# The Rényi Entropies Operate in Positive Semifields

**Francisco J. Valverde-Albacete [†]** and **Carmen Peláez-Moreno \*,[†]**

Department of Signal Theory and Communications, Universidad Carlos III de Madrid, 28911 Leganés, Spain
* Correspondence: carmen@tsc.uc3m.es; Tel.: +34-91-624-8771
† These authors contributed equally to this work.

**Abstract:** We set out to demonstrate that the Rényi entropies are better thought of as operating in a type of non-linear semiring called a positive semifield. We show how the Rényi's postulates lead to Pap's g-calculus where the functions carrying out the domain transformation are Rényi's information function and its inverse. In its turn, Pap's g-calculus under Rényi's information function transforms the set of positive reals into a family of semirings where "standard" product has been transformed into sum and "standard" sum into a power-emphasized sum. Consequently, the transformed product has an inverse whence the structure is actually that of a positive semifield. Instances of this construction lead to idempotent analysis and tropical algebra as well as to less exotic structures. We conjecture that this is one of the reasons why tropical algebra procedures, like the Viterbi algorithm of dynamic programming, morphological processing, or neural networks are so successful in computational intelligence applications. But also, why there seem to exist so many computational intelligence procedures to deal with "information" at large.

**Keywords:** shifted Rényi entropy; Pap's g-calculus; non-Newtonian calculus; positive commutative semifields; idempotent semifields; artificial intelligence; machine learning; computational intelligence

## 1. Introduction

Some non-linear algebras like the min-plus semiring [1] have wide application in applications trying to model intelligent behaviour, among other domains [2]. In this paper we fuse ideas from the theory of the means, the abstract algebra of semirings and information theory to propose the following explanation for their ubiquity: *that they are in fact the natural algebras in which each of the infinite instances of the Rényi informations [3]:*

$$H_\alpha(P_X) = \frac{1}{1-\alpha} \log\left(\sum_{i=1}^n p_i^\alpha\right) \tag{1}$$

*operate so that applications are "matched" to specific values of the order parameter α. From this starting point, the quasi-genetic, social process of scientific and technological advance selects which value of the parameter is most suited to make a technique work in a particular application.*

We will argue in this paper that these non-standard algebras are positive semifields, a special type of semiring with multiplicative inverses but no additive inverses. From the point of view of abstract algebra, recall that a semiring is an algebra $\mathcal{S} = \langle S, \oplus, \otimes, \epsilon, e \rangle$ whose additive structure $\langle S, \oplus, \epsilon \rangle$ is a commutative monoid and whose multiplicative structure $\langle S \backslash \{\epsilon\}, \otimes, e \rangle$ is a monoid with multiplication distributing over addition from right and left and an additive neutral element absorbing for $\otimes$, i.e., $\forall a \in S, \epsilon \otimes a = \epsilon$ [4]. A semiring is commutative if its product is commutative. All semirings considered in this paper are commutative, whence we will drop the qualification altogether. A semiring is zero sum-free if whenever a sum is null all summands are also null, and entire if it has no non-null

factors of zero. A semiring is positive if it is zero sum-free and entire [2]. Finally, a semiring is a semifield if there exists a multiplicative inverse for every element $a \in S$—notated as $a^{-1}$—except the null. All semifields are entire.

In this paper we concern ourselves with positive semifields, e.g., entire, zero sum-free semirings with a multiplicative inverse but no additive inverse (see Section 2.2), whose paragon is the set of non-negative reals with its standard operations $\mathbb{R}_{\geq 0} = \langle [0, \infty), +, \times, \cdot^{-1}, 0, 1 \rangle$. Their scientific and engineering interest lies in the fact that positive semifield applications abound in many areas of research. To cite but a few examples:

- Artificial intelligence (AI) [5] is an extensive field under which applications abound dealing with minimizing costs or maximizing utilities. Semifields and their dual-orderings (see Sections 2.2 and 2.2.3) provide a perspective to mix these two kinds of valuations.
- Machine learning (ML) [6] makes heavy use of Probability Theory, which is built around the positive semifield of the non-negative reals with their standard algebra $\mathbb{R}_{\geq 0}$ and negative logarithms thereof—called log-probabilities or log-likelihoods depending on the point of view—both of which are positive semifields, as shown in Sections 2.2, 3.1.1, and 3.2.1.
- Computational intelligence (CI) [7] makes heavy use of positive semirings in the guise of fuzzy semirings. Although semifields cannot be considered "fuzzy" for several technical reasons, the name is sometimes an umbrella term under which non-standard algebras are included, many of which are semifields, e.g., the morphological semifield of morphological processing and memories, a special case of the semifields in Section 3.1.
- Other applications of positive semifields not related to modeling intelligence include Electrical Network analysis and synthesis (see the example in Section 2.2), queuing theory [8] and flow shop scheduling [1].

To build a basis for our initial conjecture, in methodological Section 2 we revisit a couple of seemingly unrelated observations: First, a simple shifting $r = \alpha - 1$ of the Rényi order parameter makes the connection between these entropies and the weighted Hölder means transparent [9] (Section 2.1). Second, the Rényi procedure to justify the form of the entropy [10] dovetails seamlessly into Pap's g-calculus [11] as a construction on positive semifields (Section 2.2).

In the results Section 3 we conjoin both observations to support our hypothesis: we first briefly introduce an "entropic" semifield (Section 3.1.1) prior to proving, secondly, how the shifted Rényi entropy takes values in positive semifields, and also how this reflects back on the semifield of positive numbers interpreted as "probabilities" (Section 3.1.2). Third we use the newly defined semifield notation in two application instances: one, to provide an exception-free definition of the means for the extended positive numbers (Section 3.2.2), and two, to rewrite and provide a brief analysis into the Viterbi algorithm in terms of the semifield of max-information (Section 3.2.1). Finally we unfold the argumentation for our conjecture that AI, ML and CI applications are mostly dealing with Rényi entropies of different order (Section 3.3) and we trace some related art spanning the last 70 years (Section 3.4). We end the paper with a discussion of the issues touched upon it and some conclusions.

## 2. Materials and Methods

### 2.1. The Shifted Rényi Entropy

We first take the standpoint of information theory. Recall that the weighted power or Hölder mean of order $r$ [12] is defined as

$$M_r(\vec{w}, \vec{x}) = \left( \frac{\sum_{i=1}^{n} w_i \cdot x_i^r}{\sum_k w_k} \right)^{\frac{1}{r}} = \left( \sum_{i=1}^{n} \frac{w_i}{\sum_k w_k} \cdot x_i^r \right)^{\frac{1}{r}}. \tag{2}$$

When $r \to 0$ the geometric mean appears:

$$M_0(\vec{w}, \vec{x}) = \lim_{r \to 0} M_r(\vec{w}, \vec{x}) = \left(\Pi_{i=1}^n x_i^{w_i}\right)^{\frac{1}{\Sigma_k w_k}}.$$

Important cases of the means for historical and practical reasons and their relation to the (shifted and original) Rényi entropy are shown in Table 1.

**Table 1.** Relation between the most usual weighted power means, Rényi entropies and shifted versions of them, from [9].

| Mean Name | Mean $M_r(\vec{w}, \vec{x})$ | Shifted Entropy $\tilde{H}_r(P_X)$ | Entropy Name | $\alpha$ | $r$ |
|---|---|---|---|---|---|
| Maximum | $\max_i x_i$ | $\tilde{H}_\infty = -\log\max_i p_i$ | min-entropy | $\infty$ | $\infty$ |
| Arithmetic | $\sum_i w_i x_i$ | $\tilde{H}_1 = -\log\sum_i p_i^2$ | Rényi's quadratic | 2 | 1 |
| Geometric | $\Pi_i x_i^{w_i}$ | $\tilde{H}_0 = -\sum_i p_i \log p_i$ | Shannon's | 1 | 0 |
| Harmonic | $(\sum_i w_i \frac{1}{x_i})^{-1}$ | $\tilde{H}_{-1} = \log n$ | Hartley's | 0 | $-1$ |
| Minimum | $\min_i x_i$ | $\tilde{H}_{-\infty} = -\log\min_i p_i$ | max-entropy | $-\infty$ | $-\infty$ |

To leverage the theory of generalized means to our advantage, we start with a correction to Rényi's entropy definition: in [9] a case is made for shifting the original statement from the index that Rényi proposed $\alpha$ to $r = \alpha - 1$. Specifically, the transformation function for the average of surprisals [10] in the Rényi entropy (1) is arbitrary in the parameter $\alpha$ chosen for it and we may substitute $r = \alpha - 1$ to obtain the pair of formulas:

$$\varphi'(h) = b^{-rh} \qquad\qquad \varphi'^{-1}(p) = \frac{-1}{r}\log_b p. \tag{3}$$

Note that the basis of the logarithm is usually not theoretically important. This is what we mean when we use $\log x$ or $\log_b x$ for an unspecified $b$.

Consider the extension of the surprisal, or Hartley's function, to non-negative numbers as

$$\mathfrak{I}_*(\cdot) : [0, \infty] \to [-\infty, \infty] \qquad\qquad p \mapsto \mathfrak{I}_*(p) = -\log_b p. \tag{4}$$

This is an antitone, one-to-one function from $[0, \infty]$ onto $[-\infty, \infty]$, with inverse $(\mathfrak{I}_*)^{-1}(h) = b^{-h}$ for $h \in [-\infty, \infty]$. It is also the inverse of (3) for $r = 1$. The following definitions are thus obtained:

**Definition 1.** *Let $P_X(x_i) = p_i$ and $Q_X(y_i) = q_i$ be distributions with compatible support. Then the expression of the shifted Rényi entropy $\tilde{H}_r(P_X)$, cross-entropy $\tilde{X}_r(P_X\|Q_X)$, and divergence $\tilde{D}_r(P_X\|Q_X)$ are:*

$$\tilde{H}_r(P_X) = \mathfrak{I}_*(M_r(P_X, P_X)) \quad \tilde{X}_r(P_X\|Q_X) = \mathfrak{I}_*(M_r(P_X, Q_X)) \quad \tilde{D}_r(P_X\|Q_X) = -\mathfrak{I}_*\left(\frac{P_X}{Q_X}\right) \tag{5}$$

Several brief points are worth mentioning in this respect, although the full argument can be followed in [9]:

- The properties of the Rényi entropy, therefore, stem from those of the mean, inversion and the logarithm.
- This is not merely a cosmetic change, since it has the potential to allow the simplification of issues and the discovery of new ones in dealing with the Rényi magnitudes. For instance, since the means are defined for all $r \in [-\infty, \infty]$ there cannot be any objection to considering negative values for the index of the shifted entropy. This motivates calling $\tilde{H}_r(P_X)$ the Rényi spectrum (of entropy).

- The definition makes it also evident that the shifted cross-entropy seems to be the more general concept, given that the shifted entropy and divergence are clearly instances of it.

$$\tilde{H}_r \left( P_X \right) = \tilde{X}_r \left( P_X \| P_X \right) \qquad\qquad \tilde{D}_r \left( P_X \| Q_X \right) = -\tilde{X}_r \left( P_X \| P_X / Q_X \right)$$

- Also, the following lemma from [9] shows that the Rényi entropies can be rewritten in terms of the Shannon cross entropy and the Kullback–Leibler divergence:

**Lemma 1.** *Let* $r, s \in \mathbb{R} \cup \{\pm\infty\}$, $P_X \in \Delta^{n-1}$ *where* $\Delta^{n-1}$ *is the simplex over the support* $\mathrm{supp}\,(X)$ *and* $\tilde{q}_r(P_X) = \left\{ \frac{p_i p_i^r}{\sum_k p_k p_k^r} \right\}_{i=1}^n$ *are the escort probabilities [9,13]. Then,*

$$\tilde{H}_r \left( P_X \right) = \frac{1}{r} \tilde{D}_0 \left( \tilde{q}_r(P_X) \| P_X \right) + \tilde{X}_0 \left( \tilde{q}_r(P_X) \| P_X \right) \tag{6}$$

$$\tilde{H}_r \left( P_X \right) = \frac{-1}{r} \tilde{H}_0 \left( \tilde{q}_r(P_X) \right) + \frac{r+1}{r} \tilde{X}_0 \left( \tilde{q}_r(P_X) \| P_X \right). \tag{7}$$

- Lemma (1) rewrites the entropies in terms of the geometric means which is, by no means, the only rewriting possible. Indeed, some would say that the arithmetic mean is more natural, and this is the program of information theoretic learning [14], where it is explored under the guise of $H_2(P_X) \equiv \tilde{H}_1 \left( P_X \right)$.
- The shifting clarifies the relationship between quantities around the Rényi entropy: given (4), from every measure of information an equivalent average probability emerges. In particular:

**Definition 2.** *Let* $X \sim P_X$ *with Rényi spectrum* $\tilde{H}_r \left( P_X \right)$. *Then the equivalent probability function of* $\tilde{P}_r \left( P_X \right)$ *is the Hartley inverse of* $\tilde{H}_r \left( P_X \right)$ *over all values of* $r \in [-\infty, \infty]$

$$\tilde{P}_r \left( P_X \right) = \left( \mathfrak{I}_* \right)^{-1} \left( \tilde{H}_r \left( P_X \right) \right) = M_r(P_X, P_X). \tag{8}$$

- Similarly, the information potential $\tilde{V}_r \left( P_X \right) = E_{P_X}\{P_X^r\}$ has been independently motivated in a number of applications [14]. The next lemma is immediate using the conversion function (3) on the shifted entropy (5).

