

Article

A Dynamical Systems-Based Hierarchy for Shannon, Metric and Topological Entropy

Raymond Addabbo ^{1,*} and Denis Blackmore ² ¹ Department of Arts and Sciences, Vaughn College of Aeronautics and Technology, Flushing, NY 11369, USA² Department of Mathematical Sciences and Center for Applied and Computational Mathematics, New Jersey Institute of Technology, Newark, NJ 07102-1982, USA; denis.l.blackmore@njit.edu

* Correspondence: raymond.addabbo@vaughn.edu; Tel.: +1-718-429-6600 (ext. 261)

Received: 12 August 2019; Accepted: 20 September 2019; Published: 25 September 2019



Abstract: A rigorous dynamical systems-based hierarchy is established for the definitions of entropy of Shannon (information), Kolmogorov–Sinai (metric) and Adler, Konheim & McAndrew (topological). In particular, metric entropy, with the imposition of some additional properties, is proven to be a special case of topological entropy and Shannon entropy is shown to be a particular form of metric entropy. This is the first of two papers aimed at establishing a dynamically grounded hierarchy comprising Clausius, Boltzmann, Gibbs, Shannon, metric and topological entropy in which each element is ideally a special case of its successor or some kind of limit thereof.

Keywords: topological entropy; Shannon entropy; metric entropy; Bernoulli scheme

1. Introduction

Entropy, which can among a variety of other things, be roughly viewed as a measure of uncertainty (cf. [1]), has been and remains a fascinating and still not completely understood concept with an amazing range of applications, including quantum and ecological systems [2]. We have discovered that many mathematical, scientific and engineering colleagues share our abiding curiosity about rigorous connections among the myriad definitions of entropy. For example, what if any is the relationship between classical Clausius entropy and topological entropy or even Shannon entropy? There have been several extensive investigations of various types of entropy including historical accounts of the relevant developments and informative investigations of the linkages among the various forms of entropy such as in [3–13], but they all appear to be somewhat lacking in terms of identification of truly compelling unifying themes for the multifarious definitions.

Our intention in this and a subsequent paper is to provide a rigorous partial answer to this question by proving that there is a dynamical systems thread connecting all of the following definitions of entropy leading, with the possible addition of some mild assumptions, to the hierarchy

$$\text{Clausius} \longleftarrow \text{Gibbs} \longleftarrow \text{Boltzmann} \longleftarrow \text{Shannon} \longleftarrow \text{metric} \longleftarrow \text{topological}$$

By rigor as used here, we mean that any connections and special case identifications shall be proved. In this paper, we shall establish a dynamical systems thread connecting Shannon, metric and topological entropy; in a forthcoming paper we hope to complete the hierarchy by elucidating the connections among the first three of the above and their link to the last three. In a related note, we recommend a recent interesting functorial connection among these three entropies introduced by Delvenne [14].

The material to be presented is organized as follows. In the interest of completeness, we present the definitions and some basic properties of topological, metric and Shannon entropy. Then, in Section 3,

we prove that metric entropy is a special case of topological entropy if one adds just a few assumptions. Equality of metric and topological entropy can and shall be established in the following two forms: Constructively, namely by showing that certain measurable dynamical systems can be given a topology, which might be quite far removed from some of the more usual types, for which the two entropies in question are equal and by comparison, where one can sometimes determine when a given topology on a measurable dynamical system yields equality of the entropies. Next, in Section 4, we show that Shannon entropy is special case of metric entropy when formulated in a certain dynamical context. In particular, one can frame the underlying information theory foundation as a measurable dynamical system in which the metric and Shannon entropies are equal. As an example, we show how the result for a binary alphabet can be obtained essentially from scratch, in stark contrast to the general result due mainly to Kolmogorov and Sinai that involves very deep and extensive analysis. Finally, in Section 5 we summarize some of the conclusions reached and outline our plans for related future work on extending the hierarchy for entropies.

2. Topological, Metric and Shannon Entropy

In this section we define topological, metric and Shannon entropy, list some of their properties, and describe several basic relationships among them. We begin with the topological entropy of a discrete dynamical system on a topological space.

2.1. Topological Entropy

Let X be a nonempty Hausdorff topological space with topology \mathfrak{T} (comprising the open subsets of the space). A *topological discrete (semi-) dynamical system* (or just *discrete dynamical system* for short) is a continuous action of the form

$$F : \mathbb{N}^* \times X \rightarrow X,$$

where $F(n, x) := f^n(x)$, $f : X \rightarrow X$ is a continuous map, f^n is the n -th iterate of the map under composition \circ , with the convention that f^0 is the identity map $id = id_X$ and \mathbb{N}^* is the abelian semigroup $\{0\} \cup \mathbb{N}$ under addition. There is a standard nomenclature (see, for example [15]) that includes such definitions as (positive semi-) *orbits* $O(x) := \{f^n(x) : n \in \mathbb{N}^*\}$ and *fixed points* where $f(x) = x$, so that $O(x) = \{x\}$. For convenience, we shall denote the dynamical system by $\mathfrak{D} := (f, X)$.

Assuming X to be compact, Adler et al. [16] defined the topological entropy of the dynamical system as follows: If $\mathcal{U} = \{U\}$ is any open covering of X , the minimum cardinality $N(\mathcal{U})$ over all subcoverings must be finite owing to the compactness of X . Now, if \mathcal{U} and $\mathcal{V} = \{V\}$ are two open coverings of X , so is their *common refinement*

$$\mathcal{U} \vee \mathcal{V} := \{U \cap V : (U, V) \in \mathcal{U} \times \mathcal{V}\},$$

every set of which is contained in a set of both \mathcal{U} and \mathcal{V} , with the refinement relation being typically denoted as $\mathcal{U}, \mathcal{V} \prec \mathcal{U} \vee \mathcal{V}$. By iterating refinement, for each $n \in \mathbb{N}$ we obtain the open covering $\mathcal{U}^n := \mathcal{U} \vee f^{-1}(\mathcal{U}) \vee \dots \vee f^{-n+1}(\mathcal{U})$ for which it can readily be shown that

$$h_{\mathfrak{T}}(\mathcal{U}, f) = \lim_{n \rightarrow \infty} \frac{\log N(\mathcal{U}^n)}{n} = \inf \left\{ \frac{\log N(\mathcal{U}^n)}{n} : n \in \mathbb{N} \right\} \quad (1)$$

exists. Then, the *topological entropy* is defined as

$$h_{\mathfrak{T}}(f) := \sup \{h_{\mathfrak{T}}(\mathcal{U}, f) : \mathcal{U} \in OC(X)\}, \quad (2)$$

where $OC(X)$ is the set of all open coverings of X .

There is also another (metric-based) definition of topological entropy (see Bowen [17], Dinaburg [4,5] and [18]), which for compact metric spaces is equivalent to the Adler–Konheim–McAndrew definition above.

