Department of Applied Analysis
and Computer Science

Technical Report CS-73-19
May 1973

Why the Shannon and Hartley Entropies Are 'Natural'

by

J. Aczél, B. Forte, and C. T. Ng

Why the Shannon and Hartley Entropies Are 'Natural'

by J. Aczél, B. Forte, and C. T. Ng

Address of the authors: Faculty of Mathematics, University of Waterloo, Waterloo, Ontario, N2L 3G1, Canada.

Abstract. The following properties of entropies, as measures of expected information, seem natural. The amount of information expected from an experiment does not change if we add outcomes of zero probability (expansibility). The expected information is symmetric in the (probabilities of the) outcomes. The information expected from a combination of two experiments is less than or equal to the sum of the informations expected from the single experiments (subadditivity); equality holds here if the two experiments are independent (additivity).

In this paper it is shown that linear combinations of the Shannon and Hartley entropies and only these have the above properties. The Shannon and the Hartley entropies are also individually characterized.

Keywords: Entropies, characterization theorems, probability distributions, independence, recursivity, information functions, continuous, concave, inequalities, subadditive, additive, number theoretical functions.

Why the Shannon and Hartley Entropies Are 'Natural'

by J. Aczél, B. Forte, and C. T. Ng (Waterloo, Ontario)

1. The Shannon entropy [see, e.g., Shannon and Weaver (1949)] defined by

$$H_m(p_1,p_2,\ldots,p_m) = -\sum_{j=1}^{m} p_j \log_2 p_j \tag{1}$$

with the convention

$$0 \log_2 0 := 0 \tag{2}$$

on the sets

$$\Gamma_m := \{(p_1,p_2,\ldots,p_m) \mid \sum_{j=1}^{m} p_j = 1;\ p_j \geq 0;\ j = 1,2,\ldots,m\} \quad (m = 2,3,\ldots) \tag{3}$$

of complete probability distributions is, without doubt, the measure of (expected) information which is used most often. Certainly one reason for this is that it has so many properties which we intuitively associate with a measure of expected information.

Characterization theorems characterize the Shannon (and other) entropies by such 'natural' properties. For lists of such characterization theorems see, for instance, Aczél (1968) and Aczél and Daróczy (1974). As pointed out by Aczél (1968), however, most of the properties used in these characterizations are rather explicitly linear. The recursivity [see, a.o., Faddeev (1956), Kendall (1963), Lee (1964), Rényi (1960, 1965), Aczél (1968), Daróczy (1969), and Aczél and Daróczy (1974)]

$$H_m[p(1-q),pq,p_3,\ldots,p_m] - H_{m-1}(p,p_3,\ldots,p_m) = pH_2(1-q,q) \tag{4}$$

$$(q \in [0,1];(p,p_3,\ldots,p_m) \in \Gamma_{m-1};\ m = 3,4,\ldots)$$

is an example of our point. The right hand side is linear in p. If this is only slightly altered to p^α, i.e., we take

$$(5) \qquad H_m^\alpha[p(1-q),pq,p_3,\ldots,p_m]-H_{m-1}^\alpha(p_1p_3,\ldots,p_m) = p^\alpha H_2^\alpha(1-q,q) \quad (\alpha \neq 0)$$

instead of (4), we get [see Daróczy (1970)] entropies different from (1), the entropies of degree α

$$(6) \qquad H_m^\alpha(p_1,p_2,\ldots,p_m) = (2^{1-\alpha}-1)^{-1}\Big(\sum_{j=1}^m p_j^\alpha - 1\Big) \quad (\alpha \neq 1;\ 0^\alpha := 0).$$

These entropies go over into the Shannon entropy when $\alpha \to 1$. Another one-parameter set of entropies, the Rényi entropies or entropies of order α [see e.g. Rényi (1960, 1965)].

$$(7) \qquad {}_\alpha H_m(p_1,p_2,\ldots,p_m) = (1-\alpha)^{-1}\,\log_2 \sum_{j=1}^m p_j^\alpha \quad (\alpha \neq 1;\ 0^\alpha := 0),$$

which also go over into the Shannon entropy (1) when $\alpha \to 1$, have characterizations [see, e.g., Aczél-Daróczy (1963), Rényi (1965)] which assume that they are of a quasilinear form

$$\phi^{-1}\Big[\sum_{j=1}^m p_j\,\phi(-\log p_j)\Big].$$

On the other hand, the following conditions seem to us fairly natural for a measure of expected information and do not presuppose linearity (or quasilinearity) of its expression. Let K(P) be the information expected from an experiment P with the possible outcomes A_j (j = 1,2,...,m) of probabilities $p_j := p(A_j)$ (j = 1,2,...,m). ['Experiments' and 'outcomes' are just interpretations of partitions of a set. For formulations in those terms, see Forte (1973).] Denote now by Q a second experiment, with B_k (k = 1,2,...,n) as possible outcomes, and denote by P*Q the combination of the two experiments, i.e., the experiment $A_j \cap B_k$ (j = 1,2,...,m; k = 1,2,...,n) as possible outcomes. Then it is rather natural to expect that

(8) $$K(P*Q) \leq K(P) + K(Q),$$

that is, we cannot expect more information from a combination of two experiments than the sum of the informations which can be expected from the single experiments. We call this subadditivity.

