **2008**, *38*, 101–150. doi:10.1007/s10701-007-9186-9.
21. Kempner, A.J. A Curious Convergent Series. *The American Mathematical Monthly* **1914**, *21*, 48–50.
22. Newman, M.E.J. Power laws, Pareto distributions and Zipf's law. *Contemporary Physics* **2005**, *46*, 323–351.
23. Wikipedia contributors, the Free Encyclopedia. Hyperbolic space, 2022. [Online; accessed 25-April-2023].
24. Wikipedia contributors, the Free Encyclopedia. Inverse-square law, 2023. [Online; accessed 13-May-2023].
25. Sanderson, G. Why Is Pi Here? And Why Is It Squared? A Geometric Answer to the Basel Problem. 3Blue1Brown (Brilliant), 2018. Video.
26. Bellman, R.E. *Dynamic Programming*; Dover Books on Computer Science Series, Dover Publications, 2003.
27. McCreary, P.R.; Murphy, T.J.; Carter, C. The Modular Group. *The Mathematica Journal* **2018**, *20*. doi:10.3888/tmj.20-3.
28. Formann, A.K. The Newcomb-Benford Law in Its Relation to Some Common Distributions. *PLoS One* **2010**, *5*. doi:10.1371/journal.pone.0010541.
29. Ariew, R. *Ockham's Razor: A Historical and Philosophical Analysis of Ockham's Principle of Parsimony*; University of Illinois at Urbana-Champaign, 1976.
30. Sanchez, E. Behind Benford's Law is Inherent Scarcity. *Medium* **2020**.
31. Gisin, N. Indeterminism in Physics, Classical Chaos and Bohmian Mechanics: Are Real Numbers Really Real? *Erkenntnis* **2021**, *86*, 1469–1481. doi:10.1007/s10670-019-00165-8.
32. Bateman, P.T.; Diamond, H.G. *Analytic Number Theory - An Introductory Course*; Monographs in number theory, World Scientific, 2004.
33. Jacobsen, D. Why is finding large prime numbers difficult? — Quora, 2015.
34. Roche, J. What Is Potential Energy? *European Journal of Physics* **2003**, *24*, 185. doi:10.1088/0143-0807/24/2/359.
35. Forrest, P. The Identity of Indiscernibles. In *The Stanford Encyclopedia of Philosophy*, Winter ed.; Zalta, E.N., Ed.; Metaphysics Research Lab, Stanford University, 2020.
36. Müller, K. A Magical Theorem Was Undiscovered for Thousands of Years. *Medium* **2024**.
37. Wildberger, N.J. Real Fish, Real Numbers, Real Jobs. *The Mathematical Intelligencer* **1999**, *21*, 4–7. doi:10.1007/BF03024838.

38. Hossenfelder, S. Theory and Phenomenology of Space-Time Defects. *Advances in High Energy Physics* **2014**, 2014. doi:10.1155/2014/950672.

39. Rovelli, C. The Relational Interpretation of Quantum Physics. *arXiv [quant-ph]* **2021**, [arXiv:quant-ph/2109.09170].

40. Bosso, P.; Petrucciello, L.; Wagner, F. The Minimal Length is Physical. *Physics Letters B* **2022**, 834, 137415. doi:<https://doi.org/10.1016/j.physletb.2022.137415>.

41. Zilberstein, S. Using Anytime Algorithms in Intelligent Systems. *AI Magazine* **1996**, 17, 73. doi:10.1609/aimag.v17i3.1232.

42. Burgos, A.; Santos, A. The Newcomb-Benford Law: Scale Invariance and a Simple Markov Process Based on It. *American Journal of Physics* **2021**, 89, 851–861. doi:10.1119/10.0004957.

43. Berger, A.; Hill, T.P.; Rogers, E. Benford Online Bibliography (Last accessed 21 February 2023), 2009.

44. Miller, S.J. *Benford's Law*; Princeton University Press, 2015.

45. Berger, A.; Hill, T.P. *An Introduction to Benford's Law*; Princeton University Press, 2015.

46. de Finetti, B. *Probability, Induction, and Statistics: The Art of Guessing*; WILEY SERIES in PROBABILITY and STATISTICS, New York - John Wiley, 1972.

47. Knuth, D.E. *Art of Computer Programming, Volume 2: Seminumerical Algorithms*; Pearson Education, 2014.

48. Hill, T.P. A Statistical Derivation of the Significant-Digit Law. *Statistical Science* **1995**, 10, 354–363.

49. Havil, J. *Gamma - Exploring Euler's Constant*; Princeton University Press, 2003.

50. Hill, T.P. The Significant-Digit Phenomenon. *The American Mathematical Monthly* **1995**, 102, 322–327. doi:10.1080/00029890.1995.11990578.

