Entropy-like functionals: conceptual background and some results

Miroslav Katětov

Abstract. We describe a conceptual approach which provides a unified view of various entropy-like functionals on the class of semimetric spaces, endowed with a bounded measure. The entropy E considered in the author's previous articles is modified so as to assume finite values for a fairly wide class of spaces which fail to be totally bounded.

Keywords: entropy-like functionals, Hartley value of a piece of information, moderate E-entropy

Classification: 94A17

In the author's previous articles, various kinds of functionals have been considered which are defined on the class of semimetric spaces equipped with a bounded measure (or on the class of sets equipped with a semimetric only) and which are extensions or generalizations, in some sense, of the Shannon entropy.

In the present article, we describe an approach providing a unified view of these functionals; this is done in Section 1.

Some of the functionals considered in the author's articles do not assume finite values unless the space in question is totally bounded. In Section 2, one of the entropies examined previously, the E-entropy, is modified so that it assumes finite values for a fairly broad class including many spaces which fail to be totally bounded.

Finally, in Section 3, the exact values of two of the entropies considered are obtained for the case of an interval of reals equipped with the Lebesgue measure. The corresponding n-dimensional case remains still open.

1.

In this section, we present a fairly general approach to the concept of information or rather to that of "an individual piece of information". This approach is contained implicitly in some articles of the author; however, it has never been formulated explicitly.

We give a survey of some concepts from [K83]–[K92] in the light of the conceptual approach chosen, adding some concepts not considered in these articles. The section also contains a very short survey of some results from [K83]–[K92].

1.1. It seems that there is no sufficiently general and widely accepted definition of information which could become a base for introducing various kinds of entropies and other measures of information. We choose an approach stressing the complementarity of information and uncertainty, and introduce a definition of a piece of information. As for a definition of information, a possible version is mentioned, but is not developed.

1.2. A basic role will be played by the concept of a field of uncertainty or a field of possible outcomes; or, as we will call it, a field of virtuals, abbreviated V-field. A V-field consists of some entities, real or not, which are considered as possible in some, perhaps rather weak, sense. For instance, a V-field can consist of all actions possible or thinkable in a given situation, or of all actions considered as possible by a certain person.

It may happen that a V-field is given without a precise description of the entities belonging to it. If, however, a V-field is given as a set endowed with a structure, we will speak, as a rule, of a V-space.

We will say that V-fields S_1 and S_2 are of the same sort, if they refer to the situations of the same kind and their elements are of the same sort. This vague description can be made precise in an appropriate way whenever necessary.

1.3. Now we can introduce the following definition.

A piece of information is a pair (S, S'), where S is a V-field and S' is a subfield of S.

This definition admits various generalizations. We can call a piece of information any transformation which assigns $\psi(S) \subset S$ to every $S \in \mathcal{S}$, $\mathcal{S}$ being a collection of V-fields of the same sort. We can also speak of a (multicomponent) piece of information meaning a collection of pieces of information (S, S'). — However, in what follows, a piece of information is, as a rule, simply a couple (S, S').

Now we can define information as follows. First we define, for pieces of information J_1 and J_2, the relation "J_1 entails J_2". Then we proceed by abstraction: information is what is common to the pieces of information J_1 and J_2 such that J_1 entails and is entailed by J_2.

In this way, a concept of information is obtained which includes semantic aspects. There is also a more abstract concept devoid of these aspects and including only structural ones. Namely, we can also consider two pieces of information, $J_1 = (S_1, S_1')$ and $J_2 = (S_2, S_2')$, as equivalent, if there are V-spaces $T_1 \supset S_1$, $T_2 \supset S_2$ and an isomorphism $f : T_1 \to T_2$ such that $f(S_1) = S_2$, $f(S_1') = S_2'$. Then we define the information as what is common to J_1 and J_2, if they are equivalent either in the sense of entailment or in the structural sense just described.

However, we will not use these definitions here, and the word "information" will be used either in its intuitive meaning or in expressions like "amount of information", etc.

1.4. The problem of "measuring" the amount of information or the amount of uncertainty was given considerable attention since the fundamental work [S48] of C.E. Shannon, and even earlier, see, e.g., [H28]. In fact, rather than on measuring pieces of information, the attention was concentrated on measuring the amount of uncertainty or information in what we call V-spaces. Most measures of this kind stem, though some of them only indirectly, from the Shannon entropy $H(p_1, \dots, p_n) = \Sigma(-p_i \log p_i : i = 1, \dots, n)$ of a finite probability space. In what follows, "measures" of this kind will be called entropy-like functionals, without giving a formal definition.

1.5. It is sometimes appropriate to measure the amount of information by means of realvalued functions instead of real numbers. Namely, we can assign a certain $\varphi(\varepsilon, S) \in \overline{R}_+$ to every V-space S from a certain class $\mathcal{S}$ and every $\varepsilon > 0$. This amounts to assigning, to every $S \in \mathcal{S}$, a function $(\varepsilon \mapsto \varphi(\varepsilon, S))$, where $\varepsilon \in (0, \infty)$. A mapping of this kind will be called an $(R \to R)$-valued functional or a fluent-valued functional. A well-known instance of an entropy-like fluent-valued functional is the Kolgomorov ε-entropy, see, e.g., [KT59].

1.6. It will be seen below that entropy-like functionals can be introduced starting from a given "measure" of pieces of information. Conversely, if we have an entropy-like functional φ on a class of V-spaces, then it is often convenient to take $\psi(S, S') = \varphi(S) - \varphi(S')$ for a measure of pieces of information. On the whole, however, there is no definite one-to-one correspondence between entropy-like functionals and "measures" of pieces of information.

It seems that in order to obtain some reasonable measures of the amount of information or of uncertainty, it is necessary to introduce something like the "size" of subspaces T of a V-space S, such as $d(T)$ for $S \in \mathfrak{S}$ or $S \in \mathfrak{W}$ (see 1.7). On the other hand, it is not necessary to have anything like probability measure. However, we get far richer and often deeper theory, if we have both something like diameter and something like probability.

1.7. Recall the following V-spaces which have been examined in [K83]–[K92].

We say that $\langle Q, \varrho \rangle$ is a semimetric space (abbreviated SM-space), if Q is a non-void set and $\varrho : Q \times Q \to R_+$ is a semimetric, i.e. $\varrho(x, x) = 0$, $\varrho(x, y) = \varrho(y, x)$. We say that $\langle Q, \varrho, \mu \rangle$ is a W-space, if $\langle Q, \varrho \rangle$ is an SM-space, μ is a measure on Q, $\mu Q < \infty$, and ϱ is measurable. The class of all SM-spaces will be denoted by $\mathfrak{S}$, that of all W-spaces by $\mathfrak{W}$. If $S = \langle Q, \varrho \rangle \in \mathfrak{S}$ and $T \subset Q$, we put $d(T) = \sup(\varrho(x, y) \,:\, x, y \in T)$. If $S = \langle Q, \varrho, \mu \rangle \in \mathfrak{W}$, we put $wS = \mu Q$, and, for any $T \subset Q$, we define the diameter $d(T)$ to be equal to the infimum of all $t \in \overline{R}_+$ such that $\{(x, y) \in T \times T \,:\, \varrho(x, y) > t\}$ is of measure zero.

1.8. Remark. In the definition of SM- and W-spaces, there appear semimetrics. There are two reasons for the use of this concept, considerably broader than that of a metric. First, in many situations we need functions $\varepsilon * \varrho$ and $\varepsilon \odot \varrho$ (with $\varepsilon > 0$), which are not metrics. They are defined as follows: $(\varepsilon * \varrho)(x, y)$ is equal to 1 if $\varrho(x, y) > \varepsilon$, and to 0 if $\varrho(x, y) \leq \varepsilon$; $\varepsilon \odot \varrho = (\varepsilon * \varrho) \cdot \varrho$. Second, many results on SM- and W-spaces do not depend on whether ϱ is a metric or not.

1.9. There are useful generalizations of the SM- and W-spaces. We present the definitions; these spaces will not occur in this article, though.

