Preprints.org

Article

A Hyperbolic Sum Rule for Probability: Solving Recursive ("Chicken & Egg") Problems

Michael C. Parker [*] , Chris Jeynes , Stuart D. Walker

Article

A Hyperbolic Sum Rule for Probability: Solving Recursive ("Chicken & Egg") Problems

M.C.Parker [1,*], C.Jeynes [2] and S.D.Walker [3]

[1] School of Computer Sciences & Electronic Engineering, University of Essex, Colchester

[2] Independent scholar, Tredegar NP22 4LP;

[3] School of Computer Sciences & Electronic Engineering, University of Essex, Colchester; stuwal@essex.ac.uk

[*] Correspondence: mcpark@essex.ac.uk

Abstract: We prove that the probability of "*A* OR *B*", where *A* and *B* are events or hypotheses that are recursively dependent (OR is the logico-probabilistic operator) is given by a "Hyperbolic Sum Rule" (**HSR**) relation isomorphic to the hyperbolic-tangent double-angle formula, and that this HSR is Maximum Entropy (**MaxEnt**). The possibility of recursive probabilities is excluded by the "Conventional Sum Rule" (**CSR**) which we also prove MaxEnt (within its narrower domain of applicability). The concatenation property of the HSR is exploited to enable analytical, consistent and scalable calculations for multiple recursive hypotheses: such calculations are not conveniently available for the CSR, and are also intrinsic to current artificial intelligence and machine learning architectures presently considered intractable to analytical study and methodological validation. We also show that it is as reasonable to state that "*probability is physical*" (it is not merely a mathematical construct) as it is to state that "*information is physical*" (now recognised as a truism of communications network engineering). We relate this treatment to the physics of Quantitative Geometrical Thermodynamics which is defined in complex hyperbolic (Minkowski) spacetime, and show how the HSR is isomorphic to other physical quantities including certain components important for digital signal processing.

Keywords: QGT; entropy; Venn diagram; Bayes' Theorem; AI; ML; DSP

1. Introduction

1.1. Probability in Science

Probability has always been extraordinarily difficult to tie down. Edwin Jaynes made fundamental contributions to the subject, and in his magisterial monograph (published posthumously: "*Probability Theory: The Logic of Science*"[i]) spends two chapters on its principles and elementary applications, commenting on its 'weird' and 'complicated' history. Indeed, Jaynes' motivation was aimed at helping the interested reader who already has "*a previous acquaintance with probability and statistics*" to essentially "*unlearn*" much of what they may have previously learned!

The ubiquity (and longevity) of fallacies in the use of statistics indicates its difficulty (on the misuse of the "p-value" see for example Goodman 2008[ii] and Halsey 2019[iii]; on the persistence of fallacies see Smaldino & McElreath 2016[iv]). A significant part of the problem may be related to the fact that the "probability" of an event is apparently not a property solely of "external reality": since there must be *someone* assessing the probability, it must be a function of not only *what* information already exists about the event but also *who* knows. The fact that our estimates of the probability of some event invariably involve our prior knowledge, combined with the fact that all knowledge ultimately involves the properties of recursive statements (by Gödel's Theorem[v]; this point has recently been elaborated by Jeynes *et al.*[vi]) mean that some part of this difficulty must be due to the neglecting, in current (simplified) treatments, of recursive dependencies (such as the "Chicken and Egg" problem, on which see §2.3 below).

In 1946 R.T.Cox observed, acutely, that the concept of *probability* mixes two separate concepts ("*the idea of frequency in an ensemble and the idea of reasonable expectation*"[vii]) which are now represented

by two schools usually called the "frequentists" and the "Bayesians". We point out that Enßlin *et al*.[viii] comment that, "*The different views on probabilities of frequentists and Bayesians are addressed by … Allen Caldwell*[ix] *[who] builds a bridge between these antagonistic camps by explaining frequentists' constructions in a Bayesian language*", and we will assume that Caldwell is right and therefore that we can ignore the philosophical distinction between the frequentists and the Bayesians.

Building on Cox's fundamental work we will here derive a rigorous treatment of recursive probability which is of general applicability. In particular, we will treat probability as a *physical quantity* grounded in hyperbolic Minkowski spacetime, and obeying all the appropriate physical laws. The close relation of probability to the new Quantitative Geometrical Thermodynamics[x] (**QGT**: this constructs the quantity *info-entropy* by treating *information* and *entropy* as Hodge duals) is due to the fact that the (hyperbolic) *entropic velocity* $q'{\equiv}\mathrm{d}q/\mathrm{d}x$ in QGT is dimensionless and has properties akin to a probability ($0{\leq}q'{\leq}1$); noting that q' is isomorphic to the more familiar kinematic velocity $\dot{x} \equiv \mathrm{d}x/\mathrm{d}t$ ($0{\leq} \dot{x} {\leq}c$, where c is the speed of light) in hyperbolic (Minkowski) spacetime (see ref.[10]), and which also obeys the well-known velocity addition theorem of Special Relativity.

The treatment of probability as a *physical* quantity (and obeying *physical* laws) is as fundamental a concept as the well-known statement by Rolf Landauer that "*Information is Physical*"[xi].

We are used to treating both *probability* and *information* anthropomorphically (that is, depending on what you or I might be expected to know): here we will establish an impersonal sense for *probability*, in the same way that Landauer insisted on the (impersonal) *Shannon entropy* sense of his *Information*. Note that each of information and entropy both use and require probabilistic quantities in their fundamental definitions; and also note that the consequences of treating *information* as being a quantity as physical as is *energy* (for example) has led to important insights, and not only into basic engineering problems of the global internet (for an example of which see Parker *et al*. 2015[xii]). We expect similar advances to follow from recognising *probability* as also being an equally physical quantity.

The relationship between probability and the quantification of information using the Shannon entropy is well understood mathematically, but its interpretation as a physical theory has been convoluted. Although the Shannon entropy was quickly recognised as important, Edwin Jaynes' formulation of Maximum Entropy (**MaxEnt**) theory, where the Shannon metric plays a key role, was initially controversial and took some decades to achieve acceptance (see Jaynes' 1978 summary[xiii]). However, MaxEnt as a powerful scientific and engineering tool has helped considerably to underpin the physicality of *information*, and therefore also acts as a support to the underlying assertion of this paper (paraphrasing Landauer) that "*Probability is Physical*". We will also explore the implications of this assertion.

Since the concept of MaxEnt will be centrally important in this work, it is worth adding Jaynes' authoritative definition (1982)[xiv]: "*The MAXENT principle, stated most briefly, is: when we make inferences based on incomplete information, we should draw them from that probability distribution that has the maximum entropy permitted by the information we do have*".

1.2. Probability is Physical

Torsten Enßlin's treatment[xv] of information as a field is interesting in this context: he considers that a "physical field has an infinite number of degrees of freedom since it has a field value at each location of a continuous space" where he is specifically considering imaging problems of astrophysical datasets. Enßlin et al. [ref.8] treat information informally as an anthropomorphic concept: to the question, "What is information?" they answer "Anything that changes our minds and states of thinking!" But here we will treat information (and probability) as physical quantities, not as anthropomorphic concepts; and especially noting that infinite quantities are inimical to physical phenomena. In particular, in our QGT treatment, information is formally defined as a physical quantity (albeit in terms making full use of the properties of analytical continuation, which is itself closely related to fundamental physical precepts such as causality[xvi] and square-integrability ensuring finite physical quantities[xvii]), so that the number of degrees of freedom are finite and may be very small, as is observed for the geometrical entropy of isotopes of the helium nucleus[xviii]. Such

results for information were already pointed out by Parker & Walker (2004)[xix] who investigated the residues of a meromorphic function (that is, a function analytic nearly everywhere) due to the presence of isolated singularities in the complex plane (singularities which are entirely analogous to particles in their behaviour); and show that the information (entropy) of such a function is simply given by the sum of the residues.

This is immediately applicable to the Schrödinger equation: it is interesting that we will conclude (Eq.7) that the appropriate **Sum Rule** for recursive probabilities has a **hyperbolic form**, and we will draw out the relation of this to the (hyperbolic) *entropic velocities* in the QGT formalism[xx], in which the *entropic* Uncertainty Principle and *entropic* isomorphs of the Schrödinger equation may be derived from the *entropic* Liouville Theorem; all based on the Boltzmann constant as the relevant quantum of entropy. Of course, QGT is constructed in a hyperbolic (complex) Minkowski 4-space (see §18.4 in Roger Penrose's "*Road to Reality*"[xxi]; as another pertinent example, Maxwell's electro-magnetic field is a hyperbolic version of the Cauchy-Riemann Equations: see Courant & Hilbert[xxii] vol.II ch.III §2 Eq.8 *passim*).

The ramifications of this are very wide. Parker & Jeynes [ref.10] have already shown the relevance of QGT to the stability and structure of spiral galaxies (using the properties of black holes), and also that *entropy production* (dS/dt) is conserved even in relativistic Maximum Entropy systems[xxiii]. These issues up to now have been treated as problems in quantum gravity, and Matt Visser[xxiv] very helpfully reviews conservative entropic forces (in Euclidean space and non-relativistically, although he comments that representing general relativity entropically should be possible). Note also that Visser suggests that the negative entropies that appear in his treatment can be regarded as *information*, citing Brillouin's idea of "negentropy" which Parker & Jeynes [ref.10] have shown to be a subtle misapprehension of information and entropy (which are actually Hodge duals).

We believe that much progress may be made by using the coherent formalism of QGT (defined in hyperbolic space) which has been shown to apply to both quantum mechanical and gravitational systems (that is, at all scales from sub-atomic to cosmic: see [refs.18,10]), and since quantum mechanics is built on probabilities, the demonstration here that the general recursive sum rule for probabilities is hyperbolic is a significant conceptual regularisation. This conclusion is reinforced by Knuth's demonstration[xxv] that: "*The sum and product rules, which are familiar from, but not unique to, probability theory, arise from the fact that logical statements form a distributive (Boolean) lattice, which exhibits the requisite symmetries*". Moreover, Jaeger[xxvi] reviews a variety of treatments, some of which involve theories of generalized probability, aimed at deriving quantum mechanics from information theory.

The basic isomorphism for the hyperbolic sum rule for probabilities (that we will prove, see Eq.7) is the (purely mathematical) double-angle identity for the hyperbolic tangent function:

$$\tanh(a+b) = \{\tanh a + \tanh b\}/\{1+ \tanh a \cdot \tanh b\} \tag{1a}$$

Another interesting and very simple isomorphism is the well-known relativistic sum rule for velocities $\{u, v\}$, given by Jackson[xxvii] in his well-known textbook (in the context of a discussion of aberration and the Fizeau experiment; §11.4, Eq.11.28):

$$w \equiv \text{Sum}(u, v)$$

$$w/c = \{u/c + v/c\}/\{1 + uv/c^2\} \tag{1b}$$

where c is the speed of light. Jackson comments that if $u = c$ then also $w = c$, which is an "*explicit statement of Einstein's second postulate*" (c is a constant).

This latter (Eq.1b) is clearly physical (since c is involved) where the former (Eq.1a) is a mathematical identity. Note also that in optics the basic formula for the two-layer Fabry-Perot cavity (etalon) is well-known[xxviii]:

$$r_3 = \{r_2 + r_1\Phi\}/\{1 + r_2 r_1\Phi\} \tag{1c}$$

$$\Phi \equiv \exp(-2ik\Delta z)$$

where the overall scattering (reflectivity) coefficient r_3 is due to a pair of sequential Fresnel reflections (r_1, r_2) separated by a distance Δz for a light ray of propagation constant k; and where we note that light is the physical phenomenon *par excellence* exhibiting the physics of Special Relativity within the

context of hyperbolic (Minkowski) spacetime. Corzine *et al.*[xxix] demonstrate that this formula is closely related to the hyperbolic addition rule (Eq.1a) specifically by using a hyperbolic tangent substitution which dramatically simplifies the use of the formula in real (multilayer) cases – see Appendix F for additional and related discussion.

This approach has recently been supported in an interesting way by Skilling & Knuth[xxx], who conclude: "*it must be acknowledged, quantum theory works. So does probability. And the two are entirely mutually consistent.*" Their argument shows logical reasons why *probability* should be regarded as *physical*.

Skilling & Knuth claim not to be interested (for these purposes) in the distinction between ontology and epistemology. They say (*ibid.* §3.5):

The ontology–epistemology divide is, for quantitation at least, a distinction without a difference. A bit of information carries no flag to inform us whether it was assigned by a conscious agent or by a mechanical switch. Our job in science is to make sense of our observations, not to indulge in empty disputation between isomorphic views. Our goal here is derivation of a calculus fit for general purpose. Ontology and epistemology share the same symmetries, the same rules, and the same assignments. So they share a common calculus.

which is suggestive of Karen Barad's (2007[xxxi]) insistence that the distinction between ontology and epistemology is not a real one, and therefore speaking strictly we should refer to "*onto-epistemology*" (*ibid.* p.43). Skilling & Knuth also say (*ibid.* §1):

But, if our object can perturb a partner object, then by symmetry the partner object can also perturb our object. We could assign either role to either.

Our calculus, whatever it is, must be capable of representing such interactions … This insight that *interactions are basic* is the source of "quantum-ness."

Again, this recalls Barad's thesis that "*the primary ontological unit is the phenomenon*" (*ibid.* p.333). However, when Skilling & Knuth (*ibid.* §4) say, "*We start with an identifiable object*", this is directly contradicted by Barad, who asserts that "objects" do not have "*an inherent ontological separability*" (*ibid.* p.340); that is, strictly speaking, identifiable objects do not actually exist *per se* (since everything is entangled with everything else). But Skilling & Knuth are not aiming at philosophical precision, only at a demonstrable computability; for these purposes such fine distinctions do not matter. They are right to avoid metaphysical considerations in scientific work: although when wider social implications are important it may be necessary to consider the metaphysics (see for example [ref.6]).

However, it turns out that the inescapable human dimension appears to be especially pronounced in *probability*, in the sense that the very *idea* of a probability entails one's personal state of knowledge or ignorance (and Michael Polanyi insisted long ago that all knowledge is necessarily personal[xxxii]). Howson & Urbach[xxxiii] have carefully explained why, although Bayesian (and Maximum Entropy) methods are fully (and helpfully) rational: "*there seems to be no way of 'objectively' defining prior probabilities … this is really no weakness [since] it allows expert opinion due weight, and is a candid admission of the personal element which is there in all scientific work*". Assessing probabilities necessarily entails assessing uncertainties, and this must always involve some value judgments: although we may do our best to speak rationally about such judgments, it cannot be excluded that different people will (rationally) come to different conclusions.

Note that although scientists have a duty to argue rationally, non-scientists also normally behave rationally. Rationality is a property of humans, and only a rather small subset of humans are scientists.

1.3. Maximum Entropy

It is necessary to make a few initial simple remarks about Maximum Entropy (**MaxEnt**) methods, to clarify the discussion. Jaynes said in 1957[xxxiv]: "*The guiding principle is that the probability distribution over microscopic states which has maximum entropy subject to whatever is known, provides the most unbiased representation of our knowledge of the state of the system. The maximum-entropy distribution is the broadest one compatible with the given information; it assigns positive weight to every possibility that is not ruled out*

by the initial data." In our derivation here of the Hyperbolic Sum Rule we emphasise that the proper application of MaxEnt methods precludes the surreptitious introduction of tacit assumptions ("knowledge"). That is, *all* the "priors" (prior knowledge that conditions the data) must be stated explicitly (and our ignorance implies that some priors must sometimes be estimated, which may involve personal value judgments).

A fully Bayesian analysis requires that all prior knowledge is explicitly stated, including the "knowledge" that there is no knowledge available, in which case an "unbiased" estimate is required. This is usually stated in terms of the "Principle of Indifference" (**PI**), but unfortunately there are a set of ("Bertrand") paradoxes which appear to invalidate the PI. But Parker & Jeynes[xxxv] have resolved these paradoxes using QGT methods by supplying the missing prior information (in the form of the scale invariance condition).