51. Pietronero, L.; Tosatti, E.; Tosatti, V.; Vespignani, A. Explaining the Uneven Distribution of Numbers in Nature: the Laws of Benford and Zipf. *Physica A: Statistical Mechanics and its Applications* **2001**, 293, 297–304. doi:[https://doi.org/10.1016/S0378-4371\(00\)00633-6](https://doi.org/10.1016/S0378-4371(00)00633-6).

52. Tao, T. Benford's Law, Zipf's Law, and the Pareto Distribution. *Terrytao's Blog* **2009**.

53. Hill, T.P. The First Digit Phenomenon. *American Scientist* **1998**, 86, 358–363.

54. Simon, H.A. On a Class of Skew Distribution Functions. *Biometrika* **1955**, 42, 425–440.

55. Fang, G.; Chen, Q. Several Common Probability Distributions Obey Benford's Law. *Physica A: Statistical Mechanics and its Applications* **2020**, 540, 123129. doi:<https://doi.org/10.1016/j.physa.2019.123129>.

56. Wolfram Research. BenfordDistribution Built-in Function. *Wolfram Language and System Documentation Center* **2010**, p. Background and Context.

57. Berger, A.; Hill, T.P. The Mathematics of Benford's Law: a Primer. *Statistical Methods and Applications* **2021**, 30, 779–795. doi:10.1007/s10260-020-00532-8.

58. Jamain, A. Benford's law. Master's thesis, Imperial College of London, 2001.

59. Gonsalves, R.A. Benford's Law - A Simple Explanation. *Medium* **2020**.

60. DK. A Treatise On The Newcomb-Benford Law. *Medium* **2021**.

61. Wikipedia contributors, the Free Encyclopedia. Zipf's law, 2022. [Online; accessed 1-April-2022].

62. Hendry, R.F. Structure, Scale and Emergence. *Studies in History and Philosophy of Science Part A* **2021**, 85, 44–53. doi:<https://doi.org/10.1016/j.shpsa.2020.08.006>.

63. Williams, P.M. Bayesian Conditionalisation and the Principle of Minimum Information. *The British Journal for the Philosophy of Science* **1980**, 31, 131–144.

64. Montemurro, M.A. Beyond the Zipf-Mandelbrot Law in Quantitative Linguistics. *Physica A: Statistical Mechanics and its Applications* **2001**, 300, 567–578. doi:[https://doi.org/10.1016/S0378-4371\(01\)00355-7](https://doi.org/10.1016/S0378-4371(01)00355-7).

65. Wikipedia contributors, the Free Encyclopedia. Zeta distribution, 2021. [Online; accessed 4-April-2022].

66. Finch, S.R. *Mathematical Constants*; Encyclopedia of Mathematics and its Applications, Cambridge University Press, 2003.

67. Held, L.; Bové, D.S. *Applied Statistical Inference: Likelihood and Bayes*; Springer Berlin Heidelberg, 2013.

68. Bak, P.; Tang, C.; Wiesenfeld, K. Self-organized criticality: An explanation of the 1/f noise. *Physical Review Letters* **1987**, 59, 381–384. doi:10.1103/PhysRevLett.59.381.

69. Szendro, P.; Vincze, G.; Szasz, A. Pink-noise Behaviour of Biosystems. *European Biophysics Journal* **2001**, 30, 227–231. doi:10.1007/s002490100143.

70. Mattey, G.J. Lecture Notes on "Critique of Pure Reason": Appearance and Thing in Itself. Philosophy 175 Home Page, UC Davis Philosophy Department, 1997.

71. Stone, J.V. *Information Theory: A Tutorial Introduction*; Tutorial Introduction Book, Tutorial Introductions, 2015.

72. Fechner, G.T.; Boring, E.G.; Adler, H.E.; Howes, D.H. *Elements of Psychophysics*; Vol. 1, *Henry Holt edition in psychology*, New York - Holt, Rinehart and Winston, 1966.

73. Schatte, P. On Benford's Law to Variable Base. *Statistics and Probability Letters* **1998**, *37*, 391–397. doi:[https://doi.org/10.1016/S0167-7152\(97\)00142-9](https://doi.org/10.1016/S0167-7152(97)00142-9).

74. Rives, J. The Zero Delusion. *ResearchGate (Preprint)* **2023**, p. 60. Universe Intelligence, doi:10.13140/RG.2.2.20063.74400.