We call $\langle Q, d \rangle$ a diametric space, if Q is a non-void set and d is a diameter on Q, i.e. $d : \exp Q \to \overline{R}_+$ and (1) $X \subset Y \subset Q$ implies $d(X) \leq d(Y)$, (2) $d(X) = 0$ whenever X is a singleton. We call $\langle Q, d, \mu \rangle$ a DW-space if $\langle Q, \varrho \rangle$ is a diametric space, μ is a measure on Q, $\mu Q < \infty$, and if $X \subset Y \subset Q$, $\mu(Y \setminus X) = 0$, then $d(Y) = d(X)$.

It seems that many results on SM- and W-spaces remain true for the generalizations just described.

1.10. When examining a V-space from the standpoint of information theory, the following question is crucial: how "expensive" is the identification of elements of the space. More precisely: what is the cost (with respect to information) of systems allowing the identification (up to a prescribed ε) of any given but unknown element of the space. This cost can evidently depend on what is meant by a system.

1.11. An approach to the problem of identification will now be described.

Let a V-space S be given. Assume that we have a semimetric or a diameter (see 1.9) on S. Further, assume that M is a predicate such that $M(x)$ holds for exactly one $x \in S$, that this element x is unknown and that we have no means to decide, for a given y, whether $M(y)$ holds or not. We have to identify x up to a prescribed ε.

We can proceed as follows. Take a predicate $P^{(0)}$ such that it can be decided whether M implies $P^{(0)}$ or non-$P^{(0)}$, i.e., to which of the corresponding sets $S(1)$ and $S(0)$ does x belong. Let $x \in S(i)$, where $i = 0$ or $i = 1$. Choose another suitable predicate $P^{(1)}$, possibly defined on $S(i)$ only, and decide whether M implies $P^{(1)}$ or non-$P^{(1)}$, i.e. to which of the sets $S(i1)$, $S(i0)$ does x belong. After a finite number of such steps, we get sets $S = S(\emptyset) \supset S(i_0) \supset S(i_0 i_1) \supset \cdots \supset S(i_0 i_1 \ldots i_n)$. If the last of these is of diameter $\leq \varepsilon$, we have identified (up to ε) the element x.

In this way, we can get, for every $x \in S$, an identification procedure. However, to characterize the information content or depth of a V-space, we need rather an identification system applicable to every $x \in S$. This can be done in a similar manner. We begin with a certain $P^{(0)}$, then we choose a predicate $P_0^{(1)}$ for $S(0)$ and another one, say $P_1^{(1)}$, for $S(1)$, and proceed in this way for a finite number of steps (this number can be distinct for different "branches"). In this manner, we obtain a family of subspaces $(S(u) : u \in D)$, where $D \subset \{0,1\}^*$ is a finite binary tree (with the root $\emptyset$). As for $S(u)$, the following condition are satisfied: $S(\emptyset) = S$, and if $u, u0, u1$ are in D, then $(S(u0), S(u1))$ is a partition of $S(u)$.

A family $\mathcal{S} = (S(u) : u \in D)$ satisfying these conditions will be called a dyadic expansion of S, see, e.g., [K83, 4.3]. The following notation will be used: $D' = \{u \in D : \{u0, u1\} \subset D\}$, $D'' = D \setminus D'$, $\mathcal{S}'' = (S(v) : v \in D'')$.

1.12. We are going to express quantitatively the "expenses" of an identification (up to ε) procedure of the kind described above. To this end, we need certain valuations of pieces of information (T, T') and binary partitions (T_0, T_1) in an SM- or W-space. These valuations will be considered in 1.13–1.16. Then we will go over to valuations of "branches" of a dyadic expansion and to those of dyadic expansions themselves. Based on these valuations, we obtain various entropy-like functionals.

There is also a reverse procedure of obtaining valuations of pieces of information starting from an entropy-like functional, say φ. However, it will not be considered here. The main reason lies in the fact, already mentioned, that a quite useful valuation ψ is obtained by simply putting $\psi(S, S') = \varphi(S) - \varphi(S')$.

1.13. At first, we consider the valuations in the elementary case of a two-point space. If $S = \langle\{a, b\}, \varrho\rangle \in \mathfrak{S}$, then it is natural to assign the value $\varrho(a, b)$ to $(S, \{a\})$ and to $(S, \{b\})$ as well as to $(\{a\}, \{b\})$. If $S = \langle\{a, b\}, \varrho, \mu\rangle \in \mathfrak{W}$, then we can "forget" μ and introduce the same values as for $S = \langle\{a, b\}, \varrho\rangle \in \mathfrak{S}$. If μ is taken into account, then it seems appropriate to introduce the values $h(S, \{x\}) = d(S) \log(\mu S/\mu\{x\})$, where $x = a$ or $x = b$; expressions of this kind, i.e., containing logarithms, appear already in R.L.V. Hartley's article [H28]. Consequently, $\mu\{a\} \cdot h(S, \{a\}) + \mu\{b\} \cdot h(S, \{b\}) = d(S)H(\mu\{a\}, \mu\{b\})$ is taken as the value of the partition $(\{a\}, \{b\})$.

1.14. Consider the case of an arbitrary SM- or W-space S. To obtain various valuations of pieces of information (S, S') and of binary partitions, we first introduce various kinds of quotients of S or of a subspace $S_1 \leq S$ with respect to a binary partition (S_a, S_b) of S_1.

If $S \in \mathfrak{S}$, we consider only one sort of quotients, which happens to coincide with the quotient in $\mathfrak{S}$ understood as a part of the category of diametric spaces. Namely, we have the E-quotient defined as $\langle\{a, b\}, \sigma\rangle$, where $\sigma(a, b) = E(S_a, S_b) = d(S_a \cup S_b)$.

If $S = \langle Q, \varrho, \mu\rangle$, there are many distinct quotients. We can take, e.g., the E-quotient defined as $\langle\{a, b\}, \sigma, \nu\rangle$, where $\sigma(a, b) = E(S_a, S_b) = d(S_a \cup S_b)$, $\nu\{a\} = \mu S_a$, $\nu\{b\} = \mu S_b$. Or we can take the r-quotient, defined as $\langle\{a, b\}, \sigma_r, \nu\rangle$, where $\sigma_r(a, b) = r(S_a, S_b) = \int \varrho d\,(\mu_a \times \mu_b)/(\mu S_a \cdot \mu S_b)$, μ_a and μ_b being, respectively, the restriction of μ to S_a and S_b.

1.15. The quotients just introduced for $S \in \mathfrak{W}$ are special cases of τ-quotients which we are going to define. Let τ be a gauge functional as introduced in [K83, 3.4]; we do not restate the rather complicated definition, and recall only that E and r are gauge functionals. The τ-quotient of $S_1 \leq S$ with respect to a partition (S_a, S_b) is defined as $\langle\{a, b\}, \sigma_\tau, \nu\rangle$, where $\sigma_\tau(a, b) = \tau(S_a, S_b)$.

1.16. Now we can introduce certain valuations based on gauge functionals.

Let $S \in \mathfrak{W}$ and let τ be a gauge functional. Then $\tau(S', S \setminus S') \log(wS/wS')$ will be denoted by $h_\tau(S, S')$ and will be called the Hartley τ-value of the piece of information (S, S'); $wS_a \cdot h_\tau(S, S_a) + wS_b \cdot h_\tau(S, S_b) = \tau(S_a, S_b)H(wS_a, wS_b)$ will be denoted by $\Gamma_\tau(S_a, S_b)$ and will sometimes be called the Hartley τ-value of the partition (S_a, S_b).

In addition, we introduce Hartley $(\tau, 0)$-value $h_{\tau,0}$ of (S, S') and $\Gamma_{\tau,0}$-values of partitions: $h_{\tau,0}(S, S') = \tau(S', S \setminus S')$, $\Gamma_{\tau,0}(S_a, S_b) = \tau(S_a, S_b)$. In particular, we have $h_E(S, S') = d(S) \log(wS/wS')$, $h_{E,0}(S, S') = d(S)$, $\Gamma_E(S_a, S_b) = d(S_a \cup S_b)H(wS_a, wS_b)$, $\Gamma_{E,0}(S_a, S_b) = d(S_a \cup S_b)$.

