INFORMATION AND CONTROL 41, 275–304 (1979)

Measurement, Theory, and Information

JOHN F. CYRANSKI

Theoretical Chemistry Institute, The National Hellenic Research Foundation, 48 Vassileos Constantinou Avenue, Athens 501/1-Greece

A general scheme is presented for explaining the inter-relation of measurement, theory, and information. Ordinary measurement theory is enriched with ideas from quantum physics and coupled to the concept of informations on lattices. Algebraic properties of the class of informations are studied, and it is shown that probability theories are imbedded in the scheme. The Maximum Entropy Principle (Jaynes, 1957; 1958) is justified and generalized in the new context. Finally the Bayesian theory of inductive logic is criticized on the grounds that "attributes" and "evidences" do not belong to a common "language."

I. INTRODUCTION

The theory of information has—with few exceptions—developed as a branch of statistical communication theory. As such, its scope is intrinsically limited by Shannon's explicatum for information (Shannon, 1948) and by its dependence on the concept of probability. Recently it has been suggested that information is actually more fundamental than probability (Ingarden and Urbanik, 1961; Forte and Kampé de Fériet, 1967; Černý and Brunovský, 1974)—indeed that probabilities exist only for special cases of information measures. Also, the traditional domain of probabilities—the σ-complete Boolean lattice of "events" (Renyi, 1970)—can be extended using informations defined on *any* lattice (Sallantin, 1972).

Much older than the development of information theory has been the continual controversy over the significance of probability. There are basically two viewpoints—with a broad spectrum of subdivisions: The "frequentists" view probability as defined by the limit of the ratio of the number of "successes" to the number of "trials" as the latter becomes infinite. The "subjectivists" consider probabilities $p(a/b)$ as representing logical weights ("degrees of belief") on the relation of propositions a and b (Carnap, 1950; Jeffreys, 1961). For a more complete survey, see (Fine, 1973).

Jaynes has employed Shannon information theory to obtain "degrees of belief" on evidence in the form of expectation values of random variables (Jaynes, 1957; 1958). While enjoying much success in various applications, this algorithm has drawn fire for various reasons (Macqueen and Marschak,

0019-9958/79/060275-30$02.00/0

1975; Friedman and Shimony, 1971). Elsewhere we discuss the "MEP"—for maximum entropy principle—debate (Cyranski, 1978), and show that Jaynes' procedure actually provides a reunification of the camps: When relative frequencies are available as evidence, then the MEP yields as "degrees of belief" based on this evidence exactly the relative frequencies. But, when frequencies are not available, the MEP yields distributions consistent with the assumption that the evidence used to calculate the distributions is "total". This method has the advantage of specific reference to empirical data and the quantitative information content of the data.

Nevertheless, the MEP is based on the restricted Shannon measure of "uncertainty", its foundations and domain are not clearly defined, and its apparent basis in the "subjective" camp leads to the following difficulty:

The basis tenet of the subjective school is Bayes' rule:

$$p(A \mathbin{\&} B/C) = p(A/B \mathbin{\&} C)\, p(B/C) = p(B/A \mathbin{\&} C)\, p(A/C). \tag{1.1}$$

This is justified (for Boolean "languages") in a variety of ways (Carnap, 1950; Cox, 1946; Shimony, 1955). The difficulty with (1.1) is that *all* statements have the same logical weight. That is, if A and B are "attributes" of an object and C is "evidence" obtained from experiment about the family of attributes $\mathscr{L}$, (1.1) "mixes" A with C and B with C. In other words, Bayes' rule requires that "$A \mathbin{\&} C$" be a legitimate statement in some "language" encompassing both $\mathscr{L}$ and empirical evidence *about* $\mathscr{L}$. However, the very procedure of the MEP is such that MEP "degrees of belief" concern $\mathscr{L}$ alone, although they are chosen via C. Shimony and Friedman have tried to show by an example that the MEP is inconsistent with (1.1) (Friedman and Shimony, 1971). While their arguments fail to do so (Hestenes and Gage, 1973; Cyranski, 1978) it is still not clear whether or not (1.1) is indeed inconsistent with the MEP. In any event, the MEP raises the above issue—whether or not all propositions are "equivalent" in some language. We consider this question in Section VI.