**Lemma 2.** *Let* $X \sim P_X$. *The information potential is the* $\varphi'$ *image of the shifted Rényi entropy*

$$\tilde{V}_r \left( P_X \right) = E_{P_X}\{P_X^r\} = \sum_i \frac{p_i}{\sum_k p_k} p_i^r = b^{-r\tilde{H}_r(P_X)} = \varphi'(\tilde{H}_r \left( P_X \right)). \tag{9}$$

These three quantities—the shifted Rényi entropy, the equivalent probability function, and the information potential—stand in a relationship, as described in Figure 1a, whose characterization is the conducting thread of this paper. In [9] other formulas to convert them into each other are tabulated.
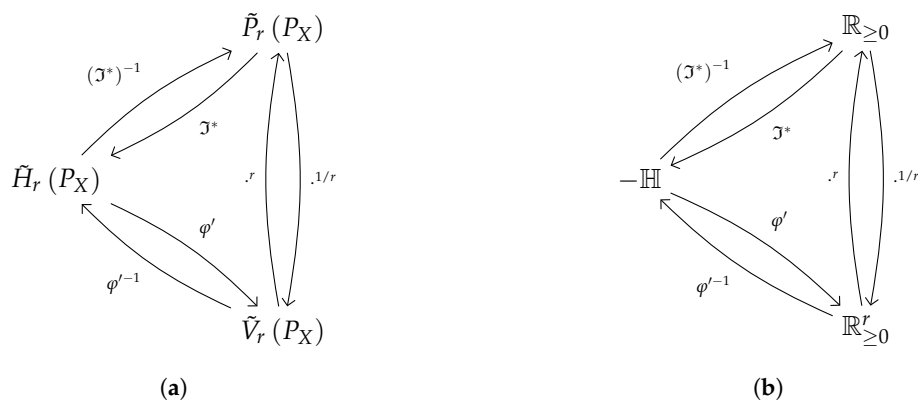
**Figure 1.** Schematics of relationship due to entropic isomorphisms. (**a**) Between shifted entropy-related quantities (from [9]). (**b**) Between entropy-related domains (see Section 3.1).

### 2.2. Positive Semifields

We now retake the point of view of abstract algebra. From the material in Section 2.1 it seems evident that non-negative quantities are important for our purposes. Non-negativity is captured by the concept of zero sum-free semiring mentioned above, but we focus in the slightly less general notion of dioid (for double monoid) [2] where there is an order available that interacts "nicely" with the operations of the algebra.

#### 2.2.1. Complete and Positive Dioids

A dioid is a commutative semiring $\mathcal{D}$ where the canonical preorder relation—$a \preccurlyeq b$ if and only if there exists $c \in D$ with $a \oplus c = b$—is actually an order $\langle D, \preccurlyeq \rangle$. In a dioid the canonical orden relation is compatible with both $\oplus$ and $\otimes$ ([2], Chapter 1, Prop. 6.1.7) and the additive zero *epsilon* is always the infimum of the dioid or bottom hence the notation $\epsilon = \inf D = \bot$. Dioids are all zero sum-free, that is, they have no non-null additive factors of zero: if $a, b \in D, a \oplus b = \epsilon$ then $a = \epsilon$ and $b = \epsilon$.

A dioid is complete if it is complete as an ordered set for the canonical order relation, and the following distributivity properties hold, for all $A \subseteq D, b \in D$,

$$\left( \bigoplus_{a \in A} a \right) \otimes b = \bigoplus_{a \in A} (a \otimes b) \qquad\qquad b \otimes \left( \bigoplus_{a \in A} a \right) = \bigoplus_{a \in A} (b \otimes a) \tag{10}$$

In complete dioids, there is already a top element $\top = \oplus_{a \in D} a$.

A semiring is entire or zero-divisor free if $a \otimes b = \epsilon$ implies $a = \epsilon$ or $b = \epsilon$. If the dioid is entire, its order properties justifies calling it a positive dioid or information algebra [2].

#### 2.2.2. Positive Semifields

A semifield, as mentioned in the introduction, is a semiring whose multiplicative structure $\langle K \setminus \{\epsilon\}, \otimes, e, \cdot^{\circledast} \rangle$ is a group, where $\cdot^{\circledast} : K \to K$ is the function to calculate the inverse such that $\forall u \in K, u \otimes u^{\circledast} = e$. Since all semifields are entire, dioids that are at the same time semifields are called positive semifields, of which the positive reals or rationals are a paragon.

**Example 1** (Semifield of non-negative reals)**.** *The nonnegative reals* $\mathbb{R}_{\geq 0} = \langle [0, \infty), +, \times, \cdot^{\circledast}, \bot = 0, e{=}1 \rangle$ *are the basis for the computations in Probability Theory and other quantities that are multiplicatively aggregated. The 0 has no inverse hence* $\mathbb{R}_{\geq 0}$ *is incomplete. Consider modeling utilities and costs with this algebra. For utilities 0 acts as a least element—a bottom—and the order is somewhat directed "away" from this element hence the underlying order is* $\langle [0, \infty), \leq \rangle$, *that is to say utilities are to be maximized. But its dual order* $\langle [0, \infty), \geq \rangle$—*intended to model costs, to be minimized—has no bottom (see below), hence applications using,*

*for instance, multiplicative costs and utilities at the same time, will be difficult to carry out in this algebra and notation.*

Figure 2 shows a concept lattice of the position of positive semifields within the commutative semirings, as well as the better known collateral families of fields like $\mathbb{R}$ and distributive lattices like $\langle [0,1], \max, \min \rangle$ (cfr. ([15], Figure 1)).
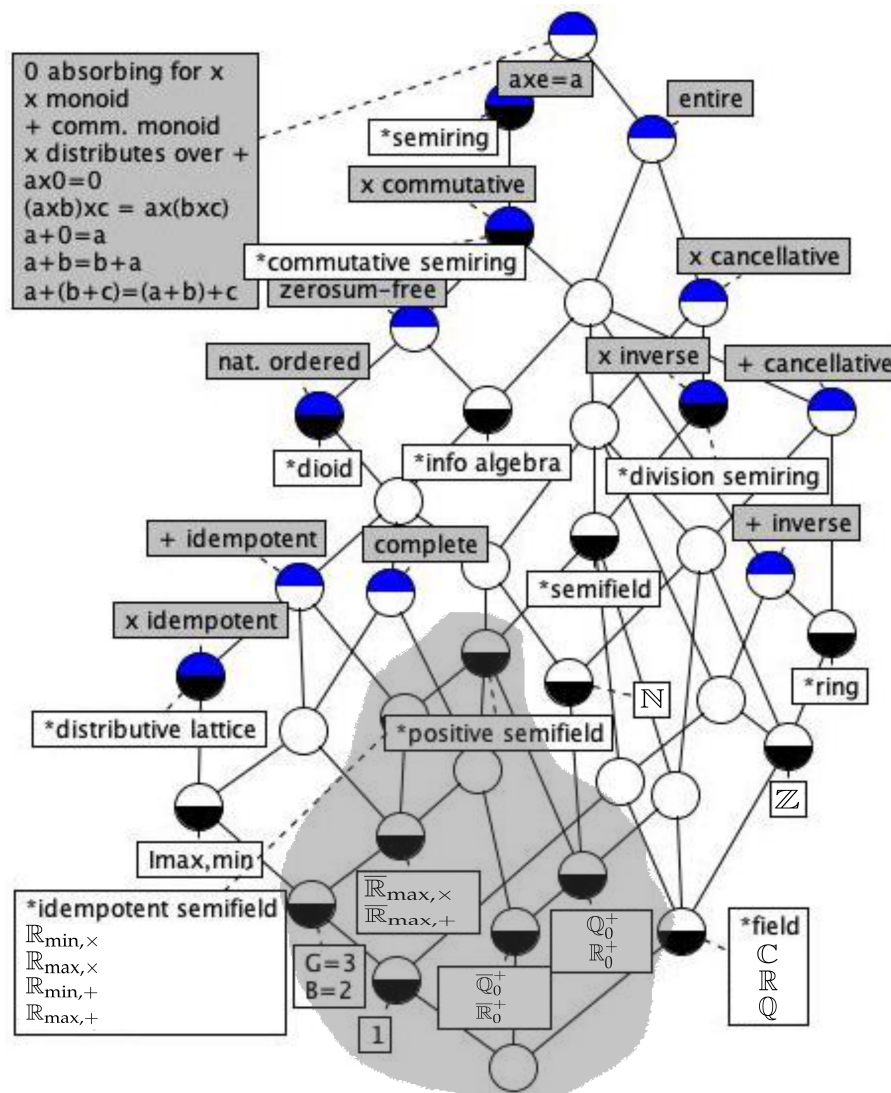


**Figure 2.** Lattice of a selection of abstract (leading asterisk, white label) and concrete (white label) commutative semirings and their properties (grey label) mentioned in the text. Adapted from [15]. Each node is a concept of abstract algebra: its properties are obtained from the gray labels in nodes upwards, and its structures from the white labels in nodes downwards. The picture is related to the chosen sets of properties and algebras and does not fully reflect the structure of the class of semirings. We have chosen to highlight positive semifields, like $\mathbb{R}_0^+$ and $\mathbb{Q}_0^+$.

In incomplete semifields like the one above, the inverse of the bottom element is the "elephant in the room" to be avoided in computations. Fortunately, semiring theory provides a construction to supply this missing element in semifields [4]. However, the problem with the dual order in the semifield mentioned above suggests that we introduce both completions at the same time through the following theorem (the need for the dotted notation will be made clear after the theorem).

**Theorem 1.** *For every incomplete positive semifield $\mathcal{K} = \langle K, \oplus, \otimes, \cdot^{\circledast}, \bot, e \rangle$*

1.　*There is a pair of completed semifields over $\overline{K} = K \cup \{\top\}$*

$$\overline{\mathcal{K}} = \langle K, \underset{.}{\oplus}, \underset{.}{\otimes}, \cdot^{\circledast}, \bot, e, \top \rangle \qquad\qquad \overline{\mathcal{K}}^{\circledast} = \langle K, \dot{\oplus}, \dot{\otimes}, \cdot^{\circledast}, \top, e, \bot \rangle \qquad (11)$$

　　*where $\top = \bot^{\circledast}$ and $\bot = \top^{\circledast}$ by definition,*

2.　*In addition to the individual laws as positive semifields, we have the modular laws:*

$$(u \underset{.}{\oplus} v) \underset{.}{\otimes} (u \dot{\oplus} v) = u \underset{.}{\otimes} v \qquad\qquad (u \underset{.}{\oplus} v) \dot{\otimes} (u \dot{\oplus} v) = u \dot{\otimes} v \qquad (12)$$

　　*the analogues of the De Morgan laws:*

$$u \underset{.}{\oplus} v = (u^{\circledast} \dot{\oplus} v^{\circledast})^{\circledast} \qquad\qquad u \dot{\oplus} v = (u^{\circledast} \underset{.}{\oplus} v^{\circledast})^{\circledast} \qquad (13)$$

$$u \underset{.}{\otimes} v = (u^{\circledast} \dot{\otimes} v^{\circledast})^{\circledast} \qquad\qquad u \dot{\otimes} v = (u^{\circledast} \underset{.}{\otimes} v^{\circledast})^{\circledast}$$

　　*and the self-dual inequality in the natural order*

$$u \underset{.}{\otimes} (v \dot{\otimes} w) \preccurlyeq (u \underset{.}{\otimes} v) \dot{\otimes} w. \qquad (14)$$

3.　*Further, if $\mathcal{K}$ is a positive dioid, then the inversion operation is a dual order isomorphism between the dual order structures $\overline{\mathcal{K}} = \langle K, \preccurlyeq \rangle$ and $(\overline{\mathcal{K}})^{\circledast} = \langle K, \succcurlyeq \equiv \preccurlyeq^{\delta} \rangle$ with the natural order of the original semifield a suborder of the first structure.*

**Proof.** For 1, consider $\overline{K} = K \cup \{\top\}$ obtained by adding a top element to $K$. The order $\langle \overline{K}, \preccurlyeq \rangle$ is extended with $x \preccurlyeq \top, \forall x \in K$, hence the notation $\top$ for this element. In this completed set the definition of the operations are (in this paper, cases in a definition-by-case should be interpreted from top to bottom: this first case to match applies):

$$u \underset{.}{\oplus} v = \begin{cases} \top & u = \top \text{ or } v = \top \\ u \oplus v & u, v \in K \setminus \{\top\} \end{cases} \qquad u \underset{.}{\otimes} v = \begin{cases} \bot & u = \bot \text{ or } v = \bot \\ \top & u = \top \text{ or } v = \top \\ u \otimes v & u, v \in K \setminus \{\bot, \top\} \end{cases}$$

with the inversion operation completed by the definition of $\top = \bot^{\circledast}$ and $\bot = \top^{\circledast}$. So $\overline{\mathcal{K}} = \langle K, \underset{.}{\oplus}, \underset{.}{\otimes}, \cdot^{\circledast}, \bot, e, \top \rangle$ is the well-known top-completion of a positive semiring ([4], p. 250).