If $S \in \mathfrak{S}$, then we introduce only $(E, 0)$-values. Namely, we put $h_{E,0}(S, S') = d(S)$, $\Gamma_{E,0}(S_a, S_b) = d(S_a \cup S_b)$.

Remark. Observe that E, defined for $\mathfrak{S}$, is a functional possessing, with respect to $\mathfrak{S}$, properties analogous to that of gauge functionals. Besides E, there is at least one other functional of this kind, namely d, defined by $d(S_a, S_b) = \sup(\varrho(x, y) : x \in S_a, y \in S_b)$. We could define, for such functionals τ, the corresponding τ-values, etc., as follows: $h_{\tau,0}(S, S') = \tau(S', S \setminus S')$, $\Gamma_{\tau,0}(S_a, S_b) = \tau(S_a, S_b)$.

1.17. From Hartley values h_τ and $h_{\tau,0}$, we obtain valuations of "branches" of a dyadic expansion and of dyadic expansions themselves.

Let h stand for h_τ or $h_{\tau,0}$. Let $\mathcal{S} = (S(u) : u \in D)$ be a dyadic expansion of S, $S \in \mathfrak{W}$ or $S \in \mathfrak{S}$. Then, for every $v \in D''$, the sum of all $h(S(u), S(ui))$, where $u \prec v$, $ui \prec v$, will be denoted by $\Gamma(\mathcal{S}, v)$ or, more precisely, by $\Gamma_\tau(\mathcal{S}, v)$ or $\Gamma_{\tau,0}(\mathcal{S}, v)$. Observe that if $S \in \mathfrak{S}$, then only $\Gamma_{E,0}(\mathcal{S}, v)$ are defined.

The weighted sum of $\Gamma(\mathcal{S}, v)$, i.e., $\Sigma(\Gamma(\mathcal{S}, v) \cdot wS(v) : v \in D'')$, will be denoted by $\Gamma(\mathcal{S})$ or, more precisely, by $\Gamma_\tau(\mathcal{S})$ or $\Gamma_{\tau,0}(\mathcal{S})$; it will sometimes be called the Γ_τ-value (the $\Gamma_{\tau,0}$-value) of $\mathcal{S}$. The maximum of all $\Gamma(S, v)$, $v \in S''$, will be denoted by $\overline{\Gamma}(\mathcal{S})$ ($\overline{\Gamma}_\tau(\mathcal{S})$ if $h = h_\tau$, and $\overline{\Gamma}_{\tau,0}(\mathcal{S})$ if $h = h_{\tau,0}$). It will be called the $\overline{\Gamma}_\tau$-value ($\overline{\Gamma}_{\tau,0}$-value) of $\mathcal{S}$. Again, only $\overline{\Gamma}_{E,0}(\mathcal{S})$ is defined if $S \in \mathfrak{S}$.

It is easy to see that we have $\Gamma_E(\mathcal{S}) = \Sigma(d(S(u))H(wS(u0), wS(u1)) : u \in D')$, the expression widely used in [K83] and [K85]. Observe that $\overline{\Gamma}_{E,0}(\mathcal{S})$ is equal to the largest of all sums $\Sigma(d(S(u)) : u \in D', u \prec v)$, $v \in D''$.

Let us note that the functionals Γ_τ have been used extensively in [K83], whereas Hartley values $h_\tau, h_{\tau,0}$ did not occur in the author's articles.

1.18. From $\Gamma(\mathcal{S})$ and $\overline{\Gamma}(\mathcal{S})$, we obtain entropy-like functionals on $\mathfrak{W}$ and on $\mathfrak{S}$ by a limit transition we are going to describe.

Let F be a functional on the class of all dyadic expansions of spaces from $\mathfrak{P}$, where $\mathfrak{P} = \mathfrak{W}$ or $\mathfrak{P} = \mathfrak{S}$. If $S \in \mathfrak{P}$ and $\varepsilon > 0$, let $\varphi_F(\varepsilon, S)$ or $[F](\varepsilon, S)$ denote the infimum of all $F(\mathcal{S})$, where $\mathcal{S} = (S_u : u \in D)$ is a dyadic expansion of S and $d(S_v) \le \varepsilon$ for all $v \in D''$. Put $[F](S) = \sup(\varphi_F(\varepsilon, S) : \varepsilon > 0)$.

In this way, we obtain various entropy-like functionals. If $F = \Gamma_\tau$, then we get $[F] = C^*_\tau$, the functional introduced, in a different manner, in [K83]. This includes the cases $\tau = E$, $\tau = r$; C^*_E is also denoted by E^* (or by E, see 1.21 below) and C^*_r is often denoted by C^*. If $F = \Gamma_{E,0}$, we obtain the functional λ considered in [K90]. If $F = \overline{\Gamma}_{E,0}$ we get the functional δ defined on $\mathfrak{W}$ and on $\mathfrak{S}$; this functional has also been introduced in [K90]. Observe that λ and δ have been introduced in [K90] in a different but equivalent way.

1.19. The functionals $\varphi_F(\varepsilon, S)$ from 1.18 also yield fluent-valued functionals. For $S \in \mathfrak{W}$ (or $S \in \mathfrak{S}$, as the case may be), let $\Phi_F(S)$ denote the realvalued (including ∞) function $\varepsilon \mapsto \varphi_F(\varepsilon, S)$. The functional obtained will be denoted by Φ_F or $\Phi[F]$, e.g., $\Phi[\Gamma_\tau]$. Clearly, $[F](S) = \sup \Phi[F](S)$ for every S.

The fluent-valued functionals $\Phi[F]$ are of real interest only if $[F](S) = \infty$.

1.20. There is another manner of obtaining fluent-valued functionals.

If $S = \langle Q, \varrho, \mu \rangle \in \mathfrak{W}$ or $S = \langle Q, \varrho \rangle \in \mathfrak{S}$ and $\varepsilon > 0$, let $\varepsilon * S$ denote the space $\langle Q, \varepsilon * \varrho, \mu \rangle$ or $\langle Q, \varepsilon * \varrho \rangle$, respectively; for $\varepsilon * \varrho$ see 1.8. If φ is an entropy-like functional on $\mathfrak{W}$ or on $\mathfrak{S}$, then G_φ will denote the fluent-valued functional $S \mapsto (\varepsilon \mapsto \varphi(\varepsilon * S))$.

In this way, we get the following fluent-valued functionals. Starting from E, we get a functional which is equivalent to the epsilon entropy examined in [PRR67]; see [K86a]. Starting from the functional (on $\mathfrak{S}$) corresponding to $\overline{\Gamma}_{E,0}$, i.e. from δ, we obtain a functional coinciding, in essence, with Kolgomorov ε-entropy examined, e.g., in [KT59].

1.21. Another kind of entropy-like functionals can be obtained by generalizing the concept of a subspace of $S \in \mathfrak{W}$. This generalization leads to a broader concept of a piece of information. Observe that, by definition (1.3), a piece of information is a pair (S, S'), S' being a subfield of S, and therefore this concept is dependent on how subfields are defined.

If $S = \langle Q, \varrho, \mu \rangle \in \mathfrak{W}$, then every $S' \in \mathfrak{W}$ of the form $S' = \langle Q, \varrho, \nu \rangle$, where $\nu \leq \mu$, will be called a subspace of S in the wide sense, abbreviated subspace (w.s.); see, e.g., [K83, 1.22]. If S' is a subspace (w.s.) of S, we write $S' \leq S$. If $S_i = \langle Q, \varrho, \mu_i \rangle$, we put $S_1 + S_2 = \langle Q, \varrho, \mu_1 + \mu_2 \rangle$. With subspaces defined in this way, we get extended notions of dyadic expansions, Hartley τ-values, etc.; observe that, in the definition of a dyadic expansion $\mathcal{S} = (S_u : u \in D)$ in the wide sense, it is required that $S_u = S_{u0} + S_{u1}$, which is equivalent to $S_{u0} \cap S_{u1} = \emptyset$, $S_{u0} \cup S_{u1} = S_u$, if S_{u0}, S_{u1} are pure. Finally, we obtain, in the manner described in 1.17 and 1.18, entropy-like functionals, in particular those corresponding to C^*_τ; they are denoted by C_τ. We often write E instead of C_E and C instead of C_r.