In order to better understand the role of "information" in measurements, the role of "theories", the domain and justification of the MEP, and in some sense to develop a universal attitude—if not methodology—towards the "scientific method", we have been led to the following considerations: A class of objects is *defined* by the relevant empirical relations that operate on the objects. These can be partially ordered and the partial order completed in a natural way to form a "logic" (a "language" governed not by its grammar but by "truth" value directly related to observation). Information is most generally an anti-homomorphism on a partially ordered set, so in particular a class of informations can be associated with the system. We consider a class of inputs from experiment that restrict the possible informations. Functions defined on the partially ordered set of empirical relations are "evaluated" according to the information subset defined by the data. The key to the synthesis of these ideas is the natural criterion that there be an *unambiguous* evaluation of any (permissible) function given the data.

In Section II we provide the basic mathematical structure needed for our viewpoint. In Sections III and V we present the basic model, while in Section IV we show that the model includes the special case of "ordinary" probability theory. We reconsider Bayes' rule in Section VI, and conclude in Section VII with suggestions for applications and open questions.

II. Mathematical Preliminaries

The basic mathematics that we require can be found in (Birkhoff, 1973). In order to clarify the significance of the results, we present our own modified proofs.

Definition 2.1. A *quasi-ordered set* (quoset) is a set S that is non-empty and has defined on it a binary relation $<$ satisfying:

$$x < x, \text{ for all } x \in S \text{ (reflexive)} \tag{2.1}$$

$$x < y \text{ and } y < z \text{ implies } x < z \text{ (transitive)}. \tag{2.2}$$

Any oriented graph with loops is a quoset. This includes, for example, computer programs (algorithms) and, more exotically, relativistic space-time ordered by "causality" (Carter, 1971).

Definition 2.2. A *partially-ordered set* (poset) is a set that is non-empty and has defined on it a transitive, reflexive binary relation $\leqslant$ satisfying:

$$x \leqslant y \text{ and } y \leqslant x \text{ implies } x = y \text{ (antisymmetric)}. \tag{2.3}$$

Starting with a quoset S ordered by $<$, define $x \sim y$ whenever $x < y$ and $y < x$. Then the quotient set $P = S/\sim$ is a poset if $E \leqslant F$ in P means $x < y$ for some $x \in E$ and $y \in F$. In other words, if one identifies all points in each loop, one can transform any quoset into a poset.

Definition 2.3. A *lattice* is a poset $\mathscr{L}$ in which for every pair $x, y \in \mathscr{L}$ $x \wedge y = \text{glb}(x, y)$ and $x \vee y = \text{lub}(x, y)$ are in $\mathscr{L}$. (Note that $z = x \wedge y$ means that for all $w \in \mathscr{L}$ such that $w \leqslant x$ and $w \leqslant y$, then $w \leqslant z$, and conversely. Dually, $z = x \vee y$ means for all $w \in \mathscr{L}$ such that $w \geqslant x$ and $w \geqslant y$ then $w \geqslant z$, and conversely.) A poset or lattice has *universal bounds* if $0 = \bigwedge_{x \in \mathscr{L}} x$ and $I = \bigvee_{x \in \mathscr{L}} x$ are in $\mathscr{L}$. A lattice is *complete* if for every subset $X \subseteq \mathscr{L}$, $\bigwedge_{x \in X} x$ and $\bigvee_{x \in X} x$ are in $\mathscr{L}$.

If Y is any set, and 2^Y is its power set, then 2^Y forms a complete lattice with universal bounds under set inclusion. The real line ordered in the usual way is only conditionally complete (every non-empty bounded subset of the real line has a lub and glb, but 0, I are not in $\mathscr{L}$).

DEFINITION 2.4. An order *morphism* between lattices $\mathscr{L}_1$ and $\mathscr{L}_2$ is a mapping $g: \mathscr{L}_1 \to \mathscr{L}_2$ such that for all $x, y \in \mathscr{L}_1$:

$$g(x \wedge y) = g(x) \wedge g(y) \tag{2.4a}$$

$$g(x \vee y) = g(x) \vee g(y). \tag{2.4b}$$

If (2.4) holds for any countable joins and meets, g is a *σ-morphism*. If (2.4) holds for arbitrary joins and meets, g is a *complete morphism*. If g is $1 - 1$ and onto, it is an (order) *isomorphism*, and if $\mathscr{L}_1 = \mathscr{L}_2$ in this case, g is an *automorphism*.

