**Universität Bielefeld/IMW**

**Working Papers**
**Institute of Mathematical Economics**

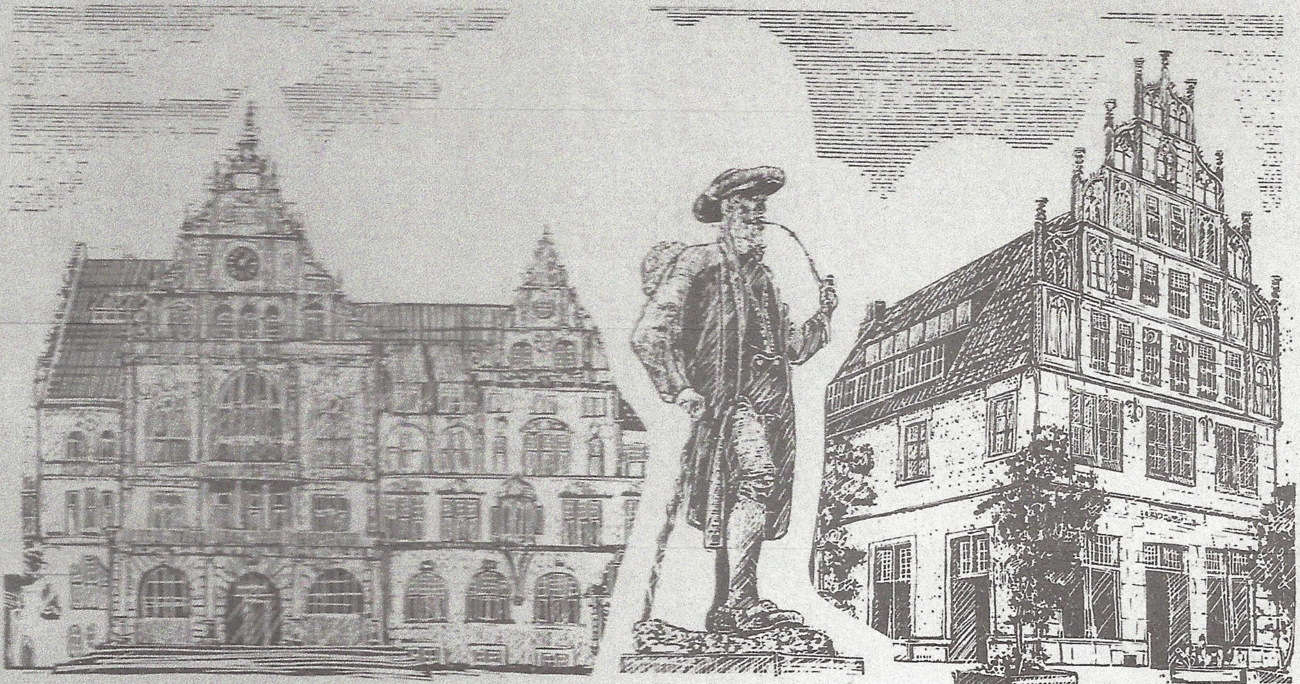
**Arbeiten aus dem**
**Institut für Mathematische Wirtschaftsforschung**

Nr. 38

Hans W. Gottinger

# Lecture Notes on Concepts and Measures of Information

October 1975



H. G. Bergenthal

Institut für Mathematische Wirtschaftsforschung
an der
Universität Bielefeld
Adresse/Address:
Universitätsstraße
4800 Bielefeld 1
Bundesrepublik Deutschland
Federal Republic of Germany

# LECTURE NOTES ON CONCEPTS AND MEASURES OF INFORMATION*

HANS W. GOTTINGER
UNIVERSITY OF BIELEFELD

* These Lectures have been held at the International Summer School on 'New Trends and Open Problems of Information Theory', International Center for Mechanical Sciences (CISM) Udine, Italy, Sept. 3-12,1975. The hospitality of this Center is much appreciated.

# LECTURE NOTES ON CONCEPTS AND MEASURES OF INFORMATION

HANS W. GOTTINGER
UNIVERSITY OF BIELEFELD
P.O.B. 8640
F.R. OF GERMANY

## INTRODUCTION.

In recent years a considerable amount of work has been done on concepts and measures of information within and beyond the fields of engineering and mathematics.

In these notes an attempt is made to trace the main sources and motivations of various approaches to conceptualize and measure information. The development so far showed that information may explain different things in different contexts, hence it will not make sense to apply a general measure of information to practical situations in which information obtains different meanings.

We will start by exhibiting the structure of the Shannon-Wiener theory of information, then, in Section 2, we turn to approaches that give axiomatizations of entropy and information measures without using probability measures. Recently, also A.N. Kolmogorov [1,2] has shown that the basic information-theoretic concepts can be formulated without recourse to probability theory. In Section 3 we outline a specific approach of Domotor in which qualitative information and entropy structures are considered and qualitative conditions are found that permit representation by suitable information or entropy measures. Clearly, this construction finds its roots in problems of model or measurement theory. In Section 4 we essentially expose our own ideas (Gottinger [3,4]) on qualitative information in which information is considered to be a 'primitive concept', separate from probability, e.g. a binary relation in an algebra of informative propositions. This approach suggests a rigorous axiomatic treatment of semantic information. Also we discuss some epistemological aspects of qualitative information, in connection with a general theory of inductive inference. In Section 5,6,7 we are concerned with 'information provided by experiments' as used in statistical decision theory. The concept originated in works of D. Blackwell [5], Blackwell and Girshick [6] and is now extensively used in the statistical decision literature as well as in the related literature in economics.

The intention of this review is to familiarize information theorists with other concepts and measures of information which do not arise from the traditional Shannon theory and are motivated from considerations to handle the many-sided concept of information beyond the engineering-technical viewpoint. It is believed that information theory itself may benefit from these considerations and may even substantially increase its potentialities toward application not restricted to engineering science.

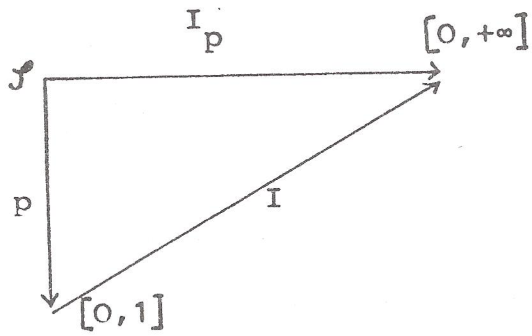
1. THE STRUCTURE OF INFORMATION THEORY.

Shannon's Problem. The abstract problem of information theory established by C.E. Shannon [7] and in a somewhat different form by N. Wiener [8] is this: Given a probability space $(S, \mathcal{S}, P)$, where S denotes the space of elementary events (basic space), $\mathcal{S}$ a $\sigma$-algebra of subsets of S and P a probability measure on $\mathcal{S}$, how much information do we receive about a (randomly selected) point $s \varepsilon S$ by being informed that s is in some subset A of S.

It is relatively easy to see that the answer depends on the dimension or measure of A, given in terms of the probability measure P attached to A. Hence an information measure is a real-valued set function on $\mathcal{S}$ defined by

$$F\,[P(A)] = IoP(A) = I_p(A) \text{ for any } A \varepsilon \mathcal{S},$$

where F denotes some appropriate monotonic transformation. Conceptually, information adopts here the nature of a surprise value or unexpectedness. In this context, note, that $I_p$ is a measurable mapping from $\mathcal{S}$ onto $[0,\infty]$ composed of the measurable mappints $P{:}\mathcal{S} \to [0,1]$ and $I{:}[0,1] \to [0,+\infty]$, with a commutative property. Hence we have the following commutative diagram:




$IoP = I_p$ with I being continuous. It is also natural to assume that $I_p$ is nonnegative and continuous.

Moreover, for any two probabilistically independent events $A, B \in \mathcal{F}$, write $A \perp\!\!\!\perp B \Leftrightarrow AB \times S \sim A \times B$, we have $A \perp\!\!\!\perp B$ $I_p(A \cap B) = I_p(A) + I_p(B)$.

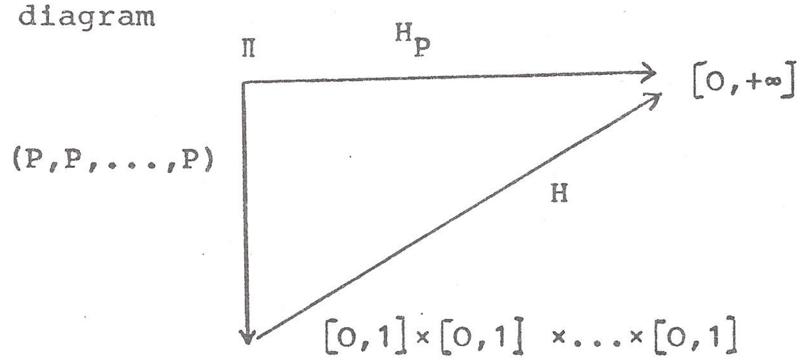
Now it has been shown by Shannon that $I_p$ satisfies the additive representation if it can be represented by $I_p(A) = -c \log_2 P(A)$, where c is any positive real constant (sometimes called Boltzmann's constant in analogy to thermodynamics).

More generally, let $\pi = \{A_i\}_{i=1}^{n}$ be an n-fold uniform partition into finitely many equiprobable events, sometimes referred to as an experiment. Then the natural question arises what would be the average amount of information, called the entropy $H_p$ with respect to a given partition $\pi$. This is computed as

$$(*) \quad H_p(\pi) = \sum_{A \in \pi} P(A) \cdot I_p(A), \text{ and } I_p(A) = -\log_2 P(A),$$

if we choose c, by convention, as unit of measurement.