The functionals C_τ have been examined in [K83] and [K85], and in some subsequent articles. Here we note only that (1) the problem of $C_\tau = C^*_\tau$ has been solved, in affirmative, only for $\tau = E$, see [K90] and [K92]; (2) generalized subspaces of $\langle Q, \varrho, \mu \rangle \in \mathfrak{W}$ correspond, roughly speaking, to measurable fuzzy subsets of Q.

1.22. We give a short survey of some concepts and results from [K83]–[K92].

In [K83] and [K85], W-spaces are examined. Extended Shannon entropies and semi-entropies are introduced. They are defined as realvalued (including ∞) functionals φ on $\mathfrak{W}$, or on a subclass of $\mathfrak{W}$ such that (1) if $S = \langle Q, 1, \mu \rangle \in \mathfrak{W}$ is finite, $\mu Q = 1$, then $\varphi(S)$ is equal to the Shannon entropy $H(\mu(q) : q \in Q)$, (2) some natural conditions are satisfied including fairly weak continuity conditions; for a semi-entropy, the conditions are somewhat weaker than for an entropy. See [K83], in particular 2.19 and 2.26.

Let us note that, in [K83], the name "extended Shannon entropy (semi-entropy) in the broad sense" was used, since the author intended to introduce a more restricted concept later. This has not been done, and we prefer the shorter name given above or simply that of a Shannon functional.

In [K83], a central role is played by certain constructions leading to C^*_τ and C_τ which are shown to be Shannon functionals; in connection with this, both [K83] and [K85] contain a lot of technicalities. The limit transitions yielding these functionals are different from but equivalent to those described above in 1.18; see [K83, 3.1–3.8, 3.15, 3.17 and 4.11].

Let us add that, in [K83] and [K85], the main attention is given to general properties of Shannon functionals, in particular C^*_τ and C_τ, and no applications are envisaged.

1.23. The main results of [K85] are as follows. It is shown that, under certain relatively mild assumptions, there are not too many Shannon functionals; see [K85, 11.4]. Sufficient conditions for $C_\tau(S)$, $C^*_\tau(S)$ to be finite are given; see [K85, 8.40 and 8.43].

It is proved that, under certain not too restrictive conditions, C_τ and C^*_τ are continuous and even Lipschitz (of an order < 1) on the space of all subspaces of an $S \in \mathfrak{W}$. See [K85], Section 9, in particular 9.37.

1.24. In [K86a], the fluent-valued functional mentioned in 1.20, which assigns $(\varepsilon \mapsto E(\varepsilon * S))$ to S, is considered and shown to be equivalent to the epsilon entropy in the sense of [PRR67].

1.25. In [K86b] and [K87], various kinds of dimension of W-spaces are examined, extending and modifying the concept considered by A. Rényi, see, e.g., [R59]. These dimensions are based on expressions $\varphi(\varepsilon * S)/|\log \varepsilon|$, where φ is a Shannon functional, and are defined as their limits (exact, upper or lower, as the case may be). The dimensions considered have various interesting properties, but their connection with problems of information is only partial. See, however, 1.26 below.

1.26. The article [K88] concerns the differential entropy. As it is well known, this concept is highly counter-intuitive, if considered as an actual measure of information, since, e.g., the differential entropy $\int(-f(x)\log f(x))\,dx$ of a probability measure on R with density f can assume arbitrary negative values. It is shown that the differential entropy, defined in a fairly general manner, coincides, roughly speaking, with the limit of $E(\varepsilon * S) - |\log \varepsilon| \cdot \dim S$, $\dim S$ being an appropriate dimension of $S \in \mathfrak{W}$.

1.27. In [K90] and [K92], the functionals δ and λ (see 1.18) and also E are examined. They are introduced by means of an extended and modified concept of a code. Roughly speaking, elements of a W-space are coded by finite sequences of elements (i, t), where $i = 0, 1$, and $t \in R_+$ represents the "length" of (i, t). — Observe that the functional E from [K90] coincides with $E = C_E$ and $E^* = C_E^*$ introduced earlier in a different manner.

The articles also contain characterization theorems. One of them: E is the largest of all functionals φ on $\mathfrak{W}$ such that, for all $S \in \mathfrak{W}$, we have

(1) $\varphi S = 0$ whenever $d(S) = 0$,
(2) $\varphi S = \sup(\varphi(\varepsilon \odot S) : \varepsilon > 0)$,
(3) $\varphi S \leq d(S)H(wS_0, wS_1) + E(S_0) + E(S_1)$ for all partitions (S_0, S_1) of S.

1.28. We conclude this section with some general remarks. The concept of a piece of information can be extended in various ways. We can consider "multicomponent" pieces of information; see 1.3. We can introduce various operations with pieces of information, such as conjuction and disjunction, possibly infinite. A change of the structure of a V-space, e.g. a transition from a probability measure to another can be considered as a piece of information.

1.29. The following extension of the concept of a piece of information seems to be important and, in a sense, indispensable. It is, however, connected with serious conceptual and technical difficulties.

We can define a piece of information, in an extended sense, as an arbitrary transition from a V-field S to a V-field S' of the same sort; in this broader setting, $S' \subset S$ is not required.

An example: we have a V-field S of possible actions, and we learn that certain other actions, not contained in S, are also possible. These actions together with those in S form a V-field $S' \supset S$. Clearly, there are good reasons for calling (S, S') a piece of information.

1.30. The semantic aspect of information is respected, in principle, in the concept of a piece of information introduced here. However, in a theory of information, with a sufficient range, many other aspects should be reflected, among them such as, e.g., engrams of information. Such aspects lie, of course, outside the scope of the present article.

2.

In this section, the entropy $\overline{E}$, a modification of E, is introduced. At first, we recall some notation and definitions.

2.1. If μ is a measure on Q, $f : Q \to R_+$ is $\overline{\mu}$-measurable non-negative, and $\int f\,d\mu < \infty$, then $f \cdot \mu$ will denote the measure μ_1 defined by $\mu_1(X) = \int_X f\,d\mu$. If $S = \langle Q, \varrho, \mu \rangle \in \mathfrak{W}$, then $f \cdot S$ denotes the space $\langle Q, \varrho, f \cdot \mu \rangle$. For the definition of a subspace (in the wide sense) of a W-space see 1.21. Recall that if $S_1 = \langle Q, \varrho, \mu_1 \rangle$ is a subspace (w.s) of $S = \langle Q, \varrho, \mu \rangle$, then there is a $\overline{\mu}$-measurable f such that $\mu_1 = f \cdot \mu$, $S_1 = f \cdot S$. If μ_1 is equal to the restriction of μ to a $\overline{\mu}$-measurable set $X \subset Q$, then we write $S_1 = X \cdot S$ and call S_1 pure. If $S_1 = f_1 \cdot S$ and $S_2 = f_2 \cdot S$ are subspaces of $S \in \mathfrak{W}$, then the sum $S_1 + S_2$ is defined by $S_1 + S_2 = (f_1 + f_2) \cdot S$.

2.2. Recall that $\mathcal{U} = (U_t : t \in T)$, where T is finite, is called a partition in the wide sense, abbreviated partition (w.s.), of $S \in \mathfrak{W}$, if all U_t are subspaces (w.s.) of S and $\Sigma(U_t : t \in T) = S$. — If all U_t are pure, $\mathcal{U}$ is called pure.

2.3. For dyadic expansions $\mathcal{S} = (S_u : u \in D)$ see 1.11 and 1.21. — If S_u are subspaces (w.s.) of $S = S_\emptyset$, then we will call $\mathcal{S}$ a dyadic expansion (w.s.); if S_u are pure, $\mathcal{S}$ will be called a pure dyadic expansion.