LEMMA 2.1. *Let $X^* = \{u \in P : u \geqslant x$, for all $x \in X\}$ and let $X^\dagger = \{r \in P : r \leqslant x$, for all $x \in X\}$. Set $\bar{X} = (X^*)^\dagger$. Then*

$$X \subseteq \bar{X} \tag{2.5}$$

$$\bar{X} = \bar{\bar{X}} \tag{2.6}$$

$$X \subseteq Y \text{ implies } \bar{X} \subseteq \bar{Y} \tag{2.7}$$

Proof. $\bar{X} = \{t \in P : t \leqslant u$, for all $u \geqslant x$, for all $x \in X\}$ is the set of lower bounds to the set of upper bounds of X. Clearly, if $x \in X$, then $x \in \bar{X}$, so (2.5) is immediate.

Now $X \subseteq Y$ implies $X^* \supseteq Y^*$, so by (2.5), $X^* \supseteq (\bar{X})^*$. But, $\bar{X} = \{t \in P : t \leqslant u$, for all $u \in X^*\}$ so $(\bar{X})^* = \{z \in P : z \geqslant y$ for all $y \leqslant u$, for all $u \in X^*\}$. Clearly, $X^* \subseteq (\bar{X})^*$, so that $X^* = (\bar{X})^*$ and $\bar{X} = \bar{\bar{X}}$.

Finally, $X \subseteq Y$ implies $X^* \supseteq Y^*$; this implies $(X^*)^+ \subseteq (Y^*)^+$ and thus $\bar{X} \subseteq \bar{Y}$. Q.E.D.

The following theorem is essential to our approach. It extends any poset to a complete lattice in a manner analogous to the celebrated extension of the rationals to the reals—Dedekind's "completion by cuts". It appears to be the most natural completion procedure valid for posets. (There are other imbedding theorems appropriate for lattices and metric lattices.) As this method parallels the imbedding of generally discrete, finite empirical results in a continuum theory, it is especially appealing for our purposes.

THEOREM 2.1. *Let P be any poset and $\mathscr{L}(P) = \{X \subseteq P : X = \bar{X}\}$. Then $\mathscr{L}(P)$ is a complete lattice under set inclusion with $\bigwedge X_a = \bigcap X_a = \overline{\bigcap X_a}$ and $\bigvee X_a = \overline{\bigcup X_a}$ for $X_a \in \mathscr{L}(P)$.*

Proof. First, if $X_a \in \mathscr{L}(P)$, $a \in A$, then for each $b \in A$ we have $\overline{\bigcap_{a\in A} X_a} \subseteq X_b$, using Lemma 2.1. Thus, $\overline{\bigcap_{a\in A} X_a} \subseteq \bigcap_{b\in A} X_b \subseteq \overline{\bigcap_{b\in A} X_b}$, again by this lemma, so finally $\bigcap X_a = \wedge X_a \in \mathscr{L}(P)$.

Now $\mathscr{L}(P)$ is obviously a poset under set inclusion. Let S be a family of sets $X \in \mathscr{L}(P)$ and let U be the family of sets $W \in \mathscr{L}(P)$ such that $X \subseteq W$, for all $X \in S$. Set $A = \bigcap_{W\in U} W$, which we have just shown belongs to $\mathscr{L}(P)$. Any $X \in S$ is a lower bound for U, so $X \subseteq A$ and A is thus an upper bound for S. If B is another upper bound for S, then $B \in U$ so $A \subseteq B$. Thus, by definition, $A = \vee_{X\in S} X$.

Finally, $\overline{\bigcup_{X\in S} X} \supseteq \bigcup_{X\in S} X$, so $\overline{\bigcup_{X\in S} X}$ is an upper bound for S. Thus, $\overline{\bigcup_{X\in S} X} \supseteq \vee_{X\in S} X \supseteq \bigcup_{X\in S} X$. Using (2.6) and (2.7) it follows that $\vee_{X\in S} X = \overline{\bigcup_{X\in S} X}$. Q.E.D.

The next theorem shows that P is actually imbedded in $\mathscr{L}(P)$.

THEOREM 2.2. *Let $I(a) = \{t \in P : t \leqslant a\}$ define a mapping I: $P \to \mathscr{L}(P)$. Then P is (lattice) isomorphic to $\{I(a) : a \in P\} \subseteq \mathscr{L}(P)$.*

Proof. Note first that $I(a) \in \mathscr{L}(P)$. Moreover, each $a \in P$ defines a unique $I(a)$. The map is clearly onto and as $I(a) = I(b)$ iff $a = b$, I is $1 - 1$. Note, in fact, that $a \leqslant b$ iff $I(a) \subseteq I(b)$.