It should be remarked that although we have confined ourselves to compact spaces, it is possible to extend the definition of topological entropy to noncompact spaces (cf. [19–21]). There is an analogous definition of topological entropy for *continuous dynamical systems* that applies to continuous (usually continuously differentiable), which we include for possible future reference. It applies to continuous actions of the type

$$\Phi : \mathbb{R} \times X \rightarrow X,$$

where \mathbb{R} is the topological abelian group of real numbers with respect to addition. These actions arise naturally from solutions of autonomous systems of ordinary differential equations satisfying conditions guaranteeing global existence and uniqueness. There is the usual notation for these systems, which includes for example, the positive semiorbit through x defined as $O_+(x) := \{\varphi_t(x) := \Phi(t, x) : t \geq 0\}$ The associated time-1 map φ_1 , which generates a discrete dynamical system, is defined as $\varphi_1(x) := \Phi(1, x)$, and the topological entropy of the continuous dynamical system is

$$h_{\mathfrak{T}}(\Phi) := h_{\mathfrak{T}}(\varphi_1) = \sup \{h_{\mathfrak{T}}(\mathcal{U}, \varphi_1) : \mathcal{U} \in OC(X)\} = \sup \{h_{\mathfrak{T}}(\mathcal{U}, f) : \mathcal{U} \in OC(X)\}. \tag{3}$$

Finally, we note in passing that the topological entropy in both the discrete and continuous cases can be intuitively viewed as the exponential growth in distinct orbits with discrete or continuous time, respectively.

Some Properties of Topological Entropy

Obviously, the topological entropy is nonnegative, and one can use the basic definitions and a bit of straightforward analysis (as in [15,16]) to prove the following properties:

- (TE1) $h_{\mathfrak{T}_X}(f) = h_{\mathfrak{T}_Y}(g)$ if $f : X \rightarrow X$ and $g : Y \rightarrow Y$ are *topologically conjugate* (there is a homeomorphism $\varphi : X \rightarrow Y$ such that $g \circ \varphi = \varphi \circ f$)
- (TE2) $h_{\mathfrak{T}}(f) \geq h_{\mathfrak{T}_Y}(f|_Y)$ if $Y \subset X$ is closed and f -invariant and \mathfrak{T}_Y is the subspace topology of \mathfrak{T} on Y , where $f|_Y$ denotes the restriction of f to Y
- (TE3) $h_{\mathfrak{T}}(f) = h_{\mathfrak{T}}(f^{-1})$ if f is a homeomorphism
- (TE4) $h_{\mathfrak{T}}(f^n) = nh_{\mathfrak{T}}(f)$ if $n \in \mathbb{N}$
- (TE5) $h_{\mathfrak{T}_X \times \mathfrak{T}_Y}(f \times g) = h_{\mathfrak{T}_X}(f) + h_{\mathfrak{T}_Y}(g)$
- (TE6) If a sequence $\{\mathcal{U}_n\}$ of open covers is *refining* - meaning that $\mathcal{U}_n \prec \mathcal{U}_{n+1}$ for every $n \in \mathbb{N}$ and for every open cover \mathcal{V} of X there exists an n such that $\mathcal{V} \prec \mathcal{U}_n$ - then $\{h_{\mathfrak{T}}(\mathcal{U}_n, f)\}$ is a nondecreasing sequence, with $h_{\mathfrak{T}}(f) = \lim_{n \rightarrow \infty} h_{\mathfrak{T}}(\mathcal{U}_n, f) = \sup\{h_{\mathfrak{T}}(\mathcal{U}_n, f) : n \in \mathbb{N}\}$.

2.2. Kolmogorov–Sinai Metric Entropy

A discrete *measurable dynamical system* consist of a nonempty set X , a σ -algebra of μ -measurable subsets \mathcal{M} of X , where μ is a probability measure (with $\mu(X) = 1$) and a measurable function $f : X \rightarrow X$ (satisfying $f^{-1}(A) \in \mathcal{M}$ whenever $A \in \mathcal{M}$) such that μ is f -invariant, which means that $\mu \circ f^{-1} = \mu$ on \mathcal{M} . The iterates of f obviously also define a discrete dynamical system, with the usual notions of orbits, fixed points and the like.

The *metric* (Kolmogorov–Sinai or just K – S) *entropy* requires a number of introductory elements for its definition. A *measurable partition* of X is a finite, pairwise-disjoint sequence $\mathcal{P} = \{Q_1, \dots, Q_m\}$ of μ -measurable sets of positive measure such that $X = \bigcup_{k=1}^m Q_k$, and it is convenient to denote the set of all such measurable partitions of X as $\mathfrak{P}(X)$. The *entropy of a measurable partition* is defined as

$$H(\mathcal{P}) := -\sum_{Q \in \mathcal{P}} \mu(Q) \log \mu(Q). \tag{4}$$

As with open coverings in the definition, we can define a common refinement of measurable partitions \mathcal{P} and $\tilde{\mathcal{P}}$ as

$$\mathcal{P} \vee \tilde{\mathcal{P}} := \{Q \cap \tilde{Q} : (Q, \tilde{Q}) \in \mathcal{P} \times \tilde{\mathcal{P}}\},$$

for which the refinement relation is analogously denoted by $\mathcal{P}, \tilde{\mathcal{P}} \prec \mathcal{P} \vee \tilde{\mathcal{P}}$.

Iterating refinement in a manner analogous to that used for topological entropy, we define measurable partitions associated to f of the form

$$\mathcal{P}^n := \mathcal{P} \vee f^{-1}(\mathcal{P}) \vee \dots \vee f^{-n+1}(\mathcal{P}),$$

and define

$$H(\mathcal{P}, f) := \lim_{n \rightarrow \infty} \frac{H(\mathcal{P}^n)}{n}, \tag{5}$$

where the limit can be readily shown to exist. Then, we define the metric entropy as

$$h_\mu(f) := \sup \{H(\mathcal{P}, f) : \mathcal{P} \in \mathfrak{P}(X)\}, \tag{6}$$

and take note of the strong similarity with the definition of topological entropy.

It should also be mentioned that just as for the case of topological entropy, we can define metric entropy for continuous measurable dynamical systems using the time-1 map of the flow as in (3). As with the topological entropy, there are intuitive characterizations of metric entropy, two of which are the following: The exponential growth rate of typical orbits; and the maximum of the rate of extractable information.

Basic Properties of K-S Entropy

The nonnegativity of K-S entropy is clear from its definition, and we have the following basic properties that are clearly analogs of those for topological entropy, with the last of which following from the concavity of $-x \log x$ (see [10,15,22–25]).

- (ME1) $h(f) = h(g)$ if $f : X \rightarrow X$ and $g : Y \rightarrow Y$ are *metrically conjugate* (there is a measurable bijection $\varphi : X \rightarrow Y$ such that $g \circ \varphi = \varphi \circ f$ and $\nu = \mu \circ \varphi^{-1}$)
- (ME2) $h(f) \geq h_{|Y}(f|_Y)$ if $Y \subset X$ is measurable and f -invariant and $|_Y$ is the restriction of μ to Y , where $f|_Y$ denotes the restriction of f to Y
- (ME3) $h(f) = h(f^{-1})$ if f is a measurable bijection
- (ME4) $h(f^n) = nh(f)$ if $n \in \mathbb{N}$
- (ME5) $h_\times(f \times g) = h(f) + h(g)$
- (ME6) If a sequence $\{\mathcal{P}_n = \{Q_1, \dots, Q_{m_n}\}\}$ in $\mathfrak{P}(X)$ is *refining* - meaning that $\mathcal{P}_n \prec \mathcal{P}_{n+1}$ for every $n \in \mathbb{N}$ and $\|\mathcal{P}_n\| := \max\{\mu(Q_k) : 1 \leq k \leq m_n\} \rightarrow 0$ as $n \rightarrow \infty$ - then $\{h_\mu(\mathcal{U}_n, f)\}$ is a nondecreasing sequence with $h_\mu(f) = \lim_{n \rightarrow \infty} h_\mu(\mathcal{U}_n, f) = \sup\{h(\mathcal{U}_n, f) : n \in \mathbb{N}\}$.

