treatment of the concept of validity and of the completeness theorem for the first order logics extending the propositional logics considered in the book through a generalization of the famous algebraic proof of Rasiowa and Sikorski of the completeness theorem of classical first order logic and of the subsequent concept derived from that proof of the canonical realization of an elementary theory. The supplement does not touch upon the theory of cylindric and polyadic algebras nor does it tackle with algebraic means any part of model theory.

An excellent article by the same author summing up much of the material of the book and developing further the first order theory of multi-valued logic has appeared in *Studies in algebraic logic* referred to above.

AUBERT DAIGNEAULT

BULLETIN OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 83, Number 2, March 1977

On measures of information and their characterizations, by J. Aczél and Z. Daróczy, Mathematics in Science and Engineering, vol. 115, Academic Press, New York, San Francisco, London, 1975, xii + 234 pp., $24.50.

The purpose of the authors cannot be stated more clearly than in the following lines of the preface (p. XI):

"We shall deal with measures of information (the most important ones being called entropies), their properties, and, reciprocally, with questions concerning which of these properties determine known measures of information, and which are the most general formulas satisfying reasonable requirements on practical measures of information. To the best of our knowledge, this is the first book investigating this subject in depth". In fact, from the 234 pages of the book, only 6 are devoted to simple applications to logical games (pp. 33 –38) and 17 to optimal coding (pp. 42–50 and pp. 156–164).

But, as the authors write (p. 29) "the problem is to determine which properties to consider as natural and/or essential". From the beginning they make a choice, which implies consequences of paramount importance: the measure of the information yielded by one event A depends only upon the probability $P(A)$ of this event; due to this choice they restrict themselves to the foundations of the classical information theory, initiated in 1948 by Claude Shannon and Norbert Wiener. Of course this classical theory has proved to be very useful indeed in many branches of science, its greatest success being the foundation and development of communication theory: the fundamental hypothesis means that the amount of information given by a message depends only on its frequency; very unexpected messages give considerable information. But, as has been often pointed, no account of the semantic content of the message, could be taken in this way. Moreover there is a subjective aspect of information, which is entirely out of the scope of the classical theory: the same event does not yield the same amount of information to all the observers.

These rather obvious remarks show that the classical information theory deals only with one very important, but particular, aspect of information.

Nevertheless the authors do not discuss the limitations of the theory based on the restrictive postulate: information = function of probability; they content themselves with a reference (p. XI) to a list of publications whose purpose is the generalization of information theory, taking account of the aspects neglected by the classical Shannon-Wiener theory.

From a much less important point of view, a restriction is made (p. 5), restraining the set of events to which the theory applies: throughout the book it is supposed that the probability algebra $\mathbb{S}$ of events A is nonatomic; this implies $P(\mathbb{S}) = [0, 1]$; due to this, the functional equations, basic to the theory, are always defined on simple domains: intervals, triangles, squares, etc . . .

The Introduction (Chapter 0) starts with an investigation of the measure $H(P)$ of the information yielded by a single stochastic event A, with the probability $P(A) = p$.

N. Wiener has been the first to attract attention to $H(p)$, a notion which has always appeared to me as the most fundamental in the theory; considering (Cybernetics, p. 75) the choice at random of a point x on $[0, 1]$ with a uniform probability, he gives $-\log_2 (b - a)$ as "the amount of information we have from the a posteriori knowledge" that $x \in [a, b]$. The authors, following A. Rényi, call $H(p)$ "the entropy of the event A"; I am afraid that this could conceal the fundamental distinction between two different situations, clearly pointed by N. Wiener: *a posteriori*, when the event A has been realized, the amount of information is measured by $H(p)$; but *a priori*, before making an experiment in which the event A could be realized or not, the only measure at our disposal is the *expectation* of this amount of information, i.e. $S(p) = pH(p) + (1 - p)H(1 - p)$; thus the name "entropy" being currently used, since Shannon, for $S(p)$, is it very proper to use it, at the same time, for $H(p)$? It must be pointed, that for the authors, except in the Introduction, *Measure of information* is always taken from the a priori point of view, as the expectation of $H(p)$ and never as the amount of information yielded a posteriori by the realization of the event A.