Let $\Pi$ be the set of all possible partitions of S. The diagram




commutes, that is $H \circ (P,P,\ldots,P) = H_p$, and H is continuous.

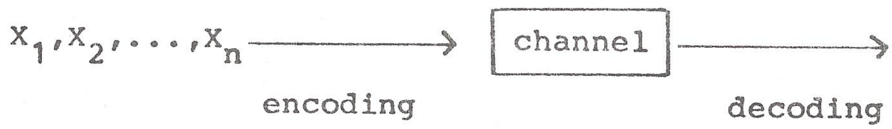
Furthermore, for every $A \in \Pi$ we have $H(\{A,\bar{A}\}) = 1$ if $P(A) = P(\bar{A})$, and $H([B|A \cap B, \bar{A} \cap B]\Pi) = H(\Pi) + P(B) \cdot H(\{A,\bar{A}\})$ if $A \perp\!\!\!\perp B$, where $A, B \in \mathcal{F}$ and $[B|A \cap B, \bar{A} \cap B]\Pi$ is the conditional experiment resulting of replacing B in the partition $\Pi$ by two disjoint events $A \cap B$, $\bar{A} \cap B$. It has been shown by D.K.Fadeev [9], using P. Erdös' [10] number-theoretic lemma on additive arithmetic functions that the only function $H_p$ satisfying the above conditions is of the form (*).

The entropy may be interpreted in various ways, either as an average measure of uncertainty removed or as an average measure of information conveyed. Which interpretation one prefers over the other is irrelevant - as will be clear in the sequel. Thus we see that there is a complete correspondence between uncertainty and information. The definition of information is here naturally tied up to probability, only the existence of the latter enables the measurement of the former.

If we say, roughly, that we have gained information when we know something now that we didn't know before, then it actually means that our uncertainty expressed in terms of probability at one instance of time has been removed at a later instance of time - according to whether the event has occurred or has not occurred. Introducing the notion of a random experiment in a statistical context we may talk about uncertainty before an experiment is carried out, at a moment where we have not yet observed anything, and we may talk about information after having performed the experiment. Sometimes Shannon's measure of information has been termed probabilistic information or selective information (Mackay [11]). There are several approaches (see Rényi [12], and Kolmogorov [2]) how to establish the measure $H_p$, either on pragmatic grounds arising from coding theory or, in an axiomatic way or by starting with the notion of an invariant from ergodic theory. Surprisingly, $H_p$ may even result from gambling theory (Kelly [13]). Shannon's original axioms for the entropy measure have been replaced several times subsequently by weaker conditions (see D.K.Fadeev [9]

A.I. Khinchin [14], H. Tveberg [15], D.G. Kendall[16] and many others). The weakest set of axioms known so far seems to be that given by P.M. Lee[17]. Mathematically, the representation of information involves a study of particular classes of functional equations.

As A.N. Kolmogorov [1] remarked the probabilistic approach seems appropriate for describing and predicting the transmission of (uniform) mass information over (physically bounded) communication channels $C|H$ as illustrated by the following scheme:




where $X_1, X_2, \ldots, X_n$ is a well-defined sequence of random variables (information source). Such kinds of problems are of fundamental importance in the engineering and physical sciences where probability measures can roughly be identified experimentally as limiting frequencies for a sufficiently long (precisely infinite) sequence of trials forming a collective in the sense of von Mises. But what sense does it make to talk about the entropy of receiving messages from a certain novel of Tolstoi, or about the experiments getting married once, twice or even three times?

In other words, can we talk about entropy in the sense of Shannon's theory if we do not have a well-established random sequence forming our information source, if events are not repeatable? Philosophers and linguists consider as a basic flaw of Shannon's measure the fact that the probability measure defined is obviously confined to a frequency interpretation. Hence, to the same extent as probability concepts gave rise to extensive discussions up to recent time probabilistic information theory is affected by these discussions concerning the adequate application of the underlying probability concept (see H.W. Gottinger [18]).

The motivation for Carnap and Bar-Hillel [19] is somewhat different from the theory of transmission of uniform mass information, e.g. the question is how can we evaluate the information provided by a sentence structure which defies representation in a random sequence. In the context of semantic information the concept of an 'ideal receiver' as one with a perfect memory plays a much similar role as that of an ideal 'rational person' assumed in the theory of logical probability due to Carnap.

As a matter of fact it turns out that semantic information theory in the sense of Carnap and Bar-Hillel leads to similar properties as Shannon's entropy measure, however, relative frequences are replaced by logical probabilities (degree of confirmation). If h represents a hypothesis, e evidence, thus c(h,e) the degree of confirmation of a hypothesis h given the evidence e, then by characterizing h as message and e as knowledge the information received from h given e is the greater the more improbable we consider h given e. This again lends itself to the interpretation of information as a surprise value, i.e. information provided by a logical true sentence is zero, and that of a logically false sentence infinity.

The question then naturally comes up as to which extent one can base a theory of prediction on a theory of information that uses a rather restrictive probability concept for real-life situations. This concept only applies to carefully prepared situations of well shuffled decks of playing cards, controlled casts of dice and in random sampling.

The problem to achieve rational predictions or making inferences from data has plagued numerous philosophers since D. Hume (see H.Jeffreys [20]), and this has been reconsidered more recently. It has given rise to a logical theory of probability predominantely based on inductive inference. However, this theory incorporated evidence as conditional probability statements, but it did not show the links between information processing (in the human mind) and probability evaluation. Information only comes in by repeatedly revising conditional probability statements as, for instance, propagated by the Bayesian approach in statistical decision theory. But treatment of information processing is essential for any theory of prediction, and it is genuine for any kind of human judgment. Of course, we cannot dispense with probabilities in a general theory of prediction, for if we can, any such theory would be either completely deterministic or arbitrary. In this I do not share the pessimism of P.J. van Heerden [21], based on the assertion 'that a number expressing the probabilities of the different possible outcomes in real life does not exist' (p.21).

In fact, what we might do is to build a theory of probability on the basis of a completely separate theory of information by generating 'qualitative information', and giving conditions under which numerical probabilities can be established. This procedure would entail a satisfactory theory of prediction.

Some approaches in this direction, including the author's, will be discussed in the next two sections.

Of course, probabilistic information theory, as it stands now, will continue to play a major role in those circumstances in which it makes sense to talk about information in a random sequence which is perfectly legitimate under conditions stated by Kolmogorov [1].

However, its value for general applications beyond those anticipated by Shannon appears to be rather limited.

## 2. INFORMATION WITHOUT PROBABILITY.

In recent years some information theorists were not completely satisfied with probabilistic information theory. The motivation for their dissatisfaction was, of course, different from that of statisticians and philosophers. Although the axiomatics of information theory was considerably refined and weakened, the natural question was raised whether one could develop an information theory without involving the concept of probability (distributions), at least in the basic axiomatic structure. The contribution by R.S. Ingarden and K. Urbanik [22], Ingarden [23,24] answered this question affirmatively. It was the first step to challenge the hitherto accepted view-point that information theory is a branch of probability theory which is also reflected in the organization of textbooks on probability theory (see Rényi [25]). Interestingly enough, the basic motivation evolved from certain experimental situations in physics where it appeared to be meaningful in some situations to talk about information regarding the state of a system (e.g. the entropy of some macroscopic system) although its probability distribution is not known.

Formally, Ingarden and Urbanik achieve to define H (the entropy) directly on a pseudo-metric space of finite Boolean rings (of events) satisfying convenient properties of monotonicity and continuity. A. Rényi [26] claimed that these results can be achieved by using Shannon's measure in terms of a uniquely defined conditional probability measure which follows from the Ingarden-Urbanik technique so that defining information without probability inevitably leads to introducing probability at a later stage. Rényi's straight-forward conclusion is that the information cannot be separated from probability. However, this misses the real point. First of all, as stated before, in some experimental situations it makes more sense to start with some basic knowledge, experience, evidence of the experimenter on the state or movement of a system (physical, biological or any system which is subject to experimental control) rather than specifying probabilities in terms of which information should be defined. Second, in a more general context of human decision making or drawing inferences from observations it is often the case that information processing precedes probability evaluations in betting on the future and making predictions on uncertain events. Most reasonable persons would deny that situations of this kind are comparable or even identical with random experiments - as probabilistic information theory does suggest.

Significant work has also been done by J.Kampé de Fériet and B. Forte (see Kampé de Fériet [27] for a summary and exposition of his results) on constructing information measures without using probabilities. Information is defined as a $\sigma$-additive, nonnegative measure (invariant with respect to translations) on a monotone class of events $\mathcal{F}$, retaining its customary nature of 'surprise value' as in probabilistic information theory. The system basically rests on three assumptions:

(1) $I: \mathcal{F} \to \bar{R}^*$, where $\bar{R}^*$ is the extended real line $[0,\infty]$.

The second is a monotonicity assumption in the form:

(2) $[(A,B) \in \mathcal{F} \times \mathcal{F},\ B \subset A] \Rightarrow I(A) \leq I(B),\ A,B \in \mathcal{F}$.

Hence, given the zero element o and the unit element S in $\mathcal{F}$, we have naturally for any $A \in \mathcal{F}$

$$0 \leq I(S) = \inf_{A \in \mathcal{F}} I(A) \leq \sup_{A \in \mathcal{F}} I(A) = I(o) \leq \infty.$$

(2) already suggests the nature of information as a 'surprise

value', in particular $I(S) = I(o) = +\infty$, so that information derived from the sure event is zero, the information derived from the impossible event, provided it happens to occur ('surprise'), is infinite.