2.3. Shannon Entropy

In contrast with topological and metric entropy, Shannon entropy (see [26,27]) has an information theory rather than a dynamical system foundation. The basic elements can be distilled as follows: Let $S := \{s_1, \dots, s_m\}$ be a nonempty finite set of symbols or messages (sometimes referred to as the alphabet) with a discrete probability p assigned to each, such that $p(s_i) \geq 0$ for all $1 \leq i \leq m$ and $p(s_1) + \dots + p(s_m) = 1$. Then, the *Shannon entropy* of the message ensemble

$$H(S) := -\sum_{i=1}^m p(s_i) \log p(s_i) \tag{7}$$

is just the average or expected value of the information content of the message ensemble X , which is very much like the entropy of a measurable partition used in the definition of metric entropy.

Several Properties of Shannon Entropy

It follows from (7) that the Shannon entropy is nonnegative, and there are several other readily verified properties such as:

- (SE1) $H(S) = H(\tilde{S})$ if there is a bijection $\varphi : S \rightarrow \tilde{S}$ such that $g \circ \varphi = \varphi \circ f$ and $\tilde{p} = p \circ \varphi^{-1}$
- (SE2) $H(S) \geq H(\tilde{S})$ if $\tilde{S} \subset S$
- (SE3) $H(S \times \tilde{S}) = H(S) + H(\tilde{S})$ if S and \tilde{S} are independent ($\text{prob}(s_i, \tilde{s}_j) = p(s_i)\tilde{p}(\tilde{s}_j)$ for all i, j)
- (SE4) $H(S) = \log n$ is maximized when $p(s_1) = \dots = p(s_n) = 1/n$.

2.4. A Relationship among the Entropies

Here we shall describe a well known relationship between topological and metric entropy obtained by Dinaburg [4,5], Goodman [8], Goodwyn [9] and Misiurewicz [28] using variational techniques (see also [12,13,15]), and often referred to as the *variational principle*. Let $f : X \rightarrow X$ be a continuous map on a compact metric space X with topology \mathfrak{T} , which defines a discrete dynamical system that we represent by the triple (f, X, \mathfrak{T}) . To establish the topological-metric entropy relation, it is convenient to define

$$\mathfrak{M}(X) := \{ \mu : \mu \text{ an invariant Borel probability measure on the } \sigma\text{-algebra } \mathcal{M}_\mu \text{ of subsets of } X \}, \quad (8)$$

where Borel means that $\mathfrak{T} \subset \mathcal{M}_\mu$ for all $\mu \in \mathfrak{M}(X)$. Then we have the following results, which is proved very efficiently in [28]:

Theorem 1. *If X is a compact metric space with topology \mathfrak{T} and $f : X \rightarrow X$ is a continuous map, then*

$$h_{\mathfrak{T}}(f) := \sup \{ h_\mu(f) : \mu \in \mathfrak{M}(X) \}. \quad (9)$$

3. When Topological Entropy Equals Metric Entropy

In this section we shall prove there is a topology on numerous discrete measurable dynamical systems (DMDS) of interest such that the corresponding topological entropy is equal to the K-S entropy. Let (f, X, \mathcal{M}, μ) be a measurable dynamical system having metric entropy $h(f)$. We shall show that in certain cases there is a topology \mathfrak{T} on X for which $f : X \rightarrow X$ is continuous and $h_{\mathfrak{T}}(f) = h(f)$. One interesting example, which is an example of type (F2) that concerns Conjecture 3 of [16], which has since been proven (see [17,18,21,29]); namely, if X is a compact separable topological group, with topology \mathfrak{T} , and $f : X \rightarrow X$ is a continuous automorphism, then the Haar measure μ is f -invariant, and $h_{\mathfrak{T}}(f) = h_\mu(f)$. What we also have in mind is a Constructive-type of problem that involves defining a topology on a given DMDS, with no pre-specified topology, such that the topological and K-S entropies coincide and this is where we shall begin.

3.1. Constructive-Type Equality

The question is when can the phase space X of a given DMDS can be endowed with a compact metric topology \mathfrak{T} such that $h_{\mathfrak{T}}(f) = h_\mu(f)$. Our main result, which follows directly from the Jewett–Krieger theorem (cf. [3,30,31]), seems not to have appeared in the literature.

Theorem 2. *Let $\mathfrak{D} = (f, X, \mathcal{M}, \mu)$ be a measurable dynamical system, with $f : X \rightarrow X$ bijective and ergodic. Then, there exists a compact metric topology \mathfrak{T} on X such that $h_{\mathfrak{T}}(f) = h_\mu(f)$.*

Proof. The Jewett–Krieger theorem implies that there is a compact metric space \hat{X} with topology $\hat{\mathfrak{T}}$ with a continuous map $\hat{f} : \hat{X} \rightarrow \hat{X}$, \hat{f} -invariant Borel measure $\hat{\mu}$, and a sigma-algebra $\hat{\mathcal{M}}$ of $\hat{\mu}$ -measurable subsets of \hat{X} defining a measure-theoretic dynamical system $\hat{\mathfrak{D}} = (\hat{f}, \hat{X}, \hat{\mathcal{M}}, \hat{\mu})$ with the following properties:

- (i) There is a *measure-theoretic dynamical system embedding* $\Phi : \mathfrak{D} \rightarrow \hat{\mathfrak{D}}$, which means there is an injective map $\phi : X \rightarrow \hat{X}$ such that $\hat{f} \circ \phi = \phi \circ f$ and $A \in \mathcal{M}$ implies that $\phi(A) \in \hat{\mathcal{M}}$, with $\hat{\mu}(\phi(A)) = \mu(A)$.

- (ii) If $\hat{\mu}_*$ is the restriction of $\hat{\mu}$ to $\phi(X)$ and $\hat{\mathfrak{T}}_*$ is the subspace topology induced by $\hat{\mathfrak{T}}$ on $\phi(X)$, then $h_{\hat{\mathfrak{T}}}(\hat{f}) = h_{\hat{\mu}_*}(\hat{f})$.

The choice of the desired topology can be inferred readily from the above; namely, we define \mathfrak{T} to be the topology induced by the injection $\phi : X \rightarrow \hat{X}$. As is well known, this topology \mathfrak{T} is simply that which is generated by the set $\{\phi^{-1}(V) : V \in \hat{\mathfrak{T}}\}$. Hence, owing to the properties described above, the systems are rendered both topologically and metrically conjugate on $\phi(X)$. This completes the proof since it follows that (ii) implies $h_{\mathfrak{T}}(f) = h_{\mu}(f)$. \square

Although the above is a simple corollary of the Jewett–Krieger result, the proof of the theorem itself is quite long and deep. This suggests that there may be a more direct proof of Theorem 1, which is something worth investigating.