From the 3 conditions, which "seem intuitively rather natural":

(a) nonnegativity: $H(p) \geqslant 0, \quad \forall p \in]0, 1]$;

(b) additivity for independent events: $H(pq) = H(p) + H(q)$;

(c) normalization: $H(1/2) = 1$, one deduces:

$$H(p) = -\log_2 p.$$

Of course, the proof is based on the uniqueness of the nonnegative (or, what is the same, nondecreasing) solution: $f(x) = cx$, $c > 0$, of the Cauchy functional equation: $f(x + y) = f(x) + f(y)$ on the interval $]0, +\infty]$. The two following sections (0.3) and (0.4) give several useful theorems on the solutions of the Cauchy functional equation and on completely additive number theoretical functions, i.e. functions satisfying $\varphi(mn) = \varphi(m) + \varphi(n)$ for all positive integers m and n. I should like to say that the lecture of these 20 pages will be a delight at the same time for the layman, who will appreciate the clarity, and for the specialist, who will note the fine improvements to several proofs.

Chapter I makes, in great details, a review of several properties of Shannon's entropy H_n $(p_1, \ldots, p_n)$ (1948), extended to "incomplete" experiments $(p_1 + \cdots + p_n < 1)$ by Rényi (1960):

$$H_n\colon \Delta_n \to R^+$$

where

$$\Delta_n = \{(p_1, \ldots, p_n)\colon 0 < \textstyle\sum p_k \leqslant 1\},$$

and

$$H_n(p_1, \ldots, p_n) = \textstyle\sum L(p_k)/\sum p_k.$$

$$\begin{aligned} L(x) &= -x \log x, \qquad && x \in]0, 1], \\ &= 0, && x = 0. \end{aligned}$$

The verification is straightforward, the properties being classified in:
(a) algebraic properties, e.g. additivity

$$\begin{aligned} H_{mn}(p_1 q_1 + \cdots + p_1 q_n, \ldots, p_m q_1 + \cdots + p_m q_n) \\ = H_m(p_1, \ldots, p_m) + H_n(q_1, \ldots, q_n); \end{aligned}$$

(b) analytic properties: e.g.

$$H_n(p_1, \ldots, p_n) \leqslant \log n$$

based on the fact that $L(x)$ is a differentiable concave function.

Chapter II is a first approach to the characterization of Shannon's information; it starts with a long list (61 lines!) of definitions, i.e. the giving of names to 26 relations verified by a sequence of functions $I_n\colon \Delta_n \to R$; e.g. the I_n are said to be n_0-expansible if:

$$\begin{aligned} I_{n_0}(p_1, \ldots, p_{n_0}) &= I_{n_0+1}(0, p_1, \ldots, p_{n_0}) \\ &= I_{n_0+1}(p_1, \ldots, p_k, 0, p_{k+1}, \ldots, p_{n_0}), \\ &\qquad\qquad\qquad k = 1, 2, \ldots, n_0 - 1, \\ &= I_{n_0+1}(p_1, \ldots, p_{n_0}, 0) \end{aligned}$$

for all $(p_1, \ldots, p_{n_0}) \in \Delta_{n_0}$. The 26 properties considered are of course not independent; many of their correlations are given in propositions like (2.3.4): If I_n is 3-recursive and 3-symmetric it is also 2-symmetric and decisive.

In the last section are given connexions with the axiomatic characterization of Hinčin (Khinchine) (1953) and Faddeev (1956); and a result somewhat stronger than that of Faddeev is proved.

The original contribution of the authors to the axiomatization of information theory is the subject matter of Chapter 3; given a probability space $(X, \mathbb{S}, P)$ to every random event $A \in \mathbb{S}$ is associated a real number $I(A)$, the "information contained in A", which depends only upon its probability $P(A)$: $I(A) = f[P(A)]$; now given an event $B \in \mathbb{S}$, to every event $A \in \mathbb{S} \cap B$ the authors define "the relative information of the event A with respect to B", by: $I(A/B) = P(B)f(P(A)/P(B))$ if $P(B) > 0$ and $= 0$ if $P(B) = 0$; the "information contained in A, B" is defined by:

$$I(A, B) = I(A) + I(B/\bar{A}) \qquad (\bar{A} = X - A).$$

Putting $P(A) = x$, $P(B) = y$, the symmetry axiom for $I(A, B)$ implies the "*fundamental equation of information*"

$$f(x) + (1 - x)f(y/(1 - x)) = f(y) + (1 - y)f(x/(1 - y))$$

(first introduced by Tverberg, 1958). An "*information function*" is any solution of this equation satisfying $f(1/2) = 1, f(0) = f(1)$; the Shannon's entropy $S(x) = L(x) + L(1 - x)$ is obviously an "information function"; but it is not unique: the same is true from all $f(x) = xh(x) + (1 - x)h(x)$ where h satisfies $h(xy) = h(x) + h(y)$; $S(x)$ is characterized by some regularity properties, which seem desirable; in 30 pages the authors make an excellent review (as could be expected from such experts on functional equations) of the most useful properties (continuity, nonnegativity, boundedness, $\cdots$), the culminating point being a beautiful proof of Lee's theorem: measurability of $f(x)$ is necessary and sufficient.