2.4. For definitions of C^*_E, C^*_r, C_E, C_r, see 1.18 and 1.21. See also, e.g., [K92, 2.24, 2.28, and 2.13] (for a definition of E by means of codes). We often write C^* and C instead of C^*_r and C_r. — Since $E^*(S)$ and $E(S)$ coincide for every $S \in \mathfrak{W}$ and are equal to $\hat{E}(S)$ introduced in [K90, 2.13], we denote their common value simply by $E(S)$ and call $E(S)$ the entropy, or the E-entropy, of S.

2.5. Notation. Let $S = \langle Q, \varrho, \mu \rangle \in \mathfrak{W}$. If S_1 and S_2 are subspaces (w.s.) of S, $S_i = f_i \cdot S$, then we put $\mathrm{md}(S_1, S_2) = \int |f_1 - f_2|\,d\mu$; see [K85, 9.12 and 9.13.2]. If $\mathrm{md}(S_n, S) \to 0$ for $n \to \infty$, we write $S_n \to S$. The collection of all subspaces (w.s.) and of all pure subspaces of S will be denoted, respectively, by $\exp S$ and by $\exp^* S$.

2.6. Definition. If $P \in \mathfrak{W}$, then a functional φ defined on $\exp P$ will be called <u>continuous from below</u> on P, if $\varphi(S_n) \to \varphi(S)$ whenever $S_n \leq S$, $n \in N$, $S_n \to S$ and S, S_n are in $\exp P$.

2.7. Recall that a W-space S is called totally bounded, if (1) $d(S) < \infty$, (2) for every $\varepsilon > 0$, there is a partition $(U_t : t \in T)$ of S such that $d(U_t) \leq \varepsilon$ for all $t \in T$.

2.8. Fact. Let $S \in \mathfrak{W}$ be totally bounded. Let $\varepsilon > 0$, let $m \in N$ and let there exist a partition $(U_t : t \in T)$ of S such that $|T| \le 2^m$ and $d(U_t) < \varepsilon$ for all $t \in T$. Then $E(\varepsilon \odot S) \le m \cdot wS \cdot d(S)$.

See [K92, 6.17]. For $\varepsilon \odot S$, see 1.8.

2.9. Fact. If (S_0, S_1) is a partition (w.s.) of $S \in \mathfrak{W}$, then $E(S) \le d(S)H(wS_0, wS_1) + E(S_0) + E(S_1)$. — See [K92, 7.11].

2.10. Proposition. *Let P be a totally bounded W-space. Then the entropy E is continuous from below on P.*

Proof: Let $S \le P$, $S_n \le S$, $n \in N$, and let $S_n \to S$. We will suppose that, for some subsequence $(T_n : n \in N)$ of (S_n), we have $\sup(E(T_n)) \le b < E(S)$ and derive a contradiction.

We can assume $wP = 1$, $d(P) = 1$. Let $\varepsilon > 0$, $\varepsilon < 1$. Let m be an integer such that there is a partition of P consisting of 2^m sets of diameter $\le \varepsilon$. Choose $p > 0$ such that $b + 2p < E(S)$, and choose $c > 0$ such that $H(c, 1-c) < p$, $mc < p$.

Choose n such that $w(S - T_n) < c$ and put $U = S - T_n$. By 2.9, we have, for every $\varepsilon > 0$,

$$E(\varepsilon \odot S) \le d(S)H(wT_n, wU) + E(\varepsilon \odot T_n) + E(\varepsilon \odot U),$$

hence

$$E(\varepsilon \odot S) \le d(S)H(1-c, c) + b + E(\varepsilon \odot U).$$

Since, by 2.8, $E(\varepsilon \odot U) \le m \cdot wU \cdot d(U)$, we get

$$E(\varepsilon \odot S) \le d(S)H(1-c, c) + b + mc \cdot d(S).$$

This implies $E(\varepsilon \odot S) \le b + 2p$, hence, due to $E(S) = \sup(E(\varepsilon \odot S) : \varepsilon > 0)$, $E(S) \le b + 2p$, which is a contradiction. □

2.11. Definition. Let $S \in \mathfrak{W}$. The supremum of all $\inf(E(T) : T \le S, w(S-T) \le \varepsilon)$, $\varepsilon > 0$, will be denoted by $\overline{E}(S)$ and will be called the $\overline{E}$-entropy (or the <u>moderate entropy</u>) of S.

2.12. Fact. If $S \in \mathfrak{W}$, then $\overline{E}(S) \le E(S)$. If $S \in \mathfrak{W}$ is totally bounded, then $\overline{E}(S) = E(S)$.

Proof: The second assertion follows easily from 2.10. □

2.13. Fact. If $S \in \mathfrak{W}$ is not totally bounded, then $E(S) = \infty$. — Remark. In Section 3, there are examples of W-spaces with $d(S) = \infty$ and $\overline{E}(S) < \infty$.

2.14. Proposition. *Let P be a W-space. Then the moderate entropy $\overline{E}$ is continuous from below on P.*

Proof: Suppose the assertion is not true. Then there are S_n and $S \le P$ such that $S_n \le S$, $S_n \to S$, but $\overline{E}(S_n) < b < \overline{E}(S)$ for all $n \in N$. Let $b < b' < \overline{E}(S)$. For every $n \in N$, there is a space $T_n \le S_n$ such that $w(S_n - T_n) \le 1/n$ and $E(T_n) < b'$. We have $T_n \le S$, $T_n \to S$, $\sup(E(T_n) : n \in N) \le b'$. This implies $\overline{E}(S) \le b'$, which is a contradiction. □

2.15. Fact. Let $S \in \mathfrak{W}$. Then (1) $\lim E(T_n) \geq \overline{E}(S)$ whenever $T_n \leq S$, $T_n \to S$; (2) there are $T_n \leq S$, $n \in N$, such that $T_n \to S$ and $E(T_n) \to \overline{E}(S)$.

Proof: I. Suppose (1) does not hold. Then there are $T_n \leq S$ such that $T_n \to S$ and $E(T_n) \to b < \overline{E}(S)$. Let $b < c < \overline{E}(S)$. Then, clearly, all the infima in 2.11 are less than c, hence $\overline{E}(S) \leq c$, which is a contradiction. — II. Put $b_n = \inf(E(T) : T \leq S, w(S-T) \leq n^{-1})$. If $b_k = \infty$ for some k, then $\overline{E}(S) = \infty$ and $E(T_n) \to \infty$ whenever $T_n \leq S$, $T_n \to S$. If $b_n < \infty$ for all $n \in N$, choose c_n such that $b_n < c_n$, $\sup(c_n : n \in N) = \sup(b_n : n \in N) = \overline{E}(S)$. Choose T_n such that $T_n \leq S$, $w(S - T_n) \leq 1/n$, $E(T_n) < c_n$. Clearly, $E(T_n) \to \overline{E}(S)$. □

2.16. Fact. If (S_0, S_1) is a partition (w.s.) of a W-space S, then $\overline{E}(S) \leq d(S)H(wS_0, wS_1) + \overline{E}(S_0) + \overline{E}(S_1)$.

Proof: By 2.15, there are, for $i = 0, 1$, spaces $T_{i,n} \leq S_i$ with $T_{i,n} \to S_i$, $E(T_{i,n}) \to \overline{E}(S_i)$. Put $T_n = T_{0,n} + T_{1,n}$. Clearly, $T_n \leq S$, $T_n \to S$ and therefore, by 2.15, $\lim E(T_n) \geq \overline{E}(S)$. For every $n \in N$, we have $E(T_n) \leq d(T_n)H(wT_{0,n}, wT_{1,n}) + E(T_{0,n}) + E(T_{1,n})$, see [K92, 7.12]. Since $wT_{i,n} \to wS_i$, $E(T_{i,n}) \to \overline{E}(S_i)$, we have $\lim E(T_n) \leq d(S)H(wS_0, wS_1) + \overline{E}(S_0) + \overline{E}(S_1)$, which proves the assertion. □

2.17. Notation. If φ is a functional on $\mathfrak{W}$, then φ^∞ denotes the functional defined as follows: $\varphi^\infty(S) = \varphi(S)$ if S is totally bounded, $\varphi^\infty(S) = \infty$ if not.