In this construction, inversion is total, injective and surjective in the completed domain $\overline{K}$, hence a bijection. It is easily seen as an involution $(x^{\circledast})^{\circledast} = x, \forall x \in \overline{K}$, in fact, the inverse for the order-dual $\overline{\mathcal{K}}^{\circledast}$. But, the operations in the inverse semifield are given by:

$$a \dot{\oplus} b = \left( a^{\circledast} \underset{.}{\oplus} b^{\circledast} \right)^{\circledast} \qquad\qquad a \dot{\otimes} b = \left( a^{\circledast} \underset{.}{\otimes} b^{\circledast} \right)^{\circledast}. \qquad (15)$$

To gain an understanding of the operations, we explore their results on a case-by-case basis. For $\dot{\oplus}$:

- If $a = \top$ then $a^{\circledast} \underset{.}{\oplus} b^{\circledast} = \bot \underset{.}{\oplus} b^{\circledast} = b^{\circledast}$, whence $\top \dot{\oplus} b = b$, and symmetrically for $b = \top$. That is, $\top$ is the neutral element of addition $\dot{\oplus}$ and $\langle \overline{K}, \dot{\oplus}, \top \rangle$ a monoid.

- If $a = \bot$, then $\bot^{\circledast} \underset{.}{\oplus} b^{\circledast} = \top \underset{.}{\oplus} b^{\circledast} = \top$, whence $\bot \dot{\oplus} b = \top^{\circledast} = \bot$ and symmetrically for $b = \bot$. This proves that $\bot$ is the maximum element of $\preccurlyeq^{\delta}$, to be defined below.

- Otherwise, for $\{a, b\} \subseteq K \setminus \{\perp\}$, we have $a \mathrel{\dot\oplus} b = \frac{1}{a^\circledast \oplus b^\circledast} = \frac{a \otimes b}{a \oplus b}$,

while for $\dot\otimes$:

- If $a = \top$ then $a^\circledast \mathrel{\dot\otimes} b^\circledast = \perp \otimes b^\circledast = \perp$, whence $\top \mathrel{\dot\otimes} b = \perp^\circledast = \top$, and symmetrically for $b$.

- If $a = \perp$ but $b \neq \top$, then $a^\circledast \mathrel{\dot\otimes} b^\circledast = \top \otimes b^\circledast = \top$, whence $a \mathrel{\dot\otimes} b = \perp$.

- Otherwise, for $\{a, b\} \subseteq K \setminus \{\perp\}$, we have $a \mathrel{\dot\otimes} b = a \otimes b = a \otimes b$.

Note that $\langle \overline{K}, \dot\oplus, \top \rangle$ is a commutative monoid, with commutativity and associativity following from those of $\oplus$. Likewise, $e^\circledast = e$ is easily proven to be the neutral element of $\langle \overline{K}, \dot\otimes, e, \cdot^\circledast \rangle$ which is a commutative group with commutativity and associativity issuing from those of $\otimes$, whose inverse is the involution $\cdot^\circledast$. So to prove that the algebra is a semifield we only need to prove the distributive of $\dot\otimes$ over $\dot\oplus$, for $u, v, z \in \overline{K}$:

$$u \mathrel{\dot\otimes} (v \mathrel{\dot\oplus} z) = \frac{1}{u^\circledast \mathrel{\dot\otimes} \left( \left( v^\circledast \mathrel{\dot\oplus} z^\circledast \right)^\circledast \right)^\circledast} = \frac{1}{u^\circledast \mathrel{\dot\otimes} (v^\circledast \mathrel{\dot\oplus} z^\circledast)} = \frac{1}{u^\circledast \mathrel{\dot\otimes} v^\circledast \mathrel{\dot\oplus} u^\circledast \mathrel{\dot\otimes} z^\circledast}$$

$$= (u^\circledast \mathrel{\dot\otimes} v^\circledast)^\circledast \mathrel{\dot\oplus} (u^\circledast \mathrel{\dot\otimes} z^\circledast)^\circledast = u \mathrel{\dot\otimes} v \mathrel{\dot\oplus} u \mathrel{\dot\otimes} z.$$

Therefore $\overline{\mathcal{K}}^\circledast = \langle K, \dot\oplus, \dot\otimes, \cdot^\circledast, \top, e, \perp \rangle$ is another completed semifield issuing from the first one.

For 2, the proof for the De Morgan-like laws is easy from the definion of $\dot\oplus$ and $\dot\otimes$: those definitions are actually one half of the laws, e.g., $a \mathrel{\dot\otimes} b = \left( a^\circledast \otimes b^\circledast \right)^\circledast$. Inverting and by the involutivity of the inversion $a^\circledast \otimes b^\circledast = (a \mathrel{\dot\otimes} b)^\circledast$, we change $a = u^\circledast$ and $b = v^\circledast$ to prove the result. The proof is analogue for the multiplicative law. The dual equalities (12) and the self-dual inequality (14) are just exercises in case analysis.

For 3 we want to find the order for the dual semifield, $\preccurlyeq^\delta = \succcurlyeq$.

- if $\{u, v\} \subseteq K \setminus \{\perp\}$, since the natural order is compatible with multiplication we multiply by $u^\circledast \otimes v^\circledast$ to obtain $u \otimes (u^\circledast \otimes v^\circledast) \preccurlyeq v \otimes (u^\circledast \otimes v^\circledast)$ whence, by cancellation, $v^\circledast \preccurlyeq u^\circledast$, or else $u^\circledast \preccurlyeq^\delta v^\circledast$, so the order is the dual on inverses.
- We have that $\perp \preccurlyeq v, \forall v \in \overline{K}$, otherwise $v \preccurlyeq^\delta \perp$ which asserts that $\perp = \top^\circledast$ is the "top" of the inverted order. Likewise we read from $u \preccurlyeq \top, \forall u \in \overline{K}$ that $\top \preccurlyeq^\delta u$, that is $\top = \perp^\circledast$ is the "bottom" in $\preccurlyeq^\delta$.

whence $\langle \overline{K}, \preccurlyeq \rangle^\circledast = \langle \overline{K}, \preccurlyeq^\delta \rangle$. $\quad \square$

This proof provides extensive guide on how to use the notation. Note that:

- The dot notation, from [16], is a mnemonic for where do the multiplication of the bottom and top go:

$$\perp \mathrel{\dot\otimes} \top = \perp \qquad\qquad \perp \mathrel{\dot\otimes} \top = \top$$

implying that the "lower" addition and multiplication are aligned with the original order in the semifield, e.g., $\perp \otimes x = \perp$ while the "upper" addition and multiplication are aligned with its dual. All other cases remain as defined for $\oplus$ in the incomplete semifield.

- The case analysis for the operators in the dual semifield allows us to write their definition-by-cases as follows:

$$
a \,\dot{\oplus}\, b =
\begin{cases}
b & a = \top \\
a & b = \top \\
\bot & a = \bot \text{ or } b = \bot \\
\frac{1}{a^{\circledast} \oplus b^{\circledast}} & \{a,b\} \subseteq K \setminus \{\bot\}
\end{cases}
\qquad
a \,\dot{\otimes}\, b =
\begin{cases}
\top & a = \top \text{ or } b = \top \\
\bot & a = \bot \text{ or } b = \bot \\
a \otimes b & \{a,b\} \subseteq K \setminus \{\bot\}
\end{cases}
.
\tag{16}
$$

This is important for calculations, but notice that $\dot{\otimes}$ and $\otimes$ only differ in the corner cases.

- The notation to "speak" about these semirings tries to follow a convention reminiscent of that of boolean algebra, where the inversion is the complement ([17], Chapter 12).

- Note that $\oplus$ and $\dot{\oplus}$ seem to operate on different "polarities" of the underlying set: if one operates on two numbers, the other operates on their inverses while this is not so for the respective multiplications. This proves extremely important to model physical quantities and other concepts with these calculi (see example below).

Regarding the intrinsic usefulness of completed positive semifields that are not fields—apart from the very obvious but degenerate case of $\mathbb{B}$, the booleans—we have the following example used, for instance, in convex analysis and electrical network theory.

**Example 2** (Dual semifields for the non-negative reals). *The previous procedure shows that there are some problems with the notation of Example 1, and this led to the definition of the following signatures for this semifield and its inverse in convex analysis [16]:*

$$
\mathbb{R}_{\geq 0} = \langle [0,\infty], +, \times, \cdot^{-1}, 0, 1, \infty \rangle
\qquad
\mathbb{R}_{\geq 0}^{-1} = \langle [0,\infty], \dot{+}, \dot{\times}, \cdot^{-1}, \infty, 1, 0 \rangle.
\tag{17}
$$

*Both of these algebras are used, for instance, in electrical engineering (EE), the algebra of complete positive reals to carry out the series summation of resistances, and its dual semifield to carry out parallel summation of conductances. With the convention that $\mathbb{R}_{\geq 0}$ semiring models resistances, it is easy to see that the bottom element, $\bot = 0$ models a shortcircuit, that the top element $\top = \infty$ models an open circuit (infinite resistance) and these conventions are swapped in the dually-ordered semifield of conductances. Since EE does not use the dotted notation explained in this paper, the formulas required for the multiplication of the extremes:*

$$
0 \,\dot{\times}\, \infty = 0
\qquad\qquad\qquad
0 \,\dot{\times}\, \infty = \infty,
$$

*are not allowed in circuit analysis. In our opinion, this strongly suggests that what is actually being operated with are the incomplete versions of these semifields, and the many problems that EE students have in learning how to properly deal with these values may stem from this fact. Other uses in Economics are detailed in [17].*

**Example 3** (Multiplicatively and additively idempotent costs and utilities). *Several pairs of such order-dual semirings are known, for instance:*

- *The completed max-times and min-times semifields.*

$$
\overline{\mathbb{R}}_{\max,\times} = \langle [0,\infty], \max, \times, \cdot^{-1}, 0, 1, \infty \rangle
\qquad
\overline{\mathbb{R}}_{\min,\times} = \langle [0,\infty], \min, \dot{\times}, \cdot^{-1}, \infty, 1, 0 \rangle.
\tag{18}
$$

- *The completed max-plus (schedule algebra, polar algebra) and min-plus semifields (tropical algebra).*

$$
\overline{\mathbb{R}}_{\max,+} = \langle [-\infty,\infty], \max, +, -\cdot, -\infty, 0, \infty \rangle
\qquad
\overline{\mathbb{R}}_{\min,+} = \langle [-\infty,\infty], \min, \dot{+}, -\cdot, \infty, 0, -\infty \rangle.
\tag{19}
$$

Note that their additions are all idempotent: a semiring with idempotent addition is simply called an idempotent semiring and it is always positive. These find usage in path-finding algorithms, and some more examples can be found in [2]. The mechanism whereby they are exposed as pairs of dually ordered semifields is explained in Section 2.2.3. For an example of their use in a particular application, see Section 3.2.1.

2.2.3. A Construction for Positive Semifields

There is a non-countable number of semifields obtainable from $\mathbb{R}_{\geq 0}$. Their discovery is probably due to Maslov and collaborators ([18], §1.1.1), but we present here the generalized procedure introduced by Pap and others [11,19,20], but see also Section 3.4.

**Construction 1** (Pap's dioids and semifields). *Let $\mathbb{R}_{\geq 0}$ be the semiring of non-negative reals, and consider a strictly monotone generator function $g$ on an interval $[a, b] \subseteq [-\infty, \infty]$ with endpoints in $[0, \infty]$. Since $g$ is strictly monotone it admits an inverse $g^{-1}$, so set*

1.  *the pseudo-addition, $u \oplus v = g^{-1}(g(u) \dotplus (g(v))$*
2.  *the pseudo-multiplication, $u \otimes v = g^{-1}(g(u) \dottimes (g(v))$*
3.  *neutral element, $e = g^{-1}(1)$*
4.  *inverse, $x^{\circledast} = g^{-1}(\frac{1}{g(x)})$,*

*then,*

1.  *if $g$ is strictly monotone and increasing increasing such that $g(a) = 0$ and $g(b) = \infty$, then a complete positive semifield whose order is aligned with that of $\mathbb{R}_{\geq 0}$ is:*

$$\overline{\mathcal{K}}_g = \langle [a, b], \dot{\oplus}, \dot{\otimes}, \cdot^{\circledast}, \bot = a, e, \top = b \rangle.$$

2.  *order-dually, if $g$ is strictly monotone and decreasing such that $g(a) = \infty$ and $g(b) = 0$, then a complete positive semifield whose order is aligned with that of $\mathbb{R}_{\geq 0}^{-1}$ is*

$$\overline{\mathcal{K}}_g^{\circledast} = \langle [a, b], \dot{\oplus}, \dot{\otimes}, \cdot^{\circledast}, \bot^{\circledast} = b, e, \top^{\circledast} = a \rangle.$$

**Proof.** See [11,19] for the basic dioid $\overline{\mathcal{K}}_g$, and ([2], p. 49) for the inverse operation and the fact that it is a semifield, hence a positive semifield. The description of the operations of $\overline{\mathcal{K}}_g^{\circledast}$ is provided by Theorem 1. □

Note how in the $g$-calculus operations are always named with a "pseudo-" prefix, but it does not agree with Semiring Theory practice, hence we drop it. Also, the effect of the type of motonicity of $g$ is to impose the polarity of the extended operations. Remember that the inversion $\cdot^{\circledast}$ is a dual isomorphism of semifields, so that $\left(\overline{\mathcal{K}}_g^{\circledast}\right)^{\circledast} = \overline{\mathcal{K}}_g$ and $\left(\overline{\mathcal{K}}_g\right)^{\circledast} = \overline{\mathcal{K}}_g^{\circledast}$. The different notation for the underlying inverse $\cdot^{-1}$ and the inverse in Pap's construction $\cdot^{\circledast}$ is introduced so that it can later be instantiated in a number of constructed inverses, as follows.