An additional assumption imposes the condition of monotone continuity, i.e. for a countable sequence of events $\{A_n, n = 1,2,\ldots\}$ we have either

$$(3)\quad A_n \subset A_{n+1} \Longrightarrow A = \bigcup_1^\infty A_n \in \mathcal{S} \text{ or}$$

$$A_{n+1} \subset A_n \Longrightarrow A = \bigcap_1^\infty A_n \in \mathcal{S}$$

which implies, for information defined, as in (1) and (2):

$$[A_n \in \mathcal{S}, A_n \uparrow A] \Longrightarrow I(A_n) \downarrow I(A)$$

$$[A_n \in \mathcal{S}, A_n \downarrow A] \Longrightarrow I(A_n) \uparrow I(A),$$ known as monotone sequential continuity.

In particular, we have a similar property (as in probabilistic information theory), i.e. for probabilistically independent pairs of events $A_n \in \mathcal{S}, n = 1,2,\ldots$:

$$I(\bigcap_1^\infty A_n) = \sum_1^\infty I(A_n).$$

An immediate consequence of assumption (2) is that information is of Inf-type, e.g. $I(A \cup B) \leq \text{Inf}[I(A), I(B)]\ A,B \in \mathcal{S}$ so that $I(A \cup B) = F[I(A), I(B)]$ where $F$ is a suitable monotone function. This motivates the introduction of a partial operation $T$ of composition $I(A)TI(B) = F[I(A), I(B)]$ which is familiar from the theory of partially ordered algebras and its representation by functional equations. Since $\mathcal{S}$ can be completed to a $\sigma$-algebra we have $I[\bigcup_1^\infty A_n] = \mathop{T}_1^\infty I(A_n), A_n \in \mathcal{S}, \bigcap_1^\infty A_n = o$, hence $T$ is $\sigma$-composable. $T$ then satisfies well-known algebraic properties for defining an abelian semigroup (see Fuchs [28], Chap. 10). It is not difficult to show the connection between information measure and the underlying algebraic structure.

Let T be defined here in terms of the union $\cup$ , and define a dual operation T* in terms of intersection $\cap$ . The presentation of partial operations in terms of the more familiar operations $\cup$ and $\cap$ proves to be convenient in case of forming ordered groups generated by T. We require that the existence of T implies a unique T*, and (T*)* = T, by definition.

The following properties hold for any $A,B \in \mathcal{S}$, T defined, and provided ATB or AT*B exist in $\mathcal{S}$ .

P.1: $A \in \mathcal{S} \Longrightarrow [AT^*A = A \& ATA = A]$ for all $A \in \mathcal{S}$.

P.2: ATB = BTA, and dually for T* for all $A,B \in \mathcal{S}$. (commutativity)

P.3: (ATB)TC = AT(BTC), for all $A,B,C \in \mathcal{S}$ , and dually for T*. (associativity)

P.4: $ATB \Longrightarrow AT^*(ATB) = AT(AT^*B)$ and dually for T* (distributivity)

Define $AT^*B \Longleftrightarrow A \leq\cdot B$ and '$\leq\cdot$' means 'not more probable than' a relation of qualitative probability.

Then it is relatively easy to see that the properties P.1 - P.4 will make $\mathcal{S}$ a lattice-ordered algebra which is also distributive, and hence a Boolean algebra endowed with a binary relation $\leq\cdot$.

Note that it presents no great difficulties to invoke continuity properties for T* so that $\mathcal{S}$ becomes a topological Boolean algebra (in the order topology). Since T is defined in terms of $\cup$ it is then easy to verify that monotone continuity is equivalent to continuity of the partial operation T and any information measure defined is σ-additive on $\mathcal{J}$ .

It is therefore interesting to note here that the introduction of a partial operation suggests certain assumptions about the qualitative ordering in a partially ordered algebra so that information measures are compatible with this ordering. In terms of this construction it would then appeal to be most natural to present information measures as Boolean homomorphisms on $\mathcal{J}$ (see Sikorski [29]).

Finally, we remark that due to the nature of information as a 'surprise value' the compatible information measure is order-reversing (antitone) w.r.t. qualitative probability rather than order-preserving (isotone).

The dual properties of T and T* can be shown in the representation of '$\leq\cdot$' in $\mathcal{J}$ by (information) measures. Since T is commutative it follows that $I(A)TI(B) = I(ATB) = T(I(A),I(B))$, let $I(A)TI(B) \Longleftrightarrow I(A) \geq I(B)$. As defined $AT^*B \Longleftrightarrow A \leq\cdot B \Longleftrightarrow A \subseteq B$ in this case yielding $AT^*B = A$, and conditions for T* are reformulations of axioms for a partial order $\leq\cdot$. Then $I(ATB) = I(A)TI(B) \Longleftrightarrow I(A) \leq I(B) \Longleftrightarrow AT^*B$, by strict compatibility with $A \leq\cdot B$.

## 3. QUALITATIVE INFORMATION AND INFORMATIVE INFERENCE.

Only in very recent time, motivated by related developments in (subjective) probability and measurement theory (see H.W. Gottinger [30]) one has shown some interest to introduce qualitative information as a primitive notion and to show whether and under which conditions we have $A \leq: B \Longleftrightarrow I(A) \leq I(B)$, for all $A,B \in \mathcal{J}$ where the binary relation '$\leq:$' stands for 'not more informative than'. Such a problem is genuine to measurement theory.

We should emphasize here that the problem of compatibility between a qualitative information structure $\langle S, \mathcal{S}, \leq: \rangle$ and a quantitative structure $\langle S, \mathcal{S}, I \rangle$ is important in view of various applications. In behavioral as well as in natural sciences we find ample experimental evidence that there is often no direct way in constructing information (or probability) measures but only indirectly via a qualitative ordering of events, propositions, statements, etc., according to qualitative information (or qualitative probability). It also comes closer to the procedure of how individuals or groups actually evaluate information. Hence it is natural to interpret qualitative information in a subjective sense.

An elaborate axiomatic system of qualitative information (consisting of 15 qualitative axioms) has been proposed recently by Z. Domotor [31]. Postulating an infinite Boolean algebra of events $\mathcal{S}$ he establishes his qualitative information structure by endowing $\mathcal{S}$ with an algebraic independence relation '$\perp\!\!\!\perp$' and the familiar order relation $\leq:$.

Hence, we may introduce a binary relation $\leq:$ on the Boolean algebra of events $\mathcal{S}$ (rather than on the set of partitions as in case of entropy) with the intended interpretation:

A $\leq:$ B means that event A does not convey more information than event B. Then one is interested in necessary and sufficient conditions to have $A \leq: B \Longleftrightarrow I_p(A) \leq I_p(B)$ for all $A, B \in \mathcal{S}$.

The results will lead to information measures in the standard information-theoretic sense. It has been the first attempt so far to derive information-theoretic notions, without recourse to probability theory at the primitive, qualitative level although the link to probability theory becomes obvious at the level of representation by information measures.

By restricting to finite structures the results will be somewhat weaker, e.g. non-unique representations are possible and we will examplify this situation next in view of defining qualitative entropy structures.

For deriving qualitative entropy structures the well-known fact is used (see O.Ore [32]) that any finite structure $\langle S, \mathcal{S}, \lesssim: \rangle$ gives rise to a finite partition - the set of possible partitions is called a class of experiments - so that for any two partitions $\pi_1, \pi_2 \epsilon \Pi$ (one of which is finer or coarser than the other) a qualitative ordering of partitions can be represented by the entropy measure, e.g.

$$\pi_1 \lesssim: \pi_2 \Longleftrightarrow H_p(\pi_1) \leq H_p(\pi_2).$$

Then $\pi_1 \lesssim: \pi_2$ means that experiment $\pi_1$ does not have more entropy than the experiment $\pi_2$.

One could consider '$\lesssim:$' as a linear ordering, e.g. being reflexive, transitive, antisymmetric and connected but these order properties are obviously insufficient to guarantee the existence of a function $H_p$.

Along the lines of Z.Domotor we develop qualitative entropy in an algebra of experiments.

First of all, we need some technicalities regarding partitions or experiments. Given two partitions $\pi_1$, $\pi_2 \epsilon \Pi$ we say that $\pi_1$ is a finer than $\pi_2 (\pi_1 \subseteq \pi_2)$ if for all $A \epsilon \pi_1$ there exists at least one $B \epsilon \pi_2$ such that $A \subseteq B$.

Example

$\{\emptyset,\bar{A},A\bar{B},AB\bar{C},ABC\} \subseteq^{\cdot} \{\emptyset,\bar{A},A\bar{B},AB\} \subseteq^{\cdot} \{\emptyset,\bar{A},A\}$.

Dually, we can define a coarser than relation.

One can introduce some lattice operations in the set of partitions.

We observe that if $\pi_1$ and $\pi_2$ are partitions of the basic space S then also $\pi_1 \cdot \pi_2$ and $\pi_1 + \pi_2$ are partitions in $\Pi$ and the binary operations '.' and '+' determine the inf (g.l.b.) and sup (l.u.b.) respectively, hence satisfy the definition af a lattice.