In terms of probabilities, by writing

(9) $$K_m(p_1,p_2,\ldots,p_m) := K(P)$$

and

(10) $$p_{jk} := p(A_j \cap B_k) \quad (j = 1,2,\ldots,m; \quad k = 1,2,\ldots,n),$$

the subadditivity (8) goes over into

(11) $$K_{mn}(p_{11},p_{12},\ldots,p_{1n},p_{21},p_{22},\ldots,p_{2n},\ldots,p_{m1},p_{m2},\ldots,p_{mn})$$

$$\leq K_m\Big(\sum_{k=1}^{n} p_{1k}, \sum_{k=1}^{n} p_{2k},\ldots, \sum_{k=1}^{n} p_{mk}\Big) + K_n\Big(\sum_{j=1}^{m} p_{j1}, \sum_{j=1}^{m} p_{j2},\ldots, \sum_{j=1}^{m} p_{jn}\Big)$$

$$[(p_{11},p_{12},\ldots,p_{mn}) \in \Gamma_{mn}; \; m,n = 2,3,\ldots].$$

Indeed, cf. (10),

$$\sum_{k=1}^{n} p_{jk} = \sum_{k=1}^{n} p(A_j \cap B_k) = p\Big[\bigcup_{k=1}^{n} (A_j \cap B_k)\Big] = p\Big[A_j \cap \Big(\bigcup_{k=1}^{n} B_k\Big)\Big] = p(A_j)$$

and, similarly,

$$\sum_{j=1}^{m} p_{jk} = p\Big[\bigcup_{j=1}^{m} (A_j \cap B_k)\Big] = p(B_k),$$

since the events A_j (or B_k) are mutually exclusive and represent all possible outcomes of the first (second) experiment ($j = 1,2,\ldots,m$; $k = 1,2,\ldots,n$).

It is also natural to expect more than (8) (or (11)) when the two experiments are independent, i.e., every outcome A_j $(j = 1,2,\ldots,m)$ of the first experiment is independent of every outcome B_k $(k = 1,2,\ldots,n)$ of the second. Indeed, it seems natural to assume that, in case of independent experiments, the information expected from the combination of two experiments is equal to the sum of informations expected from the single experiments. In this case, there is equality in (8)

$$K(P*Q) = K(P) + K(Q) \quad \text{if } P \text{ and } Q \text{ are independent.}$$

We call this additivity. With the notations (9) and $p_j := p(A_j)$ $(j = 1,2,\ldots,m)$, $q_k := p(B_k)$ $(k = 1,2,\ldots,n)$, this property can be written as

$$(12)\quad K_{mn}(p_1q_1,p_1q_2,\ldots,p_1q_n,p_2q_1,p_2q_2,\ldots,p_2q_n,\ldots,p_mq_1,p_mq_2,\ldots,p_mq_n)$$

$$= K_m(p_1,p_2,\ldots,p_m) + K_n(q_1,q_2,\ldots,q_n) \quad [(p_1,p_2,\ldots,p_m) \in \Gamma_m,$$

$$(q_1,q_2,\ldots,q_n) \in \Gamma_n;\ m,n = 2,3,\ldots,],$$

because then, cf. (10),

$$p_{jk} = p_jq_k, \quad \sum_{k=1}^{n} p_{jk} = p_j, \quad \sum_{j=1}^{m} p_{jk} = q_k \quad (j = 1,2,\ldots,m;\ k = 1,2,\ldots,n).$$

The entropies (6) of degree α $(\alpha \neq 1)$ are not additive but subadditive, while the entropies (7) of order α $(\alpha \neq 0,\ \alpha \neq 1)$ are not subadditive but additive. So the question naturally arises - and has been asked, a.o. by [Aczél (1964, 1968, 1970)] - whether additivity and subadditivity under some further reasonable suppositions, but without conditions of linearity, characterize the Shannon entropy. This has been shown by Forte (1973) [cf. Aczél and Forte (1970)].

The only further suppositions, which were made there, are the following. Expansibility

$$(13) \qquad K_{m+1}(p_1,p_2,\ldots,p_m,0) = K_m(p_1,p_2,\ldots,p_m) \quad [(p_1,p_2,\ldots,p_m) \in \Gamma_m;\ m = 2,3,\ldots]$$

means that the amount of information expected from an experiment does not change if we add outcome(s) of zero probability. The condition of symmetry

$$(14) \qquad K_m(p_1,p_2,\ldots,p_m) = K(p_{k(1)},p_{k(2)},\ldots,p_{k(m)}) \text{ for all permutations}$$

$(k(1),k(2),\ldots,k(m))$ of $(1,2,\ldots,m)$, for all $(p_1,p_2,\ldots,p_m) \in \Gamma_m$, and for all $m = 2,3,\ldots$ is self explanatory. An (unimportant) normalizing condition

$$(15) \qquad K_2\left(\frac{1}{2},\frac{1}{2}\right) = 1$$

chooses the unit of information (one bit) as the information expected from the simple alternative, that is, from the experiment with two possible outcomes which are equally probable. The last supposition is that the entropy is 'small for small probabilities', that is,

$$(16) \qquad \lim_{q \to 0+} K_2(1-q,q) = 0$$

[cf. Aczél and Daróczy (1963), Aczél (1968), and Daróczy (1969)], which gives the intuitive statement that we get very little information out of an experiment with two possible outcomes one of which is almost certain, the other almost impossible. It is easy to see [cf. Aczél (1968), Aczél-Daróczy (1974)], that the Shannon entropy (1) (with (2)) has all these 'natural' properties (11), (12), (13), (14), (15), and (16).

As P. Benvenuti has observed [see Aczél (1970)], the entropy of order 0 (cf. (7)) satisfies the conditions (11), (12), (13), (14), (15), but not (16). [In Forte (1973) there are further examples for the independence of (11), (12),

(13), (14), (15), and (16)]. One sees from (7) that these entropies of order 0 can be given as

$$(17) \qquad {}_0H_m(p_1,p_2,\ldots,p_m) = \log_2 N(P)$$

where $N(P)$ is the number of non-zeros among $(p_1,p_2,\ldots,p_m)$ (i.e., the number of outcomes, with probabilities greater than 0, of the experiment P).

Hartley (1928) was the first to introduce a measure of information. One interpretation of this measure [see Jaglom-Jaglom (1957-1969)] is given by (17). So we will call this the Hartley entropy - In this paper we are going to prove the following result [for a special case, cf. Forte (1971)].