3.2. Comparison-Type Equality

Since we start with a choice of the topology, the construction involves the possibility of noncompactness for which there are approaches in the literature such as [18–21,29,32,33]. However, we avoid this issue by selecting a compact topology for X .

We begin with the construction of the topology \mathfrak{T} for the phase space X of the DMDS $\mathfrak{D} := (f, X, \mathcal{M}, \mu)$ that we hope yields the equality of the two entropies. A rather natural choice for \mathfrak{T} is the topology $\mathfrak{T}_0(\mu)$ generated by the sets in all possible measurable partitions of X comprising subsets of positive measure, but this has inherent problems, not the least of which concerns compactness. Certainly, $(X, \mathfrak{T}_0(\mu))$ is not *a priori* compact, nor as it turns out is it suitable for proving that the corresponding topological and given metric entropies are equal. Consequently, we must be more restrictive in the definition of the chosen topology as well as the DMDS.

First, we deal with the choice of topology and then describe some of the fundamental properties of the DMDS vis a vis the topology that we will use to guarantee the equality of the metric and topological entropies. Our first assumption is that X can be compactly embedded in a metric space Y with metric d , so it may be considered as a compact metric space with metric d restricted to X , with corresponding metric topology \mathfrak{T}_d . This immediately takes care of the question of compactness, for example, and provides other desirable properties associated with metric topologies. Moreover, it applies to any finite-dimensional, smooth compact manifold. Another convenient assumption connecting the topology \mathfrak{T}_d and \mathfrak{D} is the following: The measure is Borel with respect to the topology; i.e., $\mathfrak{T}_d \subset \mathcal{M}$ and so all open sets are μ -measurable. It is worth noting that compact smooth manifolds with the usual types of measures typically satisfy these two assumptions. We shall also find it convenient to assume that the map f is continuous with respect to \mathfrak{T}_d , which implies that the map is also proper; i.e., the preimage of compact sets are compact.

3.2.1. Definitions and Additional Notation

The above assumptions are basic and conveniently serve our needs, but several more subtle properties are required to obtain equality of entropies, which are best delineated by introducing some additional simplifying notation in the context of the assumptions that have so far been made. For example, the following notion shall prove useful.

Definition 1. A (finite) open cover $\mathcal{U} = \{U_1, \dots, U_m\}$ of X is minimal with respect to a measurable partition $\mathcal{P} = \{Q_1, \dots, Q_m\}$ of X if $Q_k \subset U_k$ for every $1 \leq k \leq m$ and no proper subset of \mathcal{U} is a covering of X . For convenience, we denote this as \mathcal{U} -min- \mathcal{P} .

Definition 2. A measurable partition $\mathcal{P} = \{Q_1, \dots, Q_m\}$ is an α - β (with $0 < \alpha < \beta$) partition of X with respect to \mathfrak{T}_d and denoted as $\mathcal{P}(\alpha, \beta)$ if the following properties obtain for all the elements: (i) $\mu(Q_k) > 0$; (ii) Q_k is connected in the topology \mathfrak{T}_d ; (iii) the diameter $d(Q_k) < \beta$; and (iv) there is at least one point

$x \in Q_k$ such that the (closed ball) $\bar{B}_\alpha(x) := \{y \in X : d(x, y) \leq \alpha\}$, corresponding to the open ball $B_\alpha(x) := \{y \in X : d(x, y) < \alpha\}$, is contained in Q_k . We denote the set of all α - β partitions as $\mathfrak{P}_{\alpha\beta}(X)$.

Definition 3. The DMDS \mathfrak{D} is of L-type for \mathfrak{T}_d if it satisfies the following properties: (i) $\mu(B_\alpha(x)) > 0$ for every $\alpha > 0$; (ii) $\mu(B_\alpha(x)) \rightarrow 0$ as $\alpha \rightarrow 0$; (iii) for every element Q_k of an α - β partition of X and $\epsilon > 0$ there is a connected open set $U_k(\epsilon)$ such that $Q_k \subset U_k(\epsilon)$, $\mu(U_k(\epsilon) \setminus Q_k) < \epsilon$ and $d(U_k(\epsilon)) < \beta + \epsilon$; (iv) $d(x, \partial U_k(\epsilon)) < \epsilon/2$ for every $x \in \partial Q_k$ and (v) $U_k(\epsilon_1) \subset U_k(\epsilon_2)$ whenever $\epsilon_1 < \epsilon_2$, so that $\lim_{\epsilon \downarrow 0} \mu(U_k(\epsilon)) = \mu(Q_k)$. We note that here the L is for Lebesgue, inasmuch as these properties are well known to apply for the normalized Lebesgue measure on a compact subset of Euclidean space.

Definition 4. Let $\mathcal{P} = \{Q_1, \dots, Q_m\} \in \mathfrak{P}_{\alpha\beta}(X)$ and $\mathcal{U}(\mathcal{P}, \epsilon) = \{U_1(\epsilon), \dots, U_m(\epsilon)\}$ be a corresponding open covering in accordance with Definition 3. Then we say that open covering $\mathcal{U}(\mathcal{P}, \epsilon)$ is ϵ -tight with respect to \mathcal{P} .

The following result is a readily verified consequence of the above definitions.

Lemma 1. Let $\mathcal{P} = \{Q_1, \dots, Q_m\} \in \mathfrak{P}_{\alpha\beta}(X)$. Then, for any given integer $n > 1$ and $\epsilon > 0$ there exists a δ with $0 < \delta = \delta(\epsilon, n) < \epsilon$ such that if $\mathcal{U} = \mathcal{U}(\mathcal{P}, \delta) = \{U_1(\delta), \dots, U_m(\delta)\}$ is δ -tight and (consequently for δ sufficiently small) minimal with respect to \mathcal{P} , $\mathcal{U}^k = \mathcal{U}^k(\mathcal{P}, \epsilon) = \{U_1(\epsilon), \dots, U_m(\epsilon)\}^k$ is ϵ -tight and minimal with respect to $\mathcal{P}^k = \{Q_1, \dots, Q_m\}^k$ for all $1 \leq k \leq n$.

3.2.2. The Comparison Theorem

We now assemble some of the properties defined above for use as assumptions in the main theorem of this section.

Definition 5. The measurable dynamical system $\mathfrak{D} := (f, X, \mathcal{M}, \mu)$ is T-compatible if there exists a compact metric topology \mathfrak{T}_d on X such that the following properties obtain: (i) f is continuous with respect to \mathfrak{T}_d (ii) μ is Borel with respect to \mathfrak{T}_d ; (iii) $\mathfrak{P}_{\alpha\beta}(X) \neq \emptyset$; and (iv) \mathfrak{D} is of L-type for \mathfrak{T}_d .

The stage has now been set for the following result on comparative equality.