Define $\pi_1 \cdot \pi_2 \equiv \{A \cap B : A \varepsilon \pi_1 \;\&\; B \varepsilon \pi_2\}$, or, in general,

$\prod_{i=1}^{n} \pi_1 \equiv \{ \bigcap_{i=1}^{n} A_i : A_i \varepsilon \pi_i$ for all i}. Then $\pi_1 \cdot \pi_2$ is the

greatest experiment finer than both $\pi_1$ and $\pi_2$ such that

(i) $\pi_1 \cdot \pi_2 \subseteq^{\cdot} \pi_1$ and $\pi_1 \cdot \pi_2 \subseteq^{\cdot} \pi_2$,

(ii) $\pi \subseteq^{\cdot} \pi_1$ and $\pi \subseteq^{\cdot} \pi_2 \Longrightarrow \pi \subseteq^{\cdot} \pi_1 \cdot \pi_2$.

Dually,define the sum of experiments by

$$\pi_1 + \pi_2 \equiv \{A \cup B : A \varepsilon \pi_1 \& B \varepsilon \pi_2\},$$

or generally

$\sum_{i=1}^{n} \pi_i \equiv \{ \bigcup_{i=1}^{n} A_i$: all $A_i$ are overlapping events in

$\pi_1 \cup \pi_2 \cup \ldots \cup \pi_n\}$.

$\pi_1 + \pi_2$ is the smallest experiment coarser than both $\pi_1$ and $\pi_2$, i.e.

(iii) $\pi_1 \subseteq\cdot \pi_1+\pi_2$ and $\pi_2 \subseteq\cdot \pi_1+\pi_2$

(iv) $\pi_1 \subseteq\cdot \pi$ and $\pi_2 \subseteq \pi \Longrightarrow \pi_1+\pi_2 \subseteq\cdot \pi$.

Finally one defines the partitions $o = \{\emptyset, S\}$ as the maximal experiment, and the partition $U = \{\{s\} : s\varepsilon S\} \cup \{\emptyset\}$ is called the minimal experiment. We have $U \subseteq\cdot \pi \subseteq\cdot o$ for any $\pi\varepsilon\Pi$. The following operations are obvious:

$\pi \cdot o = \pi$ and $\pi + o = o$

$\pi \cdot U = U$ and $\pi + U = \pi$.

In order to extend the lattice structure, induced by these operations, to a full algebraic structure we need to introduce a relation of independence on experiments. Such an independence relation is certainly natural in a probabilistic context, but it is essentially an algebraic property, and applies directly to qualitative (probability or information) structures (see S. Maeda [33] ).

Let $(S, \Pi, \subseteq\cdot)$ be the lattice of experiments, then $\pi_1, \pi_2 \varepsilon \Pi$ are said to be independent, $\pi_1 \perp\!\!\!\perp \pi_2$, if $AB \times S \simeq A \times B$, where $\simeq$ means isomorphic to.

Then $\perp\!\!\!\perp$ satisfies the following properties for all $\pi, \pi_1, \pi_2 \,\varepsilon\Pi$:

(1) $o \perp\!\!\!\perp \pi$,

(2) $\pi \perp\!\!\!\perp \pi \Longrightarrow \pi = o$

(3) $\pi_1 \perp\!\!\!\perp \pi_2 \Longleftrightarrow \pi_2 \perp\!\!\!\perp \pi_1$,

(4) $\pi_1 \perp\!\!\!\perp \pi_2$ & $\pi_2 \subseteq\cdot \pi_3 \Longrightarrow \pi_1 \perp\!\!\!\perp \pi_3$

(5) $\pi_1 \perp\!\!\!\perp \pi_2$ & $\pi_2 \perp\!\!\!\perp \pi_3 \Longrightarrow (\pi_1 \cdot \pi_2 \perp\!\!\!\perp \pi_3 \Longleftrightarrow \pi_1 \perp\!\!\!\perp \pi_2 \cdot \pi_3)$,

(6) $\pi_1 \perp\!\!\!\perp \pi$ & $\pi_2 \perp\!\!\!\perp \pi \Longrightarrow \pi_1 \cdot \pi_2 \perp\!\!\!\perp \pi$, if $A \cup B = S$, $A \varepsilon \pi_1$, $B \varepsilon \pi_2$.

(7) $\pi_1 \perp\!\!\!\perp \pi_2 \cdot \pi_3$ & $\pi_1 \cdot \pi_2 \perp\!\!\!\perp \pi_3 \Longrightarrow (\pi_1 \perp\!\!\!\perp \pi_2 \Longleftrightarrow \pi_2 \perp\!\!\!\perp \pi_3)$,

(8) $\pi \perp\!\!\!\perp \pi_1$ & $\pi \perp\!\!\!\perp \pi_2 \Longrightarrow (\pi \cdot \pi_1 = \pi \cdot \pi_2 \Longleftrightarrow \pi_1 = \pi_2)$,

(9) $\pi_1 \perp\!\!\!\perp \pi_2$ & $\pi_1 \subseteq\cdot\, \pi_2 \Longrightarrow \pi_2 = o$

Furthermore, by applying the independence relation to $\Pi$, we get the set-theoretic operation of complementation out of it, e.g. we then have for any $A \varepsilon \pi$:

$\bar{A} = \{B : A \cup B = S \;\&\; A \perp\!\!\!\perp B$ for all $A, B \varepsilon \pi\}$. Extending this to partitions define

$\bar{\pi} = \prod_{\pi \cdot U = U} \{U\}$ with $\pi_1 \wedge \pi_2 = (\bar{\pi}_1 \cdot \bar{\pi}_2)^-$, $\pi \perp\!\!\!\perp U$, $\wedge$ being a 'meet-operation', in $\Pi$. Then we get a Boolean algebra for a collection of all elements of $\Pi$ for which complements do exist in this way. Therefore, in general, the introduction of $\perp\!\!\!\perp$ into $\Pi$ only induces a Boolean algebra $\mathcal{B}$ firmly embedded in $\Pi$.

Then one finds conditions under which an entropy function (quasi-entropy) exists on the entire $\Pi$.

The main characteristics of a finite qualitative quasi-entropy structure can be collected in the following properties.

Let $(S, \Pi, \leq:, \perp\!\!\!\perp)$ be a qualitative (quasi-)entropy structure, and $A, B, C \varepsilon \mathcal{S}$ be elements of any partition $\pi$ in $\Pi$.

(1) $\pi_1 \subseteq\cdot\, \pi_2 \Longrightarrow \pi_2 \leq: \pi_1$,

(2) $\pi_1 \leq: \pi_2$ and $\pi_2 \leq: \pi_3 \Longrightarrow \pi_1 \leq: \pi_3$

$\pi_1 \leq: \pi_2$ or $\pi_2 \leq: \pi_1$

$\pi_1 <: \pi_2 \Longleftarrow \pi_1 \leq: \pi_2$ and not $\pi_2 \leq: \pi_1$

$\pi_1 \ddot{=} \pi_2 \Longleftarrow \pi_1 \leq: \pi_2$ and $\pi_2 \leq: \pi_1$, $\ddot{=}$ is an equivalence relation,

(3) $o \leq: \pi$ for any $\pi \in \Pi$ and $o <: \pi$ if $\pi \ddot{=} E$, where $E = \{A, \bar{A}\}$ denotes Bernoulli experiments.

(4) $o \leq: \{A, \bar{A}\} \leq: \{A, \bar{A}B, \bar{A}\bar{B}\} \leq: \{A, \bar{A}B, \bar{A}\bar{B}C, \bar{A}\bar{B}\bar{C}\}$

$\leq: \ldots \leq: U \Longleftarrow U \subseteq \ldots \subseteq \{A, \bar{A}B, \bar{A}\bar{B}\} \subseteq \ldots \subseteq o.$

(5) $\pi_1 \leq: \pi_2 \Longleftrightarrow \pi_1 \cdot \pi \leq: \pi_2 \cdot \pi$, if $\pi \perp\!\!\!\perp \pi_1, \pi_2$,

(6) $\pi_1 \leq: U_1$ and $\pi_2 \leq: U_2 \Longrightarrow \pi_1 \cdot \pi_2 \leq: U_1 \cdot U_2$

if $\pi_1 \perp\!\!\!\perp \pi_2$ and $U_1 \perp\!\!\!\perp U_2$.

(7) $\pi_1 \ddot{=} \pi_2$ and $\pi_2 \cdot \pi_3 \leq: \pi_4 \Longrightarrow \pi_1 \cdot \pi_3 \leq \pi_4$

if $\pi_3 \perp\!\!\!\perp \pi_1, \pi_2$.

(8) <u>If for all</u> $U_o, U_1, \ldots, U_n, \pi_o, \pi_1, \ldots, \pi_n$

<u>in</u> $\Pi$ <u>it is true that</u>

$\pi_i \leq: U_i$ <u>for all</u> $i, 0 \leq i < n$ <u>and</u> $\prod_{i \leq n} \bar{\pi}_i \ddot{=} \prod_{i \leq n} \bar{U}_i$, $\pi_i \underset{i \leq n}{\perp\!\!\!\perp}$ & $U_i \underset{i \leq n}{\perp\!\!\!\perp}$.

<u>Then</u> $\pi_n :\geq U_n$.

Condition (8) is a kind of consistency condition for the representation of a qualitative structure and it is due to D. Scott [34]. This condition essentially forces $\Pi$ to be a Boolean algebra in the relevant elements being complemented.

Having stated some properties in a qualitative entropy structure one is led to represent this structure by a compatible entropy function. The situation is analogous to the problem of representing qualitative probability by compatible probability measures.