Theorem. If $K_m : \Gamma_m \to R$ $(m = 2,3,\ldots)$ satisfies (11), (12), (13) and (14), then and only then there exist nonnegative constants a, b such that

$$(18) \qquad K_m(p_1,p_2,\ldots,p_m) = aH_m(p_1,p_2,\ldots,p_m) + b\ {}_0H_m(p_1,p_2,\ldots,p_m)$$

$$[(p_1,p_2,\ldots,p_m) \in \Gamma_m, \quad m = 2,3,\ldots].$$

where H_m and ${}_0H_m$ are given by (1) (cf. (2)) and (17), respectively. In other words, every subadditive, additive, expansible and symmetric entropy is a linear combination of the Shannon and Hartley entropies. Even (11) can be replaced by the weaker condition (n = 2) of weak subadditivity

$$(19) \qquad K_{2m}(p_{11},p_{12},p_{21},p_{22},\ldots,p_{m1},p_{m2}) \le K_m(p_{11}+p_{12},p_{21}+p_{22},\ldots,p_{m1}+p_{m2})$$

$$+ K_2(p_{11}+p_{21}+\ldots+p_{m1},p_{12}+p_{22}+\ldots+p_{m2})$$

$$[(p_{11},p_{12},p_{21},\ldots,p_{m2}) \in \Gamma_{2m};\ m = 2,3,\ldots].$$

From this a new (and shorter) proof of the above mentioned characterization of the Shannon entropy follows.

Corollary 1. The Shannon entropy (1) is the only (19) weakly subadditive, (12) additive, (13) expansible, (14) symmetric, (15) normalized entropy which is (16) small for small probabilities.

The main steps of the proof that follows are essentially due to the third author. The second author has originally found another proof for a somewhat weaker theorem. The first author's role was to discover a gap in the original form of the present proof and, after its correction, to furnish minor improvements in the proof.

2. We will prove the Theorem through a sequence of Lemmas.

The first gives bounds for the differences of K_m-values if we split the first outcome in two different ways.

Lemma 1. If $K_m : \Gamma_m \to R$ $(m = 2,3,\ldots)$ is (12) additive (with $n = 2$), (14) symmetric, and (19) weakly subadditive, then

$$(20) \qquad K_2(1-q,q) - K_2[(1-p)(1-q) + p(1-r),(1-p)q + pr] \leq$$

$$K_m[p(1-q),pq,p_3,p_4,\ldots,p_n] - K_m[p(1-r),pr,p_3,p_4,\ldots,p_n] \leq$$

$$K_2[p(1-q)+(1-p)(1-r),pq+(1-p)r] - K_2(1-r,r) \quad \text{for all } q \in [0,1],$$

$$r \in [0,1] \text{ and all } (p,p_3,\ldots,p_m) \in \Gamma_{m-1} \quad (m = 3,4,\ldots).$$

Proof. By (12), (14), and (19) (in that order), we get

$$K_m[p(1-q),pq,p_3,p_4,\ldots,p_m] + K_2(1-r,r) = K_{2m}[p(1-q)(1-r),p(1-q)r,$$

$$pq(1-r),pqr,p_3(1-r),p_3r,p_4(1-r),p_4r,\ldots,p_m(1-r),p_mr] =$$

$$K_{2m}[p(1-q)(1-r),pq(1-r),p(1-q)r,pqr,p_3(1-r),p_3r,p_4(1-r),p_4r,$$

$$\ldots,p_m(1-r),p_mr] \le K_m[p(1-r),pr,p_3,p_4,\ldots,p_m] + K_2[p(1-q)+(p_3+p_4+\ldots+p_m)(1-r),$$

$$pq+(p_3+p_4+\ldots+p_m)r]$$

and so, remembering that $p+p_3+p_4+\ldots+p_m = 1$, we have

$$K_m[p(1-q),pq,p_3,p_4,\ldots,p_m] - K_m[p(1-r),pr,p_3,p_4,\ldots,p_m] \le$$

$$K_2[p(1-q)+(1-p)(1-r),pq+(1-p)r] - K_2(1-r,r).$$

This is the second inequality in (20). Interchanging q and r we get the first inequality. So Lemma 1 is proved.

It is rather well known [cf., e.g., Kendall (1963), Lee (1964), Daróczy (1969), and Aczél and Daróczy (1974)] that the information function defined by

$$f(x) := K_2(1-x,x) \qquad (x \in [0,1]) \tag{21}$$

has a fundamental role in characterization theorems. In the next lemma we derive some properties of this function from our suppositions.

Lemma 2. The inequalities (20) and the symmetry (14) (m = 2) imply that the function f, defined by (21), has the following properties.

(i) Symmetry with respect to $\frac{1}{2}$

$$f(1-q) = f(q) \qquad (q \in [0,1]).$$

(ii) Non-decreasing monotonicity on $[0, \frac{1}{2}]$ (non-increasing on $[\frac{1}{2}, 1]$).

(iii) Continuity on $]0,1[$.

(iv) Concavity on $[0,1]$, that is,

(22) $$f[(1-\lambda)q+\lambda r] \geq (1-\lambda)f(q)+\lambda f(r) \text{ for all } \lambda,q,r \in [0,1].$$

(v) The right and left derivatives D^+f and D^-f exist everywhere on $[0,1[$ and $]0,1]$, respectively - they may still be infinite at some points, but

(vi) D^+f and D^-f are finite on $]0,1[$, and, finally

(vii) $D^+f(x) \geq 0$ for all $x \in [0, \frac{1}{2}[$, $D^-f(x) \geq 0$ for all $x \in]0, \frac{1}{2}]$.

Proof. The symmetry (i) follows, of course, immediately from (14) (m = 2):

(23) $$f(1-q) = K_2(q,1-q) = K_2(1-q,q) = f(q) \qquad (q \in [0,1]).$$

Looking at the two ends of (20), we find

(24) $$K_2(1-q,q) - K_2[(1-p)(1-q)+p(1-r),(1-p)q+pr] \leq$$

$$K_2[p(1-q)+(1-p)(1-r),pq+(1-p)r] - K_2(1-r,r) \text{ for all } p,q,r \in [0,1].$$

Substituting r = 1-q into (24) and taking (14) (m = 2) and (21) into consideration, we get

(25) $$f(q) \le f[p(1-q)+(1-p)q] \quad \text{for all } p,q \in [0,1].$$

Choose $q \in [0, \frac{1}{2}]$. Now, as p runs through [0,1], the quantity $p(1-q)+(1-p)q$ runs through $[q,1-q]$, and (25) states that $f(q) = f(1-q)$ [see (i)] is the minimal value of f on the interval $[q,1-q]$. Thus f is indeed monotonic non-decreasing on $[0, \frac{1}{2}]$, non-increasing on $[\frac{1}{2}, 1]$, as asserted in (ii).