Theorem 3. Suppose that the DMDS \mathfrak{D} is T-compatible with respect to \mathfrak{T}_d on X and the following additional properties hold:

- (E1) There exists a sequence of partitions $\{\mathcal{P}(\alpha_n, \beta_n)\}$ in $\mathfrak{P}_{\alpha\beta}(X)$ such that $\mathcal{P}(\alpha_n, \beta_n) \prec \mathcal{P}(\alpha_{n+1}, \beta_{n+1})$ for all $n \in \mathbb{N}$ and $\{\alpha_n\}$ and $\{\beta_n\}$ are decreasing sequences converging to zero, which means that $\mathcal{P}(\alpha_n, \beta_n)$ is refining in the sense of (ME6).
- (E2) Moreover, the sequence in (E1) satisfies the following property: For every increasing sequence of positive integers $\{j_k\}$ there exists a dominating sequence of natural numbers $\{n_k\}$, with $n_k > j_k$ for all $k \in \mathbb{N}$, such that $\mathcal{P}^{n_k}(\alpha_k, \beta_k) = \{Q_{1(k, n_k)}, \dots, Q_{m(k, n_k)}\}$ and

$$H(\mathcal{P}^{n_k}(\alpha_k, \beta_k)) = \log(m(k, n_k)) - \sigma(k, n_k), \tag{10}$$

for all $k \in \mathbb{N}$, where $\sigma(k, n_k) > 0$ is bounded for all $(k, n_k) \in \mathbb{N} \times \mathbb{N}$.

Then, $h_{\mathfrak{T}_d}(f) = h(f)$.

Proof. A key element of our proof is the recognition that (E2) expressed in Equation (10) is tantamount to $\mathcal{P}^{n_k}(\alpha_k, \beta_k)$ being nearly equiprobable and that a sufficiently tight open covering \mathcal{U} of $\mathcal{P}(\alpha_k, \beta_k)$

yields a tight open cover \mathcal{U}^{n_k} with respect to $\mathcal{P}^{n_k}(\alpha_k, \beta_k)$ such that $\log N(\mathcal{U}^{n_k}) = \log(m(k, n_k))$. Consequently,

$$\frac{\log N(\mathcal{U}^{n_k})}{n_k} - \frac{H(\mathcal{P}^{n_k}(\alpha_k, \beta_k))}{n_k} := \Delta(\mathcal{U}, k, n_k) = \frac{\sigma(k, n_k)}{n_k} \rightarrow 0 \text{ as } n_k \rightarrow \infty.$$

Now, selecting $\{\mathcal{P}(\alpha_n, \beta_n)\}$ in $\mathfrak{P}_{\alpha\beta}(X)$ as in (E1), it follows from (ME 6) that for every $\epsilon > 0$ there exists a natural number $l = l(\epsilon)$ so large that for any integer $k \geq l(\epsilon)$ there is a $\hat{q}_k \in \mathbb{N}$ such that

$$\left| \frac{H(\mathcal{P}^{q_k}(\alpha_k, \beta_k))}{q_k} - h_\mu(f) \right| < \frac{\epsilon}{3} \tag{11}$$

whenever q_k is an integer greater than \hat{q}_k . Moreover, (E2) implies that for each $k \geq l(\epsilon)$ there are infinitely instances of Equation (11), which we refer to as \tilde{q}_k , satisfying

$$\left| \frac{H(\mathcal{P}^{\tilde{q}_k}(\alpha_k, \beta_k))}{\tilde{q}_k} - h_\mu(f) \right| = \left| \frac{\log(m(k, \tilde{q}_k)) - \sigma(k, \tilde{q}_k)}{\tilde{q}_k} - h_\mu(f) \right| < \frac{\epsilon}{3}. \tag{12}$$

There is, owing to Lemma 1, a decreasing sequence $\{\delta_n\}$ with $\delta_n \downarrow 0$ as $n \rightarrow \infty$ associated to a refining sequence of δ_k -tight open coverings $\{\mathcal{U}_k\}$ of $\{\mathcal{P}(\alpha_k, \beta_k)\}$ that is refining in the sense of (TE 6). Furthermore, the compactness of X , the continuity of $f : X \rightarrow X$, (TE 6), (E 2), Equation (10) and Lemma 1 imply that the $\{\delta_n\}$ for the sequence of tight open coverings can be chosen so that $\mathcal{U}_k^{\tilde{q}_k}$ is minimal for $\mathcal{P}^{\tilde{q}_k}(\alpha_k, \beta_k)$ for some \tilde{q}_k for which Equation (12) obtains,

$$\left| \frac{\log N(\mathcal{U}_k^{\tilde{q}_k})}{\tilde{q}_k} - h_{\mathfrak{T}_d}(f) \right| < \frac{\epsilon}{3} \tag{13}$$

and

$$\frac{\sigma(k, \tilde{q}_k)}{\tilde{q}_k} < \frac{\epsilon}{3}. \tag{14}$$

Accordingly it follows from Equations (12)–(14) that

$$h_{\mathfrak{T}_d}(f) - h_\mu(f) = |h_{\mathfrak{T}_d}(f) - h_\mu(f)|$$

$$\begin{aligned} &= \left| \left(h_{\mathfrak{T}_d}(f) - \frac{\log N(\mathcal{U}_k^{\tilde{q}_k})}{\tilde{q}_k} \right) + \left(\frac{\log N(\mathcal{U}_k^{\tilde{q}_k})}{\tilde{q}_k} - \frac{H(\mathcal{P}^{\tilde{q}_k}(\alpha_k, \beta_k))}{\tilde{q}_k} \right) + \left(\frac{H(\mathcal{P}^{\tilde{q}_k}(\alpha_k, \beta_k))}{\tilde{q}_k} - h_\mu(f) \right) \right| \\ &\leq \left| h_{\mathfrak{T}_d}(f) - \frac{\log N(\mathcal{U}_k^{\tilde{q}_k})}{\tilde{q}_k} \right| + \left| \frac{H(\mathcal{P}^{\tilde{q}_k}(\alpha_k, \beta_k))}{\tilde{q}_k} - h_\mu(f) \right| + \left| \frac{\log N(\mathcal{U}_k^{\tilde{q}_k})}{\tilde{q}_k} - \frac{H(\mathcal{P}^{\tilde{q}_k}(\alpha_k, \beta_k))}{\tilde{q}_k} \right| < \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon. \end{aligned}$$

Hence, as ϵ is arbitrary, the proof is complete. \square

We note that it appears that Theorem 2 can also be proved using the concept of an f -homogeneous invariant measure introduced in [17]. It is actually likely that the hypotheses in the above theorem are equivalent to Bowen’s f -homogeneous invariant Borel probability measure μ property for a discrete dynamical system $f : X \rightarrow X$ on a compact metric space, which is defined as follows: For every $\epsilon > 0$, there exist $\delta, c > 0$ such that

$$\mu \left(\bigcap_{k=0}^{n-1} f^{-k} \left(B_\delta \left(f^k(y) \right) \right) \right) \leq c \mu \left(\bigcap_{k=0}^{n-1} f^{-k} \left(B_\epsilon \left(f^k(x) \right) \right) \right)$$

for all $n \geq 0$ and $x, y \in X$, where B_δ and B_ϵ are the standard δ and ϵ balls, respectively, for the metric d generating the topology \mathfrak{T}_d on X . He proved that if this property holds, then

$$h_{\mathfrak{T}_d}(f) = h_\mu(f) = \lim_{\epsilon \rightarrow 0} \left\{ \limsup_{n \rightarrow \infty} \left[-\frac{1}{n} \log \mu \left(\bigcap_{k=0}^{n-1} f^{-k} \left(B_\epsilon \left(f^k(y) \right) \right) \right) \right] \right\}$$

In the same vein, it is likely that Theorem 2 can be employed to prove the equivalence of the metric and topological entropies for a DMDS comprising a continuous automorphism of a compact separable topological group leaving the Haar measure invariant. Another thing worth noting is the strong indications that this theorem is apt to have analogs for several variants of entropy, such as relative entropy, as well as some extensions to the noncompact versions of entropy in [18–21,29,32,33] and these seem like they might be an interesting topics for future research.