Chapter IV is devoted to further characterizations of Shannon's entropy: branching property, sum property, inequality, subadditivity.

To give one example: a sequence of functions $I_n(p_1, \ldots, p_n)$ is said to have the sum property if:

$$I_n(p_1, \ldots, p_n) = \sum_1^n g(p_k), \qquad g\colon [0, 1] \to R.$$

The strongest result, due to Daroczy (1971) reads:

The sequence I_n having the sum property with a function g measurable and $g(0) = 0$, then $I_n(p_1, \ldots, p_n) = H_n(p_1, \ldots, p_n)$ (Shannon's entropy) $n = 1, 2, \ldots$ iff I_n is additive and normalized.

The set of properties under scrutiny seems to me really exhaustive.

In Chapter V entropies are considered as mean values; the main results read as follows: φ being strictly monotonic and $\varphi^*\colon [0, 1] \to R$, defined by $\varphi^*(t) = t\varphi(t)$, $t \in]0, 1]$, $\varphi^*(0) = 0$, the average entropy

$$I_n(p_1, \ldots, p_n) = -\log \varphi^{-1}\left[\sum_1^n \varphi^*(p_n)\right]$$

is additive iff: $\varphi = \log t$ or $\varphi = t^{\alpha - 1}$ $(\alpha > 0, \alpha \neq 1)$, i.e. iff I_n is the Shannon's or the Rényi's entropy.

The introduction of Rényi's entropy leads to a generalization of information functions in Chapter VI: in the fundamental information equation the factors

$(1-x)$ and $(1-y)$ are replaced by $(1-x)^\alpha$ and $(1-y)^\alpha$ $(\alpha \neq 1)$: the limit of the generalized information function when α tends to 1 is the Shannon's information function.

In Chapter VII further generalizations of Rényi's entropy are introduced containing two parameters α, β: if $\beta = 1$ they reduce to Rényi's entropy.

The book of J. Aczél and Z. Daróczy represents the summing-up of a long series of fruitful researches: one has the impression that they have so thoroughly explored the field, that there is little chance for the discovery of really new properties of Shannon's entropy and eventually Rényi's entropy; perhaps this outstanding achievement, discouraging further efforts on the same line, will now stimulate explorations of neighbouring fields, taking account of all the aspects of information out of the scope of the classical theory.

J. KAMPÉ DE FÉRIET

BULLETIN OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 83, Number 2, March 1977

Differentiation of integrals in R^n, by Miguel de Guzmán, Lecture Notes in Mathematics, no. 481, Springer-Verlag, Berlin, Heidelberg, New York, 1975, xi + 225 pp., \$9.50.

Professor de Guzmán's book concerns itself with material which has come, in recent years, to play a fundamental role in the theories of real and complex analysis, Fourier analysis, and partial differential equations. Maximal functions, covering lemmas and differentiation of integrals seem to be at the core of the modern theory of singular integrals, Littlewood-Paley theory, and H^p spaces, as well as many other areas of great interest.

The starting point of the theory is the consideration of the following simple result:

Given $f \in L^1(R^n)$, we have

$$\lim_{r\to 0} \frac{1}{|B(x;r)|} \int_{B(x;r)} f(y)\,dy = f(x)$$

for a.e. $x \in R^n$. (Here $B(x;r)$ is the ball centered at x of radius r, and $|B(x;r)|$ is its Lebesgue measure.) This result, known as Lebesgue's theorem on the differentiation of the integral, is, however, just the beginning of the theory. For, in order to give their proof of this result, Hardy and Littlewood introduced the maximal operator, M, given by

$$M(f)(x) = \sup_{r>0} \frac{1}{|B(x;r)|} \int_{B(x;r)} |f(y)|\,dy, \qquad f \in L^p(R^n), \qquad 1 \leqslant p \leqslant \infty.$$

This maximal operator, which is of fundamental importance in many areas, turns out to be bounded on $L^p(R^n)$ when $p > 1$, and majorizes some of the most important operators in Fourier analysis. For example, the process of taking Cesàro means of Fourier series or Poisson integrals of functions can be