Now choose $p = \frac{1}{2}$ in (24). Then, again with (21),

(26) $$f(\tfrac{1}{2}q + \tfrac{1}{2}r) \ge \tfrac{1}{2}f(q) + \tfrac{1}{2}f(r) \quad \text{for all } q,r \in [0,1],$$

which is (22) for $\lambda = \frac{1}{2}$. By repeated use of (26) we get (22) for all $\lambda \in]0,1[$ which are diadic fractions. By the monotonicity (ii) of f (non-decreasing on $[0, \frac{1}{2}]$, non-increasing on $[\frac{1}{2}, 1]$) this extends to all real $\lambda \in]0,1[$ (if $x = (1-\lambda)q+\lambda r \ge \frac{1}{2}$, choose sequences of diadic $d_n \to \lambda$ such that $(1-d_n)q+d_n r > x$; if $x < \frac{1}{2}$, choose them so that $(1-d_n)q+d_n r < x$). For $\lambda = 0$ and $\lambda = 1$, (22) is trivially satisfied. Thus (iv) is proved.

Since, by (ii), f is monotonic on $[0, \frac{1}{2}]$ and again on $[\frac{1}{2}, 1]$, it can have at most jump discontinuities, but these are ruled out on $]0,1[$ by the concavity (iv) of f. Thus f is indeed continuous (iii) on the open interval $]0,1[$.

By (iv), f is concave on [0,1]. Therefore the difference quotients

$$\frac{f(x+h)-f(x)}{h}$$

decrease with h. Thus the right derivative D^+f (finite or infinite) exists at every $x \in [0,1[$ and, similarly, D^-f exists on $]0,1]$, which proves (v). Also, f being increasing on $[0, \frac{1}{2}]$, $D^+f \ge 0$ on $[0, \frac{1}{2}[$ and $D^-f \ge 0$ on $]0, \frac{1}{2}]$,

as asserted in (vii).

Finally, since f is concave by (iv), it is easy to see that the only places, at which D^+f or D^-f could be infinite, are $x = 0$ and $x = 1$, respectively. So (vi) holds too and this concludes the proof of Lemma 2.

Lemma 3. From (14) (m = 2) and (20) it follows also that there exist functions $J_m : \Gamma_m \to R$ (m = 2,3,...) such that

$$(27) \qquad K_m[p(1-q),pq,p_3,\ldots,p_m] = pK_2(1-q,q)+J_{m-1}(p,p_3,\ldots,p_m) \text{ for all } q \in]0,1[,\ (p,p_3,\ldots,p_m) \in \Gamma_{m-1},\ m = 3,4,\ldots,$$

Notice the similarity of (27) to the recursivity (4).

Proof. We use (20) and (21) in order to get

$$(28) \qquad \frac{f[(1-p)q+pq] - f[(1-p)q+pr]}{q-r} \le \frac{K_m[p(1-q),pq,p_3,\ldots,p_m] - K_m[p(1-r),pr,p_3,\ldots,p_m]}{q-r} \le$$

$$\frac{f[pq+(1-p)r] - f[pr+(1-p)r]}{q-r}, \text{ whenever } 0 \le r < q \le 1,$$

$$(p,p_3,\ldots,p_m) \in \Gamma_{m-1};\ m = 3,4,\ldots$$

Now let q decrease to r and observe that both extremities of (28) tend to $pD^+f(r)$, which is the right derivative both of $x \mapsto f[(1-p)q+px]$ at $q = r$ and of $x \mapsto f[px+(1-p)r]$ at r. We denote the right derivative of $x \mapsto K_m[p(1-x),px,p_3,\ldots,p_m]$ at $x = r$ [that is, the right partial derivate of $(x,p,p_3,\ldots,p_m) \mapsto K_m[p(1-x),px,p_3,\ldots,p_m]$ with respect to x at the point $(p(1-r),pr,p_3,\ldots,p_n)$] by $D^+_{x=r}\, K_m[p(1-x),px,p_3,\ldots,p_m]$ and thus get

$$D^+_{x=r}\, K_m[p(1-x),px,p_3,\ldots,p_m] = pD^+f(r)$$

(finite, if $r \in]0,1[$) or

(29) $$D^{+}_{x=r}[K_m(p(1-x),px,p_3,\ldots,p_m)-pf(x)] = 0$$

for all $r \in]0,1[$, $(p,p_3,\ldots,p_m) \in \Gamma_{m-1}$; $m = 3,4,\ldots$

Similarly, if we let r increase to q in (28) and denote the left (partial) derivative with respect to x at q by $D^{-}_{x=q}$, we get

(30) $$D^{-}_{x=q}[K_m(p(1-x),px,p_3,\ldots,p_m)-pf(x)] = 0$$

for all $q \in]0,1[$, $(p_1p_3,\ldots,p_m) \in \Gamma_{m-1}$; $m = 3,4,\ldots$

Equations (29) and (30) show that $x \mapsto K_m[p(1-x),px,p_3,\ldots,p_m)-pf(x)$ is differentiable and its derivative is 0 on $]0,1[$. Thus this function is 'constant' there, the constant depending on $p,p_3,\ldots,p_m$ (and m). So

$$K_m[p(1-x),px,p_3,\ldots,p_m]-pf(x) = J_{m-1}(p,p_3,\ldots,p_m)$$

for all $x \in]0,1[$, $(p,p_3,\ldots,p_m) \in \Gamma_{m-1}$; $m = 3,4,\ldots$

This, with (21), gives (27) and concludes the proof of Lemma 3.

3. In the following three Lemmas we establish (18) for nonzero probabilities. We consider first m = 2.

Lemma 4. If $\{K_m\}$ is (14) symmetric (m = 2,3), (12) additive (n = 2) and (19) weakly subadditive, then there exist constants $a \geq 0$ and A_2 such that

(31) $$K_2(1-q,q) = aH_2(1-q,q)+A_2 \quad \text{for all } q \in]0,1[$$

where H_2 is the Shannon entropy (1) for two complementary events.