It is essential to handle the "priors" correctly: our Eq.6 below (which is for the recursive case) asserts a hyperbolic relation for the compound "$p(A \text{ OR } B \mid C)$" where in general A may depend *both* on prior conditions C, and *also* on B (that is, involving the probabilities of $A|BC$ and $B|AC$). But if $\{A,B\}$ are *recursively* dependent, then $p(A|BC)$ simplifies to $p(A|C)$ (because, for recursive $\{A,B\}$, $p(A|B)= p(A)$ since B is ultimately dependent on A; and similarly, $p(B|A)= p(B)$ since A is ultimately dependent on B) giving the general relation describing the "*Hyperbolic Sum Rule*", **HSR** (Eq.7).

In the **Appendices** we explore the Maximum Entropy (MaxEnt) properties of the HSR, including a proof that it really is MaxEnt. We first show how to impose the MaxEnt criterion (**Appendix A**) using the Partition Function description, and then show (**Appendix B**) how other simple sum rules may be "MaxEnt" but are otherwise inadmissible. We prove explicitly that the HSR is MaxEnt (**Appendix C**) and also generalise it for multiple recursive hypotheses (**Appendix D**). We show that the Conventional Sum Rule (**CSR**) is also MaxEnt (**Appendix E**) within its own domain of applicability, and we also generalise the CSR for multiple hypotheses. Both are MaxEnt, that is, neither the HSR nor the CSR 'smuggle' in inadvertent or implicit assumptions within their specified contexts; yet, the HSR encompasses a wider domain of physical application that includes recursion between phenomena, whereas the CSR *a priori* excludes recursion (although there may still be various mutual dependencies). The HSR must also be used where the properties of the recursion are unknown; that is to say, a set of phenomena are known to be correlated, but the mechanism (or ordering) of causation is not known.

Finally in **Appendix F**, we show the immediate relevance of this treatment to digital signal processing; in particular the handling of "infinite impulse response" and "finite impulse response" filters.

It is important here to point out that the fact that an entity is Maximum Entropy does *not* mean that the entity has no structure (even though MaxEnt necessarily implies "maximum ignorance"). The reality is more nuanced. For example, we have shown that, *at a certain scale*, the alpha particle is a (MaxEnt) *unitary entity* (than which exists no simpler) [ref.18]. But of course, at a *different* scale we may see the alpha's constituent four nucleons (protons and neutrons).

Also, being MaxEnt does not preclude change: for example, free neutrons decay (with a half-life calculated *ab initio* from QGT by Parker & Jeynes 2023[xxxvi]). The most extreme known entity that is MaxEnt is the black hole, and black holes necessarily grow (proved by Parker & Jeynes in 2021 [ref.23] and confirmed in 2023[xxxvii]). Nor does the fact that some entities are unconditionally stable mean that the Second Law doesn't apply to them. On the contrary! The matter radius of the alpha is correctly calculated *ab initio* from QGT [ref.18].

The physical principle of Maximum Entropy embodies the (physical) Second Law of Thermodynamics. Both the CSR and the HSR are MaxEnt, and therefore each also embodies important aspects of the Second Law. Moreover, the *probabilities* calculated by either Rule (each with its own particular domain of applicability) refer to the probability either of the events themselves or of our reasonable expectations; these are both physical things since both we and they are physical. Either way, our conclusion is underlined that *Probability is Physical*.

2. Analysis

2.1. The Conventional Sum Rule (I)

Consider $p(A)$ which is a representation either of "*the probability of event A occurring*" or, nearly equivalently (logically speaking), "*the probability of hypothesis A being true*".   Of course, hypotheses and events are entirely different things, but for convenience we will here usually speak of "events" without any loss of generality (ignoring the philosophical frequentist/Bayesian distinction, see §1.1).

The probability that A OR B will happen is conventionally given (by the "Conventional Sum Rule", **CSR**) as the sum of the separate probabilities that A and B, respectively, will each happen, minus the probability that both of {A,B} will happen together (in general {A,B} are not independent):

$$p(A \text{ OR } B) = p(A) + p(B) - p(A \text{ AND } B) \tag{2a}$$

This is simply extended to the conditional case treating the probability of A OR B given some conditionality C:

$$p(A \text{ OR } B \mid C) = p(A \mid C) + p(B \mid C) - p(AB \mid C) \tag{2b}$$

where A and B are two events, conditional either on each other or some other event(s); in other words, in this case C represents all the possible conditionalities, including $p(A \mid B)$ etc.   Eq.2b represents the CSR.   But if {A,B} are recursively dependent (the "chicken and egg" scenario) Eq.2b is invalid since C either cannot be specified explicitly or cannot be specified at all.

Intrinsic to Eqs.2a&b is also the corresponding "Product Rule", more commonly known as Bayes' Theorem:

$$
\begin{aligned}
p(A \text{ AND } B \mid \text{C}) \quad &\equiv p(AB \mid C) \\
&= p(A \mid BC)\, p(B \mid C) \\
&= p(B \mid AC)\, p(A \mid C)
\end{aligned}
\tag{2c}
$$

Here, Bayes' Theorem appears in the latter two expressions of Eq.2c, which yield the same result since Bayes' Theorem is commutative: $p(AB) = p(BA)$.   In particular, Bayes' Theorem represents a *choice* as to which of the possible conditionalities should take priority.   The logical AND function can be calculated in different ways according to which of the conditionalities is considered to be prior.

We emphasise: although we seek here an alternative form of the Sum Rule (Eq.2b) that takes recursion into account, Bayes' Theorem (Eq.2c) remains valid.

Such recursive issues arise when we start considering which of the conditionalities is prior (an issue of causation): there appears to be a temporal dependence between the two conditionalities since one must surely occur earlier in time than the other?   But note that Special Relativity teaches us that if two *spatially-separated* events are causally connected (that is, they lie within each other's light cone) then temporal priority is always preserved whatever the frame of reference (FoR). However, if the events are not causally connected in spacetime then for any FoR where the event of A occurs before that of B, another FoR exists where event B occurs before A.   If causal connectedness cannot be determined, either due to the events being spacelike separated, or perhaps because no physical mechanism can be identified to causally connect the events, then there is an ambiguity as to which of the events is prior (and of course, correlation does not imply causation).

An example of a recursive system is a multilayer etalon represented by Eq.1c.

2.2. The Venn Diagram

Jaynes showed that the derivation of the product rule (Eq.2c) from the basic laws of Boolean logic is straightforward [ref.1].   However, performing the same derivation for the "Sum Rule" of probability is less obvious.   In particular, it is worth noting that Jaynes specifically avoids the use of the Venn diagram as a means to justify Eq.2b, preferring to interpret propositions in terms not of sets but as probability distributions carrying incomplete information.   It is useful to note in this context that there is intense current interest in how to handle incomplete (or erroneous) "information" in a formal Bayesian analysis (see for example Zhang *et al.*[xxxviii]).

Jaynes therefore calls the Venn diagram a '*useful device*' but comments that it can '*mislead*' [our emphasis].   The Venn diagram is good for representing binary logical operations (see **Figure 1**) but in general correctly represents neither temporality (including aspects of causality or recursiveness)

nor probability distributions.   We add that it is also not obvious how the Venn diagram (or the CSR of Eq.2b) could be extended to include multiple hypotheses and/or conditionalities.
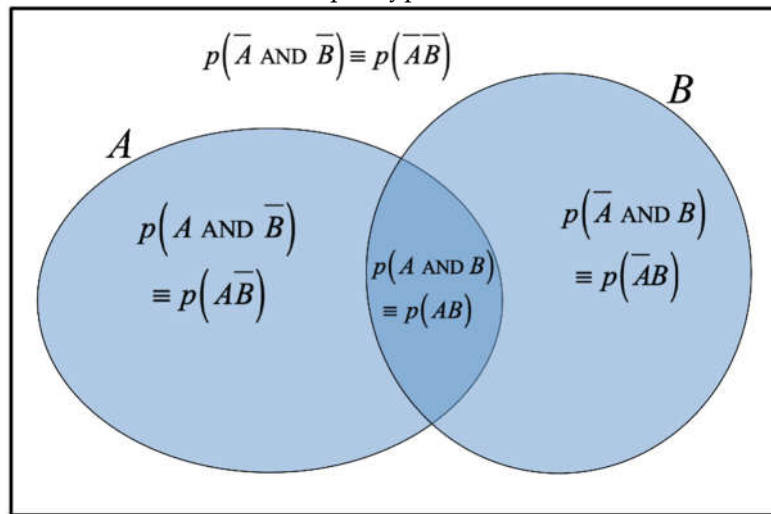

Figure 1. Classical Venn diagram showing the graphical relationship of logical quantities, with the overall enclosed (shaded) area representing the Conventional Sum Rule (CSR) for the probability of *A* OR *B*.

Jaynes' derivation [ref.1] of the "Product Rule" (Eq.2c) does not need to involve de Morgan's theorems (which follow from the rules of Boolean logic).   But his derivation of the "Sum Rule" (Eq.2b) adopts the original schema of Cox [ref.7] to also define an auxiliary function S so that $v \equiv S(u)$, where $u \equiv p(A \mid C)$ such that $v \equiv p(A \top C)$ and shows that:

$$S[S(x)] = x \quad \text{for } 0 \le x \le 1 \tag{3a}$$

that is, the function $S(x)$ is self-reciprocal:

$$S(x) = S^{-1}(x) \tag{3b}$$

Using this reciprocal property of S to calculate inverse probabilities avoids the problem that there is no quotient operator in Boolean logic.

2.3. Recursion: Chicken & Egg

Dependencies can be logical or temporal.   Causality (a temporal dependence) also implies the possibility of recursion: this is clear from the "Chicken & Egg" conundrum, on which see Simmons *et al.*[xxxix] (who consider whether hummingbirds pollinate lots of species because there are lots of hummingbirds or whether there are lots of hummingbirds because there are lots of species to pollinate) or Harvey *et al.*[xl] (who consider two forms of recursive cyclic network to show how subtle recursion relations may be).   The chicken entails a prior egg, which entails a prior chicken …

Formally: recursion is present when $p(A \mid B) > 0$ and $p(B \mid A) > 0$ both together.   Conversely, recursion is not present when C (see Eq.2b) is independent of either A or B.   This is because expressions like $p(A \mid C)$ imply that all dependencies are subsumed in C.   That is, if A depends on B then there exists a C that will say this with no loss of generality.

Here we will informally derive the "Hyperbolic Sum Rule" that we will prove formally in §4.1.

There is an apparent dilemma (when applying Bayes' Theorem) of having to make what appears to be an arbitrary decision as to which conditionality is the prior.   Indeed, when considering the classic "Chicken and Egg" dilemma, having argued in one way, we undergo a *Gestalt* change and suddenly find ourselves equally arguing in the opposite manner!   The recursiveness of this dilemma is self-evident; a different prior may be appropriate depending on where the analysis commences. Curiously, the case of the two-layer Fabry-Perot cavity (Eq.1c) involves optical events that lie *on* each other's light cone; that is to say, as a physical phenomenon it is directly on the boundary of being causally connected or disconnected.   In this case an intrinsic ambiguity of priority exists.   If either

may be prior then neither must be.   That is, for a recursive dependency of {A,B}, $p(A|B)=p(A)$ and by symmetry we also must have $p(B|A)=p(B)$.

In any case, issues of the probability of events and their priority are frequently set within a spacetime context, with the associated physical implications. One of the aims of this paper is to establish a solution of this dilemma of choice conforming to Special Relativity but still recognisably consistent with conventional probability theory.

Note that even though both expressions $p(A \text{ OR } B | C)$ and $p(A \text{ AND } B | C)$ commute for A and B, Bayes' Theorem *does* distinguish the precedence of the conditionalities associated with the primary events.

Eq.2b can be expressed in one of two ways, according to which of the primary events A and B is prior: substituting Eq.2c into Eq.2b we find two choices offering two possibilities for the logical OR function:

$$p(A \text{ OR } B) = p(A | C) + p(B | C) - p(A | BC)\, p(B | C)$$
$$= p(A | C) + p(B | C) - p(B | AC)\, p(A | C) \qquad (2d)$$

Eq.2d may be simplified by noting that the component term of Bayes' Theorem appearing at the end of each CSR expression (each such latter term representing a conditional possibility) can also be assigned a probability associated with the truth of that conditionality.   This allows us to informally re-write Eq.2d as a single statement, by expanding upon the products from Eq.2c:

$$p(A \text{ OR } B | C) = p(A | C) + p(B | C) - \{\, [p(A | BC)\, p(B | C)]_B + [p(B | AC)\, p(A | C)]_A \,\}$$
(4a)

where the subscripts {A, B} indicate the assumed prior, and the expression in each of the pair of square brackets represents the respective Bayes' Theorem choice.   Eq.4a can be regularised by including the probability associated with the prior, that is, now explicitly including a recursive element.   Note, the probability $p(A|BC)$ associated with the prior is *not* the same as the probability of the prior, which we may state informally as: $p($"A is prior to B"$)$.   Thus, we explicitly include the probabilities associated with the priors (removing the subscripts {A,B}).   In effect we now assume the possibility of a 'prior prior' and the (logical/temporal) ordering of the events {A,B} now also starts to be explicitly taken into account:

$$p(A \text{ OR } B | C) = p(A | C) + p(B | C) - \{\, [p(A | BC)\, p(B | C)]\, p(B | AC) + [p(B | AC)\, p(A | C)]\, p(A | BC) \,\}$$
(4b)

allowing a consistent (and symmetric) simplification which intrinsically includes the possibility of recursion:

$$p(A \text{ OR } B | C) = \{\, p(A | C) + p(B | C) \,\}\, \{1 - p(A | BC)\, p(B | AC) \,\} \qquad (4c)$$

In Eq.4c we have the probability of "A OR B given C" being {[the probability of A given C] OR [the probability of B given C]} reduced by a cross-probability factor F,   where:

$$F \equiv - p(A | BC)\, p(B | AC) \qquad (5a)$$

which could be thought to represent a "first level" of dependency,   that is,   the dependence of A on B (and *vice versa*).   But what if we also consider a deeper dependence between these events? This would give:

$$p(A \text{ OR } B | C) = \{p(A | C) + p(B | C) \}\{1 + F + F^2 + \dots) \} \qquad (5b)$$

or, continuing to full recursion:

$$p(A \text{ OR } B | C)$$
$$= \{p(A|C) + p(B|C)\} \sum_{m=1}^{\infty} \{1 + [-p(A|BC)p(B|AC)]^m\} \qquad (5c)$$

Given that probabilities may not be greater than unity this infinite series can be compactly written as:

$$p(A \text{ OR } B | C) = \frac{p(A|C) + p(B|C)}{1 + p(A|BC)p(B|AC)} \qquad (6)$$

However, Eq.6 may be simplified since, given that it cannot be determined which of $\{A,B\}$ is prior, we may take $p(A|B)=p(A)$ and $p(B|A)=p(B)$.   Hence:

$$p(A \text{ OR } B|C) = \frac{p(A|C) + p(B|C)}{1 + p(A|C)p(B|C)} \tag{7}$$

Eq.7 (isomorphic to Eqs.1) states the "Hyperbolic Sum Rule" (HSR) theorem to be proved where $\{A,B\}$ are recursively dependent.   In deriving the form of Eq.7, we asserted that $p(A|BC)= p(A|C)$ and $p(B|AC)= p(B|C)$; that is, due to the recursion between events $\{A, B\}$, the probability of A becomes essentially independent of B, and *vice-versa*. That is, Eq.6 can be re-expressed in the strictly hyperbolic form of Eq.7.   In effect, this assertion ($p(A|B)= p(A)$, etc.) can be justified as the most likely (MaxEnt) relationships between the conditional probabilities for A and B when we know *a priori* that A and B have a recursive dependency, but we don't know the precise (quantitative) relationship between A and B. In Appendix C we prove the MaxEnt character of Eq.7.   That is to say, these assertions can also be justified by recourse to the *Principle of Indifference* (PI) and the simplest non-informative prior in Bayesian probability (see [refs. 33, 35]).