Suppose $(S, \Pi, \leq:, \perp\!\!\!\perp)$ exists. Then there is a (quasi) entropy function $H: \Pi \to Re$, representing the structure such that for all $\pi_1, \pi_2 \in \Pi$:

(i) $\pi_1 \leq: \pi_2 \iff H(\pi_1) \leq H(\pi_2)$,

(ii) $\pi_1 \perp\!\!\!\perp \pi_2 \implies H(\pi_1 \cdot \pi_2) = H(\pi_1) + H(\pi_2)$,

(iii) $\pi_1 \subseteq \pi_2 \implies H(\pi_2) \leq H(\pi_1)$,

(iv) $H(o) = 0, H(\bar{o}) = \infty$,

(v) $H(\pi) = 1$, if $\pi \ddot{=} E$ and $E$ denotes all equiprobable experiments in $\Pi$.

Conditions (i) - (v) are appropriate adjustments of similar conditions used for representation theorems of qualitative probability, hence basically the same proof techniques can be used for proving the theorem (see D. Scott [34], H.W. Gottinger [30]). Condition (ii) constitutes the most important property of entropy, e.g. additivity. Domotor [31] provides a sufficiency proof of the theorem that is in the spirit of Scott's result, it involves an interesting technique.

Let us sketch the technique by translating everything into a geometric picture.

We associate to every partition $\pi$ a vector of partitions $\hat{\pi}$, all of these are to be considered linearly independent and $\hat{\pi} \varepsilon V(\hat{\Pi})$ where $\hat{\Pi}$ is the basis of the n-dimensional vector space $V(\hat{\Pi})$ depending on the number of linear independent $\hat{\pi}$'s. Then make $\Pi$ a finite subset of $V(\hat{\Pi})$ and extend the ordering to $V(\hat{\Pi})$. Scott's result on representation of ordered, finite-dimensional real vector spaces by linear functionals immediately applies and hence we can find a linear functional $\psi : V(\hat{\Pi}) \to R_e$ and a compatible functional $V: \Pi \to Re$ satisfying properties (i) - (v) such that $\psi(\{A,\bar{A}\}) > o$ and the quasi-entropy function is normalized by

$$H(\pi) = \psi(\pi) / \psi(\{A,\bar{A}\}).$$

If we want to establish a connection between the Shannon-Wiener entropy and entropy derived from qualitative entropy structures it would be necessary then to investigate the interrelationship between qualitative probability and qualitative information. The reason is seen in that the entropy function decomposes itself into representations of probability and information. The result obtained so far is rather weak since the prefix 'quasi' should indicate that the entropy representation is not unique. It is an open problem to find unique entropy functions derived from qualitative structures. It is rather obvious that the answer to this problem depends on imposing further structural conditions on the relations ' $\underline{\parallel}$ ' and ' $\leq\cdot$ ' as well as on the interaction between both. There also have to be interconnecting conditions between qualitative entropy and qualitative probability structures. However, the axiomatic setup for this is awfully complicated and we will not dwell upon this question here. Some hints in this direction have been provided by Domotor.

The results can be extended to conditional entropy in which the underlying qualitative structure will be a difference structure. No further technical difficulties are encountered.

## 4. QUALITATIVE SEMANTIC INFORMATION AND INFORMATIVENESS.

It has been observed by Shannon [7] that the 'semantic aspects of communication are irrelevant to the engineering problem'. As noted earlier, a semantic theory of information along Shannon's line has been provided by Carnap and Bar-Hillel[19]. Unfortunately, the latter theory does not cover aspects that are important in semantic information-processing, e.g. those related to informativeness, information-content of statements, propositions, sentences, etc. Surprisingly, not much work has been done on this over the last two decades.

R. Wells [35] made a step into the right direction and developed a semantical theory of informativeness based on qualitative comparison of propositions, sentences according to information-content. One may again point out that the nature of 'unexpectedness' covers only one aspect of information, another aspect would be to order sentences according to informativeness provided they are logically true. For example, 'the Pythagorean theorem is more informative than the proposition 7 + 5 = 12'. We will pursue the latter aspect in a modified form, substituting logical truth of a sentence by 'occurrence of an event'.

In what follows we will outline one approach (Gottinger[3,4]) toward semantic information which is quite different in spirit and method from the original Carnap-Bar-Hillel approach.

It is basically assumed that information is prior to probability, and hence is more fundamental than probability justifying a separate treatment.

This approach is built upon the following fundamentals:

(1) We deal analytically with 'information' as we deal with events as elements of an abstract algebraic structure, similarly as events are treated as 'undefined terms' in a Kolmogorov-Halmos probability algebra.

(2) We confine the notion of information to information pertaining to events (to occur or not occur) since we are interested in making predictions on future events.

(3) Qualitative information is introduced as a binary relation in an information structure and is meant as 'not more informative than' w.r.t. occurrence of the respective event.

(4) A qualitative (subjective) information structure generates a qualitative (subjective) probability structure by a Boolean homomorphism. This is to make precise the idea that a person will evaluate 'qualitative probability' only via qualitative information.

(5) The theory proposed here is entirely based on a theory of inductive inference.

(6) The ultimate goal is to measure probability in terms of information and not vice versa as suggested by standard information theory.

Note that our approach will exhibit an information measure which is nonnegative and bounded on the unit interval, and therefore is quite different from the definition of information as a surprise value. Consider quantitative information as a nonnegative normed measure on a Boolean algebra of propositions (the information structure) to be ordered according to information content. Assume the value one at the unit element and generate via a Boolean homomorphism a probability measure on a Boolean algebra of events.

Definition 1: The triple $(S, \mathcal{T}, \preceq)$ is called a qualitative (subjective) probability structure (QPS) if

(1) $\preceq$ is a partial order (p.o.) on $\mathcal{T}$.

If for $A,B,C \in \mathcal{T}$, $A \perp\!\!\!\perp B, B \perp\!\!\!\perp C$ then

(2) $A \preceq B \Longleftrightarrow A \cup C \preceq B \cup C$.

(3) $O \preceq A$, $A \in \mathcal{T}$ and $\exists$ O, $O \prec S$, O being the nullevent and S the universal event.

Now we assume that '$\preceq$' is a derived concept and that the basic concept will be a relation '$\lesssim$' (qualitative information) defined in a family of information sets or information structure $T_o$ generating an algebra of events. In order to consider a qualitative (subjective) information structure (QIS) we need the following

Definition 2: $(E, T_o, \lesssim)$ is called a QIS if

(1) $\lesssim$ is a particular order relation in $T_o$ to be specified.

(2) If $\underline{A}, \underline{B}, \underline{C}$ in $T_o$, $\underline{A} \perp\!\!\!\perp \underline{B}, \underline{B} \perp\!\!\!\perp \underline{C}$ then

$\underline{A} \lesssim \underline{B} \Longleftrightarrow \underline{A} \cup \underline{C} \lesssim \underline{B} \cup \underline{C}$

(3) For any fixed $X \in \mathcal{T}$, $\underline{O}_X \lesssim \underline{A}$ for all $\underline{A} \in T_o$,
for any fixed $Y \in \mathcal{T}$, $\underline{O}_X < E_Y$,

where $E = \bigcup(\underline{A}, \underline{B}, \underline{C}, \ldots)$, and $\underline{O}_X$ means neg-information with respect to some event $X \in \mathcal{T}$, $E_Y$ means universal information with respect to some $Y \in \mathcal{T}$.

We say a QIS H-generates ($\xrightarrow{H}$) a QPS, $T_o \ni \underline{A} \xrightarrow{H} H(\underline{A})$ , if there exists a Boolean homomorphism H that is order and structure-preserving with respect to the operations $\wedge$, $\vee$, $-$ (finite meet, join and complementation).

Suppose we want to have

$\underline{A} \lesssim \underline{B} \Longrightarrow \underline{A} \subseteq \underline{B}$ and $\underline{A} \lesssim \underline{B} \Longrightarrow H(\underline{A}) \Leftrightarrow H(\underline{B})$,

so that qualitative information may be put in terms of inclusion.

However, here we face the difficulty that $\underline{A}, \underline{B}$ may contain different semantic information, and the inclusion relation may not hold. One way out of this involves the construction of a <u>standardized</u> structure T, order-isomorphic to $T_o$. We need the following

<u>Definition</u> 3:

(1) $H : Q_X \rightarrow X \doteq o$ (neg-information generates a quasi-null event)

(2) $H : E_Y \rightarrow Y \doteq S$ (universal information generates a quasi-sure event).

The relation $\doteq$ indicates qualitative equiprobability

Then $(S, \mathcal{F}, \lesssim)$ is considered to be a qualitative probability structure which does not permit, in its representation, strictly positive measures. By this process one is able to derive a qualitative standardized information structure (QSIS) $(I, T, \lesssim)$ order-isomorphic to a QIS with L being the zero and I the unit element such that $L \lesssim I \Longrightarrow L \subseteq I$. This type of ordering corresponds to ordering of attributes according to Boole's First Law resulting in a Boolean interval algebra, and in which L, I constitute lower and upper bounds, respectively.

In the next steps we are going to show that imposing on T some weak order conditions will make T a lattice-ordered algebra being equivalent to some kind of Boolean algebra.

One interesting formal aspect of this approach reveals an exposition of binary relations that play an important role in the study of topological structures, called *topogeneous structures* according to A. Császar [36].