Our use of Construction 1 is to generate different kind of semifields by providing different generator functions:

**Construction 2** (Multiplicative-product real semifields [20]). *Consider a free parameter $r \in [-\infty, 0) \bigcup (0, \infty]$ and the function $g(x) = x^r$ in $[a, b] = [0, \infty]$ in Construction 1. For the operations we obtain:*

$$u \oplus_r v = \left( u^r \dotplus v^r \right)^{\frac{1}{r}} \qquad u \otimes_r v = \left( u^r \dottimes v^r \right)^{\frac{1}{r}} = u \dottimes v \qquad u^{\circledast} = \left( \frac{1}{u^r} \right)^{\frac{1}{r}} = u^{-1}, \quad (20)$$

*where the basic operations are to be interpreted in $\mathbb{R}_{\geq 0}$. Now,*

- *if $r \in (0, \infty]$ then $g(x) = x^r$ is strictly monotone increasing whence $\perp_r = 0$, $e_r = 1$, and $\top_r = \infty$, and the complete positive semifield generated, order-aligned with $\mathbb{R}_{\geq 0}$, is:*

$$(\mathbb{R}_{\geq 0})_r = \langle [0, \infty], \dot{\oplus}_r, \dot{\times}, \cdot^{-1}, \perp_r = 0, e, \top_r = \infty \rangle. \tag{21}$$

- *if $r \in [-\infty, 0)$ then $g(x) = x^r$ is strictly monotone decreasing whence $\perp_r = \infty$, $e_r = 1$, and $\top_r = 0$, and the complete positive semifield generated, order-aligned with $(\mathbb{R}_{\geq 0})^{-1}$, or dually aligned with $\mathbb{R}_{\geq 0}$, is:*

$$(\mathbb{R}_{\geq 0})_{-r} = (\mathbb{R}_{\geq 0})_r^{-1} = \left( \mathbb{R}_{\geq 0}^{-1} \right)_r = \langle [0, \infty], \dot{\oplus}_r, \dot{\times}, \cdot^{-1}, \perp_r^{-1} = \infty, e, \top_r^{-1} = 0 \rangle. \tag{22}$$

**Proof.** By instantiation of the basic case. See the details in [21]. □

Note that $(\mathbb{R}_{\geq 0})_1 \equiv \mathbb{R}_{\geq 0}$ and $(\mathbb{R}_{\geq 0})_{-1} \equiv \mathbb{R}_{\geq 0}^{-1}$. This suggests the following corollary:

**Corollary 1.** *$(\mathbb{R}_{\geq 0})_r$ and $(\mathbb{R}_{\geq 0})_{-r}$ are inverse, completed positive semifields.*

In particular, consider the cases:

**Corollary 2.** *In the previous Construction 2,*

$$(\mathbb{R}_{\geq 0})_1 = \mathbb{R}_{\geq 0} \qquad\qquad\qquad (\mathbb{R}_{\geq 0})_{-1} = \mathbb{R}_{\geq 0}^{-1} \tag{23}$$

$$\lim_{r \to \infty} (\mathbb{R}_{\geq 0})_r = \overline{\mathbb{R}}_{\max, \times} \qquad\qquad \lim_{r \to -\infty} (\mathbb{R}_{\geq 0})_r^{-1} = \overline{\mathbb{R}}_{\min, \times} \tag{24}$$

**Proof.** The proof of (23) by inspection. For (24) see [20]. □

All these semifields have the same product, and the same "extreme" points, $\{0, 1, \infty\}$. Their only difference lies in the addition. Sometimes, when only the product is important in an application, the addition remains in the background and we are not really sure in which algebra we are working on. Note also that instead of using the abstract notation for the inversion $\cdot^{\circledast}$, since $\mathbb{R}_{\geq 0}$ is the paragon originating all other behaviour, we have decided to use the original notation for the inversion in the (incomplete) semifield.

The case where $r = 0$ deserves to be commented. To start with, note that for $q > 0$, $(\mathbb{R}_{\geq 0})_q$ is aligned with $\mathbb{R}_{\geq 0}$ while $(\mathbb{R}_{\geq 0})_{-q}$ is aligned with $(\mathbb{R}_{\geq 0})^{\delta}$, therefore $\lim_{r \to 0^+} (\mathbb{R}_{\geq 0})_r$ and $\lim_{r \to 0^-} (\mathbb{R}_{\geq 0})_r$ are endowed with opposite orderings. Therefore the following lemma is not a surprise.

**Lemma 3.** *$(\mathbb{R}_{\geq 0})_0$ is not a semifield.*

**Proof.** Recall, that on finite operands $u \oplus_r v = (u^r + v^r)^{1/r}$. Then we may generalize, due to the associativity of the operation to a vector $\vec{v} = \{v_i\}_i$ to $\oplus_{r,i} v_i = \left( \sum_i v_i^r \right)^{1/r}$. Next, consider $\vec{w}, \vec{x} \in (0, \infty)^n$, and $q \in (0, \infty)$. From the properties of the means (see also [12], C. 2), it is easy to prove that

$$\sqrt[-q]{\sum_{i=1}^n w_i x_i^{-q}} \leq \prod_{i=1}^n x_i^{w_i} \leq \sqrt[q]{\sum_{i=1}^n w_i x_i^q}.$$

By considering $w_i = 1$ and introducing the limits as $q \to 0$ we have

$$\bigoplus_{0,i} v_i = \prod_i v_i, \tag{25}$$

whence the additive and the multiplicative structures of $(\mathbb{R}_{\geq 0})_0$ are the same, so the structure is not even a semiring. Furthermore, when $u, v \in [0, \infty]$, then $\lim_{r \to 0^-} u \oplus_r v = u \dot{\times} v$ whereas $\lim_{r \to 0^+} u \oplus_r$

$v = u \times v$. Hence, in particular when, say $u = 0$ and $v = \infty$ we have $\infty = u \,\dot{\otimes}\, v \not< u \otimes v = 0$ and the product is clearly not defined. $\square$

## 3. Results

We are now ready to start presenting a chain of results that leads to our conjecture.

### 3.1. Entropic Semifields

#### 3.1.1. The Basic Entropic Semifield

The effect of Hartley's information function is to induce from the set of positive numbers (restricted to the $[0,1]$ interval) a semifield of the extended reals $[0, \infty]$. To see this, we actually consider it acting on the whole of the non-negative reals $\mathbb{R}_{\geq 0}$ of (17) onto the algebra of entropies denoted by $\mathbb{H}$.

**Theorem 2** (Hartley's semifields). *The algebra $\langle [-\infty, \infty], \oplus, \otimes, \cdot^{\circledast}, \infty, 0 \rangle$ with*

$$h_1 \oplus h_2 = h_1 + h_2 - \ln(e^{h_1} + e^{h_2}) \qquad h_1 \otimes h_2 = h_1 + h_2 \qquad h^{\circledast} = -h, \qquad (26)$$

*obtained from that of positive numbers by Hartley's information function is a positive semifield that can be completed in two different ways to two mutually dual semifields:*

$$\mathbb{H} = \langle [-\infty, \infty], \dot{\oplus}, \dot{\otimes}, -\cdot, \bot = -\infty, e = 0, \top = \infty \rangle - \mathbb{H} = \langle [-\infty, \infty], \dot{\oplus}, \dot{\otimes}, -\cdot, -\bot = \infty, e = 0, -\top = -\infty \rangle \qquad (27)$$

*whose elements can be considered as entropic values and operated accordingly.*

**Proof.** Recall the extension of Hartley's information function to non-negative numbers in (4), with logarithm base $b = e$. Since $(\mathfrak{I}_*)^{-1}(h) = e^{-h}$ is monotone, it is a generator (function) for Construction 1, with the following addition, multiplication and inversion:

$$\begin{aligned} h_1 \oplus h_2 &= \mathfrak{I}_* \left( (\mathfrak{I}_*)^{-1}(h_1) + (\mathfrak{I}_*)^{-1}(h_2) \right) = -\ln\left( e^{-h_1} + e^{-h_2} \right) = \ln \frac{e^{h_1+h_2}}{e^{h_1}+e^{h_2}} = h_1 + h_2 - \ln\left( e^{h_1} + e^{h_2} \right) \\ h_1 \otimes h_2 &= \mathfrak{I}_* \left( (\mathfrak{I}_*)^{-1}(h_1) \times (\mathfrak{I}_*)^{-1}(h_2) \right) = -\ln\left( e^{-h_1} \cdot e^{-h_2} \right) = -\ln\left( e^{-(h_1+h_2)} \right) = h_1 + h_2 \\ h^{\circledast} &= \mathfrak{I}_* \left( \frac{1}{(\mathfrak{I}_*)^{-1}(h)} \right) = -\ln \frac{1}{e^{-h}} = -h \,. \end{aligned}$$

The "interesting" points $\{0, e, \infty\}$ are transformed as:

$$\mathfrak{I}_*(0) = \infty \qquad \mathfrak{I}_*(1) = 0 \qquad \mathfrak{I}_*(\infty) = -\infty \qquad (28)$$

The rest follows by Construction 1 and Theorem 1. $\square$

Several considerations are worth stating. First, we have not restricted this dual-order isomorphism to the sub-semiring of probabilities on purpose. Despite this, on $\mathfrak{I}_*([0,1]) = [0, \infty] \subseteq [-\infty, \infty]$ our intuitions about amounts of informations hold; we still believe that the "information" in $\mathfrak{I}^*(0) = \infty$ is the highest, whereas the probability of $(\mathfrak{I}^*)^{-1}(\infty) = 0$ is the smallest.

Second, the notation for the inverse of Hartley's semifield $-\mathbb{H}$ is a mnemonic to remind that this is a semifield in which the inversion is actually an additive inverse, and consequently the product is an addition.

Third, the order of $\mathbb{R}_{\geq 0}$ is aligned with that of $\mathbb{H}$, actually a sub-order of it in the abstract algebra sense: same properties on a subset of the original carrier set. But the order of $-\mathbb{H}$ is aligned with the dual order, that of $\mathbb{R}_{\geq 0}^{-1}$. This has to be taken into consideration in the application of Theorem 1 to the proof of Theorem 2. Corollary 3 makes this difference clear.

**Corollary 3.** *Hartley's information function is a dual-order isomorphism of completed positive semifields.*

**Proof.** In the proof of the previous theorem, note that $\mathfrak{I}_*(\cdot)$ is monotonically decreasing, entailing that the construction inverts orders in semifields, e.g., $\mathfrak{I}_*(\mathbb{R}_{\geq 0}) = -\mathbb{H}$ and $\mathfrak{I}_*\left(\mathbb{R}_{\geq 0}^{-1}\right) = \mathbb{H}$. □

Figure 3 provides a graphic depiction of the relationship of the domains described by Theorem 2 and Corollary 3, as well as the entropy domains described by Theorem 3 below. It also provides a more comprehensive picture than Figure 1b).
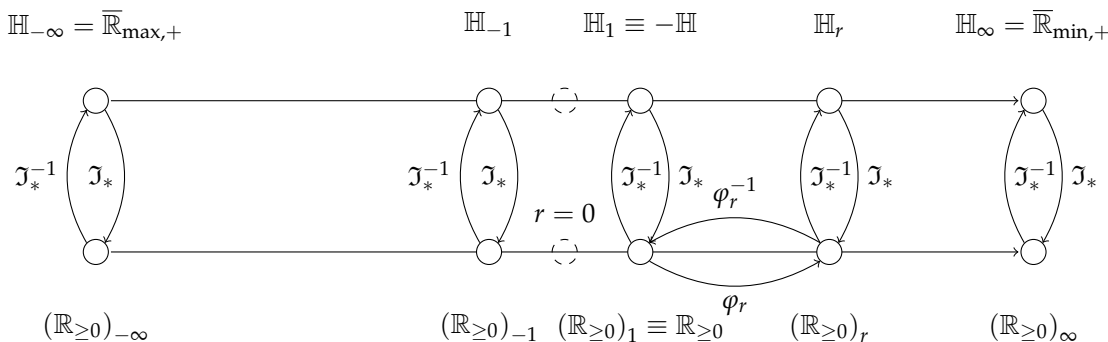


**Figure 3.** Domain diagram to interpret the Rényi transformation in the context of the expectations of probabilities and informations. The horizontal axes represent the evolution of parameter $r$. Notice that for $r = 0$ no positive semifields are found (dashed circles). Above lie the entropic semirings, below, the semirings that are deformations of $\mathbb{R}_{\geq 0}$. Between each of these pairs with the same $r$—where only a few pairs for $r \in \{-\infty, -1, 1, r, \infty\}$ have been made explicit—mediate Hartley's information function $\mathfrak{I}_*(\cdot)$ and its inverse $(\mathfrak{I}_*)^{-1}(\cdot)$. In the semifields below, only the case for the deformation between $\mathbb{R}_{\geq 0}$ and $(\mathbb{R}_{\geq 0})_r$ is shown in full.