Eq.5b expresses the case where the relation between events $\{A, B\}$ is not recursive but stops at a countable depth or iteration: the overt conditionality between them is therefore maintained such that $p(A|B) \neq p(A)$ and $p(B|A) \neq p(B)$.   For example, if one wanted to formalise the psychological aspects of Poker one might express a "bluff" as the F term in Eq.5a, a "double-bluff" as the F^2 term, and a "triple-bluff" as the F^3 term etc.   But the HSR of Eq.7 expresses the recursion needed for the etalon of Eq.1c (for example).   In contrast, it is clear that the CSR of Eq.2b is the truncated HSR case of Eq.6, where the CSR *a priori* ignores recursion and its associated higher-order dependencies (either due to specified information, or because their existence is simply not taken into account) and excludes the higher-order (logical/temporal) ordering of propositions. In any case, Bayes' Theorem (the Product Rule) remains valid.

2.4. The Conventional Sum Rule (II)

Both the HSR and the CSR are MaxEnt in each of their respective domains of applicability (proved respectively in Appendices C and E), and we need to know when to use the CSR and when the HSR (and why). In particular, the point of this paper is to emphasise the importance of the temporal (or logical) *ordering* of propositions, and therefore the possibility for recursion, which is what distinguishes between the CSR and HSR: the CSR doesn't adequately take the temporal ordering of propositions into account, whereas the HSR does.   And, this is particularly evident in the Venn diagram of Figure 1 which cannot represent the temporal or logical ordering of the propositions.

As an example, we consider the scenario of a fair coin being tossed twice.   We designate the event A as "the first toss gives a head", and the event B as "the second toss gives a head", and we wish to compute the probability of the first toss *or* the second toss giving a head.   Such a 'fair' coin does not have memory (each toss is independent of any others), so that we have $p(A) = p(B) = \frac{1}{2}$. Here there is no recursion since $p(C|A) = p(C)$ and $p(C|B) = p(C)$: that is, C is independent of both of $\{A,B\}$, and therefore we may use the simple form of the CSR (Eq.2a): $p(A \text{ OR } B) = \frac{1}{2} + \frac{1}{2} - \frac{1}{4} = \frac{3}{4}$.

Clearly, applying the HSR (Eq.7) to get $p(A \text{ OR } B|C) = 4/5$ in this context is manifestly false, precisely because *we already know* that there is no recursion present. The difference between the hyperbolic (HSR, Eq.7) and conventional (CSR, Eq.2b) sum rules is specifically the *prior information* that must be taken into account to do the calculation. The decisive question is whether or not the mutual dependence of $\{A,B\}$ is recursive.

There are other simple cases where memory is involved, such that $p(A|B) \neq p(A)$.   One such case is Cox's [ref.7] example of randomly drawing black or white balls from a bag.   The probability of a result is affected by the results of previous draws, but this is a *memory* effect, not a *recursive* one; and in this context, where the system only iterates based on the forward propagation of a previous state (i.e. no feedback), see Appendix F for a discussion of FIR (finite impulse response) filters.   In general, if $p(A|B)$ and $p(B|A)$ are expressed in non-recursive form the CSR should be used, but if not then the HSR should be used.   The CSR should be used only where recursion can be *a priori* excluded.

3. A general "Hyperbolic Sum Rule" (HSR)

3.1. Proving the HSR Theorem

Consider the function Σ:

$$\Sigma(x) = \frac{1-x}{1+x} \qquad (8a)$$

$\Sigma[\Sigma(x)] = x$, that is: Σ is self-reciprocal, or $\Sigma(x) = \Sigma^{-1}(x)$. It is also doubly differentiable and a continuous, monotonically decreasing function: $\partial\Sigma/\partial x = -2/(1+x)^2$. The inspiration for Eq.8a comes from Cox's original equation for S as another self-reciprocal function (see Eqs.3 above) which is also doubly differentiable [7]. In Cox's case, he also considered the situation where S (and x) are raised by an integer power m, such that $S^m = 1 - x^m$ for integer m. For even $m>0$ this can be factored into the geometric series:

$$(1-x^m) \equiv (1+x)(1-x+x^2-x^3+\ldots-x^{m-1}) \qquad (8b)$$

with the series in parentheses representing an infinite summation (as $m\to\infty$) of the quantity $(1+x)^{-1}$. Note that this treatment is quite general, such that x may be complex. However, for all m (with integer $m\geq1$), we can also write:

$$(1-x^m) \equiv (1-x)(1+x+x^2+x^3+\ldots+x^{m-1}) \qquad (8c)$$

Recognizing that for even m the roots of S^m lie symmetrically around the four quadrants of the unit circle centred on zero in the complex plane, allows us to additionally exploit the associated symmetry relation for the nth root, $x_n = -x_{m/2+n}$. It is important to realise that results from complex analysis underpin this analysis (as we also mention below following Eq.10c, and see Parker & Jeynes 2023 [ref.37] for a fully complexified treatment).

Therefore, substituting $x_n = -x_{m/2+n}$ for the $m-1$ terms of the series summation in Eq.8c, we can write:

$$(1-x^m) \equiv (1-x)(1-x+x^2-x^3+\ldots-x^{m-1}) \qquad (8d)$$

Applying this (counterintuitive) symmetry we see that Eq.8a is then simply the limit of Eq.8d as $m\to\infty$. In this limit the distinction between even and odd m is insignificant. In any case, it is clear that equivalent to Cox's first solution (using $m=1$) the quantity $S=1-x$, represents an important term in the factorisation of Σ, but is by no means the only available solution in the general case for the roots of the function Σ.

From de Morgan's Theorem ($\overline{A \text{ OR } B} \equiv \bar{A}\bar{B}$, hence $A \text{ OR } B \equiv \overline{\bar{A}\bar{B}}$):

$$p(A \text{ OR } B|C) = \Sigma\big[\Sigma\big(p(A|\bar{B}C)\big)\Sigma\big(p(B|C)\big)\big] \qquad (9)$$

where $\bar{B}$ (the negation of B) is required in the conditioning probability of the first probability term in order to satisfy the appropriate application of the Product Rule in de Morgan's Theorem.

Substituting Cox's *first-order expression* for Σ (that is, $\Sigma = 1 - x$) we recover the "Conventional Sum Rule" (CSR, Eq.2b), after some manipulation. However, substituting the full bilinear form for Σ (Eq.8a) into Eq.9:

$$p(A \text{ OR } B|C) = \Sigma\left[\frac{\big(1-p(A|\bar{B}C)\big)\big(1-p(B|C)\big)}{\big(1+p(A|\bar{B}C)\big)\big(1+p(B|C)\big)}\right]$$

$$= \Sigma\left[\frac{1-p(A|\bar{B}C)-p(B|C)+p(A|\bar{B}C)p(B|C)}{\big(1+p(A|\bar{B}C)\big)\big(1+p(B|C)\big)}\right] \qquad (10a)$$

and expanding the outer self-reciprocal Σ function:

$$p(A \text{ OR } B|C) = \frac{1 - \left(\dfrac{1 - p(A|\bar{B}C) - p(B|C) + p(A|\bar{B}C)p(B|C)}{\left(1 + p(A|\bar{B}C)\right)\left(1 + p(B|C)\right)} \right)}{1 + \left(\dfrac{1 - p(A|\bar{B}C) - p(B|C) + p(A|\bar{B}C)p(B|C)}{\left(1 + p(A|\bar{B}C)\right)\left(1 + p(B|C)\right)} \right)}$$

$$= \frac{2p(A|\bar{B}C) + 2p(B|C)}{2 + 2p(A|\bar{B}C)p(B|C)} \tag{10b}$$

In the recursive case (as we have previously discussed for Eq.7), the probability of *A* becomes, in effect, independent of *B*; $p(A|\bar{B}C) = p(A|C)$, such that Eq.10b can be simplified to:

$$p(A \text{ OR } B|C) = \frac{p(A|C) + p(B|C)}{1 + p(A|C)p(B|C)} \tag{10c}$$

which is Eq.7 as required.

Note that although the intermediate Eqs.8b,c,d (and the implied complex analysis) are not used in the derivation of the HSR (Eq.7 or Eq.10c), analogous complex quantities make an appearance in the discussion of how the finite impulse response (FIR) filter offers a good exemplar of the CSR (sketched in Appendix F). The standard theory of such   filters (which are important for digital signal processing applications) fundamentally relies on the theorems of complex analysis.

3.2. Some Analytical and Numerical Comparisons

Having derived a "Hyperbolic Sum Rule" (HSR) expression, it is interesting to see how much it diverges from the "Conventional Sum Rule" (CSR) in a simplified scenario that considers the function $p(A \text{ OR } B)$ for two possible and equally probable events $\{A, B\}$, that is, $p(A)=p(B)$.

The HSR is well-behaved in that it yields the same results as the CSR in the limit that neither event occurs, that is, $p(A)=0$ and $p(B)=0$.   In this case, we have $p(A|BC)= p(B|AC)=0$ (that is, $p(A \text{ AND } B|C)=0$ from Bayes' Theorem). It is straightforward to see in this specific case that $p(A \text{ OR } B) = 0$ for both sum rules.

The HSR is also well-behaved where $p(A)= p(B)=1$.   In this case, the CSR reduces to $p(A \text{ OR } B)=p(A)+p(B)-p(B) =p(A)=p(B)$.   The HSR also reduces to the same result, but via the progression $p(A \text{ OR } B|C)=2p(A|C)/2 =2p(B|C)/2=p(A|C)=p(B|C)$. For the case where we might want to include the result $p(A|B)=p(B|A)=1$, then we employ these in the denominator of Eq.6 (we must include explicit knowledge – see §2.3) to get the same (correct) result again.

For the simple system in which there are two equally probable events $\{A,B\}$ the conventional and hyperbolic forms of the sum rule converge when the events are either impossible or certain.   It is for the intermediate probabilities $0<p(A)<1$ that the two forms diverge. Figure 2 shows the behaviour of $p(A \text{ OR } B)$; calculated for the simple case of $p(A)=p(B)$ and also assuming $p(A|B)=p(B|A)=p(A)=p(B)$ for HSR. For CSR the two events $\{A,B\}$ are also assumed to be independent, that is, $p(A|B)=p(A)$ and $p(B|A)=p(B)$.
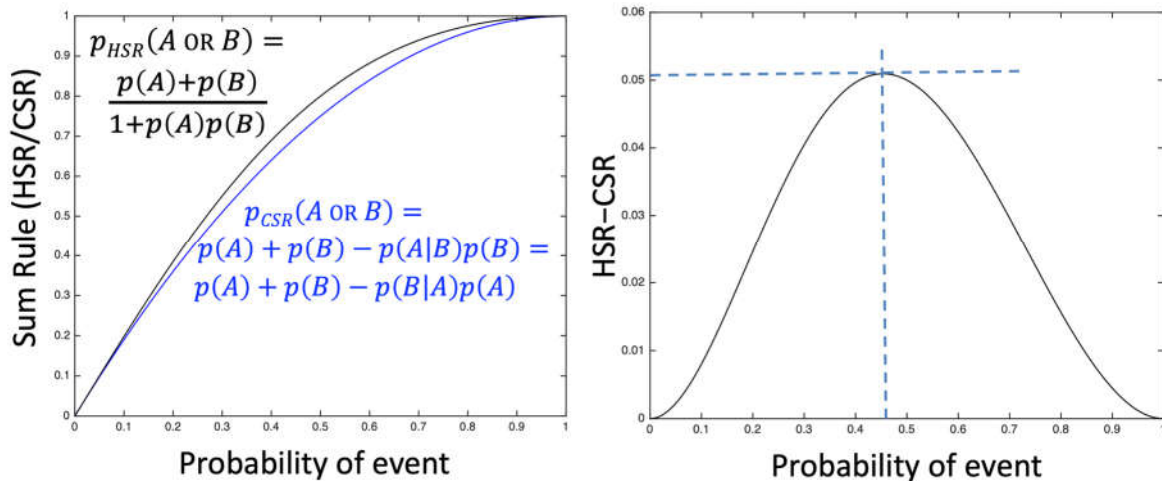
Figure 2. Comparing the HSR and CSR versions of $p(A \text{ OR } B)$ for of two equally probable hypotheses (A,B). (a) Hyperbolic Sum Rule (HSR: black) and Conventional Sum Rule (CSR: blue); (b) Difference between HSR and CSR.

From a simplistic Maximum Entropy (MaxEnt) argument, one might expect that the greatest difference between the two sum rules would occur when the events have the most uncertainty, that is, $p(A)=\frac{1}{2}$. For this situation, the CSR returns an overall probability of $p_{CSR}(A \text{ OR } B | C)=\frac{1}{2}+\frac{1}{2}-(\frac{1}{2}\times\frac{1}{2})=\frac{3}{4}$ (75%). In contrast, the HSR returns a slightly higher overall probability of $p_{HSR}(A \text{ OR } B | C)=(\frac{1}{2}+\frac{1}{2})/(1+[\frac{1}{2}\times\frac{1}{2}])=\frac{4}{5}$ (80%). The presence of recursion (mutual dependence) *increases* the probability that both events will occur.

Interestingly, however, Figure 2 shows that the maximum difference between the two sum rules is actually 5.09 percentage points, for $p=45.35\%$ in this simplified case. (The asymmetry reflects the slight asymmetry of the hyperbolic sum rule.) Thus we see that the difference between the conventional and hyperbolic sum rules may be quite minor: in most cases it would be hard to observe (less than 5 percentage points).

Being MaxEnt, an application of the HSR of Eq.7 would be to calculate the sum probability for two events where the relationship between the two events is not known, yet some correlation (including recursion) is reasonably anticipated. In contrast, the CSR (also MaxEnt) explicitly requires knowledge of the conditional probability relationship between the two events: this may be estimated or simply assumed not to exist (since independence is often a reasonable approximation). Note that the CSR explicitly excludes recursion, an assumption which should be justified. Thus, the HSR of Eq.7 obviates the need for any unknown conditional probability, and also avoids any unjustified assumptions relating to recursion.

For example, imagine a scenario of elections occurring in neighbouring countries "A" and "B", with leading candidates "Alice" and "Bob" in each, respectively. Opinion polls consistently suggest that both candidates each has a good chance of being elected: 45% in each case. We pose the question (to which we would like a Bayesian answer, if possible): what is the probability that at least one of the two leading candidates will be elected?

In order to apply the conventional sum rule we require the conditional probability that Alice will be elected given Bob is elected, or alternatively the probability that Bob is elected given Alice is successful (given by Bayes' Theorem, Eq.2c). In the CSR context (in the absence of any other information to the contrary) the best we can assume is these are essentially independent events such that:

$p(A)=p(B)=p(A|B)=p(B|A)=0.45$

The CSR can then be applied to offer the MaxEnt probability of either Alice or Bob being elected (given that the two events are mutually independent) as:

$p(A \text{ OR } B)=p(A)+p(B)-p(A)p(B)=0.698$

However, the HSR allows us to include the possibility of a mutual (recursive) relationship between these two events (without making any other assumptions which could skew the calculation), giving a MaxEnt (most likely) probability in the absence of any other information (or any implicit assumptions):

$$p(A \text{ OR } B) = \{p(A) + p(B)\} / \{1 + p(A)p(B)\} = 0.748$$

Thus, the HSR probability estimation of at least one of the candidates being successful is noticeably higher than that anticipated from the CSR. This is because there are a variety of ways the two elections could potentially influence each other, and the HSR accounts for the possible existence of these mutual interactions, while (in the absence of specific information as given by the appropriate conditional probabilities) the CSR ignores such influences.