3.3. Examples of Topological and Metric Entropy Equivalence

We mentioned the case of automorphisms on compact, separable topological groups and now we want to show some examples that follow from Theorems 1 and 2. Most of these examples have already been covered in the literature, so they are mainly meant for illustrative purposes.

Example 1. Let $f : \mathbb{S}^1 \rightarrow \mathbb{S}^1$ be a rotation that is an irrational multiple of 2π , where \mathbb{S}^1 is the unit circle with the metric d and measure μ both based on the normalized arclength. Then the associated DMDS is ergodic, so Theorem 1 implies that $h_{\mathfrak{T}_e}(\mathbb{S}^1) = h(\mathbb{S}^1)$, which is readily shown to be zero as is that of every rotation of the circle. It is also a simple matter to obtain the same result using Theorem 2. On the other hand, if we chose a rational multiple of 2π rotation, the DMDS is no longer ergodic, so Theorem 1 cannot be used, but the result can be obtained by employing Theorem 2.

Example 2. Consider the standard tent map $f = \Lambda : [0, 1] \rightarrow [0, 1]$ defined as

$$\Lambda(x) := \begin{cases} 2x, & 0 \leq x < 1/2 \\ 2(1-x), & 1/2 \leq x \leq 1 \end{cases} .$$

combining this with the Lebesgue measure space on the unit interval and the euclidean topology \mathfrak{T}_e , we obtain a DMDS on the compact unit interval. It is easy to see that by successively bisecting the unit interval, we obtain a refining sequence of partitions in satisfying Theorem 2 for all $\{n_k\}$, from which we conclude that $h_{\mathfrak{T}_e}(\Lambda) = h(\Lambda) = \log 2$.

Example 3. The same reasoning shows that the $2x \pmod{1}$ map $f_{2/1} : [0, 1] \rightarrow [0, 1]$ defined as

$$f_{2/1}(x) := \begin{cases} 2x, & 0 \leq x < 1/2 \\ 2x - 1, & 1/2 \leq x \leq 1 \end{cases} ' .$$

with respect to the Lebesgue measure and the euclidean topology \mathfrak{T}_e , is case where $h_{\mathfrak{T}_e}(f_{2/1}) = h(f_{2/1}) = \log 2$.

Example 4. The previous example can be readily generalized to all $mx \pmod{1}$ maps $f_{m/1} : [0, 1] \rightarrow [0, 1]$ for $m \in \mathbb{N}, m \geq 2$, defined as

$$f_{m/1}(x) := \begin{cases} mx, & 0 \leq x < 1/m \\ mx - 1, & 1/m \leq x < 2/m \\ \dots, & 2/m \leq x < (m-1)/m \\ mx - (m-1), & (m-1)/m \leq x \leq 1 \end{cases} ' .$$

with the same measure and topological structure above for the $2x \pmod{1}$ map. The result is $h_{\mathfrak{T}_e}(f_{m/1}) = h(f_{m/1}) = \log m$.

Example 5. We can neatly recast the $mx \pmod{1}$ map examples as smooth maps on a smooth manifold and extend them to higher dimensions with ease. To begin, we again let \mathbb{S}^1 be the unit circle in the complex plane \mathbb{C} with the standard arclength measure μ normalized by a factor of $1/2\pi$, so that it is a probability measure, with the usual Riemannian arclength metric d also scaled by a factor of $1/2\pi$. Then, the $mx \pmod{1}$ maps can be identified with the smooth maps $z \rightarrow z^m$ restricted to the unit circle; namely, $F_m : \mathbb{S}^1 \rightarrow \mathbb{S}^1$ defined by $(z^m)|_{\mathbb{S}^1}$ so that $F_m(e^{i\theta}) := e^{im\theta}$, and this combination defines a DMDS on a compact space. By mimicking the construction in the previous example, it can be readily shown the Theorem 2 implies that $h_{\mathfrak{s}_e}(F_m) = h(F_m) = \log m$. Similarly, given an n -tuple $m = (m_1, \dots, m_n) \in \mathbb{N}^n$, we define the smooth self map of the n -torus $\Phi_m : \mathbb{T}^n \rightarrow \mathbb{T}^n$ by

$$\Phi_m(e^{i\theta_1}, \dots, e^{i\theta_n}) := (e^{im_1\theta_1}, \dots, e^{im_n\theta_n}),$$

where as usual, \mathbb{T}^n is just the cartesian product of n copies of the unit circle with the product topology and product measure associated to the single circle. This then comprises a smooth DMDS on the compact n -torus, which can be easily shown to satisfy the hypotheses of Theorem 2 and yield $h_{\mathfrak{s}_e}(\Phi_m) = h(\Phi_m) = \log(m_1 \cdots m_n)$.

4. Shannon Entropy as a Special Case of Metric Entropy

We shall prove that Shannon’s entropy may be formulated in the context of a Bernoulli scheme; whence, it is equal to the corresponding metric entropy owing to the Kolmogorov–Sinai entropy theorem [23,25], which is also covered in [12,13,15,24]. What we undertake in this section was more or less observed in the process of the development of K–S entropy, although not actually proved in detail as in what follows. In this regard, the work of Frigg [34] should also be noted. To recast Shannon entropy (described above) as a measurable dynamical system called a *Bernoulli scheme* or *Bernoulli shift*, we define the phase space X as the set of all bi-infinite sequences of symbols; namely,

$$X := S^{\mathbb{Z}} := \{\zeta : \zeta : \mathbb{Z} \rightarrow S\} = \{(\dots, \zeta_{-2}, \zeta_{-1}, \zeta_0, \zeta_1, \zeta_2, \dots) : \zeta_k \in S = \{s_1, \dots, s_n\} \forall k \in \mathbb{Z}\},$$

and the (bijective) map $\mathfrak{s} : X \rightarrow X$ to be the (Bernoulli) shift

$$\mathfrak{s}(\zeta)(k) = \zeta(k + 1)$$

for all $k \in \mathbb{Z}$ together with the following probability distribution (p_1, \dots, p_n) for the symbol set S , where $p_k := p(s_k) > 0$ for all $k \in \mathbb{Z}$ and $p_1 + p_2 + \dots + p_n = 1$. This definition fits rather nicely with the information foundation of Shannon entropy inasmuch as it corresponds to reading the string of symbols from left to right one at a time.

To complete the reformulation as a measurable dynamical system, it remains to define a σ -algebra of μ -measurable subsets of X , denoted as \mathcal{M} , where μ is an \mathfrak{s} -invariant probability measure on X . This can be done in many ways, so the trick is to find a definition that yields a metric entropy equal to the Shannon entropy. Toward this end, we start by defining *cylinder sets* of the form

$$C(F, \psi) := \{\zeta \in X : \zeta|_F = \psi\},$$

where F is a finite set of integers, $\psi \in S^F$ and $\zeta|_F$ is the restriction of ζ to $F \subset \mathbb{Z}$, and let \mathcal{C} be the collection of all cylinder sets together with the null set and all of X .