Proof. Using (27) with $m = 3$ and $p = p_1+p_2$, $q = p_2/(p_1+p_2)$ $(p_1 > 0, p_2 > 0)$ we get

$$(32) \qquad K_3(p_1,p_2,p_3) = (p_1+p_2)K_2\left(\frac{p_1}{p_1+p_2}, \frac{p_2}{p_1+p_2}\right) + J_2(p_1+p_2,p_3)$$

whenever $p_1 > 0$, $p_2 > 0$, $p_3 \geq 0$, $p_1 + p_2 + p_3 = 1$.

If 0 probabilities are excluded [as was not the case in (3)] then we write

$$(33) \qquad \Gamma_m' := \{(p_1,p_2,\ldots,p_m) \mid \sum_{j=1}^{m} p_j = 1;\ p_j > 0;\ j = 1,2,\ldots,m\} \quad (m = 2,3,\ldots).$$

So, (32) is certainly true on Γ_3'.

Now we bring in the functions f and g, defined by (21) and by

$$g(x) := J_2(1-x,x) \qquad (x \in [0,1]),$$

respectively. In view of (14), $(m = 3)$, we get from (32)

$$(p_1+p_2)f\left(\frac{p_2}{p_1+p_2}\right) + g(p_3) = K_3(p_1,p_2,p_3) = K_3(p_1,p_3,p_2) = (p_1+p_3)f\left(\frac{p_3}{p_1+p_3}\right) + g(p_2) \text{ on } \Gamma_3'.$$

With $x = p_3$, $y = p_2$ this goes over into

$$(34) \qquad (1-x)f\left(\frac{y}{1-x}\right) + g(x) = (1-y)f\left(\frac{x}{1-y}\right) + g(y) \quad \text{whenever } x \in]0,1[,\ y \in]0,1[,$$

$$x+y < 1.$$

This equation is a generalization of the so called 'fundamental equation of information' [see e.g. Kendall (1963), Lee (1964), Daróczy (1969), and Aczél and Daróczy (1974)] where $g = f$ and can be solved in a very similar way, [see also Kannappan and Ng (1973)]. We include here the proof for completeness' sake. We have proved in Lemma 2 (iii) that f is continuous on $]0,1[$. By (34), also g is continuous on $]0,1[$ and so both are locally

integrable there.

Now, take an arbitrary $y \in]0,1[$ and let λ, μ be such that

$$0 < y < y + \lambda < y + \mu < 1.$$

So, whenever $x \in [\lambda, \mu] \subset]0,1[$, both $x/(1-y)$ and $y/(1-x)$ fall into closed subintervals on $]0,1[$ on which f and g are integrable. Indeed,

$$0 < \lambda < \frac{\lambda}{1-y} \le \frac{x}{1-y} \le \frac{\mu}{1-y} < 1, \quad 0 < y < \frac{y}{1-\lambda} \le \frac{y}{1-x} \le \frac{y}{1-\mu} < 1. \tag{35}$$

Evidently, also $x+y \le y+\mu < 1$. Thus, integrating (34) with respect to x from λ to μ, we get

$$\begin{aligned}(\mu-\lambda)g(y) &= \int_\lambda^\mu g(y)\,dx = \int_\lambda^\mu g(x)\,dx + \int_\lambda^\mu (1-x) f\left(\frac{y}{1-x}\right) dx - (1-y)\int_\lambda^\mu f\left(\frac{x}{1-y}\right) dx = \\ &= \int_\lambda^\mu g(x)\,dx + y^2 \int_{y/(1-\lambda)}^{y/(1-\mu)} s^{-3} f(x)\,dx - (1-y)^2 \int_{\lambda/(1-y)}^{\mu/(1-y)} f(t)\,dt.\end{aligned} \tag{36}$$

Here we have used the substitution

$$s = \frac{y}{1-x}, \qquad t = \frac{x}{1-y}. \tag{37}$$

As we have seen in (35), when $x \in]0,1[$, $y \in]0,1[$, then $s \in [\frac{y}{1-\lambda}, \frac{y}{1-\mu}] \subset]0,1[$, $t \in [\frac{\lambda}{1-y}, \frac{\mu}{1-y}] \subset]0,1[$. On the other hand, for every $s \in]0,1[$, $t \in]0,1[$ there exist x,y such that

$$x \in]0,1[, \quad y \in]0,1[, \ x+y < 1 \text{ and } s = \frac{y}{1-x}, \quad t = \frac{x}{1-y}. \tag{38}$$

Indeed,

$$x = \frac{t-st}{1-st} \in]0,1[, \quad y = \frac{s-st}{1-st} \in]0,1[$$

satisfies all these conditions, since (37) is evidently satisfied and

$$\frac{y}{1-x} = s < 1, \qquad \text{so} \quad x+y < 1.$$

Therefore s and t, as defined by (38), run through]0,1[. This changes (34) into

$$(39) \qquad f(t) = \frac{1-t}{1-s} f(s) + \frac{1-st}{1-s} \left[g\left(\frac{t-st}{1-st}\right) - g\left(\frac{s-st}{1-st}\right)\right] \quad \text{for all } s \in]0,1[,\ t \in]0,1[.$$

We return now to (36). Since f is continuous on]0,1[, the right hand side of (36) is differentiable in y. Thus the left hand side is differentiable too and so, since $\mu-\lambda \neq 0$, g is differentiable on]0,1[. By (39), f is differentiable when g is differentiable. But then the right hand side of (36) is twice differentiable, so also g and, by (39), also f.

Now differentiate (34) with respect to x and then differentiate the resulting equation

$$\frac{y}{1-x} f'\left(\frac{y}{1-x}\right) - f\left(\frac{y}{1-x}\right) + g'(x) = f'\left(\frac{x}{1-y}\right)$$

with respect to y in order to obtain

$$\frac{y}{(1-x)^2} f''\left(\frac{y}{1-x}\right) = \frac{x}{(1-y)^2} f''\left(\frac{x}{1-y}\right)$$

or, with the substitution (37),

$$s(1-s)f''(s) = t(1-t)f''(t) = c(\text{constant}) \text{ on }]0,1[.$$

By successive integrations we get

$$f(t) = c[t(\ell n\ t-1) + (1-t)(\ell n(1-t)-1)] + bt+C$$

or, with $a = -c\ \ell n\ 2$, $A_2 = C-c$,

(40) $$f(t) = a[-(1-t)\log_2(1-t) - t\log_2 t] + bt + A_2.$$

By the (23) symmetry we have $b = 0$ and, by the non-decreasing monotonicity of f on $[0, \frac{1}{2}]$, proved in Lemma 2 (ii), we have also $a \geq 0$. So (21) and (1) carry (40) over into (31) and Lemma 4 is proved.