### 3.1.2. Constructed Entropic Semifields

The next important fact is that Rényi's modified averaging function $\varphi(p) = \frac{-1}{r}\log_b p$ and its inverse are also dual isomorphisms of positive semifields. To prove this we make said functions appear in the construction of the semifields as generators.

**Theorem 3** (Additive-product real semifields or Entropy semifields). *Let $r \in [-\infty, \infty] \setminus \{0\}$ and $b \in (1, \infty)$. Then the algebra $\langle [-\infty, \infty], \oplus_r, \otimes_r, \cdot^{-1}, \bot = \infty, e = 0 \rangle$ whose basic operations are:*

$$u \oplus_r v = u + v - \log_b \left(b^{ru} + b^{rv}\right)^{\frac{1}{r}} \qquad u \otimes_r v = u + v \qquad u^{\circledR} = -u \tag{29}$$

*can be completed to two dually-ordered positive semifields*

$$\mathbb{H}_r = \langle [-\infty, \infty], \underset{\cdot\, r}{\oplus}, \underset{\cdot\, r}{\otimes}, -\cdot, \bot = -\infty, e = 0, \top = \infty \rangle \tag{30}$$

$$-\mathbb{H}_r = \langle [-\infty, \infty], \dot{\oplus}_r, \dot{\otimes}_r, -\cdot, -\bot = \infty, e = 0, -\top = -\infty \rangle \tag{31}$$

*whose elements can be considered as emphasized, entropic values and operated accordingly.*

**Proof.** We build these semifields with a composition of results. The first one is the well known result from the theory of functional means that we choose to cast into the framework of Pap's $g$-calculus: the power mean of order $r$ is the pseudo arithmetic-mean with generator $\varphi_r(x) = x^r$ and inverse $\varphi_r^{-1}(y) = y^{1/r}$. This was used in Construction 2 to build the semifields of (21) and (22). The second result is Corollary 3, where $\mathfrak{I}_*(\cdot)$ is proven a dual order isomorphism of semirings.

We next use the composition of functions $\varphi_r' = \mathfrak{I}_* \circ \varphi_r$ and its inverse $(\varphi_r')^{-1} = \varphi_r^{-1} \circ \mathfrak{I}_*^{-1}$. The latter exists, since it is a composition of isomorphisms and it is a dual order isomorphism, since

$\mathfrak{I}_* (\cdot)$ is order-inverting while $\varphi_r$ is not. That composition is precisely Rényi's function $\varphi'_r(h) = b^{-rh}$ with inverse $(\varphi'_r)^{-1}(p) = \frac{-1}{r} \log_b p$, whence:

$$
\begin{aligned}
u \oplus_r v &= (\varphi'_r)^{-1}(\varphi'_r(u) + \varphi'_r(v)) = \frac{-1}{r} \log_b (b^{-ru} + b^{-rv}) = \frac{-1}{r} \log_b \left( \frac{b^{ru} + b^{rv}}{b^{r(u+v)}} \right) = u + v - \log_b (b^{ru} + b^{rv})^{\frac{1}{r}} \\
u \otimes_r v &= (\varphi'_r)^{-1}(\varphi'_r(u) \times \varphi'_r(v)) = \frac{-1}{r} \log_b (b^{-ru} \times b^{-rv}) = \frac{-1}{r} \log_b \left( b^{-r(u+v)} \right) = u + v \\
u^\circledast &= (\varphi'_r)^{-1} \left( \frac{1}{\varphi'_r(u)} \right) = \frac{-1}{r} \log_b \left( \frac{1}{b^{-ru}} \right) = -u
\end{aligned}
$$

This composition is strictly increasing when $r \in [-\infty, 0)$ and strictly decreasing when $r \in (0, \infty]$, hence, when applying Construction 1 with it:

- for $r \in (0, \infty]$ we get $-\mathbb{H}_r = \mathfrak{I}_* ((\mathbb{R}_{\geq 0})_r) = \langle [-\infty, \infty], \dot{\oplus}_r, \dot{\otimes}_r, -\cdot, -\bot = \infty, e = 0, -\top = -\infty \rangle$, and

- for $r \in [-\infty, 0)$ we obtain $\mathbb{H}_r = \mathfrak{I}_* \left( (\mathbb{R}_{\geq 0})_r^{-1} \right) = \langle [-\infty, \infty], \underset{\cdot}{\oplus}_r, \underset{\cdot}{\otimes}_r, -\cdot, \bot = -\infty, e = 0, \top = \infty \rangle$,

with the extended operations:

$$-\bot = -(-\infty) = \infty = \top \qquad\qquad -\top = -\infty = \bot$$

$$
u \underset{\cdot\, r}{\oplus} v = \begin{cases} \infty & \text{if } u = \infty \text{ or } v = \infty \\ u \oplus_r v & \text{otherwise} \end{cases} \qquad\qquad u \underset{\cdot\, r}{\otimes} v = \begin{cases} -\infty & \text{if } u = -\infty \text{ or } v = -\infty \\ u + v & \text{otherwise} \end{cases}
$$

$$
u \dot{\oplus}_r v = \begin{cases} -\infty & \text{if } u = -\infty \text{ or } v = -\infty \\ u \oplus_r v & \text{otherwise} \end{cases} \qquad\qquad u \dot{\otimes}_r v = \begin{cases} \infty & \text{if } u = \infty \text{ or } v = \infty \\ u + v & \text{otherwise} \end{cases}
$$

$\square$

Note that in these semirings we have: $-\infty \underset{\cdot\, r}{\otimes} \infty = -\infty$ and $-\infty \dot{\otimes}_r \infty = \infty$ for $r \in [-\infty, \infty] \setminus \{0\}$.

Moreau's original proposals [16] are found for $r = 1$ $-\infty \underset{\cdot}{\times} \infty = -\infty$ and $-\infty \dot{\times} \infty = \infty$. Further details for compositions of generating functions and other averaging constructions can be found in [22].

Similar to Corollary 2 we have the following:

**Proposition 1.** *In the semifield construction of Theorem 3*

$$\mathbb{H}_{-1} = \mathbb{H} \qquad\qquad\qquad \mathbb{H}_1 = -\mathbb{H} \qquad\qquad (32)$$

$$\lim_{r \to -\infty} \mathbb{H}_r = \overline{\overline{\mathbb{R}}}_{\max,+} \qquad\qquad \lim_{r \to \infty} \mathbb{H}_r = \overline{\overline{\mathbb{R}}}_{\min,+} \qquad\qquad (33)$$

**Proof.** The proof of (32) is done by inspection. For (33), well-known, see [2,20]. $\square$

The apparent incongruity of $\mathbb{H}_1 = -\mathbb{H}$ stems from the different origins of each notation: $\mathbb{H}_1$ comes from the theory of semirings and is due to the fact that we have forced it to be aligned with $-\mathbb{H}$ which is ultimately motivated by Shannon's choice of sign for the entropy function [23]. Note that in thermodynamics, this choice of sign is the inverse, that is, negentropy is the privileged concept.

The following corollary is the important result we announced at the beginning of this section.

**Corollary 4.** *The Rényi entropies in the Rényi spectrum $\tilde{H}_r (P_X)$ take values in the semifields $\mathbb{H}_r$.*

**Proof.** We notice that the generating function to use Construction 1 in Theorem 3 is none other than the function to calculate the Rényi non-linear average $\varphi'$ of (3). Hence the values resulting from Rényi's entropies belong in that semifield of Theorem 3 with the respective $r$ parameter. $\square$

Therefore Theorem 3 and Corollary 4 state that we have to use semifield algebra when carrying out calculations with entropies, otherwise said, they jointly confer meaning to the statement entropies operate on positive semifields in the title of this paper. An example follows in Section 3.2.1.

*3.2. Applications*

We would like to clarify further how entropies operate in positive semifields, and, in general, the importance of positive semifields for entropy. For this purpose we present two different applications of the theory of entropic semifields: one shows how to write entropy-processing algorithms in linear (semifield) algebra and the second shows why we cannot avoid positive semifields in entropy definitions, if we want to consider extreme values.

3.2.1. Rewriting the Viterbi Algorithm in Semifields

In this application note we analyze rewriting the Viterbi algorithm [24] in semifield notation, following recent work presented in [25].

Introduction Initially devised as a teaching aid for convolutional codes [24], the "Viterbi" was soon proven to be an optimal algorithm for shortest-path decoding in a specific type of network [26]. When this network comes from the unfolding over time of a Markov chain, it can also be used to recover an "optimal" sequence of states and transition over a generative model for a given sequence of observations [27]. In this natural generalization, it has been applied to text and speech recognition and synthesis, among other cognitively-relevant applications that used to be considered part of "classical" AI but are modernly better tackled with more specialized ML or CI techniques.

Development. Consider a weighted finite automaton of $n$ states over a finite alphabet $\Sigma$, $(\vec{p}, W, \vec{b})$ defined by

- A starting distribution $\vec{p}$, where $\vec{p}_i, 1 \leq i \leq n$, is the probability of starting at state $i$.
- A transition matrix $W$, where $w_{ij} = (W)_{ij}, 1 \leq i, j \leq n$ is the probability of a transition from state $i$ to state $j$, and
- An emission distribution $\vec{b}$, where $\vec{b}(\alpha)_i$ is the probability of emitting symbol $\alpha \in \Sigma$ from state $i, 1 \leq i \leq n$.

It is also possible to assign non-null probabilities for definite accepting states, but these can be assimilated to the more general framework used here.

Let $\sigma \in \Sigma^+$ be a non-empty string of length $T$ from the free monoid built over $\Sigma$ that we call the sequence of observations and call $\sigma_t$ the $t$-th observation. Supposing the sequence of observations $[\sigma_t], 1 \leq t \leq T$ to have been generated by the automaton, the purpose of the Viterbi algorithm is to find the best path score $\vec{q}(T)$ over the set of possible generation paths.

This may be done using Bellman's Principle of Optimality [28] leading to the following recursion for $0 \leq t \leq T$

$$\vec{q}_i(0) = \vec{p}_i \tag{34}$$
$$\vec{q}_i(t) = \max_j \left( w_{ij} \cdot \vec{q}_j(t-1) \right) \cdot \vec{b}_i(\sigma_t)$$

It is easy to see that this equation can be written in $(\mathbb{R}_{\geq 0})_\infty \equiv \overline{\mathbb{R}}_{\max,\times}$ as introduced in (18). First, for any vector $\vec{p}$, let $\Lambda(\vec{p})$ be a square matrix whose main diagonal is built with the components of $\vec{p}$, and whose off-diagonal entries are zero. Then we may write (34) in matrix form as

$$\vec{q}(t) = \Lambda(\vec{b}(\sigma_t)) \underset{\cdot}{\times} W^{\mathrm{T}} \underset{\cdot}{\times} \vec{q}(t-1) \tag{35}$$

where we have co-opted the scalar lower multiplication to represent also matrix multiplication, as customary in the theory of linear algebra over semirings [4].

To actually solve (35), Hartley's function is customarily invoked on all the probabilities appearing in (34) (cfr. [25])—often citing the advantages of working on log-probabilities for numeric computing—to obtain

$$-\log \vec{q}_i(t) = \min_j \left( -\log w_{ij} \dot{+} (-\log \vec{q}_j(t-1)) \right) \dot{+} (-\log \vec{b}_i(\sigma_t)). \tag{36}$$

Two remarks are pertinent here. First, a realization already put forward by many researchers [2]: this expression describes a non-linear computation—a minimum score in this case—by means of a linear equation in $\overline{\mathbb{R}}_{\min,+}$ (19). Second, one of the main ideas put forward in this paper: due to the fact that the original equations dealt with probabilities and the nature of Hartley's function, (37) is a linear equation on $\mathbb{H}_\infty$ entropies.