Next, we introduce the set function $\mu_0 : \mathcal{C} \rightarrow \mathbb{R}$ defined as follows:

$$\mu_0(E) := \begin{cases} 0, & E = \emptyset \\ 1, & E = X \\ \prod_{k \in F} p(\psi(k)), & E = C(F, \psi) \end{cases}.$$

Now, there is a theorem of Kolmogorov [22] (see also, for example [24], p. 628) of a rather technical nature stating that the above set function can be uniquely extended to a complete probability measure $\mu : \mathcal{M} \rightarrow \mathbb{R}$, called the *product measure determined by the distribution* (p_1, \dots, p_n) , where \mathcal{M} is the smallest σ -algebra containing \mathcal{C} and $p_k := p(s_k)$ for all $1 \leq k \leq n$.

Finally, to show that this measure is invariant under the shift map, it suffices to verify this for cylinder sets, and this can readily be seen by first observing that

$$\mathfrak{s}^{-1}(C(F, \psi)) = C(\tilde{F}, \tilde{\psi}),$$

where $\tilde{F} := F + 1 := \{k + 1 : k \in F\}$ and $\tilde{\psi} : \tilde{F} \rightarrow S$ with $\tilde{\psi}(k + 1) = \psi(k)$ for all $k \in F$. Consequently,

$$\mu(\mathfrak{s}^{-1}(C(F, \psi))) = \mu(C(\tilde{F}, \tilde{\psi})) = \prod_{j \in \tilde{F}} \mu_0(\tilde{\psi}(j)) = \prod_{k \in F} \mu_0(\psi(k)) = \mu(C(F, \psi)).$$

Thus we have reformulated the Shannon communication system with alphabet S as the measurable dynamical system $(\mathfrak{s}, X := S^{\mathbb{Z}}, \mathcal{M}, \mu)$, where \mathcal{M} and μ are as defined above, which is called a Bernoulli scheme and often denoted as $B(p_1, \dots, p_n)$. The details necessary to prove that the metric entropy of the Bernoulli scheme equals the Shannon entropy of the corresponding communication system are quite extensive and covered rather neatly and completely in [24], so we shall only try to summarize the results in what we hope is a readily understood fashion, at least from an intuitive standpoint.

First, the Kolmogorov–Sinai entropy theorem can be roughly stated as follows: *Let $\mathcal{D} = (f, X, \mathcal{M}, \mu)$ be a measurable dynamical system with f invertible, as it is in the case of $B(p_1, \dots, p_n)$, such that there is a measurable partition \mathcal{P} for which \mathcal{M} is the smallest σ -algebra containing the union of the unions of all sets of the form*

$$f^{-m}\mathcal{P} \vee f^{-m+1}\mathcal{P} \vee \dots \vee f^{m-1}\mathcal{P} \vee f^m\mathcal{P},$$

as m ranges over the natural numbers \mathbb{N} . Then $h_\mu(f) = H(\mathcal{P}, f)$.

Now, it is not difficult to show (see, e.g., [24]) that the measurable partition

$$\mathcal{P} := \{C_k : 1 \leq k \leq n\},$$

where $C_k := \{\zeta \in X = S^{\mathbb{Z}} : \zeta_0 = s_k\}$, satisfies the requirements of the Kolmogorov–Sinai theorem. This follows primarily from the fact that $\mathfrak{s}^{-m}(C_k) = \{\zeta \in X : \zeta_m = s_k\}$ for all $m \in \mathbb{N}$. Now the entropy of the partition \mathcal{P} is

$$H(\mathcal{P}) = -\sum_{k=1}^n \mu(C_k) \log \mu(C_k) = -\sum_{k=1}^n p_k \log p_k.$$

Moreover, we find that

$$H(\mathcal{P}^m) = -\sum_{k_0, \dots, k_{m-1}} \prod_{j=0}^{m-1} p_{k_j} \log \left(\prod_{j=0}^{m-1} p_{k_j} \right) = -m \sum_{k=1}^n p_k \log p_k.$$

Therefore, it follows from the two expressions above and the entropy theorem of Kolmogorov–Sinai that

$$h_\mu(\mathfrak{s}) = H(\mathcal{P}, \mathfrak{s}) = \lim_{m \rightarrow \infty} \frac{H(\mathcal{P}^m)}{m} = -\sum_{k=1}^n p_k \log p_k,$$

which is the desired result.

A Simple Bernoulli Scheme Example

To illustrate the result just obtained, we shall verify it by direct calculation for a binary Bernoulli scheme. In particular, consider $X := B(p_1, p_2)$, with $S := \{0, 1\}$, in which case it follows readily from

the definitions above that the phase space of the dynamical system comprises all bi-infinite binary sequences; namely,

$$X = 2^{\mathbb{Z}} = \{\zeta : \zeta : \mathbb{Z} \rightarrow \{0, 1\}\} = \{\cdots a_{-2}a_{-1}a_0.a_1a_2a_3 \cdots : a_k \in \{0, 1\} \forall k \in \mathbb{Z}\},$$

and the map $s : X \rightarrow X$ is the shift map defined as $s(\zeta)(k) := \zeta(k + 1)$ or equivalently

$$s(\cdots a_{-2}a_{-1}a_0.a_1a_2a_3 \cdots) := \cdots a_{-2}a_{-1}a_0.a_1a_2a_3 \cdots.$$

We know from the above that for the measurable partition

$$\mathcal{P} := \{C_0 := C(\{0\}, 0), C_1 := C(\{0\}, 1)\},$$

where $C_k := \{\zeta \in X = \{0, 1\}^{\mathbb{Z}} : \zeta_0 = k\}$ for $k = 0, 1$, has the following entropy

$$H(\mathcal{P}) = -p_1 \log p_1 - p_2 \log p_2.$$

Our intention is to show that this is equal to the entropy with respect to the shift map s , that is

$$h_{\mu}(s) = H(\mathcal{P}).$$

Owing to the readily verified fact that every member of $\bigvee_{k=-m}^m s^{-k}\mathcal{P}$ is of the form $C(\mathbb{Z}_{-m}^m, \varphi)$ with $\varphi : \mathbb{Z}_{-m}^m \rightarrow \{0, 1\}$ and the Kolmogorov–Sinai entropy theorem (see e.g., [24]), it suffices to prove that every cylinder $C(F, \psi)$ satisfies

$$C(F, \psi) \subset \bigcup \{C(\mathbb{Z}_{-m}^m, \varphi) : \varphi : \mathbb{Z}_{-m}^m \rightarrow \{0, 1\}\},$$

where $\mathbb{Z}_{-m}^m := \{l \in \mathbb{Z} : |l| \leq m, m \in \mathbb{N}\}$. This is obvious since F is finite, which proves the desired result.