We give now a representation, similar to (31), also for $m > 2$.

Lemma 5. If $\{K_m\}$ is (14) symmetric, (12) additive ($n = 2$) and (19) weakly subadditive, then there exist constants $a \geq 0$ and $A(m)$ ($m = 2,3,\ldots$) such that

(41) $$K_m(p_1,p_2,\ldots,p_m) = aH_m(p_1,p_2,\ldots,p_m) + A(m) \text{ for all } (p_1,p_2,\ldots,p_m) \in \Gamma_m' \text{ and all } m = 2,3,\ldots,$$

where H_m ($m = 2,3,\ldots$) is the Shannon entropy.

Proof. Again put into (27) $p = p_1+p_2$, $q = p_2/(p_1+p_2)$ ($p_1 > 0$, $p_2 > 0$) and then apply (31) in order to get

(42) $$K_m(p_1,p_2,\ldots,p_m) = (p_1+p_2)K_2\left(\frac{p_1}{p_1+p_2}, \frac{p_2}{p_1+p_2}\right) + J_{m-1}(p_1+p_2,p_3,\ldots,p_m) =$$

$$= (p_1+p_2)\left[aH_2\left(\frac{p_1}{p_1+p_2}\right) + A_2\right] + J_{m-1}(p_1+p_2,p_3,\ldots,p_m) = (p_1+p_2)aH_2\left(\frac{p_1}{p_1+p_2}, \frac{p_2}{p_1+p_2}\right)$$

$$+ \Psi_{m-1}(p_1+p_2,p_3,\ldots,p_m) \text{ for all } (p_1,p_2,p_3,\ldots,p_m) \in \Gamma_m' \ (m = 3,4,\ldots),$$

where the new functions Ψ_m ($m = 2,3,\ldots$) are defined by

$$\Psi_{m-1}(p,p_3,\ldots,p_m) := J_{m-1}(p,p_3,\ldots,p_m) + pA_2, \quad (m = 3,4,\ldots).$$

With K_m also J_{m-1} and Ψ_{m-1} are symmetric in $(p_3,p_4,\ldots,p_m)$.

The symmetry (14) and the equation (42), just obtained, imply

$$(43)\quad (p_1+p_2)aH_2\left(\frac{p_1}{p_1+p_2}, \frac{p_2}{p_1+p_2}\right) + \Psi_{m-1}(p_1+p_2,p_3,p_4,\ldots,p_m) = K_m(p_1,p_2,p_3,p_4,\ldots,p_m) =$$

$$= K_m(p_1,p_3,p_2,p_4,\ldots,p_m) = (p_1+p_3)aK_2\left(\frac{p_1}{p_1+p_3}, \frac{p_3}{p_1+p_3}\right) + \Psi_{m-1}(p_1+p_3,p_2,\ldots,p_m)$$

for all $(p_1,p_2,\ldots,p_m) \in \Gamma_m'$ $\quad (m = 3,4,\ldots)$.

On the other hand, the (1) Shannon entropy $\{H_m\}$ is recursive and symmetric, that is, it satisfies (4) and (14). Therefore ($p = p_1+p_2$, $q = p_2/(p_1+p_2)$ again)

$$(44)\quad (p_1+p_2)aH_2\left(\frac{p_1}{p_1+p_2}, \frac{p_2}{p_1+p_2}\right) + aH_{m-1}(p_1+p_2,p_3,p_4,\ldots,p_m) = aH_m(p_1,p_2,p_3,p_4,\ldots,p_m) =$$

$$aH_m(p_1,p_3,p_2,p_4,\ldots,p_m) = (p_1+p_3)aH_2\left(\frac{p_1}{p_1+p_3}, \frac{p_3}{p_1+p_3}\right) + aH_{m-1}(p_1+p_2,p_3,p_4,\ldots,p_m)$$

on Γ_m'.

Subtracting (44) from (43) we get, with the notation

$$(45)\quad \Phi_{m-1}(p,q,p_4,\ldots,p_m) := \Psi_{m-1}(p,q,p_4,\ldots,p_m) - aH_{m-1}(p,q,p_4,\ldots,p_m),$$

the equation

$$\Phi_{m-1}(p_1+p_2,p_3,p_4,\ldots,p_m) = \Phi_{m-1}(p_1+p_3,p_2,p_4,\ldots,p_m) \text{ for all}$$

$$(p_1,p_2,\ldots,p_m) \in \Gamma_m' \quad (m = 3,4,\ldots).$$

We now write $p_1+p_2+p_3 = r_3 \in]0,1[$ ($r_3 = 1$ if $m = 3$) and obtain

$$\Phi_{m-1}(r_3-p_3,p_3,p_4,\ldots,p_m) = \Phi_{m-1}(r_3-p_2,p_2,p_4,\ldots,p_m)$$

so that $q \mapsto \Phi_{m-1}(r_3-q,q,p_4,\ldots,p_m)$ is constant. The functions $(p_1p_3,p_4,\ldots,p_m)$ $\mapsto \Phi_{m-1}(p_1p_3,p_4,\ldots,p_m)$ ($m = 3,4,\ldots$) are symmetric in $(p_3,p_4,\ldots,p_m)$, by (45), because the Ψ_{m-1} and H_{m-1} were symmetric in these variables. Therefore

$$\Phi_{m-1}(p_1+p_2,p_3,p_4,\ldots,p_{k-1},p_k,p_{k+1},\ldots,p_m) = \Phi_{m-1}(p_1+p_2,p_k,p_4,\ldots,p_{k-1},p_3,\ldots,p_m)$$

so that $p \mapsto \Phi_{m-1}(r-p,p_3,\ldots,p_{k-1},p,p_{k+1},\ldots,p_m)$ is also constant ($k = 3,4,\ldots,m$). Thus