Let $(I,T,\underset{\sim}{<})$ be a QSIS. Then $\underset{\sim}{<}$ satisfies the properties of a semi-topogeneous order (STO) if

(1) $L\underset{\sim}{<}L$, $I\underset{\sim}{<}I$ for $L,I\in T$, hence $\underset{\sim}{<}$ is reflexive,

(2) $\alpha\underset{\sim}{<}\beta \Longrightarrow \alpha\subseteq\beta$ , $\alpha,\beta\in T$

(3) $\alpha\subseteq\alpha'\underset{\sim}{<}\beta'\subseteq\beta \Longrightarrow \alpha\underset{\sim}{<}\beta$, $\alpha',\beta'\in T$

(4) Furthermore, $\underset{\sim}{<}$ is supposed to be symmetric, i.e. there exists a complementary order $\underset{\sim}{<}^{-}$ with $(\underset{\sim}{<}^{-})^{-}=\underset{\sim}{<}$ such that (1) - (3) can be reformulated in terms of $\underset{\sim}{<}^{-}$, for example $\alpha\underset{\sim}{<}^{-}\beta \Longrightarrow I-\beta\underset{\sim}{<}I-\alpha$.

Conditions (1) and (2) are rather weak, (3) is a kind of transitivity, (4) has far-reaching structural consequences. Finally, it is shown in Gottinger [3] that a STO in $(I,T,\underset{\sim}{<})$ can naturally be extended to a *topogeneous* order (TO) provided the following condition is satisfied:

(5) $\alpha_i\underset{\sim}{<}\beta_j \Longrightarrow [\bigwedge_{i=1}^{n} \alpha_i \underset{\sim}{<} \bigwedge_{j=1}^{m}\beta_j \,\&\, \bigvee_{i=1}^{n}\alpha_i \underset{\sim}{<} \bigvee_{j=1}^{m}\beta_j]$, i.e. the ordering is preserved under lattice operations.

The main result can be obtained after several intermediate steps (see Gottinger [3,4]).

*Let $\underset{\sim}{<}$ be a STO generated by a semi-topogeneous structure* T. *For $\underset{\sim}{<}$ to be a symmetric TO it is necessary and sufficient that* T *is a Boolean algebra.*

Rough sketch of proof. $\Longleftarrow$ is obvious since every Boolean algebra implies a symmetric TO via its partial order. $\Longrightarrow$ 1. Complementarity. It is known that to every STO$\lesssim$ generated by T there is associated a complementary order $\lesssim^-$ generated by $T^- \ni I-\tau\,(\tau \varepsilon T)$. Let $\alpha \lesssim^- \beta, \alpha' \lesssim^- \beta'$, for a symmetric TO$\lesssim$ each of the following statements is true:

(1) $[\alpha \lesssim^- \beta \ \& \ \alpha' \lesssim \beta'] \Longrightarrow \alpha \wedge \alpha' \lesssim \beta \wedge \beta'$

(2) $|\alpha \lesssim^- \beta \ \& \ \alpha' \lesssim \beta'| \Longrightarrow \alpha \vee \alpha' \lesssim \beta \vee \beta'$

if T is a lattice. T has a L and I element and therefore via finite join and meet operations T is relatively complemented and hence complemented.

2. Distributivity. It can also be shown that T is distributive since because of symmetry and (1) we have:

(3) $I - (\beta \vee \beta') = (I-\beta) \wedge (I-\beta') \lesssim (I-\alpha) \wedge (I-\alpha')$
$= I-(\alpha \vee \alpha')$

and therefore $\alpha \vee \alpha' \lesssim \beta \vee \beta'$ which coincides with the conclusion in (2).
Applying de Morgan's Law we have
$I-(\alpha \vee \alpha') = (I \wedge \alpha^-) \wedge (I \wedge \alpha'^-) = \alpha^- \wedge \alpha'^-$,
likewise for $I-(\beta \vee \beta')$. Consequently, $\beta^- \wedge \beta'^- \lesssim \alpha^- \wedge \alpha'^-$ and there exists $\lesssim^-$ such that we get

$I-(\alpha^- \wedge \alpha'^-) = I \wedge (\alpha \vee \alpha') = (I \wedge \alpha) \vee (I \wedge \alpha') = \alpha \vee \alpha'$,

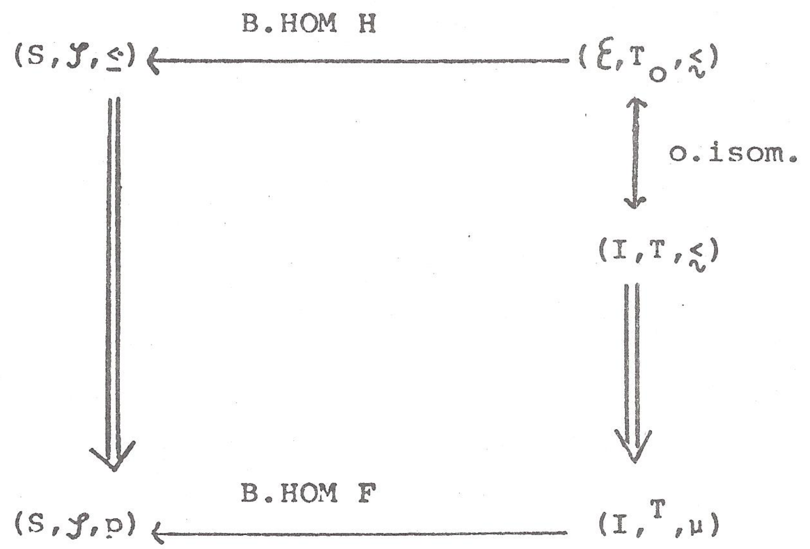
analogously for $\beta$, hence $\alpha \vee \alpha' \lesssim \beta \vee \beta'$.
Similarly, (2) implies (1), namely again by applying de Morgan's Law

(4) $I-(\beta \wedge \beta') = I \wedge (\beta^- \wedge \beta'^-) = (I-\beta) \vee (I-\beta') \lesssim (I-\alpha) \vee (I-\alpha')$
$= I-(\alpha \wedge \alpha')$.

All remaining properties can easily be derived, by analogy, hence T is a complemented, distributive lattice, i.e. a Boolean algebra.

On the basis of this and similar results it is possible to obtain representation theorems in order to construct finitely additive information measures strictly compatible with the QIS. Standard measure-theoretic results by Horn and Tarski [37], Kelley [38] and D. Kappos [39] can be used in this regard.

Then every measure space (I,T,μ) that exists by the representation will induce a probability space via a Boolean homomorphism F. By the analogy of measures and Boolean homomorphisms one can construct a mesure F oμ = P whose properties are shown in the diagram below:



## 5. INFORMATION IN STATISTICAL DECISION THEORY.

The structure of a statistical game can be outlined as follows. The statistician plays a game against nature, at his disposal is a class A of possible actions which he can take (or decisions he can make) in view of the unknown state of nature (nature's pure strategy) $s \epsilon S$. (By a quick change of notation we now consider S as the set of states of nature.)

He may decide to take an action without experimentation (e.g. without 'spying' on nature's strategies) and for doing this may incur a numerical loss $L(s,a)$. The possibility of performing experiments does exists, thus reducing the loss by gaining at least partial information about s. Therefore the concept of information in this context is naturally tied up with payoff-relevance, any bits of information that do not reduce the loss are considered irrelevant.

What prevents the statistician of getting full knowledge of s is the cost of experiments. This cost may assume specific functional forms, but, in general, is considered to be proportional to the number of experiments. Technical definitions are needed in order to look at the general structure of a statistical game. Let Z be the space of outcomes of an experiment, then a function p is defined on $Z \times S$ such that for a fixed $s \epsilon S$ $p_s$ is a probability distribution. The triple $\mathcal{Z} = (Z,S,p)$ is sometimes referred to as the sample space, in general, one does not distinguish between $\mathcal{Z}$ and Z and both may refer to sample spaces. For every subset $A \subset Z$, the probability of the event A is given by

$$P_s(A) = \sum_{z \epsilon A} p_s(z),$$

and P satisfies all properties of a probability measure. A function $d \epsilon D$, defined on Z mapping $Z \xrightarrow{\text{into}} A$ is called a decision function such that $d(z) = a$.

A risk function is represented by expected loss, i.e. a function R on $S \times D$:

$$R(s,d) = \sum_{z \in Z} L(s,d(z))p_s(z) = E[L(s,d(z))].$$

Now a <u>mixed</u> or <u>randomized</u> strategy for nature is the same as a prior probability distribtuion (for the statistician) on the set of states of nature S, denoted by $\mu \in \mathcal{M}$.
The problem of collecting information in a statistical game may be generally posed as follows: Does there exist a partition of Z such that every possible risk attainable with a complete knowledge of $z \in Z$ is also attainable with only the information that z belongs to a set of this partition? Such partitions, if they exist, are as informative as the entire sample space. They are given by the principle of sufficiency.

We are concerned with information provided by an experiment. An experiment X is completely described by a random variable associated with the sample space (Z,S,p) giving rise to a set of conditional probability distributions for every possible parameter (state of nature) $s \in S$. X might be of fixed sample size or of a sequential type where the experimenter may collect observations finitely many times. To set up the problem assume you (the experimenter, the statistician or generally the decision-maker) are confronted with an uncertain situation where you wish to know about the true value of a parameter (state of nature) $s \in S$. Of course you can make some wild guesses, but you can only gain knowledge about the true state by experimentation. Let $\mu$ be some prior probability distribution of the true state s which indicates the amount of uncertainty or ignorance on your part. (Adopt a Bayesian viewpoint that such $\mu$ always exists and is non-null.)