It is perhaps surprising that such a difference between the two sum rules has not been noted up to now: surely empirical observation would have shown a discrepancy from the (wrong) current theoretical calculation based purely on the CSR? That no-one has suspected that the CSR is faulty may be due to the fact that for most simple cases (e.g. card games, gambling etc.) recursion is generally absent (and the CSR is correct). Of course, probability theory originated in the analysis of games of chance where recursion is excluded. For more complex cases there is usually no possibility of re-running the observations (elections in this case) sufficiently to gain a statistically significant estimate of the associated empirical probabilities. Most real cases have complex (and recursive) conditionalities and dependencies, and most are unrepeatable, meaning that the associated probabilities cannot be determined by conventional methods. Thus, any systematic error in the underlying probability theory is probably unobservable, and it is hardly surprising that it has been unobserved and unsuspected until now.

3.3. Concatenation Rules for Multiple Hypotheses

In **Appendix D** we prove the following relations:

A version of Bayes Theorem for multiple summed hypotheses $\{X\}_{OR}$ as the number of these hypotheses grows large:

$$p(A)p(\{X\}_{OR}|A) = p(\{X\}_{OR})p(A|\{X\}_{OR}) \qquad \text{as } \{X\} \to \infty \qquad \text{(see App. D Eq. D4b)}$$

We also exploit the hyperbolic tangent function to prove a Hyperbolic Sum Rule for multiple hypotheses with multiple conditionalities:

$$p(A \text{ OR } B | D \text{ OR } E) = \frac{p(A|D \text{ OR } E) + p(B|D \text{ OR } E)}{1 + p(A|B \text{ OR } D \text{ OR } E)p(B|A \text{ OR } D \text{ OR } E)} \qquad \text{(see App.D Eq.D6)}$$

Finally, a generalised Hyperbolic Sum Rule with N multiple hypotheses, which sorts odd-ordered conjunctions into the numerator and even-ordered conjunctions into the denominator:

$$p(A_1 \text{ OR } A_2 \text{ OR } \dots \text{ OR } A_N) = \frac{\sum_{l=1,\text{odd}}^{N} \sum_{m=1}^{m=\frac{N!}{l!(N-l)!}} \prod_{n=P\{l,1:N\}}^{l} p(A_n | \{A_{n'}, l-1\})}{\sum_{l=0,\text{even}}^{N} \sum_{m=1}^{m=\frac{N!}{l!(N-l)!}} \prod_{n=P\{l,1:N\}}^{l} p(A_n | \{A_{n'}, l-1\})} \qquad \text{(see App. D Eq. D7)}$$

This may also be written in a form such that the probability distributions are represented (as MaxEnt decaying exponentials probabilities, see Jaynes 1982 [ref.14]) by the appropriate Lagrange multiplier β (noting that different distributions are generated by different choices of β, with an example of this given in a recent treatment of the Wine/Water paradox [ref.35]):

$$p(A_1 \text{ OR } A_2 \text{ OR } \dots \text{ OR } A_N) = \frac{\sum_{l=1,\text{odd}}^{N} \frac{N!}{l!(N-l)!} \left\{ \left(\frac{1-e^{-\beta}}{1-e^{-N\beta}} \right) e^{\beta} \right\}^{l} \exp\{-\beta \sum_{n=P\{l,1:N\}}^{l} n\}}{\sum_{l=0,\text{even}}^{N} \frac{N!}{l!(N-l)!} \left\{ \left(\frac{1-e^{-\beta}}{1-e^{-N\beta}} \right) e^{\beta} \right\}^{l} \exp\{-\beta \sum_{n=P\{l,1:N\}}^{l} n\}} \qquad \text{(see App. D Eq. D9d)}$$

additionally showing that the Hyperbolic Sum Rule is well-behaved for multiple hypotheses as $N \to \infty$.

In **Appendix E** we prove the following relation for the Conventional Sum Rule for multiple hypotheses:

$$p(A_1 \text{ OR } A_2 \text{ OR } \dots \text{ OR } A_N) =$$

$$1 + \sum_{l=1,\text{odd}}^{N} \sum_{m=1}^{m=\frac{N!}{l!(N-l)!}} \prod_{n=P\{l,1:N\}}^{l} p(A_n|\{A_{n'}, l-1\}) \;-\; \sum_{l=0,\text{even}}^{N} \sum_{m=1}^{m=\frac{N!}{l!(N-l)!}} \prod_{n=P\{l,1:N\}}^{l} p(A_n|\{A_{n'}, l-1\}) \quad \text{(see App. E Eq. E6)}$$

4. Discussion

4.1. Probability is Physical

In simple cases, the distinction between the Conventional Sum Rule (CSR) and the Hyperbolic Sum Rule (HSR) is probably unobservable, given the attainable empirical precision. The case of the two parallel elections considered in §3.2 has 80% for the HSR compared to only 75% for the CSR, when the two election probabilities involved are both 50% ($p(A)=p(B)=0.5$). But the CSR cannot model recursion.

However, the derivation of the HSR uncovers a relation that is also observed in diverse physical settings. In particular, the HSR is isomorphic not only to the hyperbolic tangent double-angle identity (Eq.1a) but also to Einstein's famous velocity addition formula (Eq.1b) and to the reflectivity expression for an optical etalon (Eq.1c). Thus, the general (hyperbolic) probability sum rule conforms to a 'template' seen in the physical world; that is to say, although being apparently a *logical* relation it behaves just as many physical phenomena do. Therefore we can say that the etalon function (Eq.1c: intrinsically probabilistic in nature, being a wave scattering phenomenon) represents at least a *physical embodiment* of the HSR. In particular, the physical structure of the etalon with its two scattering interfaces (*A* and *B*, say) suggests that the reflection coefficient calculated from the solution to Eq.1c can also be interpreted as representing the physico-logical answer to the question: what is the probability that a quantum-mechanical particle (that is, a photon) is reflected from interfaces "*A* OR *B*"? It is clear that the basic two-faceted etalon physically expresses the logical OR operation.

This raises the interesting philosophical question: is the hyperbolic nature of the HSR a result of the intrinsic hyperbolic character of the natural universe? Affirming that it is implies that probability theory is firmly empirical (and not merely mathematical): both the theory of Special Relativity (Eq.1b) and the quantum-mechanical scattering probability (Eq.1c) are consistent with the hyperbolic nature of spacetime. But denying it would imply that it is simply *coincidental* that the HSR shares fundamental characteristics of the natural universe.

This question actually goes to the root of a deep philosophical debate as to the intrinsic nature of probability theory: is it purely an empirical construct where probabilistic variables only acquire meaning when a statistical measurement takes place? This view is implied by the so called 'frequentist' fraternity, who do not ascribe an independent reality to the "laws" of probability; rather they consider that the probability of a hypothesis depends solely upon the frequency of success over a given number of trials. But Bayesians would say that the hyperbolic nature of the HSR implies that probability theory and the underlying theoretical equations form a component part of the full set of universal natural laws; that is to say, they are an intrinsic aspect of the (hyperbolic) universe with their own independent existence.

We should add that the result that the MaxEnt nature of both the HSR of Eq.7 (recursive, see Appendix C) and the CSR (non-recursive, see Appendix E) indicates that the *equilibrium* state of such physical systems (corresponding to a stable and most-likely configuration) obeys both the *kinematical* Principle of Least Action as well as the *entropic* Principle of Least Exertion (ref.[10]). These fundamental principles underly all known physical phenomena and therefore indicate that both the HSR and CSR are consistent with the requirements of the physical world. The fact that the HSR is MaxEnt also implies that all the parts of such systems recursively condition all the other parts. That is, in the general case everything is entangled with everything else just as Karen Barad implies [ref.31]. But here we wish to avoid metaphysical discussion (for which see [ref.6]).

4.2. Recursion

Why has the issue of recursion apparently not previously been explicitly considered in probability theory? This may simply be an accidental result of history, with probability theory having developed from the gaming scenarios associated with cards and dice, where events are discretely independent and are characterised with shallow conditionalities that do not reach far back into the past.   For example, the game of Poker has explicit calculable conditionalities based on cards revealed to the table, cards in your hand, and the hidden cards remaining; the calculable conditionalities do not involve the (fascinating) issues of bluffing, double bluffing, and multiple bluffing.   These could in principle be expressed logically (using the formalism sketched above), but are currently treated as psychology.

However, it is interesting to note that Cox (1946 [7]) derives the general result $S^m = 1 - x^m$ for the self-reciprocal function $S(x)$, stating that the value for m (assumed an integer) is "*arbitrary*" and "*purely conventional*"; indeed, Jaynes writes that the value of m is "*actually irrelevant*" [ref.1].   For "*simplicity of notation*" Cox arbitrarily chooses $m = 1$, which is both convenient and intuitive, since it immediately leads to the Conventional Sum Rule associated with the Venn diagram of Figure 1. Thus Cox's analysis allows for recursion conceptually, although he apparently appreciated neither this nor the physical interpretation of his function S^m.

But exploiting the hyperbolic insight described here, we can now understand m to represent the degree of mutual dependency leading to recursion; that is, it has both a logical and physical (spatio-temporal) meaning, with real physical implications which are not at all "arbitrary".   In particular, $m=0$ is a trivial case, $m=1$ is the simple CSR case, $m^{-1}=0$ is the HSR case.   We consider this further in §4.3.

Another reason for ignoring recursion is that the appropriate calculus of probabilities requires substantial computing power which has only become available in recent decades.   Up to the point of the emergence of artificial intelligence (AI) and machine learning (ML), approximately at the turn of the millennium, there hasn't been a good reason to develop a theory of recursive probabilities, since before this time the necessary computations were practically intractable.   However, as large-scale computing power has evolved over recent decades, this has also enabled the development of the sophisticated (Bayesian) algorithms underpinning AI and ML programming.   Such algorithms allow a dispassionate approach to making probabilistic inferences and decision-making in an environment of uncertainty, within a mathematical framework for updating beliefs and reasoning about uncertain events.   In addition, recursion provides AI with additional flexibility to handle complex and dynamic problems by recursively decomposing them, while leveraging self-referential architectures and utilizing iterative processes.

Recursion can be exploited relatively simply in AI as a means to break down complex problems into smaller, more manageable subproblems.   Alternatively, the complexity and power of modern computing also enables the employment of recursive data structures that can be used to represent hierarchical or interconnected relationships.   Recursion can also be seen in the inductive reasoning aspect of AI, which involves inferring general rules or patterns from specific observations, so as to learn iteratively from data in order to improve the accuracy of predictions or classifications.

In AI (and the associated field of neural networking) recursion has been employed empirically: the complexity associated with the algorithms and feed-back routines aren't generally amenable to analytical (closed-form or tractable mathematical) analysis, such that currently only 'simulations' can be used to analyse AI behaviour.   It is in this context of recursive AI that the Hyperbolic Sum Rule described here allows the systematic, reliable and consistent handling of multiple recursive hypotheses (or events).   We therefore expect the HSR to enable a more analytic (closed-form and tractable) analysis of complex AI algorithms to give additional insights into their form of operation and anticipated outputs.

4.3. Two Distinct Sum Rules

We have shown that in different   situations different rules apply for summing the probabilities of events (or, nearly equivalently, the probabilities of certain hypotheses being true).   We wish to

find a relation that allows us to compute $p(A \text{ OR } B | C)$, that is, the probability of (A OR B) given C, where {A,B} are events (or hypotheses) and {C} are conditionalities.

There are two distinct cases depending on whether or not {A,B} are recursively dependent. If {A,B} are independent (such as coin flipping, see §2.4) then the "*Conventional Sum Rule*" (**CSR**) applies. But if {A,B} are recursively dependent (such as reflection from a multilayer stack, see Eq.1c) then the "*Hyperbolic Sum Rule*" (**HSR**) applies. Concatenation formulae for either or both of multiple events {$A,B, …$} or multiple conditionalities {$C,D, …$} are readily derived for both the HSR (Appendix D), and the CSR (Appendix E; also see the summary in §3.3 of important results for both HSR and CSR). Therefore there is no loss of generality from treating only the simple cases.

If $p(A|B)>0$ and $p(B|A)>0$, with both probabilities being valid independent of any temporal (that is, causal) or logical constraints, then {A,B} are recursively dependent and the HSR applies. Otherwise the CSR applies. Conversely, the CSR applies if the probability relation {$p(C|A)=p(C)$ OR $p(C|B)=p(C)$} holds.

Note that in simple cases (see Figure 2) the difference between CSR and HSR is probably too small to be easily observed: even in the fair coin or election cases (§2.4, §3.2) the CSR gives 75% where the HSR gives 80%.

It is clear that the CSR and HSR are formally related (see Eqs.8). It also seems fairly clear that the "truncated expansion" of Eq.8d has properties related to (for example) the "finite impulse response" ("FIR") filter that is important in digital signal processing (DSP) applications (further details in Appendix F). The case represented by finite m where $m>1$ and $m^{-1}>0$ (analogous to the FIR filter) offers an engineering approach to the probabilistic treatment and control of system noise within a finite (and specified) timeframe - an issue of practical importance for stochastic systems. Such practical probabilistic cases, formally CSR (see Appendix E), blur the distinction between CSR and HSR, since it seems that such "FIR filtering" may be represented approximately as a "truncated HSR", with a formalism that is also approximately MaxEnt in its own terms (see Appendix C). However, the derivation of the HSR is directly from Eq.8a: the discussion of finite values of m in Eqs.8b,c,d is presented only by way of comment.

5. Conclusions

We have shown how a *Hyperbolic Sum Rule* (**HSR**) for probability can be derived that is consistent with the mathematical and logical requirements as laid out by Cox [ref.7] and Jaynes [ref.1]. In particular, we have proved that the HSR is Maximum Entropy (**MaxEnt**), and thereby is also consistent with the Second Law of Thermodynamics.

For the simple case of only two equally probable hypotheses (or events) the conventional and hyperbolic versions of the sum rule yield similar (but not identical) results (see Figure 2). However, when large numbers of hypotheses are being considered we have shown (Appendix E) that properties derived from the HSR are needed for an appropriate Conventional Sum Rule (**CSR**) to behave reasonably. That is, although the CSR is applicable to certain simple cases, the HSR must be used where recursion cannot be ruled out (and where a MaxEnt solution is desirable).

Apart from its mathematical consistency, we should also note empirically that numerous other physical phenomena involving probabilities also obey the HSR, as one might expect considering the (widely acknowledged) hyperbolic nature of the spacetime metric for the universe (see for example Penrose 2004 [ref.21]). Therefore we conclude that the Hyperbolic Sum Rule implies that *probability is physical* (also supported by the fact that the HSR is MaxEnt).

Probability is physical because *probabilities* always refer to *phenomena* (however idealised). Phenomena belong in the real world, and their value (including measurement) is necessarily assessed by people for their own purposes. It is therefore satisfactory that the general MaxEnt case is represented by a formalism isomorphic to formalisms representing probabilities incorporated within general physical systems (such as the etalon, Eq.1c).

In terms of application, we expect that these methods will make those calculations analytically tractable that involve many hypotheses featuring multiple dependencies (both spatial and temporal). For example, disentangling the highly complex interactions of the different genes within a single

genome represents an extremely difficult computational challenge, yet one with a glittering set of potential medicinal benefits. Another entirely different example regards the handling of certain important filters used in modern digital signal processing.   Doubtless, others will also emerge.

Another application which exploits the handling of multiple hypotheses featuring highly interrelated, highly complex and recursive dependencies (currently thought either intractable or impracticable for explicit analytical evaluation) is in the emerging field of rule-based expert learning systems for inference. Such applications depend on the use of Bayesian reasoning, and exploit modern computational resources to handle a high degree of multi-modal inputs and outputs, each requiring a very high number of explicit (and implicit) variables (hypotheses) with associated conditional dependencies. However, issues of reliability, predictability and engineering-control of AI technology are becoming more salient, since these are a critical prerequisite for the systemic, safe, and validated use of AI technology as a powerful tool in modern society. We are confident that the Hyperbolic Sum Rule and the insights associated with the recognition that *probability is physical* will contribute towards encouraging AI's development as a scalable and high-value technological asset working for the full benefit of humanity.