$$\Phi_{m-1}(r_3-p_3,p_3,p_4,\ldots,p_m) = \Phi_{m-1}(r_3-a_3,a_3,p_4,\ldots,p_m) =$$

$$\Phi_{m-1}(r_4-a_3-a_4,a_3,a_4,p_5,\ldots,p_m) = \ldots = \Phi_{m-1}(r_m-a_3-a_4-\ldots-a_m,a_3,a_4,\ldots,a_m)$$

$$= A(m) \quad \text{(constant, i.e., depending only upon m)},$$

where the a_k ($k = 3,4,\ldots,m$) are constants and $r_k = p_1+p_2+p_3+\ldots+p_k$ $= p_1+p_2+a_3+\ldots+a_k$ ($k = 3,4,\ldots,m$), in particular, $r_m = 1$. Thus, by (42), (45), and (44),

$$K_m(p_1,p_2,\ldots p_m) = (p_1+p_2)\ aH_2\left(\frac{p_1}{p_1+p_2}, \frac{p_2}{p_1+p_2}\right) + aH_{m-1}(p_1+p_2,p_3,\ldots,p_m)$$

$$+ A(m) = aH_m(p_1,p_2,\ldots,p_m) + A(m) \text{ on } \Gamma_m'$$

for all $m = 3,4,\ldots$, while (31) gave the same for $m = 2$ ($A(2) := A_2$). We had also $a \geq 0$ in Lemma 4. So we have (41) and this concludes the proof of Lemma 5.

Now we determine A(m) which figures in (41).

Lemma 6. Under the conditions of the Theorem, there exists a nonnegative constant b such that the function A in Lemma 5 is given by

$$A(m) = b \log_2 m \quad \text{for all} \quad m = 2,3,... \tag{46}$$

Proof. If we substitute (41) into the additivity condition (12), we get

$$A(mn) = A(m) + A(n) \quad \text{for all } m, n = 2,3,... \tag{47}$$

(if we define A(1) : = 0, then (47) holds also if m = 1 or n = 1). That is, A is a completely additive number theoretical function [cf. Kátai (1967), Aczél and Daróczy (1974)].

Now we go back to the inequality (20), with (m+1) instead of m, and substitute $q = \frac{1}{2}$, $r = 0$, $p = p_3 = \ldots = p_{m+1} = \frac{1}{m}$, in order to get, with aid of (ii) in Lemma 2 and of (21) and (13),

$$0 \le f(\tfrac{1}{2}) - f(\tfrac{1}{2}\,\tfrac{m-1}{m}) \le K_{m+1}(\tfrac{1}{2}\,\tfrac{1}{m}, \tfrac{1}{2}\,\tfrac{1}{m}, \tfrac{1}{m}, \tfrac{1}{m}, \ldots, \tfrac{1}{m}) - K_m(\tfrac{1}{m}, \tfrac{1}{m}, \ldots, \tfrac{1}{m}).$$

With (41), this goes over into

$$0 \le aH_{m+1}(\tfrac{1}{2}\,\tfrac{1}{m}, \tfrac{1}{2}\,\tfrac{1}{m}, \tfrac{1}{m}, \tfrac{1}{m}, \ldots, \tfrac{1}{m}) - aH_m(\tfrac{1}{m}, \tfrac{1}{m}, \ldots, \tfrac{1}{m}) + A(m+1) - A(m) =$$

$$= a\,\tfrac{1}{m}\,H_2(\tfrac{1}{2}, \tfrac{1}{2}) + A(m+1) - A(m) = a\,\tfrac{1}{m} + A(m+1) - A(m) \quad (m = 2,3,\ldots), \tag{48}$$

because the Shannon entropy $\{H_m\}$ is recursive and normalized, that is, it satisfies (4) and (15). From (48) we get

$$\liminf_{m \to \infty} [A(m+1) - A(m)] \ge \liminf_{m \to \infty} (-a\,\tfrac{1}{m}) = 0. \tag{49}$$

It has been proved by Kátai (1967), that every additive (and, a fortiori, every completely additive) number theoretical function which satisfies (49) is of the form (46). If we put (46) into (41), we get

$$(50) \qquad K_m(p_1,p_2,\ldots,p_m) = aH_m(p_1,p_2,\ldots,p_m) + b \log_2 m \,[(p_1,p_2,\ldots,p_m) \in \Gamma_m'; \; m = 2,3,\ldots].$$

We have seen in Lemma 4 that $a \geq 0$. We still have to prove that $b \geq 0$.

By (50) and (21),

$$(51) \qquad f(q) = K_2(1-q,q) = aH_2(1-q,q) + b.$$

For the Theorem (and this Lemma 6) we did not suppose that K_2 satisfies (16) but we know that the Shannon entropy is 'small for small probabilities', that is, H_2 satisfies (16). So (51) implies

$$(52) \qquad \lim_{q \to 0+} f(q) = a \lim_{q \to 0+} H_2(1-q,q) + b = b.$$

Also, from the additivity (12) we get, with $m = n = 2$, $p_1 = q_1 = 1$, $p_2 = q_2 = 0$ and with aid of the expansibility (13), $K_2(1,0) = K_4(1,0,0,0) = K_2(1,0) + K_2(1,0)$ or, cf. (21),

$$(53) \qquad f(0) = K_2(1,0) = 0.$$

But in Lemma 2 (ii) we have shown that f is non-decreasing on $[0, \frac{1}{2}]$ so (52) and (53) imply $b \geq 0$. This concludes the proof of Lemma 6.

4. With (50) (and $a \geq 0$, $b \geq 0$) we have proved the statement of our Theorem for Γ_m' ($m = 2,3,\ldots$), see (33). All what is left to do is to

extend our result to Γ_m (see (3)).