Then the information provided by X may be verbally expressed as the difference between the amount of uncertainty you attach to the prior distribution and that amount of your expected uncertainty of the posterior distribution (after having performed X), i.e. it reflects the residual value of your uncertainty (reduced).

More technically, let $\mathcal{M}$ be the set of prior probability distributions over S (i.e. the space of randomized strategies for nature), define u as a nonnegative, real-valued measurable function on $\mathcal{M}$ which for obvious reasons should be concave, i.e. decreasing with increasing observations. Then $u(\mu)$ represents the amount of your uncertainty (before experimentation) when your distribution over S is $\mu$. In some cases the uncertainty function u is just equivalent to a risk function in a statistical decision problem, in other cases it can be directly assigned. Under specific circumstances it can be identified with the entropy function or can assume some other form that is compatible with its properties of nonnegativity and concavity. Now performing X and observing values of X you may specify a posterior distribution $\phi(X)$, then your measure of information I is determined by

$$I(X,\mu,u) = u(\mu) - E\left[u(\phi(X))\,|\,\mu\right]$$

where $\phi(X)$ is usually obtained by an appropriate application of the well-known Bayes' theorem to get the posterior distribution (see Blackwell & Girshick [6],Chap.3, de Groot [40,41]).

It is usually assumed for reasons of non-triviality that most experiments provide information and that any experiment being more informative than another is also preferable to the other. Therefore, for any given uncertainty function, I is nonnegative, and also for reasons of convenience, continuous.

As we clearly recognize this measure of information provided by an experiment relative to the specification of u and μ naturally evolves from a model of statistical decision. Usually the determination of the uncertainty function hinges upon the loss structure in a statistical game, this becomes clear when we describe comparisons of experiments according to informativeness.

## 6. COMPARISON OF INFORMATION.

In the relevant literature of statistical decision theory comparison of experiments are sometimes confined to those which can be represented by Markov matrices (in this context as information matrices). This is very natural in terms of viewing it in the context of a statistical game. Assume the experiment E produces N distinct value $e_1,\ldots,e_N$ (signals, observations) and let be $S = (s_1,\ldots,s_n)$. Then the experiment E can be represented by an $n\times N$ Markov matrix $\underline{P} = (p_{ij})$ associated to the sample space (Z,S,p) such that $p_{ij} = \underline{P}_{s_i}(\{s:E(z) = e_j\}), p_{ij}\geq 0$ and $\sum_j p_{ij} = 1$ for each i.

Henceforward, for reasons of simplicity, we adopt an earlier definition of D.Blackwell [5] in terms of defining experiments as random variables. Incidentally, Blackwell's definition is one of the first better known definitions of 'comparative informativeness' in the context of statistical decision theory. Every experiment E associated to a sample space generates a risk function. Then, according to Blackwell, an experiment E is more informative than another experiment E' (the set of experiments being partially ordered) if the risk obtained from E is at least obtainable also from E'. $E > E' \Longleftrightarrow R(E) \leq R(E')$. In other words, the numerical risk for E' is at least as large as that of E. If the distribution over S is known then comparative informativeness hinges upon the loss structure of the decision problem. This immediately gives rise to an economic view on the evaluation of information. There were further generalizations and improvements of Blackwell's results in recent years, in particular, in connection with the foundations of statistical inference ('informative inference') (see A. Birnbaum [42]); these results give the main motivations for economic studies on the subject (see C.B.McGuire [43] and Gottinger [44]). This view has been originated and consistently pursued by J. Marschak*). His 'value of information' $V(\eta)$ (attached to experiment $\eta$) w.r.t. any probability distribution $\mu$ over S (his space of events uncontrollable to the decision maker) and his benefit function $b : S \times A \to Re$ is just the converse value of Blackwell's risk $R(E)$, this is due to the fact that economists prefer to talk about benefit or utility, whereas statisticians are more pessimistic and talk about losses. Note, again, that the risk function is completely specified by a probability distribution over S and a loss function on $S \times A$.

One can easily see the strong agreement between information provided by an experiment and the value of information by considering an experiment as a Markov matrix. In this case null-information corresponds to identical rows in the Markov matrix, i.e. any observations made through an experiment is independent of any state of nature.

Accordingly, the risk function obtained by the less informative experiment is larger in value than the risk function obtained by the more informative experiment. It is obvious that the dual statement holds if we deal with an economist's benefit function instead of a statistician's loss function.

---

*) See next Section

We have learned of different, but interelated characterizations of an experiment, either as a partition of Z, as a random variable and in particular as a Markov matrix. Comparative informativeness in terms of partitions of Z given by the principle of sufficiency has also been studied by Blackwell & Girschick [6] (Chap.8).

Let X and Y be two experiments whose values are in the sample spaces, denoted by $Z_X$ and $Z_Y$, respectively.

Then experiment Y is sufficient for experiment X if there exists a nonnegative function h on the product space $Z_X \times Z_Y$ satisfying the following relations

(i) $f_X(x|s) = \int_{Z_Y} h(x,y) f_Y(y|s) d\mu(y)$ for $s \in S$ and $x \in Z_X$,

(ii) $\int_{Z_X} h(x,y) d\mu(x) = 1$ for $y \in Z_Y$

(iii) $0 < \int_{Z_Y} h(x,y) d\mu(y) < \infty$ for $x \in Z_Y$.

h is a stochastic transformation from Y to X. For each fixed value $y \in Z_Y$ the function $h(\cdot, y)$ is a generalized probability density function on $Z_X$. Since this function does not involve the parameter s, a point $x \in Z_X$ could be generated according to the generalized probability density function by means of an auxiliary randomization.

Thus, Y is sufficient for X, if regardless to the value of the parameter s, an observation on Y and an auxiliary randomization make it possible to generate a random variable which has the same distribution as X. The integrability condition on h in (iii) is introduced for technical convenience only.

If Y is sufficient for X then the statistician is strongly advised not to perform the experiment X when Y is available. In fact, one can prove that the sufficient experiment Y must be at least as informative as the experiment X.

Suppose that experiment Y is sufficient for the experiment X. Then, for any uncertainty function u and any posterior distribution $\phi$

$$E|u(\phi(X) \geq E|u(\phi(Y))|$$

The proof of this result is straight-forward and can be found in M. de Groot [45].

## 7. ECONOMIZING INFORMATION.

Despite strong trends in economics and related behavioral science in recent years to use basic results of information theory for their purpose some serious doubts have been expressed concerning the usefulness of $H_p$ for application in economics and for decision-making in general.*) Among others, J. Marschak [47] argues that Shannon's entropy does not tell us anything about the benefit of transmitting information since it assumes equal penalty for all communication errors. What he instead has in mind is a concept of behavioral information processing in an economic system, in particular as

---

*) R.A. Howard [46] has put it this way: '...If losing all your assets in the stock market and having whale steak for supper, then the information assciated with the occurrence of either event is the same. Attempts to apply Shannon's information theory to problems beyond communications have, in large, come to grief. The failure of these attempts could have been predicted because no theory that involves just the probabilities of outcomes without considering their consequences could possibly be adequate in describing the importance of uncertainty to a decision maker.'

related to an economic theory of teams (Marschak and Radner[48]). From an economic viewpoint information may be regarded as a particular kind of commodity which will be traded at a certain price yielding benefits for consumers and causing costs for producers. Hence, the economic theory of information (still in its infancy) is an appropriate modification of the approach used in statistical decision theory. To put it in other terms, here we are interested in the economic aspect of usefulness of information (based on some kind of utility or loss function) rather than in the (original) engineering viewpoint of transmitting and controlling information flows through a large (noisy or noiseless) communication channel. As a digression more recently some information theorists tried to remedy the flaw or restrictiveness of 'equal penalty of all communication errors' by weighing entropy in terms of utility. For any partition $\pi$ of Z they attach to every $A\epsilon\pi$ a utility such that the entropy is given by $H(\pi) = \sum_{A\epsilon\pi} U(A)\cdot P(A)\cdot I_p(A)$ (see Belis & Guiasu [49] and for a further elaboration Guiasu[50]). U(A) satisfies well-known properties of expected utility, i.e. it is for a given preference pattern on $\pi$, unique up to positive linear transformations. It is clear that this proposed measure makes sense if the amount of information to be transmitted through a channel exceeds its upper (physical) bound so that a subjective evaluation procedure (via a utility function) reduces irrelevant information. In this approach there is no obvious relationship between the utility of the message and the information derived from the message, and therefore both should be measured separately. Let $p_1$ be the message 'you will receive five dollars', and let $p_2$ be 'you will receive five dollars and you will be shot'. Clearly, $p_2$ is at least as informative as $p_1$ but $p_2$ is hardly as desirable as $p_1$. One could even attach utility to the sources so that the encoder could select only those sources which are

useful to the encoder (see Longo [51]). The approach has been generalized by introducing explicitly a cost function, that is really a tradeoff function being dependent on the length of code words associated to the message (letter) and on the utility of the message. Clearly, the cost is increasing in the first variable but decreasing in the second. The tradeoff function is uniquely fixed as soon as the utility and the cost of coding are determined.
Assessment of the tradeoff and utility function is treated separately. An optimization principle is involved by minimizing the expected tradeoff so determined.