Appendices: MaxEnt Properties of the Sum Rules for Probabilities

We investigate the Maximum Entropy (MaxEnt) properties of both the Hyperbolic Sum Rule (HSR) and the Conventional Sum Rule (CSR).   First (Appendix A) we establish a convenient criterion for proving that a probability function is MaxEnt, then (by way of example: Appendix B) we explore some simple functions.   We prove (Appendix C) that the hyperbolic tangent sum rule is MaxEnt and therefore that the (isomorphic) HSR is also.   In Appendix D we derive sum rules for multiple hypotheses together with multiple conditionalities: these cases are complicated and require use of the HSR, being intractable on the assumptions of the CSR. In Appendix E we show that the CSR is also MaxEnt for the limited cases where it applies, and show how the CSR should be concatenated to enable the handling of multiple hypotheses.

Finally (Appendix F) we discuss how the HSR may find application in the engineering design of infinite impulse response (IIR) filters used in digital signal processing (DSP).

Appendix A: Maximum Entropy Criterion for the HSR

The Hyperbolic Sum Rule (HSR) for probabilities is intended to express the most-likely (Maximum Entropy) sum of probabilities where the conditional probabilities associated with the conventional sum rule are not available and where the recursive properties are not explicitly known. That is to say, in the absence of the additional information (associated with the conditional probabilities as well as recursion) required to define the conventional sum rule, we can still define the most likely addition of the probabilities using Maximum Entropy.

The following function (which is equivalent to the Partition Function in statistical mechanics) is known to be MaxEnt (see Appendix I of Jaynes 1982 [ref.14]).   We first discuss here why it is MaxEnt. The function $f(z_0, z_1, z_2 \ldots z_i \ldots z_N)$ consists of a series of $N+1$ terms, whose summation is given by the quantity Z:

$$Z = \sum_{i=0}^{N} z_i \left(\equiv \sum_{i=0}^{N} e^{-\lambda i} \right) \tag{A1a}$$

or

$$\sum_{i=0}^{N} \frac{z_i}{Z} = 1 \tag{A1b}$$

The free parameter λ is the essential controlling variable of the function f, expressing the 'shape' (or rate of exponential decay) of the function, while the index i is constrained to be an integer, with the maximum value N representing another free parameter of the function. The entropy H of the function f is given by term-by-term summation of the Shannon entropy of each ith component:

$$H = -\sum_{i=0}^{N} \frac{z_i}{Z} \ln \frac{z_i}{Z} \tag{A2}$$

Each ith component z_i is associated with an ith state of the system f; that is to say, the function f can also be thought of as describing the relative 'probability' (or propensity for existence) of each of the components comprising a system of N different (distinguishable) states. In which case, the ith component z_i can also be considered to represent a measure of the 'relative strength' of the ith component of the system. The fact that the system may be characterised with a finite entropy H means that we can assume that the number of states of the system is finite and has a granularity (that is, it is discretised, or quantised).

Consequently, this also means that there exists a quantised number E, representative of the overall system, where the discretisation (granularity) is represented by the quantum ΔE, such that the overall total number of quanta Q existing within the system is $Q = E/\Delta E$. By inspection, it is also clear that $Q \equiv Z$. How these Q quanta are distributed across the function f represents a constraint to the system as we maximise its entropy. We therefore associate a quantity E_i with each of the component terms of f such that the product of z_i and E_i for each component represents a measure of the strength of the ith state, and by summing across all the $N+1$ states we obtain the overall quantity E:

$$\sum_{i=0}^{N} z_i E_i = E \tag{A3a}$$

which can be re-written as:

$$\sum_{i=0}^{N} \frac{z_i}{Z} E_i = \frac{E}{Z} = \frac{E}{Q} = \Delta E \tag{A3b}$$

Following the standard treatment of Caticha[xli] we employ Lagrange multipliers α and λ, such that we set the independent equations to zero, we can use Eqs.A1b, A2 & A3b to consider the variation of the system entropy H with z_i so as to find its maximum:

$$\frac{\partial}{\partial z_i} \left\{ \sum_{i=0}^{N} -\frac{z_i}{Z} \ln \frac{z_i}{Z} + \alpha \left(1 - \sum_{i=0}^{N} \frac{z_i}{Z} \right) + \lambda \left(\Delta E - \sum_{i=0}^{N} \frac{z_i}{Z} E_i \right) \right\} = 0 \tag{A4}$$

We can swap the order of the summation and differentiation operations to evaluate the derivative of each of the i terms in Eq.A4:

$$\sum_{i=0}^{N} \left(-\frac{1}{Z} \ln \frac{z_i}{Z} - \frac{1}{Z} - \frac{\alpha}{Z} - \frac{\lambda}{Z} E_i \right) = 0 \tag{A5}$$

To maximise the system entropy H the ith component z_i of the function f must be simply given by:

$$z_i = \frac{Z}{e^{1+\alpha}} e^{-\lambda E_i} \tag{A6}$$

That is to say, a distribution f is Maximum Entropy if its constituent terms follow a decaying exponential profile (see Eq.A1a).

Appendix B: Inadmissible "MaxEnt" Sum Rules

The most basic sum rule function $S_0(A,B) = A+B$ is Maximum Entropy (MaxEnt), since it is easy to demonstrate that the summation of any two terms can be made to follow the decaying exponential template of Eq.A6. For example:

$$S_0(A, B) = A + B = A\left(1 + {}^B\!/\!_A\right) \equiv \frac{Z}{e^{1+\alpha}}\left(1 + e^{-\lambda}\right) \tag{B1a}$$

which is simply the two terms of a $N=1$ series, for $i=0$ and $i=1$. That is to say, we can express the ratio B/A as being equivalent to $\exp(-\lambda)$ for an appropriate value of λ, while A is equivalent to $Z/\exp(1+\alpha)$ for an appropriate value of α. Of course, the converse is also true:

$$S_0(A, B) = A + B = B\left(1 + {}^A\!/\!_B\right) \equiv \frac{Z}{e^{1+\alpha}}\left(1 + e^{-(-\lambda)}\right) \tag{B1b}$$

where for consistency, we employ the same value for λ as in Eq.B1a, but the values for α and Z are as appropriate. The simplest sum rule of Eqs.B1 is therefore easily found to be MaxEnt: are there any other sum rules that are also MaxEnt?

Clearly, any such sum rule has to obey the template of Eq.A6 in order to be MaxEnt. The next obvious sum rule that could obey Eq.A6, and which only employs the two parameters A and B is given by the following geometric progression of $N+1$ terms:

$$S_1(A, B) = A + B = A\left(1 + \frac{B}{A} + \left(\frac{B}{A}\right)^2 + \left(\frac{B}{A}\right)^3 + \cdots + \left(\frac{B}{A}\right)^N\right) \equiv \frac{Z}{e^{1+\alpha}}\left(1 + e^{-\lambda} + e^{-2\lambda} + e^{-3\lambda} + \cdots + e^{-N\lambda}\right) \tag{B2a}$$

There are two aspects to this sum rule that need to be noted: it only converges for $B<A$ as $N\to\infty$, and it is clear that the overall summation can easily sum to greater than unity for various values of A and B. In particular, this latter aspect can be seen by re-writing Eq.B2a as:

$$S_1(A, B) = A\left(1 + \frac{B}{A} + \left(\frac{B}{A}\right)^2 + \left(\frac{B}{A}\right)^3 + \cdots + \left(\frac{B}{A}\right)^N\right) = A + B + A\left(\frac{B}{A}\right)^2 + A\left(\frac{B}{A}\right)^3 + \cdots + A\left(\frac{B}{A}\right)^N \tag{B2b}$$

That is to say, the first two terms of the series come from the most basic sum rule, and there then follow potentially an infinity of additional terms, such that the overall summation could well be greater than unity. Thus as a general probabilistic sum rule, it would appear be inadmissible. Of course, the case where $A<B$, requires the converse sum rule:

$$S_2(A, B) = B\left(1 + \frac{A}{B} + \left(\frac{A}{B}\right)^2 + \left(\frac{A}{B}\right)^3 + \cdots + \left(\frac{A}{B}\right)^N\right) \equiv \frac{Z}{e^{1+\alpha}}\left(1 + e^{-\lambda} + e^{-2\lambda} + e^{-3\lambda} + \cdots + e^{-N\lambda}\right) \tag{B2c}$$

These geometric progressions offer a closed form analytic solution:

$$S_1(A, B) = A\frac{1 - \left(\frac{B}{A}\right)^N}{1 - \frac{B}{A}} = \frac{A^2}{A - B} \quad \text{as} \quad N \to \infty \quad \text{for } A > B \tag{B3a}$$

and

$$S_2(A, B) = B\frac{1 - \left(\frac{A}{B}\right)^N}{1 - \frac{A}{B}} = \frac{B^2}{B - A} \quad \text{as} \quad N \to \infty \quad \text{for } B > A \tag{B3b}$$

Clearly, both forms for the sum rule have a pole (singularity) when $A=B$, such that the overall summation will be greater than unity. These are indeed 'acceptable' MaxEnt solutions as per Eq.A6, but are clearly deficient as *probabilistic* sum rules, since the two rules each need to be judiciously selected (i.e., disjoint application of the two rules is required according to which of A and B is the greater), and they can also sum to greater than one. As such, we can exclude the sum rules S_1 and S_2 as an appropriate general MaxEnt probabilistic sum rule. Indeed, the initial sum rule $S_0(A,B)$ can also be excluded as an appropriate probabilistic sum rule for this same reason, since it will also frequently sum to greater than unity.

It is interesting to note that the conventional probabilistic sum rule (employing conditional probabilities) always lies within the range of zero and one:

$$S_{CSR}(A, B) = A + B - A \times B_{|A} = A + B - B \times A_{|B} \tag{B4}$$

That is to say, the first two terms of the conventional sum rule (CSR) are equivalent to the basic sum rule function of Eqs.B1, but there is an additional third term to the rule, which ensures that the overall quantity behaves 'correctly' as a probabilistic quantity, such that it never exceeds unity. However, this is at the cost of introducing an additional term (indeed, additional information), alongside A and B: the conditional probabilities $B_{|A}$ or $A_{|B}$ (depending on which probability is considered prior). Such a conditional probability is essentially an additional (third) parameter to the

sum rule, such that the function in effect becomes a sum rule requiring three input arguments, rather than just the two parameters (*A* and *B*) that we have been discussing above. From this perspective, the conventional probabilistic sum rule of Eq.B4 cannot be MaxEnt (apart from the fact that it doesn't follow the template of Eq.A6), because it requires additional information (based on the conditional probabilities); whereas the rationale for a Maximum Entropy sum rule is to sum (in the most likely manner) the two parameters *A* and *B* in such a way that the minimal assumptions as to how *A* and *B* might depend upon or be related to each other is made; that is say, the MaxEnt sum rule requires the least additional information over and above the values for *A* and *B* so as to provide their *most likely* sum value in the absence of any other information.

In addition, it is worth noting that when the events *A* and *B* are mutually exclusive, then the conventional probabilistic sum rule of Eq.B4 degenerates to the basic sum rule of Eqs.B1. However, although the basic sum rule is MaxEnt (as we have previously discussed, Eqs.B1) the underlying reason for the degeneracy in this case of Eq.B4 means that we now *know* the exact (mutually exclusive) relationship between *A* and *B*. In general, the conventional sum rule of Eq.B4 is *not* MaxEnt (since it requires additional knowledge or a best-guess estimate of the (conditional) relationship between *A* and *B*); but the knowledge that *A* and *B* are mutually exclusive causes the conventional sum rule to degenerate to the basic sum rule (with its MaxEnt properties), but which is still inadmissible as a general probabilistic sum rule since it frequently leads to probabilities greater than one. However, Appendix E shows that in the limit where the conditional probabilities are assumed identical to the independent probabilities, then the CSR also becomes MaxEnt; however, this is a special case of the CSR where Bayes' Theorem (as a form of the Product Rule) becomes trivial.

Appendix C: Hyperbolic Sum Rule is MaxEnt

However, there still exists another potential probabilistic sum rule, which offers still better behaviour. This function emerges as the generalised expression arising from Cox's treatment. Fortunately, the MaxEnt aspect now becomes apparent as we can show here that the Hyperbolic Sum Rule (HSR) is a well-behaved *maximum entropy* function.   We start by expressing it in its basic form:

$$S_{HSR}(A, B) = \frac{A + B}{1 + AB} \qquad (C1a)$$

We first notice that the hyperbolic sum rule is isomorphic to the hyperbolic tangent double-angle identity:

$$\tanh(a + b) = \frac{\tanh a + \tanh b}{1 + \tanh a \tanh b} \qquad (C1b)$$

Just for clarity, we remind ourselves that we use *A* and *B* here as shorthand for *p(A)* and *p(B)*; that is to say, *A* and *B* each represent an event, and *p(A)* and *p(B)* each represent the respective probabilities for events *A* and *B* to occur. Thus, when we employ the symbols *A* and *B* in Eq.C1a, this is actually a shorthand for the probability functions *p(A)* and *p(B)* associated with each of the events *A* and *B*. Thus, when in Eq.C1b we explicitly employ the hyperbolic tangent (tanh) functions of *a* and *b*, we can immediately see the isomorphism between *p(A)* and tanh(*a*), as well as *p(B)* and tanh(*b*). This means that we can therefore also assume that $S_{HSR}(A,B)$ is isomorphic to the function tanh(*a+b*). This is a longwinded way of saying that the probabilistic hyperbolic sum rule S_{HSR} we are seeking to analyse must be identical to the hyperbolic tangent (tanh) function.

Using $\tanh \theta = \sinh \theta / \cosh \theta$, as well as $\sinh \theta = \frac{1}{2}\{\exp\theta - \exp(-\theta)\}$ and $\cosh \theta = \frac{1}{2}\{\exp\theta + \exp(-\theta)\}$ then it is straightforward to re-write Eq.C1b as:

$$\tanh(a + b) = \frac{\exp(a + b) - \exp(-(a + b))}{\exp(a + b) + \exp(-(a + b))} = \frac{1 - \exp(-2(a + b))}{1 + \exp(-2(a + b))}$$

$$= \frac{1 - e^{-\lambda}}{1 + e^{-\lambda}} \qquad (C2)$$

where in this case $\lambda = 2(a + b)$. We can immediately see that the hyperbolic tangent sum rule identity is composed of the quotient of two functions, each of which conforms to the template of Eq.A6, so that each of the two functions composing the quotient is MaxEnt. However, this immediately poses the question, is the resulting (quotient) function therefore also MaxEnt? We can binomially expand the denominator function of Eq.C2 as an infinite series as follows:

$$\tanh(a + b) = \left(1 - e^{-\lambda}\right)\left(1 + e^{-\lambda}\right)^{-1}$$

$$= \left(1 - e^{-\lambda}\right)\left(1 - e^{-\lambda} + e^{-2\lambda} - e^{-3\lambda} + e^{-4\lambda} - \cdots\right) \quad (C3)$$

Multiplying out the brackets and grouping the terms:

$$\tanh(a + b) = 1 - 2e^{-\lambda} + 2e^{-2\lambda} - 2e^{-3\lambda} + 2e^{-4\lambda} - 2e^{-5\lambda} + \cdots \quad (C4)$$

Eq.C4 closely resembles the template MaxEnt function of Eq.A6, except for its initial term, which is 'wrong' by the constant amount of one. That is to say, we can re-express Eq.C4 as:

$$\tanh(a + b) = -1 + 2\sum_{m=0}^{\infty}\left(-e^{-\lambda}\right)^{m} \quad (C5)$$

It's interesting to note that since the discrepancy is a *constant* (−1 in the case of Eq.C5) then all the results emerging by employing $Z \equiv \tanh(a + b)$ as a moment generating function (that is, to calculate quantities such as the mean and variance) are the same as if Z really were a MaxEnt function according to Eq.A6. This can be alternatively understood by approximating Eq.C5 to a finite series of N+1 terms, and dividing the constant overall 'discrepancy' of −1 amongst all the N+1 terms as follows:

$$Z = 2\sum_{m=0}^{N}\left(-e^{-\lambda}\right)^{m} - 1$$

$$= 2\sum_{m=0}^{N}\left(\left(-e^{-\lambda}\right)^{m} - \frac{1}{2N}\right) \quad (C6)$$

We then take the limit as $N \to \infty$, such that $Z \to \tanh(a + b)$, where the limiting value for each term z_m of the infinite series is:

$$z_i = 2\left(-e^{-\lambda}\right)^{m} - \frac{1}{N} \to 2\left(-e^{-\lambda}\right)^{m} \quad \text{as} \quad N \to \infty \quad (C7)$$

This makes it clear that in the limit as $N \to \infty$ the limiting value for each term z_m does conform to the MaxEnt template of Eq.A6, hence $\tanh(a + b)$ is MaxEnt and the isomorphic HSR $S_{HSR}(A,B)$ is also MaxEnt.