5. Concluding Remarks

After a brief review of the main features and fundamental properties of topological, metric (Kolmogorov–Sinai) and Shannon entropy, we embarked on our effort to establish dynamical systems theory as their principal connective thread. We began with comparisons between the topological entropy, which may be considered the most general manifestation of the multifarious forms, and the metric entropy, perhaps the most real-world applicable of the embodiments of entropy. For example, we recounted the variational principle that the topological entropy of a discrete dynamical system is the supremum of the metric entropies over all possible corresponding discrete measurable dynamical systems, which naturally leads to the question of whether or not the supremum is actually assumed. In that vein, we proved our main theorem giving sufficient conditions for the equality of the topological and Kolmogorov–Sinai entropies and provided several illustrative examples. Finally, we showed that Shannon’s information entropy is a special case of metric entropy, inasmuch as a dynamical information system can be identified with a Bernoulli scheme for which the Kolmogorov–Sinai entropy theorem provides a formula for the entropy identical to that of Shannon. In summary then, our project aimed at providing a rigorous underlying dynamical system theme for several of the more important entropy definitions.

Author Contributions: Both authors contributed equally to the research presented, with R.A. focusing mainly on metric and Shannon entropy and D.B. working principally on topological and metric entropy.

Funding: This research received no external funding.

Acknowledgments: The authors would like to thank John Tavantzis for many useful, detailed discussions of the research in this paper. In addition, thanks are due Anatolij Prykarpatski and Jörn Dunkel for some interesting conversations and insightful suggestions that led us to a better understanding of entropy in several of its many manifestations and improvement of the quality of the work presented here. Thanks are also due Vaughn College

and the Center for Applied Mathematics and Statistics for some very helpful support. Finally, the authors thank the reviewers for their very helpful suggestions and insightful constructive criticism, which substantially improved the original version of this manuscript. In this regard, comments concerning the Jewett–Krieger theorem were especially useful.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Robinson, D. Entropy and uncertainty. *Entropy* **2008**, *10*, 493–506. [[CrossRef](#)]
2. Kirwan, A., Jr. Quantum and ecosystem entropies. *Entropy* **2008**, *10*, 58–70. [[CrossRef](#)]
3. Denker, M.; Grillenberger, C.; Sygmond, K. *Ergodic Theory on Compact Spaces*; Lecture Notes in Math.; Springer: Berlin, Germany, 1976; Volume 527, pp. 300–309.
4. Dinaburg, E. The relation between topological entropy and metric entropy. *Dokl. Akad. Nauk SSSR* **1970**, *100*, 19–22.
5. Dinaburg, E. On the relations among various entropy characteristics of dynamical systems. *Izv. Akad. Nauk SSSR* **1971**, *35*, 324–366. [[CrossRef](#)]
6. Downarowicz, T. *Entropy in Dynamical Systems*; Cambridge University Press: Cambridge, UK, 2011.
7. Frigg, R.; Werndl, C. Entropy—A guide for the perplexed. In *Probabilities in Physics*; Beisbart, C., Hartman, S., Eds.; Oxford University Press: Oxford, UK, 2011.
8. Goodman, T. Relating topological entropy and measure entropy. *Bull. Lond. Math. Soc.* **1971**, *3*, 176–180. [[CrossRef](#)]
9. Goodwyn, L. The product theorem for topological entropy. *Trans. Am. Math. Soc.* **1971**, *158*, 445–452. [[CrossRef](#)]
10. Greven, A.; Keller, G.; Warnecke, G. (Eds.) *Entropy*; Princeton University Press: Princeton, NJ, USA, 2003.
11. Kawan, C. Topological Entropy—A Survey. Available online: https://www.math.uni-augsburg.de/de/prof/appa/Mitarbeiter/ehemalige/christoph_kawan/Publikationen/Entropy_v30.pdf (accessed on 19 September 2019).
12. Katok, A. Fifty years of entropy in dynamics: 1958–2007. *J. Modern Dyn.* **2007**, *1*, 545–596. [[CrossRef](#)]
13. Young, L.-S. Entropy in dynamical systems. In *Entropy*; Greven, A., Keller, G., Warnecke, G., Eds.; Princeton University Press: Princeton, NJ, USA, 2003.
14. Delvenne, J.-C. Category theory for autonomous and networked dynamical systems. *Entropy* **2019**, *21*, 302. [[CrossRef](#)]
15. Katok, A.; Hasselblatt, B. *Introduction to the Modern Theory of Dynamical Systems*; Cambridge University Press: Cambridge, UK, 1995.
16. Adler, R.; Konheim, A.; McAndrew, M. Topological entropy. *Trans. AMS* **1965**, *114*, 309–319. [[CrossRef](#)]
17. Bowen, R. Entropy for group endomorphisms and homogeneous spaces. *Trans. AMS* **1971**, *15*, 401–414. [[CrossRef](#)]
18. Dikranjan, D.; Sanchis, M.; Virili, S. New and old facts about entropy in uniform spaces and topological groups. *Topol. Its Appl.* **2012**, *159*, 1916–1942. [[CrossRef](#)]
19. Bowen, R. Topological entropy for noncompact sets. *Trans. AMS* **1973**, *184*, 125–136. [[CrossRef](#)]
20. Patrão, M. Entropy and its variational principle for non-compact sets. *Ergod. Theory Dyn. Syst.* **2010**, *30*, 1529–1542. [[CrossRef](#)]
21. Walters, P. *An Introduction to Ergodic Theory*; Springer: Berlin/Heidelberg, Germany, 1981.
22. Kolmogorov, A. *Grundbegriffe der Wahrscheinlichkeitsrechnung*; Springer: Berlin/Heidelberg, Germany, 1933; pp. 33–50.
23. Kolmogorov, A. Entropy per unit time as a metric invariant of automorphism. *Dokl. Russ. Acad. Sci.* **1959**, *124*, 754–755.
24. McDonald, J.; Weiss, N. *A Course in Real Analysis*, 2nd ed.; Academic Press: Waltham, MA, USA, 2013; pp. 618–630.
25. Sinai, Y.G. On the notion of entropy of a dynamical system. *Dokl. Russ. Acad. Sci.* **1959**, *124*, 768–771.
26. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [[CrossRef](#)]
27. Shannon, C.; Weaver, W. *The Mathematical Theory of Communication*; University of Illinois Press: Urbana, IL, USA, 1963.

28. Misiurewicz, M. A short proof of the variational principal for a \mathbb{Z}_+^n action on a compact space. *Astérisque* **1976**, *40*, 147–157.
29. Ward, T. Entropy of Compact Groups Automorphisms. Available online: www.academia.edu/1736474 (accessed on 19 September 2019).
30. Jewett, R. The prevalence of uniquely ergodic systems. *J. Math. Mech.* **1970**, *19*, 717–729.
31. Krieger, W. On unique ergodicity. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*; University of California: Berkeley, CA, USA, 1970; pp. 327–346.
32. Liu, L. Pre-image entropy for maps on noncompact topological spaces. *Acta Math. Univ. Comen.* **2013**, *82*, 219–230.
33. Nitecki, Z.; Przytycki, F. Preimage entropy for mappings. *Int. J. Bifur. Chaos* **1999**, *9*, 1815–1840. [[CrossRef](#)]
34. Frigg, R. In what sense is the Kolmogorov–Sinai entropy a measure for chaotic behaviour?—Bridging the gap between dynamical systems theory and communication theory. *Brit. J. Philos. Sci.* **2004**, *55*, 411–434. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).