75. Foster, J.E. A Number System without a Zero-Symbol. *Mathematics Magazine* **1947**, *21*, 39–41.

76. Wikipedia contributors, the Free Encyclopedia. Differential entropy, 2024. [Online; accessed 10-September-2024].

77. Baillie, R. Sums of Reciprocals of Integers Missing a Given Digit. *The American Mathematical Monthly* **1979**, *86*, 372–374.

78. Wildberger, N.J. Finite versus Infinite and Number Systems. In *Sociology and Pure Mathematics*; Insights into Mathematics, YouTube Channel, 2022.

79. Dowek, G. Real Numbers, Chaos, and the Principle of a Bounded Density of Information. Theory and Applications. - Computer Science Symposium in Russia; A. A., B.; Shur, A.M., Eds.; , 2013; Vol. 7913, *LNTCS*, [[978-3-642-38535-3](#)].

80. Santo, F.D., Undecidability, Uncomputability, and Unpredictability; The Frontiers Collection, Springer International Publishing: Cham, 2021; chapter Indeterminism, Causality and Information: Has Physics Ever Been Deterministic?, pp. 63–79. doi:10.1007/978-3-030-70354-7_5.

81. Bailey, B. The Cost of Accuracy. *Semiconductor Engineering (Low Power - High Performance)* **2018**.

82. Schmelzer, T.; Baillie, R. Summing a Curious, Slowly Convergent Series. *The American Mathematical Monthly* **2008**, *115*, 525–540.

83. Farhi, B. A Curious Result Related to Kempner's Series. *The American Mathematical Monthly* **2008**, *115*, 933–938. doi:10.2307/27642640.

84. Lubeck, B.; Ponomarenko, V. Subsums of the Harmonic Series. *The American Mathematical Monthly* **2018**, *125*, 351–355.

85. Howson, C., The Logic of Bayesian Probability. In *Foundations of Bayesianism*; Corfield, D.; Williamson, J., Eds.; Springer Netherlands: Dordrecht, 2001; Vol. 24, pp. 137–159. doi:10.1007/978-94-017-1586-7_6.

86. Joyce, J. Bayes' Theorem. In *The Stanford Encyclopedia of Philosophy*, Spring ed.; Zalta, E.N., Ed.; Metaphysics Research Lab, Stanford University, 2019.

87. Walker, S.; Cronin, L. Time Is an Object. *Aeon* **2023**.

88. Wikipedia contributors, the Free Encyclopedia. Odds algorithm, 2021. [Online; accessed 4-June-2021].

89. Girdhar, Y.; Dudek, G. Optimal Online Data Sampling or How to Hire the Best Secretaries. Canadian Conference on Computer and Robot Vision, 2009, pp. 292–298. doi:10.1109/CRV.2009.30.

90. Wikipedia contributors, the Free Encyclopedia. Doppler effect, 2024. [Online; accessed 4-February-2024].

91. Tipler, P.A.; Llewellyn, R. *Modern Physics*; W. H. Freeman, 2012.

92. Milne, J.J. *An Elementary Treatise on Cross-ratio Geometry: With Historical Notes*; The University Press, 1911.

93. Krakauer, D. Complexity, Agency, and Information. *Sean Carroll's Mindscape* **2023**.

94. Young, N.J. Linear fractional transformations in rings and modules. *Linear Algebra and its Applications* **1984**, *56*, 251–290. doi:[https://doi.org/10.1016/0024-3795\(84\)90131-9](https://doi.org/10.1016/0024-3795(84)90131-9).

95. Stillwell, J. Modular Miracles. *The American Mathematical Monthly* **2001**, *108*, 70–76. doi:10.1080/00029890.2001.11919727.

96. Poincaré, H. *Science and Hypothesis* (1905); Read Books Limited, 2016.

97. Wikipedia contributors, the Free Encyclopedia. Poincaré disk model, 2023. [Online; accessed 20-January-2023].

98. Wikipedia contributors, the Free Encyclopedia. Real projective line, 2023. [Online; accessed 29-January-2024].

99. Trillas, E. Sobre Funciones de Negacion en la Teoria de Conjuntos Difusos. *Stochastica* **1979**, *3*, 47–60.

100. Karapetoff, V. Restricted Theory of Relativity in Terms of Hyperbolic Functions of Rapidities. *The American Mathematical Monthly* **1936**, *43*, 70–82.

101. Russell, N. Framing Lorentz Symmetry. *CERN Courier* **2004**.

102. Abramowitz, M.; Stegun, I.A. *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables*; Applied mathematics series, Dover Publications, 1965.