Conclusion of the proof of the Theorem. Let $p_{j_1}, p_{j_2}, \ldots, p_{j_N}$ be the non-zeros among $p_1, p_2, \ldots, p_m$ [$N = N(P)$]. If we apply (the symmetry (14) and) the expansibility (13) both of the Shannon entropy $\{H_m\}$ [cf. (1), (2)] and of the entropy $\{K_m\}$, which we have to determine, then, in view of (50), we get

$$K_m(p_1,p_2,\ldots,p_m) = K_N(p_{j_1},p_{j_2},\ldots,p_{j_N}) = aH_N(p_{j_1},p_{j_2},\ldots,p_{j_N}) + b \log_2 N$$

$$= aH_m(p_1,p_2,\ldots,p_m) + b \log_2 N(P) \text{ for all } (p_1,p_2,\ldots,p_m) \in \Gamma_m \text{ and all } m = 2,3,\ldots$$

This, cf. (17), is (18) and so the Theorem is proved.

Proof of Corollary 1. The property (16) implies $b = 0$ in (18), because ${}_0H_m(p_1,p_2,\ldots,p_m) = \log_2 N(P)$ is not 'small for small probabilities', while the Shannon entropy H_m is, and K_m ($m = 2,3,\ldots$) is now supposed to be. After that, the normalization (15) gives $a = 1$ and the Corollary 1 is proved too.

It is clear now that H_m is the continuous and ${}_0H_m$ the discontinuous part of K_m in (18). So, while the Shannon entropy is characterized in Corollary 1 by (16), which is a weaker form of continuity, the Hartley entropy could be characterized by adding to the other conditions, instead of (16), a condition which assures that the H_m's are step functions.

When Hartley (1928) has introduced his measure of information he has strongly emphasized [cf. Jaglom and Jaglom (1957-1969)] as rationale that it depends only upon the number of outcomes, not upon their probabilities (as long as they are different from 0). A weaker form of this statement would be

$$K_2(1-t,t) = c \text{ (constant)} \qquad (t \in]0,1[)$$

and this has been used by Forte (1971) to characterize the Hartley entropy. An even weaker condition is that f [cf. (21)] is not strictly monotonic on $]0, \frac{1}{2}]$ or

(54) $K_2(1-p_1,p_1) = K_2(1-p_2,p_2)$ <u>for at least one pair</u> p_1,p_2 <u>with</u>

$0 < p_1 < p_2 \leq \frac{1}{2}$.

These are conditions of <u>insensitivity</u>.

<u>Corollary 2</u>. <u>If an entropy is (19) weakly subadditive, (12) additive, (13) expansible, (14) symmetric, (15) normalized and (54) insensitive, then and only then it is the Hartley entropy (17)</u>.

The <u>proof</u> is again obvious. The conditions (54) and (15), if applied to (18), give a = 0 and b = 1, respectively.

As mentioned earlier, the normalization (15) is unimportant in the Corollaries. Without it they would still be true up to multiplicative constants.

This research has been supported in part by National Research Council of Canada grants A-2972, A-7677, and A-8212.

References

Aczél, J. (1964), Some Unsolved Problems in the Theory of Functional Equations. Arch. Math. (Basel) 15, 435-444.

Aczél, J. (1968), On Different characterizations of Entropies. Probability and Information Theory. Proc. Internat. Symp. McMaster Univ., Canada, 1968. Springer-Verlag, Berlin-Heidelberg and New York, 1969, pp. 1-11.

Aczél, J. (1970), Problems 6 (P51, P52, P53, P54, P51S1). Aequationes Math. 4, 242-243.

Aczél, J. and Daróczy, Z. (1963), Sur la caractérisation axiomatique des entropies d'ordre positif, y comprise l'entropie de Shannon. C. R. Acad. Sci. Paris 257, 1581-1584.

Aczél, J. and Daróczy, Z. (1974), On Measures of Information and Their Characterizations. Academic Press, New York.

Aczél, J. and Forte, B. (1970), A System of Axioms for the Measure of the Uncertainty. Notices Amer. Math. Soc. 17, 202.

Daróczy, Z. (1969), On the Shannon Measure of Information (Hungarian).[1] Magyar Tud. Akad. III. Oszt. Közl. 19, 9-24.

Daróczy, Z. (1970), Generalized Information Functions. Information and Control 16, 36-51.

Faddeev, D. K. (1956), On the Concept of Entropy of a Finite Probabilistic Scheme (Russian). Uspechi Mat. Nauk 11, No. 1 (67), 227-231.

Forte, B. (1971), Functional Inequalities in Information Theory. Aequationes Math. 6, 102-103.

Forte, B. (1973), Why Shannon's Entropy. Convegno Inform. Teor., Ist Naz. Alta Mat., Roma 1973. Symposia Math. Vol. XI, Academic Press, New York 1974.

Hartley, R. V. (1928), Transmission of Information. Bell System Tech. J. 7, 535-563.

[1]) English translation in Selected Translations in Mathematical Statistics and Probability Vol. 10, Inst. of Math. Stat. - Amer. Math. Soc., Providence, R.I., 1972, pp.193-210.

References Continued

Jaglom, A. M. and Jaglom, I. M. (1957-1969), Probability and Information (Russian). GITTL, Moscow, 1957. 2nd ed., Fizmatigiz, Moscow 1960. French translation, Probabilité et Information, Collection Sigma Vol. 17, Dunod, Paris 1969.

Kannappan, PL. and Ng, C. T. (1973), Measurable Solutions of Functional Equations Related to Information Theory. Proc. Amer. Math. Soc. 38.

Kátai, I. (1967), A Remark on Additive Arithmetical Functions. Ann. Univ. Sci. Budapest. Eötvös Sect. Math. 12 (1967), 81-83.

Kendall, D. G. (1963), Functional Equations in Information Theory. Zeit. Wahrscheinlichkeitsth. 2, 225-229.

Lee, P. M. (1964), On the Axioms of Information Theory. Ann. Math. Statist. 35, 414-418.

Rényi, A. (1960), On Measures of Entropy and Information. Proc. 4th Berkeley Sympos. Math. Statist. and Prob. 1960. Univ. of Calif. Press, Berkeley, Calif. 1961 Vol. I, pp. 547-561.

Rényi, A. (1965), On the Foundations of Information Theory, Rev. Int. Statis. Inst. 33, 1-14.

Shannon, C. F. and Weaver, W. (1949), The Mathematical Theory of Communication. Univ. of Ill. Press, Chicago.