Thus we have proved that the Hyperbolic Sum Rule represents the MaxEnt (most likely) value of the probabilistic addition of two probabilities (i.e. the probabilistic OR function) in the absence of *any* additional information regarding the relationship between the two events A and B represented by the two probabilities to be summed. This HSR S_{HSR} is well-behaved, since it always unconditionally yields results between zero and one for the sum of any conventional probability quantities (lying between zero and one).

Appendix D: MaxEnt Treatment of HSR Featuring Multiple Hypotheses

D.1 'OR' Treatment of Multiple Conditional Hypotheses

Since the Hyperbolic Sum Rule is isomorphic to the hyperbolic tangent double-angle identity it immediately suggests how additional hypotheses can now be added to a given probabilistic scenario, while continuing to ensure a logical and consistent treatment (exploiting the method discussed by Cox in [ref.7]), and maintaining its intrinsically MaxEnt characteristic (see Appendix C).   In particular, considering an additional hypothesis D (with its own set of dependencies upon the

hypotheses A and B) being added to the present scenario, the following identity result can now be exploited:

$$\tanh\{(A + B) + D\} \equiv \frac{\tanh(A + B) + \tanh(D)}{1 + \tanh(A + B)\tanh(D)}$$

$$= \frac{\tanh A + \tanh B + \tanh D + \tanh A \tanh B \tanh D}{1 + \tanh A \tanh B + \tanh A \tanh D + \tanh B \tanh D} \quad \text{(D1)}$$

Since the addition (OR) function is associative, it is clear that all the terms in Eq.D1 must be mutually symmetrical, such that any other pair could have been chosen as the first argument, with an equivalent result. Therefore, we can write the probability for a set of three hypotheses A, B and D as:

$$p(A \text{ OR } B \text{ OR } D|C) = \frac{p(A|C) + p(B|C) + p(D|C) + p(A|C)p(B|C)p(D|C)}{1 + p(A|C)p(B|C) + p(A|C)p(D|C) + p(B|C)p(D|C)} \quad \text{(D2a)}$$

It should be noted that the conditionalities C associated with each probability are generic; but the probabilities employed in Eq.D2a for the explicit events A, B, and D are the *independent* probabilities for each of those events. That is to say, Eq.D2a is the MaxEnt expression for the HSR featuring three hypotheses, where no explicit knowledge about the dependencies between the three hypotheses is known.

Considering the RHS of Eq.D2a, the first group of (single) terms in the numerator represent the independent probabilities $\{p(A_n)\}$, while from a logical perspective, it is clear that the final (3rd-order) term of the numerator represents the product of these independent probabilities. Likewise, in the denominator, the three 2nd-order product terms (i.e. the 2nd-order permutations of the three hypotheses) each represent the symmetric probability products, as previously discussed. Clearly, if the third hypothesis is null, i.e. $p(D)=0$, then from Bayes' Theorem, all the terms containing D as a hypothesis also become zero, since $p(D|H)=p(D)*p(H|D)/p(H)$, where H may represent any of the combinations of the hypotheses A and B etc.; in which case, Eq.D2a reverts back to the two-hypotheses form of the equations Eqs.7 & 10c in the main text.

It is noteworthy that Eq.D1 also shows that multiple conditional hypotheses are to be treated using the OR function. For example, keeping the logical pair (A OR B) distinct from D, and using the form of Eq.7, if we have explicit knowledge of the conditional probabilities, we can re-write Eq.D1 as:

$$p\big((A \text{ OR } B) \text{ OR } D\big) = \frac{p(A \text{ OR } B) + p(D)}{1 + p(A \text{ OR } B|D)p\big(D|(A \text{ OR } B)\big)} \quad \text{(D3a)}$$

Expanding out the paired (A OR B) terms using the *tanh* identity of Eq.1a as template, we can further write:

$$p\big((A \text{ OR } B) \text{ OR } D\big) = \frac{\dfrac{p(A) + p(B)}{1 + p(A|B)p(B|A)} + p(D)}{1 + \dfrac{p(A|D) + p(B|D)}{1 + p\big(A|(B \text{ OR } D)\big)p\big(B|(A \text{ OR } D)\big)}p\big(D|(A \text{ OR } B)\big)} \quad \text{(D3b)}$$

In order to simplify the above equation we need the conditionalities of the product terms to be identical whenever they appear in Eq.D3b; this ensures that all the respective terms are interchangeable as required by considerations of symmetry and maximum entropy. Therefore we re-write Eq.D3b:

$$p(A \text{ OR } B \text{ OR } D) = \frac{\dfrac{p(A) + p(B)}{1 + p(A|B \text{ OR } \boldsymbol{D})p(B|A \text{ OR } \boldsymbol{D})} + p(D)}{1 + \dfrac{p(A|D) + p(B|D)}{1 + p\big(A|(B \text{ OR } D)\big)p\big(B|(A \text{ OR } D)\big)}p\big(D|(A \text{ OR } B)\big)} \quad \text{(D3c)}$$

where the inclusion of the additional conditionalities is indicated in bold for clarity. These additions here are reasonable, since clearly the hypothesis D is "in the vicinity" of the expression,

and it also symmetrically aligns the terms with similar terms in the rest of the equation. We then simplify:

$$p(A \text{ OR } B \text{ OR } D) = \frac{p(A) + p(B) + p(D) + p(D)p\big(A|(B \text{ OR } D)\big)\big(B|(A \text{ OR } D)\big)}{1 + p\big(A|(B \text{ OR } D)\big)p\big(B|(A \text{ OR } D)\big) + [p(A|D) + p(B|D)]p\big(D|(A \text{ OR } B)\big)} \text{(D3d)}$$

or, imposing the necessary symmetries:

$$p(A \text{ OR } B \text{ OR } D)$$

$$= \frac{p(A) + p(B) + p(D) + p(D|(\boldsymbol{A} \text{ OR } \boldsymbol{B}))p\big(A|(B \text{ OR } D)\big)p\big(B|(A \text{ OR } D)\big)}{1 + p\big(A|(B \text{ OR } D)\big)p\big(B|(A \text{ OR } D)\big) + [p\big(A|(D \text{ OR } \boldsymbol{B})\big) + p\big(B|(\boldsymbol{A} \text{ OR } D)\big)]p\big(D|(A \text{ OR } B)\big)}$$

(D3e)

Clearly, if the conditional probabilities are *not* known, then the appropriate MaxEnt version of Eq.D3e (where no extraneous assumptions or inadvertent information is added) is required, whereby only the independent probabilities are employed as appropriate, and as seen in Eq.D2a.

The probability of a hypothesis being true only requires at least one of the multiplicity of conditional hypotheses to be also true. This means that the more the number of conditional hypotheses, say the set {X}, that a given hypothesis A is conditional upon, then the higher is the conditional probability that A will be true. Indeed, as the number of conditional hypotheses tends towards infinity, it's clear that the conditional probability of A given the set {X} should tend to the intrinsic probability of A itself. In this situation the independent probability of A should assert itself, since it's almost certain that at least one of the many conditional hypotheses that are being summed (with the OR function) will be satisfied: $p(A|\{X\}_{\text{OR}}) \to p(A)$ as $\{X\} \to \infty$. The converse is also true: as the number of conditional hypotheses increases, they 'crowd out' the lone hypothesis A, and the probability of at least one of the conditional hypotheses {X} being true is increasingly independent of A: $p(\{X\}_{\text{OR}}|A) \to p(\{X\}_{\text{OR}})$ as $\{X\} \to \infty$. Interestingly, we see here the emergence of a form of Bayes' Theorem, since we can write:

$$\frac{p(\{X\}_{\text{OR}}|A)}{p(\{X\}_{\text{OR}})} \to 1 \quad \text{and} \quad \frac{p(A|\{X\}_{\text{OR}})}{p(A)} \to 1 \quad \text{as } \{X\} \to \infty \qquad \text{(D4a)}$$

Equating the two expressions offers us a type of Bayes' Theorem, based upon multiple contingent hypotheses:

$$p(A)p(\{X\}_{\text{OR}}|A) = p(\{X\}_{\text{OR}})p(A|\{X\}_{\text{OR}}) \qquad \text{as } \{X\} \to \infty \qquad \text{(D4b)}$$

Unfortunately, it's not intuitively obvious that the same is true for a low number of hypotheses {X}; however, in the case for just two hypotheses, A and X, it would indeed correspond to the conventional form of Bayes' Theorem.

D.2 Generalisation of the Concatenated "Hyperbolic Sum Rule" for Multiple Hypotheses

For a more convenient overview of the emerging 'pattern' as more hypotheses are concatenated, we re-write Eq.D2a using only the logical hypotheses (and making the conditional hypotheses {H} implicit) thus:

$$p(A \text{ OR } B \text{ OR } D) = \frac{A + B + D + ABD}{1 + AB + AD + BD} \qquad \text{(D2b)}$$

Adding a fourth hypothesis, E, we once again employ the hyperbolic tangent identity thus:

$$\tanh\{(A + B + D) + E\} \equiv \frac{\tanh(A + B + D) + \tanh(E)}{1 + \tanh(A + B + D)\tanh(E)} \qquad \text{(D5a)}$$

Again, only writing the symbol of the hypotheses for brevity and overview, we can therefore derive the associated hyperbolic sum rule for four such hypotheses as follows:

$$p(A \text{ OR } B \text{ OR } D \text{ OR } E) = \frac{\frac{A + B + D + ABD}{1 + AB + AD + BD} + E}{1 + \frac{A + B + D + ABD}{1 + AB + AD + BD}E} = \frac{A + B + D + E + ABD + ABE + ADE + BDE}{1 + AB + AD + BD + AE + BE + DE + ABDE} \quad \text{(D5b)}$$

For clarity, we remind ourselves that the final term of the denominator of Eq.D5b is to be understood as: $ABDE \equiv p(A|(B \text{ OR } D \text{ OR } E))p(B|(A \text{ OR } D \text{ OR } E))p(D|(A \text{ OR } B \text{ OR } E))p(E|(A \text{ OR } B \text{ OR } D))$, where we assume here the conditional probabilities are known. If, however, the conditional probabilities are not known, then the independent probabilities need to be employed as per the MaxEnt requirements.

D.3 Sum Rule for Multiple Hypotheses with Multiple Conditionalities

We can also use the hyperbolic tangent function to analyse the issue of calculating the MaxEnt probability of a set of multiple hypotheses that are conditionally dependent on an additional set of hypotheses. For example, the simplest such 'multiple' scenario is given by the probability quantity $p(A \text{ OR } B|D \text{ OR } E)$, which represents the probability of hypotheses A or B being correct, given that they are conditionally dependent on the hypotheses D or E. Using Eq.7 and the probability calculus of the section §D.1, this can be expressed as:

$$p(A \text{ OR } B|D \text{ OR } E) = \frac{p(A|D \text{ OR } E) + p(B|D \text{ OR } E)}{1 + p(A|B \text{ OR } D \text{ OR } E)p(B|A \text{ OR } D \text{ OR } E)} \tag{D6}$$

Clearly, the complexity of the overall expression increases rapidly with increasing number of hypotheses and conditional dependencies, with the attendant multitude of dependency terms and cross-products. However, we have shown here that such a calculation is indeed analytically possible, and the MaxEnt method for analysing higher numbers of hypotheses with their conditionalities is clearly apparent. This simple and generalisable result is not available from the conventional sum rule.

D.4 Generalised Hyperbolic Sum Rule for Multiple Hypotheses

What is becoming apparent is that the hyperbolic sum rule sorts out all the odd and even combinations of the hypotheses, such that the odd-ordered conjunction products are located in the numerator, while the even-ordered products are in the denominator. It is also clear that each of the elements in the numerator and denominator corresponds to an element of the binomial expansion of the expression $(a+b)^N$, where N represents the number of hypotheses; there being a total of 2^N different and unique combinations of the various hypotheses, and there being $C_n^N \equiv N!/n!(N-n)!$ combinatorial elements associated with each order of conjunction. In which case, the general expression for the Hyperbolic Sum Rule involving N different hypotheses each with their various mutual Bayesian dependencies is given by:

$$p(A_1 \text{ OR } A_2 \text{ OR } \dots \text{ OR } A_N) = \frac{\sum_{l=1,\text{odd}}^{N} \sum_{m=1}^{m=\frac{N!}{l!(N-l)!}} \prod_{n=P\{l,1:N\}}^{l} p(A_n|\{A_{n'}, l-1\})}{\sum_{l=0,\text{even}}^{N} \sum_{m=1}^{m=\frac{N!}{l!(N-l)!}} \prod_{n=P\{l,1:N\}}^{l} p(A_n|\{A_{n'}, l-1\})} \tag{D7}$$

The term in the curly brackets $\{A_{n'}, l-1\}$ of Eq.D7, forms the conditionality part of the probability $p(A_n|\{A_{n'}, l-1\})$, and indicates that there are l-1 conditional hypotheses (from the set $\{A_{n'}\}$, where the prime in the subscript indicates that the hypothesis A_n is itself excluded from the set) associated with the probability of hypothesis A_n. These conditional hypotheses (which, of course, exclude the hypothesis under question A_n) must be selected in the appropriate permutations from the set of N-1 probabilities making up the rest of the set $\{p(A_{n'})\}$. It is also clear that as the number of conditional hypotheses l−1 becomes large, then it also reasonable to start making the simplifying approximation $p(A_n|\{A_{n'}, l-1\}) \approx p(A_n)$ as $l \to \infty$, as previously discussed (§D.1).

We note that the case for l=0 (the first even term) in the denominator of Eq.D7 is simply the constant term of unity ('1') always found in the denominator of the HSR equation, as already seen in Eq.D3e & Eq.D6. In addition, the function $P\{l,1:N\}$ associated with the product (Π) terms in Eq.D7 represents the selection of the complete set of all possible (and distinct) permutated product terms (of overall degree l), each consisting of the product of l distinct probability terms from the set of N probabilities $\{p(A_n)\}$.

doi:10.20944/preprints202407.1660.v1

25

We note the close functional isomorphism of Eq.D7 with the transfer function (defined by the *Z*-transform) of an infinite impulse response (IIR) filter used in digital signal processing (see further in Appendix F), which is given by a quotient of summations in both the numerator and denominator:

$$H(z) = \frac{\sum_{l=0}^{L} b_l z^{-l}}{1 + \sum_{n=1}^{N} a_n z^{-n}} \qquad (D8)$$

The parameter *z*, is given by $z \equiv \exp(i\omega)$, where ω is a frequency in normalised units, such that *z* is therefore located on the unit circle in the complex plane (see the discussion in §3.1 for the self-reciprocal function S^m, whose roots also lie on the unit circle.) Of particular interest is to note that the conventional IIR filter generally has the first co-efficient of the denominator as unity, just like the HSR of Eq.7.