103. Richeson, D.S. *Euler's Gem: The Polyhedron Formula and the Birth of Topology*; Princeton Science Library, Princeton University Press, 2019.
104. Jaynes, E.T. Information Theory and Statistical Mechanics. In *Statistical Physics*; Ford, K.W., Ed.; Benjamin, 1963.
105. Bekenstein, J.D. Black Holes and the Second Law. *Lettere al Nuovo Cimento (1971-1985)* **1972**, *4*, 737–740. doi:10.1007/BF02757029.
106. Wikipedia contributors, the Free Encyclopedia. Tetration, 2024. [Online; accessed 31-January-2024].
107. Maier, J.F.; Eckert, C.M.; Clarkson, P.J. Model Granularity in Engineering Design, Concepts and Framework. *Design Science* **2017**, *3*, e1. doi:10.1017/dsj.2016.16.
108. Barbeau, E.J. *Polynomials*; Problem Books in Mathematics, Springer New York, 2003.
109. Lidl, R.; Niederreiter, H. *Finite Fields*; Number v. 20,pt. 1 in EBL-Schweitzer, Cambridge University Press, 1997.
110. Seiferth, D.; Sollich, P.; Klumpp, S. Coarse Graining of Biochemical Systems Described by Discrete Stochastic Dynamics. *Physical Review E* **2020**, *102*. doi:10.1103/physreve.102.062149.
111. Weinan E. *Principles of Multiscale Modeling*; Cambridge University Press, 2011.
112. Clenshaw, C.W.; Olver, F.W.J. Beyond Floating Point. *Journal of the Association for Computing Machinery* **1984**, *31*, 319–328.
113. Peltzer, A.R. The Riemann Zeta Distribution. phdthesis, University of California, Irvine, 2019.
114. Mitzenmacher, M.; Upfal, E. *Probability and Computing: Randomization and Probabilistic Techniques in Algorithms and Data Analysis*; Cambridge University Press, 2017.
115. Ken-Iti, S. *Lévy Processes and Infinitely Divisible Distributions*; Cambridge studies in advanced mathematics, Cambridge University Press, 1999.
116. Castellanos, D. The Ubiquitous Pi. *Mathematics Magazine* **1988**, *61*, 67–98.
117. Billingsley, P. *Probability and Measure*; Wiley Series in Probability and Statistics, Wiley, 1995.
118. Mathar, R.J. Survey of Dirichlet Series of Multiplicative Arithmetic Functions, 2011.
119. Ornstein, D. Bernoulli Shifts with the Same Entropy Are Isomorphic. *Advances in Mathematics* **1970**, *4*, 337–352. doi:https://doi.org/10.1016/0001-8708(70)90029-0.
120. Sinharay, S. Discrete Probability Distributions. In *International Encyclopedia of Education (Third Edition)*, Third Edition ed.; Peterson, P.; Baker, E.; McGaw, B., Eds.; Elsevier: Oxford, 2010; pp. 132–134. doi:10.1016/B978-0-08-044894-7.01721-8.
121. Alzer, H.; Choi, J. Four Parametric Linear Euler Sums. *Journal of Mathematical Analysis and Applications* **2020**, *484*, 123661. doi:https://doi.org/10.1016/j.jmaa.2019.123661.
122. Grove, L.C.; Benson, C.T., Finite Groups in Two and Three Dimensions. In *Finite Reflection Groups*; Springer New York: New York, NY, 1985; pp. 5–26. doi:10.1007/978-1-4757-1869-0_2.
123. Wikipedia contributors, the Free Encyclopedia. Triangle group, 2022. [Online; accessed 19-May-2023].
124. Pellikaan, R., Codes, Curves, and Signals. In *Codes, Curves, and Signals: Common Threads in Communications*; Vardy, A., Ed.; Springer US: Boston, MA, 1998; chapter The Klein Quartic, the Fano Plane and Curves Representing Designs, pp. 9–20.
125. Baez, J.C. Klein's Quartic Curve. *John Baez's Blog* **2013**.
126. Baez, J.C. The Octonions. *Bulletin of the American Mathematical Society* **2002**, *39*, 145–205.
127. Wilson, R.A. The Monster Is a Hurwitz Group. *Journal of Group Theory* **2001**, *4*, 367–374. doi:doi:10.1515/jgth.2001.027.
128. Witten, E. Magic, Mystery, and Matrix. *Notices of the American Mathematical Society* **1998**, *45*, 1124–1129.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.