Let us now sketch the basic ingredients of Marschak's approach as discussed in detail in Marschak [47], (see also Marschak [52,53,54]). Information processing is defined as $P = \langle X, Y, \eta, \kappa, \tau \rangle$, where X is a set of inputs, Y is a set of outputs, $\eta$ a transformation from X to Y, $\kappa$ is a cost function on X, and $\tau$ a time-delay function on X. If you consider, as Marschak does, information as an economic good there is sufficient motivation for looking at the economic system as a mechanism (machine); producing and processing information over time which involves costs and delays. In this respect, information processing is indistinguishable from the processing of physical commodities in, say, a transportation network. ( A convenient theoretical framework for studying such processes would be the well-established algebraic theory of sequential machines, and Marschak's approach motivates a study of these machines for economic information processing, please see in this regard Gottinger [55].)

On the other hand, Marschak's approach is firmly embedded in the general statistical decision model. We might then conceive $\eta$ as a stochastic transformation from the random set X (the space of events, non-controllable to the decision-maker) to the random set Y (the space of available and feasible decision acts), $\eta : X \rightarrow Y$ then establishes a strategy (action). Up to now we have described the particular case of a one-link processing chain, more generally, we may conceive a (time) sequence of information processing $P^1, \ldots, P^N$ such that for every $n = 1, \ldots, N$ we write $P^n = \langle X^n, X^{n-1}, \eta^n, \ldots \rangle$ and $\eta^1 : X^1 \rightarrow X^2, \ldots, \eta^{n-1} : X^{N-1} \rightarrow X^N$ describe experiments with nature, for example. Hence $X^2, \ldots, X^N$ may be referred to as sets of observations or data, whereas $\eta^n : X^N \rightarrow X^{N+1}$ forms a strategy of the decision-maker. A chain of information processing is an information system à la Marschak. Some extensions of this viewpoint with respect to particular organizational forms (such as multi-person control systems, extensive games and dynamic teams) have been recently given by the author [56].

ACKNOWLEDGEMENT:
I am deeply indebted to Professor G. Longo (Trieste) for encouragement and to Mrs. Goergel (IMW-Bielefeld) for superb typing of the manuscript.

## REFERENCES


1. Kolmogorov,A.N., 'Three Approaches to the Definition of the Concept "Quantity of Information"', Problemy Poredaci Informacii 1, 1965, 3 (in Russian)

2. Kolmogorov, A.N., 'Logical Basis for Information Theory and Probability Theory', IEEE Trans. Information Theory IT-14,1967,662

3. Gottinger, H.W., 'Qualitative Information and Comparative Informativeness', Kybernetik 13,1973,81

4. Gottinger, H.W., 'Konstruktion subjektiver Wahrscheinlichkeiten', Math. Operationsforschung u.Statist. 5, 1974,50

5. Blackwell, D., 'Comparison of Experiments', Proc.2nd Berkeley Symp.Math. Statist. and Probability, 1953,93

6. Blackwell, D., Girshick,M.A., Theory of Games and Statistical Decisions, Wiley:New York 1954

7. Shannon, C.E., Weaver, W., The Mathematical Theory of Communication, Illinois Uni.Press:Urbana, Ill. 1949

8. Wiener, N., Cybernetics, M.I.T. Press:Cambridge(Mass.) 1948

9. Fadeev, D.K., 'On the Concept of the Entropy for a Finite Probability Model', Uspehi Mat. Nauk 11, 1958,227 (in Russian)

10. Erdös, P., 'On the Distribution Function of Additive Functions', Ann. of Math. 47, 1946, 1

11. MacKay, D.M., Information, Mechanism and Meaning,M.I.T. Press: Cambridge, 1969

12. Rényi, A., 'On Measures of Entropy and Information', Proc. 4th Berkeley Symp.Math. Statist. and Probability, 1961, 547

13. Kelly, J.L., 'A new Interpretation of Information Rate', Bell System Technical Jour. 35,1956,917

14. Khinchin, A.I., Mathematical Foundations of Information Theory, Dover Publ.:New York 1957

15. Tveberg, H., 'A New Derivation of the Information Function', Math. Scand. 6, 1958,297

16. Kendall, D.G., 'Functional Equations in Information Theory', Zeitschr. Wahrscheinlichkeitstheorie 2, 1964,225

17. Lee, P.M., 'On the Axioms of Information Theory', Ann. Math. Statist. 35,1964,415

18. Gottinger, H.W., 'Review of Concepts and Theories of Probability', Scientia (Rivista di Scienza), 109,1974,83

19. Bar-Hillel, Y. and Carnap R., 'An Outline of a Theory of Semantic Information', Tech.Rep.No.247, Research Lab. of Electronics, M.I.T. 1952, in Language and Information, Bar-Hillel,Y. ed.,Addison-Wesley, Reading, Mass. 1964,221

20. Jeffreys, H.,Theory of Probability, 3rd ed., Clarendon Press: Oxford 1961

21. Heerden van, P.J., 'The Foundation of Empirical Knowledge', Uitgeverij Wistik, Wassenaar, Holland 1968

22. Ingarden,R.S. and Urbanik,K., 'Information without Probability', Colloq.Math. 9, 1962, 131

23. Ingarden, R.S., 'A Simplified Axiomatic Definition of Information', Bull.Acad.Sci. Polonaise, Ser. math.,astr. et phys., 11, 1963, 209

24. Ingarden, R.S., 'Simplified Axioms for Information without Probability', Prace Matematyczne 9, 1965,273

25. Rényi, A., Wahrscheinlichkeitsrechnung (mit einem Anhang über Informationstheorie), VEB Deutscher Verlag der Wissenschaften: Berlin 1962

26. Rényi, A., 'On the Foundations of Information Theory', Bull. International Statistical Institute 33, 1965

27. Kampé de Fériet, J., 'Mesure de l'information fournie par un évènement', Les Probabilités sur les Structures Algébriques, Paris: Symp. CNRS, 1970, 191

28. Fuchs, L., Partially Ordered Algebraic Systems, Pergamon: London 1963

29. Sikorski, R., Boolean Algebras, (2nd ed.), Springer, Göttingen-Berlin 1964

30. Gottinger, H.W., 'Konstruktion subjektiver Wahrscheinlichkeiten', Math. Operationsforschung u.Statist.5, 1974, 509

31. Domotor, Z., 'Probabilistic Relational Structures and Applications', Tech. Report 144, Inst. Math. Studies Social Sciences, Stanford University, 1969

32. Ore, O., 'Theory of Equivalence Relations', Duke Mathematical Jour, 9, 1942, 573

33. Maeda, S., 'A Lattice-Theoretic Treatment of Stochastic Independence', Jour.Sci.Hiroshima Univ. 27, 1963,Ser.A-1,1

34. Scott, D., 'Measurement Structures and Linear Inequalities', Jour. Math. Psychology 1, 1964, 233

35. Wells, R., 'A Measure of Subjective Information', Proc. Symp.Appl. Math.(Amer. Math. Soc.) 12, 1961, 237

36. Császar, A., Foundations of General Topology, Pergamon, London 1963

37. Horn, A. and Tarski, A.,'Measures in Boolean Algebras', Trans. Amer. Math. Soc. 64 1948, 467

38. Kelley, J.L., 'Measures in Boolean Algebras', Pacific Jour. Math. 9, 1959, 1165

39. Kappos, D.A., Strukturtheorie der Wahrscheinlichkeitsfelder und -räume, Springer: Berlin 1960

40. Groot de, M., 'Uncertainty, Information and Sequential Experiments', Ann. Math. Statist. 33, 1962,404

41. Groot de, M., 'Optimal Statistical Decisions', McGraw-Hill: New York 1970

42. Birnbaum, A., 'On the Foundations of Statistical Inference', Ann. Math. Statist. 32, 1961, 414

43. McGuire,C.B., 'Comparison of Information Structures', in Decision and Organization, Radner, R. and McGuire, C.B. eds., North Holland,Amsterdam 1972, 101

44. Gottinger, H.W., 'Some Measures of Information arising in Statistical Games', Kybernetik 15, 1974, 111

45. Groot de, M.,'Optimal Statistical Decisions', McGraw-Hill: New York 1970

46. Howard, R.A.,'Information Value Theory', IEEE Trans.Syst. Science and Cyber. SSC-2, 1966, 22

47. Marschak, J., 'Economics of Information Systems', in Frontiers of Quantitative Economics,Intrilligator, M.D. ed., North-Holland, Amsterdam 1971, 31

48. Marschak, J. and Radner, R., Economic Theory of Teams, Yale Univ.Press, New Haven 1972

49. Belis, M. and Guiasu, S., 'A Quantitative-Qualitative Measure of Information in Cybernetic Systems', IEEE Trans. Information Theory IT-14, 1967, 593

50. Guiasu, S., Mathematical Structure of Finite Random Cybernetic Systems, CISM Courses and Lectures No.86, Wien:Springer-Verlag 1971

51. Longo, G., Quantitative-Qualitative Measure of Information, CISM Courses and Lectures No. 138, Wien:Springer-Verlag 1972

52. Marschak, J., 'Optimal Systems for Information and Decision', Techniques of Optimization, Academic Press: New York 1972, 355

53. Marschak, J., 'Limited Role of Entropy in Information Economics', in 5th Conference on Optimization Techniques, Part II, Conti, R. et al. eds., Springer, New York 1973, 264

54. Marschak, J., 'Value and Cost of Information Systems', Working Paper 51, Wirtschaftstheoretische Abt., Universität Bonn, September 1973

55. Gottinger, H.W., 'Computable Organizations - Representation by Sequential Machine Theory', Ann. Systems Research 3, 1973, 81

56. Gottinger, H.W., 'Information Structures in Dynamic Team Decision Problems', to appear in Economic Computation and Economic Cybernetic Studies and Research (Bucharest), 1975.