It is interesting to note that for each hypothesis being equally possible (a MaxEnt scenario), $p(A_n)=1/N$, and indeed, with the assumption that each conditional dependency is also equal and similarly given by $p(A_n|\{A_{n'}\})=1/N$, then we can see that the concatenated hyperbolic sum rule offers the following overall probability:

$$p(A_1 \text{ OR } A_2 \text{ OR } \dots \text{ OR } A_N) = \frac{\sum_{l=1,\text{odd}}^{N} \frac{N!}{l!\,(N-l)!} \prod_{n=1}^{l} p(A_n)}{\sum_{l=0,\text{even}}^{N} \frac{N!}{l!\,(N-l)!} \prod_{n=1}^{l} p(A_n)} = \frac{\sum_{l=1,\text{odd}}^{N} \frac{N!}{l!\,(N-l)!} \frac{1}{N^l}}{\sum_{l=0,\text{even}}^{N} \frac{N!}{l!\,(N-l)!} \frac{1}{N^l}} \qquad (D9a)$$

We make use of the following binomial expansion identity to solve for Eq.D9a:

$$\left(1 + \frac{1}{N}\right)^N = \sum_{l=0}^{N} \frac{N!}{l!\,(N-l)!} \frac{1}{N^l} \qquad (D9b)$$

In which case, the summations appearing in Eq.D9a, respectively consisting of the odd and even terms only, therefore each represent approximately half the amount on the LHS of Eq.D9b; this approximation being more accurate as *N* becomes large. In which case we can write:

$$\sum_{l=1,\text{odd}}^{N} \frac{N!}{l!\,(N-l)!} \frac{1}{N^l} \approx \sum_{l=1=0,\text{even}}^{N} \frac{N!}{l!\,(N-l)!} \frac{1}{N^l} \approx \frac{1}{2}\left(1 + \frac{1}{N}\right)^N \quad \text{for} \quad N \gg 1 \qquad (D9c)$$

So that substituting Eq.D9c into Eq.D9a we find:

$$p(A_1 \text{ OR } A_2 \text{ OR } \dots \text{ OR } A_N) = \frac{\frac{1}{2}\left(1 + \frac{1}{N}\right)^N}{\frac{1}{2}\left(1 + \frac{1}{N}\right)^N} = 1 \qquad (D9d)$$

It's also worth noting the identity $(1 + 1/N)^N \to e$ as $N \to \infty$, so that for large *N* we can see that we can re-write Eq.D9d as:

$$p(A_1 \text{ OR } A_2 \text{ OR } \dots \text{ OR } A_N) = \frac{0.5e}{0.5e} = 1 \quad \text{as } N \to \infty \qquad (D9e)$$

Clearly, the hyperbolic sum rule is well-behaved as $N \to \infty$, since the probability for at least one of the full set of *N* possible hypotheses being true is, as might be expected, unity.

D.5 Maximum Entropy Analysis for "Hyperbolic Sum Rule" of Multiple Hypotheses

In the previous paragraphs we showed that the Hyperbolic Sum Rule (HSR) is well-behaved for multiple hypotheses as $N \to \infty$, assuming a uniform (equal) MaxEnt distribution for the probabilities of the set of *N* possible hypotheses. However, we can also undertake a similar analysis for a finite temperature-based distribution of probabilities, also based on MaxEnt principles. In particular, we assume that we can arrange the probabilities of the hypotheses as a set of probabilities in a descending order of probability that conforms to a negative exponential distribution; this is the classic MaxEnt distribution of probabilities (see Jaynes 1982 [ref.14]).

Thus we assume that the independent probability of the nth hypothesis $p(A_n)$ is given by:

$$p(A_n) = \left\{ e^\beta \left(\frac{1 - e^{-\beta}}{1 - e^{-N\beta}} \right) \right\} e^{-\beta n} \tag{D10a}$$

where β acts as the appropriate Lagrange multiplier for the MaxEnt distribution. In passing, we note that Eq.D10a therefore conforms to the following implicit results:

$$\sum_{n=1}^{N} e^{-\beta n} = e^{-\beta} \left(\frac{1 - e^{-N\beta}}{1 - e^{-\beta}} \right) \quad \text{such that} \quad \sum_{n=1}^{N} p(A_n) = 1 \tag{D10b}$$

Substituting Eq.D10a into Eq.D9a, we obtain:

$$p(A_1 \text{ OR } A_2 \text{ OR } \dots \text{ OR } A_N) = \frac{\sum_{l=1,\text{odd}}^{N} \frac{N!}{l!\,(N-l)!} \left\{ \left(\frac{1 - e^{-\beta}}{1 - e^{-N\beta}} \right) e^\beta \right\}^l \prod_{n=P\{l,1:N\}}^{l} e^{-\beta n}}{\sum_{l=0,\text{even}}^{N} \frac{N!}{l!\,(N-l)!} \frac{1}{N^l} \left\{ \left(\frac{1 - e^{-\beta}}{1 - e^{-N\beta}} \right) e^\beta \right\}^l \prod_{n=P\{l,1:N\}}^{l} e^{-\beta n}} \tag{D10c}$$

which can be re-written:

$$p(A_1 \text{ OR } A_2 \text{ OR } \dots \text{ OR } A_N) = \frac{\sum_{l=1,\text{odd}}^{N} \frac{N!}{l!\,(N-l)!} \left\{ \left(\frac{1 - e^{-\beta}}{1 - e^{-N\beta}} \right) e^\beta \right\}^l \exp\{-\beta \sum_{n=P\{l,1:N\}}^{l} n\}}{\sum_{l=0,\text{even}}^{N} \frac{N!}{l!\,(N-l)!} \left\{ \left(\frac{1 - e^{-\beta}}{1 - e^{-N\beta}} \right) e^\beta \right\}^l \exp\{-\beta \sum_{n=P\{l,1:N\}}^{l} n\}} \tag{D10d}$$

The key thing to point out here is that the final product term (now transformed into a summation) at the end of both the numerator and denominator, corresponds to the permutated choices of all the l probability terms being considered for the lth-order conjunction. We can make an approximation for this by assuming that the average value of n for any given choice is simply $N/2$. In which case, choosing l such instances means that the average (expected) summation is then simply $\bar{n} = lN/2$. Thus we can approximate Eq.D10d as follows:

$$p(A_1 \text{ OR } A_2 \text{ OR } \dots \text{ OR } A_N) = \frac{\sum_{l=1,\text{odd}}^{N} \frac{N!}{l!\,(N-l)!} \left\{ \left(\frac{1 - e^{-\beta}}{1 - e^{-N\beta}} \right) e^\beta \right\}^l e^{-\frac{\beta l N}{2}}}{\sum_{l=0,\text{even}}^{N} \frac{N!}{l!\,(N-l)!} \left\{ \left(\frac{1 - e^{-\beta}}{1 - e^{-N\beta}} \right) e^\beta \right\}^l e^{-\frac{\beta l N}{2}}} \tag{D10e}$$

which can again be simplified as follows:

$$p(A_1 \text{ OR } A_2 \text{ OR } \dots \text{ OR } A_N) = \frac{\sum_{l=1,\text{odd}}^{N} \frac{N!}{l!\,(N-l)!} \left\{ \left(\frac{1 - e^{-\beta}}{1 - e^{-N\beta}} \right) e^{-\beta\left(\frac{N}{2}-1\right)} \right\}^l}{\sum_{l=0,\text{even}}^{N} \frac{N!}{l!\,(N-l)!} \left\{ \left(\frac{1 - e^{-\beta}}{1 - e^{-N\beta}} \right) e^{-\beta\left(\frac{N}{2}-1\right)} \right\}^l} \tag{D10f}$$

Using the approximation result of Eq.D9c for large N, we can approximate the numerator as:

$$\sum_{l=1,\text{odd}}^{N} \frac{N!}{l!\,(N-l)!} \left\{ \left(\frac{1 - e^{-\beta}}{1 - e^{-N\beta}} \right) e^{-\beta\left(\frac{N}{2}-1\right)} \right\}^l \approx \frac{1}{2} \left\{ 1 + \left(\frac{1 - e^{-\beta}}{1 - e^{-N\beta}} \right) e^{-\beta\left(\frac{N}{2}-1\right)} \right\}^N$$

$$\approx \frac{1}{2} \left\{ 1 + \frac{e^{-\frac{\beta}{2}}}{e^{-\frac{N\beta}{2}}} \left(\frac{e^{\frac{\beta}{2}} - e^{-\frac{\beta}{2}}}{e^{\frac{N\beta}{2}} - e^{-\frac{N\beta}{2}}} \right) e^{-\beta\left(\frac{N}{2}-1\right)} \right\}^N = \frac{1}{2} \left\{ 1 + \left(\frac{\sinh \beta/2}{\sinh N\beta/2} \right) e^{\frac{\beta}{2}} \right\}^N \tag{D10g}$$

Clearly for $\beta = 0$, we can assume that $\sinh(\beta/2)/\sinh(N\beta/2) = 1/N$, so that we revert to the initial uniform distribution of Eq.D9c. However, for $\beta \neq 0$ and $N \to \infty$, but $\beta \ll N$, then we can approximate Eq.D10g to:

$$\frac{1}{2} \left\{ 1 + \left(\frac{\sinh \beta/2}{\sinh N\beta/2} \right) e^{\beta/2} \right\}^N \approx \frac{1}{2} \left\{ 1 + \frac{e^{\beta/2}}{N} \right\}^N \approx \frac{1}{2} \exp\{e^{\beta/2}\} \tag{D10h}$$

We can assume the same for the even-ordered series of the denominator, such that substituting into Eq.D9a we have:

$$p(A_1 \text{ OR } A_2 \text{ OR } \dots \text{ OR } A_N) \approx \frac{0.5\exp\{e^{\beta/2}\}}{0.5\exp\{e^{\beta/2}\}} = 1 \quad \text{as} \quad N$$
$$\rightarrow \infty \qquad\qquad \text{(D10i)}$$

Equation Eq.D10i represents the MaxEnt generalisation for the Hyperbolic Sum Rule (HSR) as N gets very large. It can be seen that the HSR still remains very well behaved as the Lagrange multiplier (inverse-temperature parameter) β is varied for the distributions of the set $\{p(A_n)\}$, with β representing distributions associated with different effective temperatures.

Appendix E: Conventional Sum Rule is MaxEnt

We explore here how the *conventional sum rule* (CSR) behaves as the number of hypotheses considered increases.   In particular, Eq.2b (together with the HSR formalism of Eq.D7) indicates a possible (but actually, naïve and erroneous, as we shall see) concatenation rule for the CSR:

$$p(A_1 \text{ OR } A_2 \text{ OR } \dots \text{ OR } A_N) = \sum_{n=1}^{N} p(A_n) - \sum_{l=2}^{N} \sum_{m=1}^{m=\frac{N!}{l!(N-l)!}} \prod_{n=P\{l,1:N\}}^{l} p(A_n | \{A_{n'}, l-1\}) \qquad \text{(E1)}$$

where the final double-summation and product term is recognised to consist of all the possible combinations (at all orders) for two or more conjuncting terms (that is, for $l \geq 2$).

However, for the case when each hypothesis is equally possible and each conditional dependency is equal ($p(A_n|\{A_{n'}\})=1/N$; and $p(A_n)=1/N$) then:

$$p(A_1 \text{ OR } A_2 \text{ OR } \dots \text{ OR } A_N) = \sum_{n=1}^{N} \frac{1}{N} - \sum_{l=2}^{N} \frac{N!}{l!\,(N-l)!}\frac{1}{N^l} \qquad\qquad \text{(E2)}$$

The second summation of Eq.E2 (which only starts at l=2) may be rewritten:

$$\sum_{l=2}^{N} \frac{N!}{l!\,(N-l)!}\frac{1}{N^l} \equiv \sum_{l=0}^{N} \frac{N!}{l!\,(N-l)!}\frac{1}{N^l} - \sum_{n=1}^{N} \frac{1}{N} - 1 = \sum_{l=0}^{N} \frac{N!}{l!\,(N-l)!}\frac{1}{N^l} - 2 \qquad \text{(E3)}$$

and therefore:

$$p(A_1 \text{ OR } A_2 \text{ OR } \dots \text{ OR } A_N) = 1 - \sum_{l=2}^{N} \frac{N!}{l!\,(N-l)!}\frac{1}{N^l} = 1 - \left(\sum_{l=0}^{N} \frac{N!}{l!\,(N-l)!}\frac{1}{N^l} - 2 \right)$$
$$= \quad 3 - \sum_{l=0}^{N} \frac{N!}{l!\,(N-l)!}\frac{1}{N^l} = 3 - \left(1 + \frac{1}{N}\right)^N = 3 - e \quad \text{as } N \rightarrow \infty \qquad \text{(E4)}$$

Eq.E4 is therefore not well behaved as a reliable probability quantity (and consequently Eq.E1 is also unsatisfactory), since Eq.E4 tends to a non-unity result (3−e=0.282) as N gets large.   In which case, it is clear that the 'possible' CSR concatenation rule of Eq.E1 is deficient. However, in the hyperbolic sum rule case (Eq.D7) the odd- and even-ordered conjunctions are treated differently, with the even terms all sorted into the denominator and the odd terms located in the numerator. (This is the reason why Jaynes suggested that the Venn diagram of Figure 1 can 'mislead' since such respective groupings of the odd- and even-ordered overlaps isn't obvious.) Applying this alternative (HSR-inspired) model to Eq.E1, and then subsequently, to Eq.E4, such that the even-ordered terms are negative and the odd-ordered terms add positively, we have:

$$p(A_1 \text{ OR } A_2 \text{ OR } ... \text{ OR } A_N) = 1 - \sum_{l=2,\text{even}}^{N} \frac{N!}{l!\,(N-l)!}\frac{1}{N^l} + \sum_{l=3,\text{odd}}^{N} \frac{N!}{l!\,(N-l)!}\frac{1}{N^l}$$

$$= \sum_{l=1,\text{odd}}^{N} \frac{N!}{l!\,(N-l)!}\frac{1}{N^l} - \sum_{l=2,\text{even}}^{N} \frac{N!}{l!\,(N-l)!}\frac{1}{N^l}$$

$$= \frac{1}{2}\left(1+\frac{1}{N}\right)^N - \left[\frac{1}{2}\left(1+\frac{1}{N}\right)^N - 1\right] = 1 - 0.5e + 0.5e = 1 \quad (\text{E5})$$

Here, we recognise that the first term of the CSR (see the second line of Eq.E5) is equivalent to the l=1 term of the odd series of the first line: this term can be transferred to the odd series, extending its range. We also recognise that the even series is defined from l=2, that is to say, it omits the l=0 component. However, in using the approximations of Eq.D9c to calculate a reasonable closed-form for the series, we use the even series summation, and take away 1 in order to appropriately observe the omission of the l=0 component. Note the subtle ambiguity for why '1' in the first line of Eq.E5 (on the RHS) is present: is it due to the l=1 term of the odd series or is it an artefact of the l=0 term of the even series (as seen in the final line of Eq.D7)?

In any case, the CSR is now well behaved as the number of hypotheses increases. It seems that the general form of the CSR should be given by:

$$p(A_1 \text{ OR } A_2 \text{ OR } ... \text{ OR } A_N) = \sum_{l=1,\text{odd}}^{N} \sum_{m=1}^{m=\frac{N!}{l!(N-l)!}} \prod_{n=P\{l,1:N\}}^{l} p(A_n|\{A_{n'},l-1\}) - \sum_{l=2,\text{even}}^{N} \sum_{m=1}^{m=\frac{N!}{l!(N-l)!}} \prod_{n=P\{l,1:N\}}^{l} p(A_n|\{A_{n'},l-1\})$$

$$= 1 + \sum_{l=1,\text{odd}}^{N} \sum_{m=1}^{m=\frac{N!}{l!(N-l)!}} \prod_{n=P\{l,1:N\}}^{l} p(A_n|\{A_{n'},l-1\}) - \sum_{l=0,\text{even}}^{N} \sum_{m=1}^{m=\frac{N!}{l!(N-l)!}} \prod_{n=P\{l,1:N\}}^{l} p(A_n|\{A_{n'},l-1\}) \quad (\text{E6})$$

But it is fascinating that, in order to make the CSR behave reasonably for multiple hypotheses, key results have had to be borrowed from the HSR (i.e. that the odd and even terms take different signs). The top line of Eq.E6 can therefore be considered to represent the *arithmetic* version of the HSR (Eq.D7, apart from the fact that the even summation only runs from $l \geq 2$). Thus the {CSR, HSR} can be mutually considered as {arithmetic, geometric} analogues.