Obviously, $\mathbb{H}_\infty \equiv \overline{\mathbb{R}}_{\min,+}$, hence, call $\vec{x}(t) = -\log \vec{q}(t)$, $H_b(\sigma) = -\log \Lambda(b(\sigma))$ and $H_W = -\log W$, so that (36) can be rewritten in matrix form as

$$\vec{x}(t) = H_b(\sigma_t) \dot{\otimes} H_W{}^{\mathsf{T}} \dot{\otimes} \vec{x}(t-1). \tag{37}$$

It is also illustrative to see the inclusion of a pruning threshold $\eta$ in this computation by $\overline{\mathbb{R}}_{\min,+}$-linear means [25]. By definition $\vec{x}(t)^{\mathsf{T}} \dot{\otimes} \vec{x}(t) = 2 \cdot \min_i x_i(t)$ so consider the vector $\vec{\eta}_i = \eta + \frac{1}{2}\vec{x}(t)^{\mathsf{T}} \dot{\otimes} \vec{x}(t), 1 \le i \le n$ and $X(t) = \Lambda(\vec{x}(t))$. The problem of thresholding the path computations can be stated as finding $\vec{y}$ such that $X(t) \dot{\otimes} \vec{y} \ge \vec{\eta}$. The solution to this set of equations is well-known (see e.g., [1], Theorem 3.2.3 of on p. 58): for generic and conformant matrix $A$ and vector $\vec{b}$ we have

$$A \dot{\otimes} \vec{y} \ge \vec{b} \iff \vec{y} \ge A^{\circledast} \dot{\otimes} \vec{b}.$$

Note that for a matrix $A$ with $\overline{\mathbb{R}}_{\min,+}$ entries, the Cuninghame-Green conjugate $A^{\circledast} = -A^{\mathsf{T}}$ is the analogue of the complex conjugate [29], and that the maximal solution takes an analytical form in the dual semifield, in this case $\mathbb{H}_{-\infty} \equiv \overline{\mathbb{R}}_{\max,+}$.

Then the negative elements of $\vec{y} = X(t)^{\circledast} \otimes \vec{\eta}$ indicate which components need to be pruned. That is, if the individual equations read $\vec{y}_i = -x_i(t) + \eta + \min_j x_j(t)$, setting in $x_j(t) = -\infty$ at step $t$ for those indices of $\vec{y}$ which are negative actually carries out the pruning.

Discussion What we want to highlight with this application are several facts:

- Actual non-linear computations in the Viterbi algorithm take the form of linear operations over a particular semifield, used here to minimize costs in $\overline{\mathbb{R}}_{\min,+}$. This takes the form of a linear matrix equation, an instance of "linear processing" in a non-linear algebra.
- Secondly and more importantly, because of Theorem 3 we know that the values in which the log-probabilities are being operated in the Viterbi are actually the Rényi entropies with index $r = -\infty$. Hence we conclude that the Viterbi is a quantitative, information entropy-processing algorithm.
- Even such a non-linear process as pruning using a threshold can be characterized and carried out by linear processing in the semifield. This is an instance of the fact that linear operations in $\mathbb{H}_\infty$ process information in ("standard" algebraic) non-linear ways.

Already the authors of [25] provide a geometric construction to characterize the search space associated to the Viterbi as described by $\overline{\mathbb{R}}_{\min,+}$ equations in terms of half-spaces and their tropical polytopes. Our own work has proven that the spectral spaces for matrices like $H_b(\sigma_t)$ and $H_W$ are actually complete lattices instead of the better known orthogonal subspaces of the Singular Value Decomposition of matrices in "standard" algebra [15,30]. In fact, the left and right spectral spaces issued form the $\overline{\mathbb{R}}_{\min,+}$ Singular Value Decomposition show a relation akin to that of extents and intents in Formal Concept Analysis [31]. Further exploration of these issues is left for future work.

Finally, a crucial issue is to realize that the "optimality" of the decoding strategy is prescribed by the algebra being used in the decoding—in the language of this paper, Viterbi is $\overline{\mathbb{R}}_{\min,+}$- or $\mathbb{H}_\infty$-optimal—and this is often based in a-priori considerations on the problem. It follows that several other algorithms can be built with the template of the Viterbi by changing the underlying semifield, but all require that the addition be idempotent, that is $\forall a \in S, a \oplus a = a$ ([2], Chapters 1 and 8).

### 3.2.2. Rewriting the Hölder Means and Rényi Entropies with Semifields

Corollary 4 and Definition 1 suggest that the mean $M_r(\vec{w}, \vec{x})$ in (2) actually takes values in the positive semifield $(\mathbb{R}_{\geq 0})_r$. The surprisal function would then transform these values into the entropy semifield $\mathbb{H}_r$. However, the expression in Definition 1 does not easily lend itself to accept the extreme values $\infty$ and 0. Therefore, we propose the following definition for the means in positive semifields:

**Definition 3.** *Let $\vec{x}, \vec{w} \in [0, \infty]^n$ and $r \in [-\infty, \infty]$. Then in the dual completed semifields issuing from $\mathbb{R}_{\geq 0}$,*

$$
\hat{M}_r(\vec{w}, \vec{x}) = \begin{cases} \left( \sum_i^\bullet \left( w_i \mathbin{\dot{/}} \sum_k^\bullet w_k \right) \dot{\otimes} x_i^r \right)^{\frac{1}{r}} & r < 0 \\ e^{\hat{M}_1(\vec{w}, \ln \vec{x})} & r = 0 \\ \left( \sum_{\bullet i} \left( w_i \mathbin{/_\bullet} \sum_{\bullet k} w_k \right) \otimes_\bullet x_i^r \right)^{\frac{1}{r}} & r > 0 \end{cases}
\tag{38}
$$

*where the case for $r = 0$ is only valid when there are no $i \neq j$ with $x_i^{w_i} = 0$ and $x_j^{w_j} = \infty$.*

**Justification.** To justify this definition, we need to translate the painstakingly developed description of the means in ([12], Chapter II) into complete positive semifield notation to obtain the rewritten means $\hat{M}_r(\vec{w}, \vec{x})$. For that purpose, we will carry out a case-based analysis based in the following set definitions: let $n$ be the dimension of $\vec{x}$ and $\vec{w}$ and $\bar{n} = \{1, \ldots, n\}$ the full set of indices on all the components, and let:

$$
B_{\vec{x}} = \{i \in \bar{n} \mid x_i = 0\} \qquad T_{\vec{x}} = \{i \in \bar{n} \mid x_i = \infty\} \qquad F_{\vec{x}} = \{i \in \bar{n} \mid x_i \in (0, \infty)\}
\tag{39}
$$

be, respectively, the set of indices of zero, finite, and infinite components, and notice that $\bar{n} = B_{\vec{x}} \cup F_{\vec{x}} \cup T_{\vec{x}}$.

Note that the corner cases when $\vec{x} = 0$ are catered to by the reflexivity of the means [12].

In particular, when $r > 0$, if $\vec{x} = 0$, then $B_{\vec{x}} = \bar{n}$ and we require that $\hat{M}_r(\vec{w}, \vec{x}) = 0$. This entails that the multiplication in (38) for $r > 0$ must be the lower multiplication so that

$$
\hat{M}_r(\vec{w}, \vec{x}) = \left( \sum_{\bullet i} \tilde{q}_0(\vec{w}) \otimes_\bullet 0 \right)^{\frac{1}{r}} = 0.
$$

In [12], Chapter II, the existence of $w_i = 0$ is directly disallowed, and, we take the same is the case for $w_i = \infty$. However, the authors later take into consideration the second possibility, so we have decided to include both from the start. For this purpose, consider the case where $B_{\vec{w}} = \bar{n}$. Let us first define $W = \sum_{\bullet i} w_i$, which is the only possible definition consistent with our intuitions of an overall weight $W$ for $r > 0$, and notice that $B_{\vec{w}} = \bar{n} \iff W = 0 \iff W^\circledast = \infty$. This means we could, in principle, define the normalized weight distribution $\tilde{q}_0(\vec{w})$ in two possible ways:

1.  either $\tilde{q}_0(\vec{w}) = \{w_i \mathbin{/_\bullet} W\} = \{w_i \dot{\otimes} W^\circledast\}$;

2.  or $\tilde{q}_0(\vec{w}) = \{w_i \mathbin{\dot{/}} W\} = \{w_i \otimes_\bullet W^\circledast\}$.

In the first case, if all the weights are null, that is $B_{\vec{W}} = \bar{n}$ then $\forall i, w_i \dot{\otimes} W^\circledast = 0 \dot{\otimes} \infty = \infty$. For zero coordinates of $\vec{x}$ this poses no problem, but for $j \in \overline{B_{\vec{x}}} \neq \varnothing$ we find that $(w_j \dot{\otimes} W^\circledast) \otimes x_j^r = \infty \otimes x_i^r = \infty$ whence the whole summation is infinite. This does not seem reasonable. On the other hand, in the second case, we have that $(w_i \otimes W^\circledast) \otimes x_i^r = (0 \otimes \infty) \otimes x_i^r = 0 \otimes x_i^r = 0$, as expected.

However, this has too much "annihilating power" for if $\dot{T}_{\vec{w}} \neq \varnothing \iff W = \infty \iff W^\circledast = 0$, considering an $i \in T_{\vec{w}}$ such that $w_i = \infty$, then every factor is erased since for $j \in \overline{B_{\vec{w}}}$ is $(w_j \otimes W^\circledast) \otimes x_j^r = (w_j \otimes 0) \otimes x_j^r = 0$, and the alternative is even less intuitive than the preceding one. Therefore we are forced to choose alternative 1 $\tilde{q}_0(\vec{w}) = \{w_i \,/\, W\}$ with the caveat that when all of the weights are null (a strange situation indeed) this does not make sense.

Two points are worth stating:

- The case where $r < 0$ is reasoned out by duality, with $T_{\vec{x}}$ being dual to $B_{\vec{x}}$, $r < 0$ being the dual to $r > 0$, $\infty$ dual to 0, and the upper multiplication and addition duals to the lower multiplication and addition. In the following we just use this "duality" argument to solve the case for $r < 0$.
- The rest of the cases to analyze essentially have $B_{\vec{x}} \neq \bar{n}$ and $B_{\vec{w}} \neq \bar{n}$ whence their complements are non-null $\overline{B_{\vec{x}}} \neq \varnothing, \overline{B_{\vec{w}}} \neq \varnothing$. This means that the actual summation in (38) is extended to $\overline{B_{\vec{x}}} \cap \overline{B_{\vec{w}}}$, with $\overline{B_{\vec{x}}} = F_{\vec{x}} \cup T_{\vec{x}}$ and $\overline{B_{\vec{w}}} = F_{\vec{w}} \cup T_{\vec{w}}$.

Due to Theorem 1 and Construction 1 we know that if $F_{\vec{x}} = \overline{B_{\vec{x}}}$ and $F_{\vec{w}} = \overline{B_{\vec{w}}}$—that is, all non-zero coordinates are finite—then the expressions reduce to those of the classical definition in (2) so that $\hat{M}_r(\vec{w}, \vec{x}) = M_r(\vec{w}, \vec{x})$ for $r \in [-\infty, \infty]$, that is including $r \leq 0$.

The only cases left to analyze for $r > 0$ are those where $\infty$ appears in either the weights $\vec{w}$ or the quantities $\vec{x}$ on subindices $i \in \overline{B_{\vec{x}}} \cap \overline{B_{\vec{w}}}$. Notice that if $i \in T_{\vec{x}}$ then $(w_i \dot{\otimes} W^\circledast) \otimes \infty = \infty$ since $w_i \dot{\otimes} W^\circledast$ cannot be zero, therefore $\hat{M}_r(\vec{w}, \vec{x}) = \infty$. Likewise, if $i \in T_{\vec{w}}$ then $W = \infty \iff W^\circledast = 0$ and $w_i \dot{\otimes} W^\circledast = \infty \dot{\otimes} 0 = \infty$ whence the factor $(w_i \dot{\otimes} W^\circledast) \otimes x_i^r = \infty \otimes x_i^r = \infty$ and $\hat{M}_r(\vec{w}, \vec{x}) = \infty$. We have collected all cases for $r > 0$, in Figure 4, along with those for $r < 0$ obtained by duality.

| $\hat{M}_{r<0}(\vec{w}, \vec{x})$ | $T_{\vec{x}} = \bar{n}$ | $\overline{T_{\vec{x}}} \neq \varnothing$ | |
|---|---|---|---|
| | | $F_{\vec{x}} = \bar{n}$ | $B_{\vec{x}} \neq \varnothing$ |
| $T_{\vec{w}} = \bar{n}$ | $\infty$ | 0 | 0 |
| $\overline{T_{\vec{w}}} \neq \varnothing$   $F_{\vec{w}} = \bar{n}$ | $\infty$ | $M_r(\vec{w}, \vec{x})$ | 0 |
| $\overline{T_{\vec{w}}} \neq \varnothing$   $B_{\vec{w}} \neq \varnothing$ | $\infty$ | 0 | 0 |

| $\hat{M}_{r>0}(\vec{w}, \vec{x})$ | $B_{\vec{x}} = \bar{n}$ | $\overline{B_{\vec{x}}} \neq \varnothing$ | |
|---|---|---|---|
| | | $F_{\vec{x}} = \bar{n}$ | $T_{\vec{x}} \neq \varnothing$ |
| $B_{\vec{w}} = \bar{n}$ | 0 | $\infty$ | $\infty$ |
| $\overline{B_{\vec{w}}} \neq \varnothing$   $F_{\vec{w}} = \bar{n}$ | 0 | $M_r(\vec{w}, \vec{x})$ | $\infty$ |
| $\overline{B_{\vec{w}}} \neq \varnothing$   $T_{\vec{w}} \neq \varnothing$ | 0 | $\infty$ | $\infty$ |

(**a**) For $r < 0$            (**b**) For $r > 0$

**Figure 4.** A summary of cases for $\hat{M}_r(\vec{w}, \vec{x})$ (**a**) for $r < 0$ and (**b**) for $r > 0$

Only the case $r = 0$ is left for analysis, and recall this is $M_0(\vec{w}, \vec{x}) = \left( \Pi_i x_i^{w_i} \right)^{\frac{1}{\sum_k w_k}}$. In [12] this case is treated exceptionally throughout the treatise, and one of the often used alternative expressions for it is $M_0(\vec{w}, \vec{x}) = \exp\{\sum_i \frac{w_i}{W} \ln x_i\}$ leading immediately to $M_0(\vec{w}, \vec{x}) = \exp\{M_1(\vec{w}, \ln \vec{x})\}$ where the logaritm has to be interpreted entry-wise.