2.18. Fact. $E^\infty = E$, $\overline{E}^\infty = E$.

Proof: The first assertion follows from 2.13. As a consequence of 2.10, we have $\overline{E}(S) = E(S)$ whenever S is totally bounded. If S is not totally bounded, then $E(S) = \infty$, by 2.13, $\overline{E}^\infty(S) = \infty$ by definition. □

2.19. Fact. Let a functional φ on $\mathfrak{W}$ satisfy the following condition $(*)$ $\varphi(S) \leq d(S)H(wS_0, wS_1) + \varphi(S_0) + \varphi(S_1)$ for every partition (w.s.) of $S \in \mathfrak{W}$. Then the condition $(*)$ is also satisfied with φ replaced by φ^∞.

Proof: If S is totally bounded, then so are S_0 and S_1 and therefore γ^∞ and φ coincide on S_0, S_1 and S. If $d(S) = \infty$, then the inequality in $(*)$ is trivial. If $d(S) < \infty$ and S is not totally bounded, then either S_0 or S_1 fails to be and therefore either $\varphi^\infty(S_0) = \infty$ or $\varphi^\infty(S_1) = \infty$. □

2.20. In [K92], the following characterization theorem for E has been proved.

The functional E is the largest of all functionals φ on $\mathfrak{W}$ satisfying, for every $S \in \mathfrak{W}$, the following conditions:

(1) $\varphi(S) = 0$ whenever $d(S) = 0$,
(2) $\varphi(S) = \sup(\varphi(\varepsilon \odot S) : \varepsilon > 0)$,
(3) $\varphi(S) \leq d(S)H(wS_0, wS_1) + \varphi(S_0) + \varphi(S_1)$ for all pure partitions (S_0, S_1) of S.

2.21. Characterization theorem for $\overline{E}$. *The functional $\overline{E}$ is the largest of all functionals φ on $\mathfrak{W}$ satisfying, for every $S \in \mathfrak{W}$, the following conditions:*

(1) $\varphi(S) = 0$ *whenever* $d(S) = 0$,

(2) $\varphi^\infty(S) = \sup(\varphi^\infty(\varepsilon \odot S) : \varepsilon > 0)$,
(3) $\varphi(S) \le d(S)H(wS_0, wS_1) + \varphi(S_0) + \varphi(S_1)$ *for every pure partition* (S_0, S_1) *of* S.
(4) φ *is continuous from below on* S.

Proof: I. Evidently, $\overline{E}$ satisfies (1). By 2.20 and 2.18, $\overline{E}$ satisfies (2). By 2.16, $\overline{E}$ satisfies (3). Finally, $\overline{E}$ satisfies (4), by 2.14. — II. Let φ satisfy (1)–(4). Put $\psi = \varphi^\infty$. It is evident that ψ satisfies (1). Clearly, φ satisfies the condition (2) from 2.20. By 2.19, ψ satisfies (3). By 2.20, we obtain $\psi \le E$, hence $\varphi \le E$. Since φ is continuous from below, we have, for every $S \in \mathfrak{W}$, $\inf(\varphi(T) : T \le S, w(S-T) \le \varepsilon) \to \varphi(S)$ for $\varepsilon \to 0$. By the definition of $\overline{E}$, this implies $\varphi(S) \le \overline{E}(S)$. $\square$

2.22. Example. If $P \in \mathfrak{W}$ is totally bounded, then E is continuous from below on P, see 2.10. However, E can fail to be continuous on P.

For $k \in N$, put $A_k = \{n \in N : \exp(k^3) < n \le \exp(k+1)^3\}$, where we write $\exp x$ instead of 2^x. For $n \in A_k$, $k > 0$, put $x_n(n) = 1/k$, $x_n(p) = 0$ for $p \in N$, $p \ne n$; put $x_n = (x_n(p) : p \in N) \in [0,1]^\omega$. For $k \in N$, $k > 0$, put $X_k = \{x_n : n \in A_k\}$; put $X = \bigcup(X_k : k > 0)$. Let ϱ be the metric on X obtained by embedding X into the space $[0,1]^\omega$ with the sup-metric. Clearly, we have $\varrho(x_p, x_q) = 1/u$ whenever $p \in A_u$, $q \in A_v$, $u \le v$, $p \ne q$. Let μ be the measure on X defined by $\mu\{x_n\} = k^{-2}/|A_k|$ for $n \in A_k$, $k > 0$. We have $\mu X_k = k^{-2}$, hence $\mu X < \infty$. Clearly, $P = \langle X, \varrho, \mu\rangle$ is a metric W-space. It is easy to see that P is totally bounded and that $E(X_k \cdot P) \to 1$ for $k \to \infty$. On the other hand $\mu X_k \to 0$ for $k \to \infty$.

2.23. We are going to formulate a condition sufficient for $\overline{E}(S) < \infty$ in the case of $S = \langle R, \varrho, \mu\rangle$, where ϱ is the usual metric. It seems that the result we obtain can be extended, in a modified form, to W-spaces of the form $\langle R^n, \varrho, \mu\rangle$. However, neither this question nor that of weakening the condition we give for $\langle R, \varrho, \mu\rangle$ will be considered here.

2.24. Proposition. *Let* $S = \langle [0,1], \varrho, \mu\rangle \le \mathfrak{W}$, *where* $\varrho(x,y) = |x-y|$, $\mu[0,1] = 1$. *Then* $C(S) \le 1$, $C^*(S) \le 1$, $E(S) \le 2$.

Proof: Put $S_\emptyset = S$. If, for some $u \in \{0,1\}^*$, we have already defined $S_u = T_u \cdot S$, where T_u is an interval with endpoints a_u, b_u, we put $c_u = (a_u + b_u)/2$, $T_{u0} = T_u \cap [a_u, c_u]$, $T_{u1} = T_u \cap (a_u, b_u]$. In this way, we define $S_u = T_u \cdot S$ for all $u \in \{0,1\}^*$. For $n \in N$, $n > 0$, put $D_n = \bigcup(\{0,1\}^k : k \le n)$, $\mathcal{S}_n = (S_u : u \in D_n)$. Clearly, $d(S_v) = 2^{-n}$ for $v \in D''_n$. If $n \in N$, $n > 0$ and $k < n$, we have $\Sigma(H(wS_{u0}, wS_{u1}) : u \in D'_n, |u| = k) = 1$. Hence, for every $n \in N$, $n > 0$, $\Gamma(\mathcal{S}_n) = 1 - 2^{-n-1}$, $E(\mathcal{S}_n) = 2 - 2^{-n}$. This proves the proposition. $\square$

2.25. Notation. The set of all integers is denoted by Z. We put $L(0) = 0$, $L(x) = -x \log x$ for $x > 0$.

2.26. Fact. Let $S \in \mathfrak{W}$. Let $T_n \le S$, $n \in N$, and let $w(S - T_n) \to 0$ for $n \to \infty$. Then $\overline{E}(S) \le \sup(E(T_n) : n \in N)$.