The HSR of Eq.D7 is consistent with many other physical phenomena, and it has also allowed us to find Eq.E6. The HSR elegantly employs *all* the terms of the complete series summation without exception, whereas the CSR has to omit the first $l = 0$ term from the even series (unless the additional 'unity' term is introduced) as seen in the last line of Eq.E6. In addition, the hyperbolic sum rule doesn't require a change in sign of the conjunction terms, since they are all positive as they appear in Eq.D7; this is in contrast to Eq.E6 where a change of sign for the odd and even conjunction terms is necessary. Without the insights of the hyperbolic nature of Eq.D7, and in particular, the properties of the multi-angle hyperbolic tangent function, it would not be possible to derive the form of Eq.E6 so easily. Note also that the generalised CSR for multiple hypotheses (Eq.E6) still offers results relatively close to the HSR, e.g. as seen in Figure 2.

Finally, a comment on whether the CSR of Eq.E6 is MaxEnt. To decide this, we exploit the fact that the CSR is an arithmetic version of the (MaxEnt) HSR, such that given the HSR of Eq.D7 is a quotient quantity (i.e. HSR$\equiv K/L$), then it is clear from Eq.E6 (when comparing it to Eq.D7 for the HSR) that the CSR is simply given by CSR$\equiv K-L+1$ for the same quantities K and L. However, in the previous Appendix C we assumed that HSR$\equiv$tanh($a+b$) which is already an intrinsically quotient quantity, such that it is equally clear that we can write: $K\equiv$sinh($a+b$) and $L\equiv$cosh($a+b$). We already demonstrated in Appendix C that the HSR is MaxEnt, but now with CSR$\equiv K-L+1$, then using the same hyperbolic identities in the vicinity of equation Eq.C2, we can simply write: CSR=$1-e^{-\lambda/2}$, where $\lambda\equiv2(a+b)$, which conforms to the MaxEnt template of Eq.A1a discussed in Appendix A.

Thus it is clear that the Eq.E6 for the CSR is therefore also MaxEnt, albeit within its narrower domain of applicability.

Appendix F: A Sum Rule for Finite and Infinite Impulse Response Filters

Filtering is an important aspect of most modern electronic processing systems, and is generally described using the theory of "digital signal processing" (**DSP**). Such DSP filters are divided into two main classes, which may be considered either as an "infinite impulse response" (**IIR**) or a "finite impulse response" (**FIR**) filter. Photonic device technologies using in digital telecommunications have also borrowed heavily from the DSP theory which has been applied to develop important optical components used in optical fibre communications[xlii]. In particular, Parker *et al*. state (citing [ref.42]): "*it is known that the 'Arrayed Waveguide Grating' (AWG) is a FIR filter, in contrast to the 'Fiber Bragg Grating' (FBG) which is an IIR device*", and they also show that the AWG and FBG filters (both important photonic devices) can be shown to be isomorphic to each other in a particularly simple mathematical relationship [xliii].

Considering only the first summation of the CSR of Eq.E6, we can see that the CSR is isomorphic to the general equation for a "finite impulse response" (**FIR**) filter:

$$H(z) = \sum_{n=0}^{N} b_n z^{-n} \qquad (F1)$$

where H is the transfer function: that is, the (complex-valued) frequency response of the filter. The mathematical form of Eq.F1 is equivalent to a discrete Fourier transform (**DFT**) since, as before for Eq.D8, the (complex) parameter z is given by $z \equiv \exp(i\omega)$, in normalised frequency units, and the parameters b_n (which subsume the other summation terms of Eq.E6) act as Fourier coefficients. The transfer functions of FBG and AWG photonic devices are related through the hyperbolic tangent (*tanh*) function ([ref.43]):

$$H_{FBG}(\omega) \equiv \tanh[H_{AWG}(\omega)] \qquad (F2)$$

What this means is that the IIR response of a FBG can be generated by taking the hyperbolic tangent of the FIR response of an equivalent AWG. In the case of a 'weak' IIR filter response (that is, a filter with little feedback) the small argument approximation (tanh $x \simeq x$ as $x \to 0$) means that the IIR and FIR become indistinguishable. This is analogous to the "First" (high energy) Born Approximation of quantum mechanics in which the spectral response of a scattering process is closely given by the Fourier transform of the spatial distribution describing a weak potential variation.

It is interesting that the IIR function of Eq.F2 is described using a hyperbolic tangent function in the same way as seen in Appendix C, Eqs.C1. This has the implication that there is an IIR filter *sum rule* consisting of the simple concatenation of different FIR filters within an overall hyperbolic tangent function, which is also of DSP engineering interest. For example, consider two different FIR functions H_{FIR1} and H_{FIR2}, each associated with a different IIR function via:

$$H_{IIRn} = \tanh(H_{FIRn}) \qquad \text{for } n = 1,2 \qquad (F3)$$

For the concatenation of the two FIR functions it is clear that we can write:

$$H_{IIR(1+2)} = \tanh(H_{FIR1} + H_{FIR2}) = \frac{\tanh(H_{FIR1}) + \tanh(H_{FIR2})}{1 + \tanh(H_{FIR1})\tanh(H_{FIR2})} \qquad (F4)$$

$$H_{IIR(1+2)} = \frac{H_{IIR1} + H_{IIR2}}{1 + H_{IIR1}H_{IIR2}} \qquad (F5)$$

Thus, the generic IIR filter (e.g. as similarly seen in Eq.D8) is seen to be also isomorphic to the HSR.

The interpretation and application of Eq.F5 for DSP engineering and filter design remains the work of future research. In the same way that the CSR is *not* a quotient function (in contradistinction to the HSR of Eq.7) the same applies to the generic mathematical form for a FIR Z-transform transfer function given by Eq.F1. And while DSP design of complex yet stable filters will often employ concatenation of multiple FIR filters, as shown in Appendix E so also can the CSR be concatenated to enable multiple hypotheses to be successfully and analytically combined into a single, well-behaved probabilistic sum rule expression.

References

i   E.T.Jaynes, *Probability Theory: The Logic of Science* (Cambridge University Press, 2003; edited by G.Larry Bretthorst); https://doi.org/10.1017/CBO9780511790423

ii   Stev Goodman, A Dirty Dozen: Twelve P-Value Misconceptions, *Seminars in Hematology* **45** (2008) 135-140; http://dx.doi.org/10.1053/j.seminhematol.2008.04.003

iii   Lewis G. Halsey, The reign of the p-value is over: what alternative analyses could we employ to fill the power vacuum? *Biology Letters* **15** (2019) 20190174 (8pp); http://dx.doi.org/10.1098/rsbl.2019.0174

iv   Paul E. Smaldino, Richard McElreath, The natural selection of bad science, *Royal Society Open Science* **3** (2016) 160384 (17pp); http://dx.doi.org/10.1098/rsos.160384

v   Kurt Gödel, Über formal unentscheidbare Sätze der *Principia Mathematica* und verwandter Systeme I (On Formally Undecidable Propositions of *Principia Mathematica* and Related Systems I), *Monatshefte für Mathematik und Physik* **38** (1931) 173–198; http://dx.doi.org/10.1007%2FBF01700692 (reprinted in English in: Stephen Hawking, *God Created the Integers*, Running Press, 2005)

vi   Chris Jeynes, Michael C.Parker, Margaret Barker, The Poetics of Physics, *Philosophies* **8** (2023) 3 (54pp); http://dx.doi.org/10.3390/philosophies8010003

vii   R.T. Cox, Probability, frequency and reasonable expectation, *American Journal of Physics* **14** (1946) 1-13; https://doi.org/10.1119/1.1990764

viii   Torsten A. Enßlin, Jens Jasche, John Skilling, The Physics of Information, *Annalen der Physik* **531** (2019) 1900059 (2pp); https://doi.org/10.1002/andp.201900059

ix   Allen Caldwell, Aspects of Frequentism, *Annalen der Physik* **531** (2019) 1900057 (2pp); https://doi.org/10.1002/andp.201700457

x   M.C.Parker, C.Jeynes, Maximum Entropy (Most Likely) Double Helical and Double Logarithmic Spiral Trajectories in Space-Time, *Scientific Reports* **9** (2019) 10779 (10 pp); http://dx.doi.org/10.1038/s41598-019-46765-w

xi   Rolf Landauer, Information is Physical, *Physics Today* **44** (1991) 23-29; http://dx.doi.org/10.1063/1.881299

xii   M.C. Parker, P. Wright, A. Lord, Multiple Fiber, Flexgrid Elastic Optical Network Design using MaxEnt Optimization [Invited], *Journal of Optical Communications and Networking* **7** (2015) B194-B201; http://dx.doi.org/10.1364/JOCN.7.00B194

xiii   E. T. Jaynes, *Where Do We Stand on Maximum Entropy?* (1978: in R.D.Rosenkrantz, ed., *Papers on Probability, Statistics and Statistical Physics*. Synthese Library, vol 158. Springer, Dordrecht, 1983; reprinted 1989); https://doi.org/10.1007/978-94-009-6581-2_10

xiv   E.T. Jaynes, On the Rationale of Maximum Entropy Methods, *Proceedings of the IEEE*, **70** (1982) 939-952; (invited paper) https://doi.org/10.1109/PROC.1982.12425

xv   Torsten A. Enßlin, Information Theory for Fields, *Annalen der Physik* **531** (2019) 1800127 (18pp); http://dx.doi.org/10.1002/andp.201800127

xvi   J.S. Toll, Causality and the dispersion relation: logical foundations, *Phys. Rev.* **104** (1956) 1760-1770; https://doi.org/10.1103/PhysRev.104.1760

xvii   H. Primas, *Time, Temporality, Now: The Representation of Facts in Physical Theories* (Springer Verlag, Berlin, 1997), p. 241

xviii   M.C.Parker, C.Jeynes, W.N.Catford, Halo Properties in Helium Nuclei from the Perspective of Geometrical Thermodynamics, *Annalen der Physik* **534** (2022) 2100278 (11pp); http://dx.doi.org/10.1002/andp.202100278

xix   M.C. Parker, S.D. Walker, Information transfer and Landauer's principle, *Optics Communications*, 229 (2004) 23-27; https://doi.org/10.1016/j.optcom.2003.10.019

xx   M.C. Parker, C. Jeynes, Entropic uncertainty principle, partition function and holographic principle derived from Liouville's Theorem, *Physics Open* **7** (2021) 100068 (11pp); https://doi.org/10.1016/j.physo.2021.100068

xxi   Roger Penrose, *The Road to Reality: a complete guide to the laws of the Universe* (Jonathan Cape, London, 2004)

xxii   R.Courant, D.Hilbert, *Methods of Mathematical Physics* (Interscience - Wiley, New York, 1962) p.178

xxiii   Michael C. Parker, Chris Jeynes, A Relativistic Entropic Hamiltonian–Lagrangian Approach to the Entropy Production of Spiral Galaxies in Hyperbolic Spacetime, *Universe* **7** (2021) 325 (15pp); https://doi.org/10.3390/universe7090325

xxiv   Matt Visser, Conservative entropic forces, *Journal of High Energy Physics* (2011) Issue 10, 140 (21pp); http://dx.doi.org/10.1007/JHEP10(2011)140

xxv   Kevin H. Knuth, Lattices and Their Consistent Quantification, *Annalen der Physik* **531** (2019) 201700370; https://doi.org/10.1002/andp.201700370

xxvi   Gregg Jaeger, Information and the Reconstruction of Quantum Physics, *Annalen der Physik* **531** (2018) 201800097; https://doi.org/10.1002/andp.201800097

xxvii   J.D.Jackson, *Classical Electrodynamics* (John Wiley, New York, 1962)

xxviii   J. Buus, M.-C. Amann, D.J. Blumenthal, *Tunable Laser Diodes* (SPIE Press and Wiley-IEEE Press, 1998) 2nd Edition 2005, Appendix E: Theory of General Reflectors

xxix   S.W. Corzine, R.H. Yan, L.A. Coldren, A *tanh* substitution technique for the analysis of abrupt and graded interface multilayer dielectric stacks, *IEEE J. of Quant. Electr.*, **27** (1991) 2086-2090; https://doi.org/10.1109/3.135163

xxx   John Skilling, Kevin H. Knuth, The Symmetrical Foundation of Measure, Probability, and Quantum Theories, *Annalen der Physik* **531** (2018) 201800057; https://doi.org/10.1002/andp.201800057

xxxi   Karen Barad, *Meeting the Universe Halfway: quantum physics and the entanglement of matter and meaning* (Duke University Press 2007)

xxxii   Michael Polanyi, *Personal Knowledge: towards a post-critical philosophy* (University of Chicago Press, 1958)

xxxiii   C. Howson, P. Urbach, Bayesian reasoning in science, *Nature* **350** (1991) 371-374; https://doi.org/10.1038/350371a0

xxxiv   E.T.Jaynes, Information Theory and Statistical Mechanics II, *The Physical Review* **108** (1957) 171-190; https://doi.org/10.1103/PhysRev.108.171

xxxv   M.C.Parker, C.Jeynes, A Maximum Entropy Resolution to the Wine/Water Paradox, *Entropy* **25** (2023) 1242 (10pp); https://doi.org/10.3390/e25081242

xxxvi   M.C.Parker, C.Jeynes, *ab initio* thermodynamics calculation of beta decay rates, *Annalen der Physik* **535** (2023) 2300259 (11pp); http://dx.doi.org/10.1002/andp.202300259

xxxvii   M.C.Parker, C.Jeynes, Relating a system's Hamiltonian to its Entropy Production using a Complex-Time approach, *Entropy* **25** (2023) 629 (19pp); http://dx.doi.org/10.3390/e25040629

xxxviii   Yu Zhang, Fanghui Huang, Xinyang Deng and Wen Jiang,   A New Total Uncertainty Measure from A Perspective of Maximum Entropy Requirement, *Entropy* **23** (2021) 1061 (20pp); https://doi.org/10.3390/e23081061

xxxix   Benno I. Simmons, Jeferson Vizentin-Bugoni, Pietro K. Maruyama, Peter A. Cotton, Oscar H. Marín-Gómez, Carlos Lara, Liliana Rosero-Lasprilla, María A. Maglianesi, Raul Ortiz-Pulido, Márcia A. Rocca, Licléia C. Rodrigues, Boris A. Tinoco, Marcelo F. Vasconcelos, Marlies Sazima, Ana M. Martín González, Jesper Sonne, Carsten Rahbek, Lynn V. Dicks, Bo Dalsgaard, William J. Sutherland, Abundance drives broad patterns of generalisation in plant–hummingbird pollination networks, *Oikos* **128** (2019) 1287-1295; https://doi.org/10.1111/oik.06104

xl   Nicholas J.A. Harvey, Robert Kleinberg, Chandra Nair, Yunnan Wu, A "Chicken & Egg" Network Coding Problem, *2007 IEEE International Symposium on Information Theory* (24-29 June 2007); https://doi.org/10.1109/ISIT.2007.4557215

xli   Ariel Caticha, *Lectures on Probability, Entropy, and Statistical Physics*, Invited lectures at MaxEnt 2008, the 28th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering (July 8-13, 2008, Boraceia Beach, Sao Paulo, Brazil); https://doi.org/10.48550/arXiv.0808.0012 (accessed 15th June 2023)

xlii   G. Lenz, B. J. Eggleton, C. K. Madsen, C. R. Giles, G. Nykolak, Optimal dispersion of optical filters for WDM systems, *IEEE Photonics Technology Letters*, **10** (1998) 567–569;   https://doi.org/10.1109/68.662596

xliii   Michael C. Parker, Stuart D. Walker, Robert J. Mears, An Isomorphic Fourier Transform Analysis of AWGs and FBGs, *IEEE Photonics Technology Letters*, **13** (2001) 972-974; https://doi.org/10.1109/68.942664