Recalling that $\log(\cdot)$ and $\exp(\cdot)$ are order isomorphisms of semifields, this suggests that we are trying to solve the hard problem for $r = 0$ in the semifield $\mathbb{H}$ by using $r = 1$ thanks to the properties of the logarithm. However, the addition in $\mathbb{H}$ is a multiplication so the rules for calculating the weighted means there might be different, considering that $\log x_i$ might be negative, or, specially, that factors like $w_i \cdot \log x_i$ are actually $\mathbb{H}$ exponentiations, i.e., they come from factors $\log(x_i^{w_i})$.

In fact, the clue to suggest the form for this mean comes from information theory and the requirement that $0 \cdot \log \frac{1}{0} = 0$, necessary to ignore impossible events with $p_i = 0$ with $-p_i \cdot \log p_i = 0$ in

the entropy calculation. This can be modeled in our framework by demanding that this multiplication be $0 \otimes \log(1 / 0) = 0 \otimes \infty = 0$ and to write the geometric mean as:

$$\hat{M}_0(\vec{w}, \vec{x}) = \exp\{\hat{M}_1(\vec{w}, \ln \vec{x})\} \tag{40}$$

Note that since the logarithm changes the base semifield for $\hat{M}_1(\vec{w}, \log \vec{x})$ the Figure 4b has to be reinterpreted, along with the sets of subindices: if the components of a vector $\vec{x}$ belong to a semifield $\mathcal{K}$ with carrier set $K$, then the index sets have to be defined with respect to the semifield in it. Since the carrier set of $\mathbb{H}$ is $\mathbb{R}_{\pm\infty} = [-\infty, \infty]$ then they read as:

$$B_{\vec{x}} = \{i \in \bar{n} \mid x_i = \perp_{\mathbb{R}_{\pm\infty}} = -\infty\} \quad T_{\vec{x}} = \{i \in \bar{n} \mid x_i = \top_{\mathbb{R}_{\pm\infty}} = \infty\} \quad F_{\vec{x}} = \{i \in \bar{n} \mid x_i \in (\perp_{\mathbb{R}_{\pm\infty}}, \top_{\mathbb{R}_{\pm\infty}})\}$$

So a table similar to Figure 4b for $\hat{M}_1(\vec{w}, \log \vec{x})$ would have $-\infty$ instead of 0. Nevertheless, the exponentiation would bring the mean back to $\mathbb{R}_{\geq 0}$, and entailing that (40) actually follows Figure 4b. □

This concludes our justification of the casting of the weighted means into semifield algebra. Note that a single $\vec{x}$ may have $x_i = 0$ and $x_j = \infty$ for $i \neq j$, a case explicitly addressed in [12]: $\hat{M}_r(\vec{p}, \vec{x}) = \infty \otimes u(r)$ where $u(r)$ is the step function, undefined at $r = 0$

$$u(r) = \begin{cases} 0 & r < 0 \\ 1 & r > 0 \end{cases}.$$

With this formulation we are now capable of describing the Rényi entropies in semifield notation:

**Definition 4.** *Let $P_X(x_i) = p_i$ and $Q_X(y_i) = q_i$ be two distributions with compatible support. Then in the dual completed semifields issuing from $\mathbb{R}_{\geq 0}$, the expression of the Rényi cross-entropy, entropy and divergence are:*

$$\tilde{H}_r(P_X) = \mathfrak{I}_* \left( \hat{M}_r(P_X, P_X) \right) \quad \tilde{X}_r(P_X \| Q_X) = \mathfrak{I}_* \left( \hat{M}_r(P_X, Q_X) \right) \quad \tilde{D}_r(P_X \| Q_X) = -\mathfrak{I}_* \left( \hat{M}_r(P_X, \tfrac{P_X}{Q_X}) \right) \tag{41}$$

Notice that when $F_{P_X} = \bar{n}$ and $F_{Q_X} = \bar{n}$ the definition of entropy based on the mean and in the semifield expressed are the same, whereas in the other cases there is an extension in the definition which is in agreement with several arbitrary choices made in the definition of the new means, e.g., $0 \otimes \log 0 = 0$ to comply with the convention extant for entropies. This supports our claim that entropies can be written and operated in semifields.

Since the equivalent probability function is a mean and the means can be expressed in a complete semifield, we have an expression of the former in the complete semifields of reals. Therefore, the expressions for the equivalent probability function and the information potential are:

$$\tilde{P}_r(P_X) = \hat{M}_r(P_X, P_X) \qquad\qquad \tilde{V}_r(P_X) = \hat{M}_r(P_X, P_X)^r. \tag{42}$$

However, in the definition of the equivalente probability function, and the entropies, we have that $w_i = x_i$ which entails some of the cases in Figure 4 are not visited.

### 3.3. Discussion: A Conjecture on the Abundance of Semifields in AI, ML and CI Applications

We are now in a position to better sustain our main conjecture about the abundance of semifields in science domains that try to model intelligent behaviour:

**Conjecture** (Pervasiveness of positive semifields in modelling intelligence). *Models and applications in Machine Intelligence—whether AI, ML or CI—operate with information, equivalent probability or proxies thereof, and those calculations are better conceptualized and successfully operated with the adequate dual pairs of positive semifields, especially entropy semifields.*

The argument for this conjecture is as follows:

1.  First, the shifting in definition of the Rényi entropy by $r = \alpha - 1$ in [9] leads to a a straightforward relation (5) between the power means of the probability distribution and the shifted Rényi entropy. For a given probability function or measure $P_X$ the evolution of entropy with $r \in [-\infty, \infty]$ resembles an information spectrum $\tilde{H}_r(P_X)$. In a procedure reminiscent of defining an inverse transform, we may consider an equivalent probability $\tilde{P}_r(P_X) = b^{-\tilde{H}_r(P_X)}$, which is the Hölder path of $P_X$, $\tilde{P}_r(P_X) = M_r(P_X, P_X)$.

2.  The function used by Rényi to define the generalized entropy, when shifted, is the composition of two functions: Hartley's information function and the power function of order $r$, which are monotone and invertible in the extended non-negative reals $[0, \infty]$. They are also bijections:

    *   The power function is a bijection over of the extended non-negative reals, and
    *   Hartley's is a bijection between the extended reals and the extended non-negative reals.

3.  But in Construction 1 both the power function and Hartley's prove to be isomorphisms of positive semifields, which are semirings whose multiplicative structure is that of a group, while the additive structure lacks additive inverses. Positive semifields are all naturally ordered and the power function respects this order within the non-negative reals, being an order isomorphism for generic power $r$. Importantly, positive semifields come in dually-ordered pairs and the expressions mixing operations from both members in the pair are reminiscent of boolean algebras.

    (a)  The power function $g(x) = x^r$ with $r \in [-\infty, \infty] \setminus \{0\}$ actually generates a whole family of semifields $(\mathbb{R}_{\geq 0})_r$ related to emphasizing smaller (with small $r$) or bigger values (with big $r$) in the non-negative reals $\mathbb{R}_{\geq 0}$. Indeed, the traditional weighted means are explained by the Construction 2 as being power-deformed aritmetic means, also known as Kolmogorov-Nagumo means with the power function as generators. These, semirings come in dually-ordered pairs for orders $r$ $(\mathbb{R}_{\geq 0})_r$ and $-r$ $(\mathbb{R}_{\geq 0})_{-r}$ whose orders are aligned or inverted with respect to that of $\mathbb{R}_{\geq 0}$. Indeed, $\mathbb{R}_{\geq 0} \cong (\mathbb{R}_{\geq 0})_1$ (Corollaries 1 and 2).

    (b)  However, Hartley's function is a dual-order isomorphism, entailing that the new order in the extended reals is the opposite of that on the non-negative reals (Corollary 3). It actually mediates between the (extended) probability semifield $\mathbb{R}_{\geq 0}$ and the semifield of informations, notated as a homage to Hartley as $-\mathbb{H}$ (Theorem 2).

4.  Since the composition of the power mean and Hartley's information function produces the function that Rényi used for defining his information measures, and this is a dual-order semifield isomorphism, we can see that entropies are actually operated in modified versions of Hartley's semifields $\mathbb{H}_r$ which come in pairs, as all completed positive semifields do (Theorem 3).

5.  Many of the $(\mathbb{R}_{\geq 0})_r$ and $\mathbb{H}_r$ semifields appear in domains that model intelligent behaviour. Among a list of applications we list the following:

    *   In AI, maximizing utilities and minimizing costs is used by many applications and algorithms, e.g., heuristic search, to mimic "informed" behaviour ([5], Chapter 3), decision theory ([5], Chapter 16), uncertainty and probability modelling ([5], Chapters 13–15). In most applications $(\mathbb{R}_{\geq 0})_1$, for multiplicatively-aggregated costs and utilities, and $\mathbb{H}_1$, for additively aggregated ones are being used. Note that both a semifield and its order-dual are needed to express mixed utility-cost expressions, as in electrical network analysis with resistances and conductances.
    *   In ML, $\mathbb{R}_{\geq 0}$ itself is used to model uncertainty as probabilities and $\mathbb{H}$ as log-probabilities. Sometimes the idempotent versions of these spaces $\lim_{r \to \pm \infty} (\mathbb{R}_{\geq 0})_r$, e.g., $\overline{\mathbb{R}}_{\min, \times}$ and $\overline{\mathbb{R}}_{\max, \times}$, and $\lim_{r \to \pm \infty} \mathbb{H}_r$, e.g., $\overline{\mathbb{R}}_{\min, +}$ and $\overline{\mathbb{R}}_{\max, +}$, are used, e.g., for A* stack decoding to find best candidates [32]. The Viterbi algorithm operates in $\mathbb{H}_\infty$, the semifield of max entropies,

as required by the application of decoding Markov models ([25], and Section 3.2.1). Although many of the problems and solutions cited above for AI can also be considered as part of ML, a recent branch of ML is solely based upon the Rényi entropy with $\alpha = 2$, $H_2(P_X) = \tilde{H}_1(P_X)$ [14]. Importantly, recall that every possible Hölder mean can be expressed as the arithmetic mean of a properly exponentiated kernel, whence the importance of this particular Rényi entropy would come.

- In CI, the sub-semifield obtained by the restriction of the operations to $\{\bot, e, \top\}$ appears as a ternary sub-semifield distinct from the Boolean semifield, which in turn appears as a binary sub-semifield of every complete semifield by restricting the carrier set to $\{\bot, \top\}$. As an important example, this ternary sub-semifield, as seen in Proposition 22 and Theorem 3 is pivotal in Spohn's logical rank theory [33] that essentially leverages the isomorphism of semifields between $\mathbb{R}_{\geq 0}$ and the $\overline{\mathbb{R}}_{\min,+} \equiv \mathbb{H}_\infty$ in logical applications.

  Finally—and also related to signal processing—mathematical morphology and morphological processing need to operate in the dual pair $(\overline{\mathbb{R}}_{\max,+}, \overline{\mathbb{R}}_{\min,+}) \equiv (\mathbb{H}_{-\infty}, \mathbb{H}_\infty)$ for image processing applications [34].

In our opinion, these facts provide a substantive grounding for our conjecture.

### 3.4. Historical Notes: The Rise of Positive Semifields

Our first noted appearance of the addition operation of (20) in the context of information is by Barnard [35]. From a system of axioms—typical of the stage of development of Information Theory in those days—it was concluded that for the accumulation of probabilities $p(x \oplus y) = f(p(x), p(y))$ the only possible functions are the (non-weighted) Kolmogorov-Nagumo (KN) means with $g(x) = x^r$, therefore the addition operation is $f(a, b) = (a^q + b^q)^{\frac{1}{q}}$. A special note both prescribed that $q \geq 0$, since otherwise the function $f$ is not strictly increasing, but cautioned against supposing that $q = 1$ is the only possible value.