2.27. Proposition. *Let $S = \langle R, \varrho, \mu \rangle$, where $\varrho(x,y) = |x-y|$, be a W-space. For $n \in Z$, put $p_n = \mu[n, n+1]$. If $\Sigma(|n|L(p_n) : n \in Z) < \infty$, then $\overline{E}(S) < \infty$. In particular, $\overline{E}(S) < \infty$ whenever there is a $c > 0$ such that $p_n \le |n|^{-2-c}$ provided $|n|$ is large.*

Proof: For $n \in N$ put $A_n = [-n-1, -n) \cup (n, n+1]$, $B_n = [-n, n]$, $a_n = \mu A_n$, $b_n = \mu B_n$. Put $g_n = E(B_n \cdot S)$. To prove the first assertion, it is sufficient, by 2.26. to show that $\sup(g_n : n \in N) < \infty$. Since $B_{n+1} = B_n \cup A_{n+1}$, we have, by 2.20,

$$g_{n+1} \le 2(n+1)H(b_n, a_{n+1}) + g_n + E(A_{n+1} \cdot S).$$

By 2.24, $E([n, n+1] \cdot S) \le 2p_n$ for all $n \in Z$, and therefore, by 2.20,

$$\begin{aligned} E(A_n \cdot S) &\le 2(n+1)H(p_{-n-1}, p_n) + 2E([-n-1, -n) \cdot S) + 2E((n \cdot n+1] \cdot S), \\ E(A_n \cdot S) &\le 2(n+1)(p_{-n-1} + p_n) + 2(p_{-n-1} + p_n). \end{aligned}$$

We have shown that

$$g_{n+1} \le g_n + (2n+6)(p_{-n-2} + p_{n+1}) + 2(n+1)H(b_n, a_{n+1}).$$

It is easy to see that, for a fixed $x > 0$, we have $H(x,y) \le 2L(y)$ for all y sufficiently small. Since $H(b_n, a_{n+1}) \le H(\mu R, a_{n+1})$ and $a_{n+1} \to 0$ for $n \to \infty$, it follows that $H(b_n, a_{n+1}) \le 2L(a_{n+1})$ for large $n \in N$. Hence, for large n, we have

$$g_{n+1} - g_n \le (2n+6)(p_{-n-2} + p_{n+1}) + 4(n+1)L(a_{n+1}).$$

Since $a_{n+1} \le p_{-n-2} + p_{n+1}$, we have $L(a_{n+1}) \le L(p_{-n-2}) + L(p_{n+1})$. It follows that, for large $|n|$,

$$g_{n+1} - g_n \le (2n+6)(p_{-n-2} + p_{n+1}) + 4(n+1)(L(p_{-n-2}) + L(p_{n+1})).$$

Since $\Sigma(|n|L(p_n) : n \in Z) < \infty$, this implies $\Sigma(g_{n+1} - g_n : n \in N) < \infty$, hence $\sup g_n < \infty$. This proves the first assertion.

If $p_n \le |n|^{-2-c}$ whenever $|n|$ is large, then $|n|L(p_n) \le (2+c)|n|^{-1-c} \cdot \log|n|$ provided $|n|$ is large. Hence $\Sigma(|n|L(p_n) : n \in Z) < \infty$, which proves the second assertion. □

3.

In this section, we obtain the values of $C(S), C^*(S)$ and $E(S)$ for W-spaces S of the form $\langle T, \varrho, \lambda \rangle$, where T is a bounded interval of reals, $\varrho(x,y) = |x-y|$, and λ is the Lebesgue measure. — Note that Proposition 3.11 below, which concerns $C(S)$ and $C^*(S)$, has been already announced; see [K80a, 3.14].

3.1. Notation. If $S = \langle Q, \sigma, \mu \rangle \in \mathfrak{W}$ and $S_i = f_i \cdot S \le S$, $i = 1, 2$, then we put $\hat{r}(S_1, S_2) = \int \varrho f_1 f_2 \, d(\mu \times \mu)$, $r(S_1, S_2) = \hat{r}(S_1, S_2)/(wS_1 \cdot wS_2)$; cf. 1.14. If $S_0 = f_0 \cdot S \subset S$, we put $\hat{r}(S_0) = \hat{r}(S_0, S_0)$, $r(S_0) = \hat{r}(S_0)/(wS_0)^2$. — If $\mathcal{U} = (U_t : t \in T)$ is a partition of $S \in \mathfrak{W}$, we put $d(\mathcal{U}) = \max(d(U_t) : t \in T)$.

3.2. Fact. For every $S \in \mathfrak{W}$, $r(S) \le d(S)$.

3.3. Fact. For every $S \in \mathfrak{W}$, $S_1 \le S$, $S_2 \le S$, we have $\hat{r}(S_1, S_2) = (\hat{r}(S_1 + S_2) - \hat{r}(S_1) - \hat{r}(S_2))/2$.

The proof is elementary and can be omitted.

3.4. Notation. If $T \subset R$ is an interval, we put $J_T = \langle T, \varrho, \lambda \rangle$. We put $J_1 = J_T$, where $T = [0,1]$.

3.5. Lemma. *Let $S \le J_1$. Put $a = wS$. Then $\hat{r}(S) \ge \hat{r}([0,a] \cdot J_1)$.*

Proof: Let $S = f \cdot J_1$. As it is easy to see, we can assume that $f(t) > 0$ for all $t \in [0,1]$. For $x \in [0,1]$ put $\psi(x) = \int_0^x f \, d\lambda$. Clearly, ψ is a bijection of $[0,1]$ onto $[0,a]$. For $u, v \in [0,a]$, put $\sigma(u,v) = |\psi^{-1}u - \psi^{-1}v|$; evidently, σ is a metric on $[0,a]$. It is easy to see that ψ is an isomorphic (i.e., preserving distance and measure) mapping of S onto $U = \langle [0,a], \sigma, \lambda \rangle$. Hence $\hat{r}(S) = \hat{r}(U)$. Since $\varrho \le \sigma$, we have $\hat{r}([0,a] \cdot J_1) = \hat{r}\langle [0,a], \varrho, \lambda \rangle \le \hat{r}(U) = \hat{r}(S)$. $\square$

3.6. Fact. If $S \le J_1$, then $r(S) \ge wS/3$, $\hat{r}(S) \ge (wS)^3/3$.

Proof: By 3.5, $\hat{r}(S) \ge \hat{r}([0,a] \cdot J_1)$, where $a = wS$. By an easy calculation, we get $\hat{r}([0,a] \cdot J_1) = a^3/3$, which proves the assertion. $\square$

3.7. Fact. $H(p,q) \ge 4pq/(p+q)$ whenever $p \in R_+$, $q \in R_+$.

This is well known; see, e.g., [K83, 2.16.1].

3.8. Lemma. *Let $P = \langle Q, \sigma, \mu \rangle \in \mathfrak{W}$ and let $r(T) \ge wT/3$ for every $T \le P$. Let $\mathcal{P} = (P_x : x \in D)$ be a dyadic expansion of P and assume that $wP_x > 0$ for all $x \in D$. Then*

$$\Gamma(\mathcal{P}) \ge (wP)^2 - \Sigma((wP_x)^2;\ x \in D'') - s,$$

where $s = \Sigma(\hat{r}(P_{y0})/wP_y + \hat{r}(P_{y1})/wP_y : y \in D', \{y0, y1\} \subset D'')$.

Proof: For $x \in D$, put $a_x = wP_x$, $b_x = \hat{r}(P_x)$. For $y \in D$, $y \ne \emptyset$, let $\overline{y}$ be defined by $y = \overline{y} \cdot (i)$, $i = 0$ or $i = 1$; put $\overline{a}_y = a_{\overline{y}}$.

By 3.3 and 3.7, we have, for every $x \in D'$,

$$r(P_{x0}, P_{x1}) = (b_x - b_{x0} - b_{x1})/2a_{x0}a_{x1},$$
$$\Gamma(P_{x0}, P_{x1}) \ge H(a_{x0}, a_{x1}) r(P_{x0}, P_{x1}) \ge 2(b_x - b_{x0} - b_{x1})/a_x$$

and therefore

$$\Gamma(\mathcal{P}) \ge 2\Sigma((b_x - b_{x0} - b_{x1})/a_x : x \in D').$$

It is easy to see that the right side is equal to

$$(*) \qquad 2b_\emptyset/a_\emptyset + 2\Sigma(b_x(1/a_x - 1/\overline{a}_x) : x \in D',\ x \ne \emptyset) - t$$

where $t = 2\Sigma(b_y/\overline{a}_y : y \in D'')$.

Since $b_x \ge a_x^3/3$ the expression $(*)$ is not less than

$$2a_\emptyset^2/3 + 2\Sigma(a_x^3(1/a_x - 1/\overline{a}_x)/3 : x \in D',\ x \ne \emptyset) - t,$$

which is easily seen to be equal to

$$2\Sigma((a_x^3 - a_{x0}^3 - a_{x1}^3)/3a_x : x \in D, \{x0, x1\} \subset D')+$$
$$+ \Sigma(a_x^2/3 : x \in D', \{x0, x1\} \subset D'') - t.$$

Consequently, $\Gamma(\mathcal{P}) \geq 2\Sigma(a_{x0}a_{x1} : x \in D') - t$.