Rényi explicitly cites the work of Barnard when introducing his information measure [3], and his approach is well-known to have enlightened the relationship between the KN means and entropy. This introduces two considerations:

- From Barnard's approach, it seems natural to use only $\alpha \geq 0$ in the KN means, and this is what Rényi chose.
- It is more difficult to guess why he decided to place the "origin" of his entropies at Hartley's ($\alpha = 0$) implying the harmonic average, instead of at Shannon's $\alpha = 1$, related to the geometric average.

  He must have known about the theory of the means [12], but may have been doubtful about allowing negative values of his parameter, which he never discussed, e.g., in [10]. Since Hartley's was already established as a measure of entropy, he may have felt that it was the "lowest" measure of entropy.

Notice there are no modern objections to using $\alpha < 0$ [36], and indeed the max-entropy $\alpha = -\infty$ is thoroughly used for specific applications. This led the authors of the present paper to define the shift in the order parameter that clarifies the relationship between the positive semifields in which they are defined vis-à-vis Pap's g-calculus [9].

In what looks like yet another case of almost simultaneous discovery, the deformed averages of Barnard had been looked into as alternate models of algebra for the modelling of physical phenomena already by Grossman and Katz ([37], and references therein) and Burgin [38]. Due to our inability to gather more than the fundamental mathematics from Burgin's paper, in Russian, we will review here the former work.

The non-Newtonian Calculus is an apt name by Grossman and Katz [37] to describe the behaviour of a systematic "deformation" of the standard algebra of the real numbers $\mathbb{R}$. In this formalism, cast in

the notation of this paper, an arithmetic is a complete ordered field $\mathcal{A}$ whose carrier set is a subset of $\mathbb{R}$, and a generator $\alpha$ is a one-to-one function whose domain is $\mathbb{R}$ and whose range is a subset of it. Then the $\alpha$-arithmetic is the complete ordered field obtained from the transformations:

- $\alpha$-zero: $0_\alpha = \alpha^{-1}(0)$
- $\alpha$-unit: $1_\alpha = \alpha^{-1}(1)$
- $\alpha$-addition: $u \oplus_\alpha v = \alpha\left(\alpha^{-1}(u) + \alpha^{-1}(v)\right)$
- $\alpha$-subtraction: $u \ominus_\alpha v = \alpha\left(\alpha^{-1}(u) - \alpha^{-1}(v)\right)$
- $\alpha$-multiplication: $u \otimes_\alpha v = \alpha\left(\alpha^{-1}(u) \times \alpha^{-1}(v)\right)$
- $\alpha$-division: $u \oslash_\alpha v = \alpha\left(\alpha^{-1}(u)/\alpha^{-1}(v)\right)$ for $v \neq 0_\alpha$
- $\alpha$-order : $u \leq_\alpha v \iff \alpha^{-1}(u) \leq \alpha^{-1}(v)$

The properties of such $\alpha$-arithmetics, and in particular the integral–differential calculus emanating from them, resemble formally the standard ones, since they operate in complete fields. The avowed intention of the authors is that "they may also be helpful in developing and understanding new systems of measurement that could yield simple physical laws ([37], p. 33)."

Indeed, this has been expanded upon and applied to alternate descriptions of physical phenomena by other authors [39]. The fundamental idea of this application is that the use of standard algebra to provide the "grounding" of integral-differential calculus for the description of physical phenomena is one symmetry (of nature) susceptible of "re-normalization", e.g., for fractal description in a non-Newtonian calculus [40].

We can see two main differences the approach in [37] as compared to our own.

- First, the motivation for our construction, viz. the modelling of operations on information, is clearly more specialized thant Grossman and Katz's, as purported in the above-mentioned quote.
- Second, and more importantly, in their framework, as in Barnard's paper, $\alpha$ seems to be assumed monotone, and so the properties of antitone generators and their results (leading to the second half of the dual pair) in Pap's g-calculus are downplayed.

However, regarding the second point, the order properties of dioids interact very tightly with their operations (see Section 2.2) and we believe this is the reason why Pap found it necessary to distinguish this case in his analysis on generator functions and their domains. As it happens, the resulting structure lacks an additive inverse which dovetails into the intuition that entropies and informations cannot be "(abstractly) subtracted", hence we consider the construction in this paper more adequate for our purposes.

In any case it would be interesting to follow up on Grossman and Katz's approach to particular "positive fields". Since we have already proven that the geometries of their vector spaces are quite different in the, admittedly, extreme case of the idempotent (positive) semifields [15,30], it seems that some reinterpretation of this particular case should be undertaken in their formalism. We leave this for future work.

## 4. Conclusions

In the context of information measures, we have reviewed the notion of positive semifield—a positive semiring with a multiplicative group structure—distinct from that of the more usual fields with an additive group structure: in positive semirings there are no additive inverses, but there is a "natural order" compatible with addition and multiplication.

Through Pap's g-calculus and Mesiar and Pap's semifield Construction, we have related the Hölder means to the shifted Rényi measures of information $\tilde{H}_r(P_X)$ for pmf $P_X$, which appear as just the logarithm of the Kolmogorov-Nagumo means in different semifields obtained by ranging the $r$ parameter in $[-\infty, \infty]$. As a fundamental example, we provide the rewriting of the Hölder means in $(\mathbb{R}_{\geq 0})_r$ and its dual, which provides the basis for the shifted Renyi entropy, cross-entropy and divergence.

Our avowed intention with this exploration was to provide a conjecture, from an information theoretic point of view, about the abundance of semifield valued quantities in a variety of machine learning and computational intelligence tasks. Namely, that such semifield-valued quantities are being used either directly as Rényi information measures—including Shannon's—or indirectly as proxies of such. Experimental corroboration of this conjecture would seem to entail some form of constructive theory for intelligence.

## Abbreviations

The following abbreviations are used in this manuscript:

| | |
|---|---|
| AI | Artificial intelligence |
| CI | Computational intelligence |
| KN | Kolmogorov–Nagumo |
| ML | Machine learning |

## References

1. Butkovič, P. *Max-linear Systems. Theory and Algorithms*; Monographs in Mathematics; Springer: Heidelberg, Germany, 2010.

2. Gondran, M.; Minoux, M. *Graphs, Dioids and Semirings. New Models and Algorithms*; Operations Research Computer Science Interfaces Series; Springer: Heidelberg, Germany, 2008.

3. Renyi, A. On measures of entropy and information. In *Proceedings of the Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics, Berkeley, CA, USA, 20 June–30 July 1960*; University of California Press: Berkeley, CA, USA, 1961; pp. 547–561.

4. Golan, J.S. *Semirings and Their Applications*; Kluwer Academic: Dordrecht, The Netherlands, 1999.

5. Russell, S.J.; Norvig, P. *Artificial Intelligence—A Modern Approach*, 3rd international ed.; Artificial Intelligence; Prentice Hall: Upper Saddle River, NJ, USA, 2010.

6. Murphy, K.P. *Machine Learning. A Probabilistic Perspective*; MIT Press: Cambridge, MA, USA, 2012.

7. Engelbrecht, A.P. *Computational Intelligence. An Introduction*; Wiley: Hoboken, NJ, USA, 2002.

8. Baccelli, F.; Cohen, G.; Olsder, G.; Quadrat, J. *Synchronization and Linearity*; Wiley: Hoboken, NJ, USA, 1992.

9. Valverde-Albacete, F.J.; Peláez-Moreno, C. The case for shifting the Renyi entropy. *Entropy* **2019**, *21*, 46. [CrossRef]

10. Renyi, A. *Probability Theory*; Courier Dover Publications: Mineola, NY, USA, 1970.

11. Pap, E. g-calculus. In *Zbornik Radova Prirodno-Matematichkog Fakulteta. Serija za Matematiku. Review of Research. Faculty of Science. Mathematics Series*; University of Novi Sad: Novi Sad, Serbia, 1993; pp. 145–156.

12. Hardy, G.H.; Littlewood, J.E.; Pólya, G. *Inequalities*; Cambridge University Press: Cambridge, UK, 1952.

13. Beck, C.; Schögl, F. *Thermodynamics of Chaotic Systems: An Introduction*; Cambridge University Press: Cambridge, UK, 1995.

14. Principe, J.C. *Information Theoretic Learning*; Information Science and Statistics; Springer: New York, NY, USA, 2010.

15. Valverde-Albacete, F.J.; Peláez-Moreno, C. The Spectra of irreducible matrices over completed idempotent semifields. *Fuzzy Sets Syst.* **2015**, *271*, 46–69. [CrossRef]

16. Moreau, J.J. Inf-convolution, sous-additivité, convexité des fonctions numériques. *J. Math. Pures Appl.* **1970**, *49*, 109–154. (In French)

17. Ellerman, D.P. *Intellectual Trespassing as a Way of Life. Essays in Philosophy, Economics and Mathematics*; Rowman & Littlefield Publishers Inc.: Lanham, MD, USA, 1995.

18. Maslov, V.; Volosov, K. *Mathematical Aspects of Computer Engineering*; Mir: Moscow, Russia, 1988.

19. Pap, E.; Ralević, N. Pseudo-Laplace transform. *Nonlinear Anal. Theory Methods Appl.* **1998**, *33*, 533–550. [CrossRef]

20. Mesiar, R.; Pap, E. Idempotent integral as limit of g-integrals. *Fuzzy Sets Syst.* **1999**, *102*, 385–392. [CrossRef]

21. Valverde-Albacete, F.J.; Peláez-Moreno, C. Towards Galois Connections over Positive Semifields. In *Information Processing and Management of Uncertainty in Knowledge-Based Systems*; Springer International Publishing: Heidelberg, Germany, 2016; *CCIS* Volume 611, pp. 81–92.

22. Grabisch, M.; Marichal, J.L.; Mesiar, R.; Pap, E. Aggregation functions: Construction methods, conjunctive, disjunctive and mixed classes. *Inf. Sci.* **2011**, *181*, 23–43. [CrossRef]

23. Shannon, C.E. A mathematical theory of Communication. *Bell Syst. Tech. J.* **1948**, *XXVII*, 379–423. [CrossRef]

24. Viterbi, A.J. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Trans. Inf. Theory* **1967**, *13*, 260–269. [CrossRef]

25. Theodosis, E.; Maragos, P. Analysis of the Viterbi algorithm using tropical algebra and geometry. In Proceedings of the IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC—18), Kalamata, Greece, 25–28 June 2018; pp. 1–5.

26. Forney, G. The Viterbi algorithm. *Proc. IEEE* **1973**, *61*, 268–278. [CrossRef]

27. Rabiner, L.R. A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE* **1989**, *77*, 257–286. [CrossRef]

28. Bellman, R.; Kalaba, R. On the role of dynamic programming in statistical communication theory. *IRE Trans. Inf.* **1957**, *3*, 197–203. [CrossRef]

29. Cuninghame-Green, R. *Minimax Algebra*; Number 166 in Lecture notes in Economics and Mathematical Systems; Springer: Heidelberg, Germany, 1979.

30. Valverde-Albacete, F.J.; Peláez-Moreno, C. The spectra of reducible matrices over complete commutative idempotent semifields and their spectral lattices. *Int. J. Gen. Syst.* **2016**, *45*, 86–115. [CrossRef]

31. Ganter, B.; Wille, R. *Formal Concept Analysis: Mathematical Foundations*; Springer: Berlin/Heidelberg, Germany, 1999.

32. Paul, D.B. An efficient A* stack decoder algorithm for continuous speech recognition with a stochastic language model. In Proceedings of the workshop on Speech and Natural Language—HLT '91, Harriman, NY, USA, 23–26 February 1992; Association for Computational Linguistics: Morristown, NJ, USA, 1992; p. 405.

33. Spohn, W. *The Laws of Belief: Ranking Theory and Its Philosophical Applications*; Oxford University Press: Oxford, UK, 2012.

34. Ronse, C. Why mathematical morphology needs complete lattices. *Signal Process.* **1990**, *21*, 129–154. [CrossRef]

35. Barnard, G. The theory of information. *J. R. Stat. Soc. Ser. B* **1951**, *13*, 46–64. [CrossRef]

36. Harremoës, P. Interpretations of Rényi entropies and divergences. *Phys. A Stat. Mech. Appl.* **2005**, *365*, 57–62. [CrossRef]

37. Grossman, M.; Katz, R. *Non-Newtonian Calculus*; Lee Press: Pigeon Cove, MA, USA, 1972.

38. Burgin, M. Nonclassical models of the natural numbers. *Uspekhi Matemat. Nauk.* **1977**, *32*, 209–210.

39. Czachor, M. Relativity of arithmetic as a fundamental symmetry of physics. *Quantum Stud. Math. Found.* **2016**, *3*, 123–133. [CrossRef]

40. Czachor, M. Waves along fractal coastlines: from fractal arithmetic to wave equations. *Acta Phys. Pol. B* **2019**, *50*, 813–831. [CrossRef]