Clearly, $2\Sigma(a_{x0}a_{x1} : x \in D') = (wP)^2 - \Sigma((wP_x)^2 : x \in D'')$, $t = 2\Sigma(\hat{r}(P_{y0})/wP_y + \hat{r}(P_{y1})/wP_y : y \in D, \{y0, y1\} \subset D'') = s$. This proves the lemma. □

3.9. Lemma. *Let $P \in \mathfrak{W}$ and let $r(T) \geq wT/3$, $d(T) \geq w(T)$ for every $T \leq P$. Let $\mathcal{P} = (P_x : x \in D)$ be a dyadic expansion of P. Then $\Gamma(\mathcal{P}) \geq (wP)^2 - 3wP \cdot d(\mathcal{P}'')$.*

Proof: Due to $wP_x \leq d(P_x)$, $\Sigma((wP_x)^2 : x \in D'') \leq wP \cdot d(\mathcal{P}'')$. As for s in 3.8, we easily obtain $s \leq 2\Sigma(r(P_z)wP_z : z \in D'') \leq 2d(\mathcal{P}'') \cdot wP$. Consequently, by 3.8, $\Gamma(\mathcal{P}) \geq (wP)^2 - wP \cdot d(\mathcal{P}'') - 2wP \cdot d(\mathcal{P}'')$. □

3.10. Fact. If $P = \langle Q, \sigma, \mu\rangle$ is a bounded W-space, σ is a metric and the topological weight of $\langle Q, \sigma\rangle$ is countable, then $C(P) \leq C^*(P)$.

This is a special case of 8.38 in [K85].

3.11. Proposition. *Let $[a, b] \subset R$ be a bounded interval. The C-entropy and the C^*-entropy of $\langle [a, b], \varrho, \lambda\rangle$ are equal to $(b - a)^2$.*

This follows from 3.6, 3.9, 3.10 and 2.24.

3.12. Lemma. *Let $\mathcal{P} = (P_u : u \in D)$ be a dyadic expansion of J_1; let $wP_u > 0$ for all $u \in D$. Then there exists a pure dyadic expansion $\mathcal{S} = (S_u : u \in D)$ of J_1 such that (1) for every $u \in D$, we have $wS_u = wP_u$, $S_u = A_u \cdot J_1$ for some interval A_u, (2) $E(\mathcal{S}) \leq E(\mathcal{P})$, $\Gamma(\mathcal{S}) \leq \Gamma(\mathcal{P})$.*

Proof: We proceed by induction on $n = |D''|$. If $n = 2$, put $A_0 = [0, wP_0]$, $A_1 = (wP_0, 1]$, $S_i = A_i \cdot J_1$. It is clear that (1) and (2) are satisfied.

Assume that the assertion is valid for $|D''| \leq n$. Let $\mathcal{P} = (P_u : u \in D)$ be a dyadic expansion of P, $|D''| = n + 1$. Choose an $x \in D'$ with $\{x0, x1\} \subset D''$. Put $\hat{D} = D \setminus \{x0, x1\}$, $\hat{\mathcal{P}} = (P_u : u \in \hat{D})$. Since $|\hat{D}''| \leq n$, there is a dyadic expansion $\hat{\mathcal{S}} = (S_u : u \in \hat{D})$ satisfying (1) and (2) with respect to $\hat{\mathcal{P}}$. Let $S_x = A_x \cdot J_1$, where A_x is an interval with endpoints a and b. Since $wP_{x0} + wP_{x1} = wP_x$, $wS_x = wP_x$, $wS_x = b - a$, we have $a + wP_{x0} < b$. Put $c = a + wP_{x0}$, $A_{x0} = A_x \cap [a, c]$, $A_{x1} = A_x \cap (c, b)$, $S_{xi} = A_{xi} \cdot J_1$, $\mathcal{S} = (S_u : u \in D)$. It is easy to see that $\mathcal{S}$ satisfies (1) and (2) with respect to $\mathcal{P}$. □

3.13. Fact. Let $\mathcal{P} = (P_u : u \in D)$ be a dyadic expansion of J_1 and let every P_u be of the form $A_u \cdot J_1$, where A_u is an interval. Then $E(\mathcal{P}) = 2\Gamma(\mathcal{P})$.

Proof: It is easy to show that, for every $u \in D'$, $r(P_{u0}, P_{u1}) = E(P_{u0}, P_{u1})/2 = d(P_u)/2$, hence $d(P_u)H(wP_{u0}, wP_{u1}) = 2\Gamma(P_{u0}, P_{u1})$. This implies $E(\mathcal{P}) = 2\Gamma(\mathcal{P})$. □

3.14. Proposition. *Let $[a,b] \subset R$ be a bounded interval. Then the E-entropy of $\langle [a,b], \varrho, \lambda \rangle$ is equal to $2(b-a)^2$.*

Proof: It is sufficient to show that $E(J_1) = 2$. Put $c = E(J_1)$. By 2.24, $c \leq 2$. By the definition of E, there exists, for every $\varepsilon > 0$, a dyadic expansion $\mathcal{P} = (P_u : u \in D)$ of J_1 such that $d(\mathcal{P}'') \leq \varepsilon$, $E(\mathcal{P}) < c + \varepsilon$. By 3.12, there is a dyadic expansion $\mathcal{S} = (S_u : u \in D)$ of J_1 with $E(\mathcal{S}) \leq E(\mathcal{P})$ such that every S_u is of the form $A_u \cdot S_1$, where A_u is an interval. By 3.13, $E(\mathcal{S}) = 2\Gamma(\mathcal{S})$, hence $\Gamma(\mathcal{S}) < c/2 + \varepsilon/2$. This implies $C^*(J_1) \leq c/2$, which proves, by 3.11, that $c \geq 2$. $\square$

References

[H28] Hartley S.V.L., *Transmission of information*, Bell System Tech. J. **7**, no. 3 (1978), 535–563.

[K80a] Katětov M., *Extensions of the Shannon entropy to semimetrized measure spaces*, Comment. Math. Univ. Carolinae **21** (1980), 171–192.

[K80b] ______, *Correction to "Extensions of the Shannon entropy to semimetrized measure spaces"*, Comment. Math. Univ. Carolinae **21** (1980), 825–830.

[K83] ______, *Extended Shannon entropies I*, Czechoslovak Math. J. **33** (108) (1983), 546–601.

[K85] ______, *Extended Shannon entropies II*, Czechoslovak Math. J. **35** (110) (1985), 565–616.

[K86a] ______, *On extended Shannon entropies and the epsilon entropy*, Comment. Math. Univ. Carolinae **27** (1986), 519–534.

[K86b] ______, *On the Rényi dimension*, Comment. Math. Univ. Carolinae **27** (1986), 741–753.

[K87] ______, *On dimensions of semimetrized measure spaces*, Comment. Math. Univ. Carolinae **28** (1987), 399–411.

[K88] ______, *On the differential and residual entropy*, Comment. Math. Univ. Carolinae **29** (1988), 319–349.

[K90] ______, *On entropy-like functionals and codes for metrized probability spaces I*, Comment. Math. Univ. Carolinae **31** (1990), 49–66.

[K92] ______, *On entropy-like functionals and codes for metrized probability spaces II*, Comment. Math. Univ. Carolinae **33** (1992), 79–95.

[KT59] Kolgomorov A., Tihomirov V., *ε-entropy and ε-capacity of sets in function spaces* (in Russian), Uspehi Mat. Nauk **14**, no. 2 (1959), 3–86.

[PRR67] Posner E.C., Rodemich H., Rumsey H., Jr., *Epsilon entropy of stochastic processes*, Ann. Math. Statist. **38** (1967), 1000–1020.

[R59] Rényi A., *On the dimension and entropy of probability distributions*, Acta Math. Acad. Sci. Hungar. **10** (1959), 193–215.

[S48] Shannon C.E., *A mathematical theory of communication*, Bell System Tech. J. **27** (1948), 375–423, 623–656.

Department of Mathematics and Physics, Charles University, Sokolovská 83, 186 00 Praha 8, Czechoslovakia

(*Received* June 11, 1992)