Now if $a = \wedge_{x\in X} x \in P$, then $b \leqslant a$ iff $b \leqslant x$ for all $x \in X$. Thus, $I(a) = \bigcap_{x\in X} I(x)$. If $a = \vee_{x\in X} x \in P$, $\vee_{x\in X} I(x) = \{t \in P : t \leqslant u$, for all $u \geqslant v$ for all $v \leqslant x$, for some $x \in X\}$. This is readily seen to reduce to $I(a)$ Q.E.D.

The following apparently new result will be of considerable importance.

THEOREM 2.3. $\bar{A} = \vee_{a\in A} I(a)$ *for* $A \subseteq P$.

Proof. $A = \bigcup_{a\in A} \{a\} \subseteq \bigcup_{a\in A} I(a)$, since $I(a) = \overline{\{a\}}$. Thus, $\bar{A} \subseteq \vee_{a\in A} I(a)$. But, $\{a\} \subseteq A$ implies $I(a) \subseteq \bar{A}$ and thus $\bigcup_{a\in A} I(a) \subseteq \bar{A}$. Hence, $\vee_{a\in A} I(a) \subseteq \bar{A}$ as well. Q.E.D.

We now turn to the concept of information:

DEFINITION 2.5. An *information* on a poset P is a mapping J_P: $P \to [0, \infty]$ such that

$$a \leqslant b \text{ implies } J_P(b) \leqslant J_P(a). \tag{2.8}$$

If P has universal bounds 0 and I,

$$J_P(0) = \infty; \qquad J_P(\mathrm{I}) = 0. \tag{2.9}$$

Intuitively we can understand this definition as saying that if "proposition" a "implies" b, then the information associated with b cannot exceed that associated with a. The amount of information needed to fix b, in other words, will be generally less than that of the "cause" a of b. The boundary values (2.9) are chosen mostly for mathematical convenience, but are natural enough, for the "proposition" I is "certain" in the sense that it is implied by all other "propositions" of P. Hence, its information value is null. Conversely, 0 implies *all* "propositions" of P, so the information inherent in 0 must be infinite. We will be particularly interested in informations defined on $\mathscr{L}(P)$. If J is an information on $\mathscr{L}(P)$, $J(I(a)) = J_P(a)$ obviously defines a related information on P.

As we noted in the introduction, the general concept of informations on lattices was first proposed by Jean Sallantin in 1972 with specific application to quantum physics in mind. The approach has been developed by Sallantin, Comyn and Losfeld with the general goals of applications to pattern recognition problems, statistical analyses, and questionnaire theory, within the general approach of Kampé de Fériet, Forte, and Aczél in characterizing information measures. Among the many articles and texts containing these developments we single out (Kampé de Fériet and Picard, 1974), (Picard, 1975, 1976, 1977), (Picard and Sallantin, 1977) and the articles by Comyn, Losfeld, and Sallantin in (Picard, 1978). The general notion of informations on lattices is thus not original, but is quite new. This is not to dismiss other work, eg. that of Carnap and Bar Hillel (Bar Hillel, 1964) who as early as 1952 defined a "semantic information" on "languages". Their approach, however, is based on "subjective" probability and has apparently not led to significant progress.

Our interest in and application of the informations defined by Definition 2.5 leads us to study in particular the properties of such informations as a class, i.e., $Z(P)$ the family of all informations defined on a poset (lattice) P.

THEOREM 2.4. *$Z(P)$ is a complete lattice.*

Proof. We define $J_1 \leqslant J_2$ by:

$$J_1 \leqslant J_2 \quad \text{iff} \quad J_1(a) \leqslant J_2(a) \quad \text{for all} \quad a \in P. \tag{2.10}$$

Clearly $Z(P)$ is a poset under this operation. If Q is a subset of $Z(P)$, we certainly can define:

$$\bigwedge_{J \in Q} J(a) = \inf\{J(a) \colon J \in Q\} \tag{2.11a}$$

$$\bigvee_{J \in Q} J(a) = \sup\{J(a) \colon J \in Q\}. \tag{2.11b}$$

The question is whether the maps defined by (2.11) from P to $[0, \infty]$ are informations. If $0, \mathrm{I} \in P$, clearly $\wedge_{J\in Q} J(\mathrm{I}) = 0 = \vee_{J\in Q} J(\mathrm{I})$ and $\vee_{J\in Q} J(0) = \infty$. If $a \leqslant b$, then $J(b) \leqslant J(a)$, for all $J \in Q$. Thus, $\wedge_{J'\in Q} J'(b) \leqslant J(b) \leqslant J(a)$

for all $J \in Q$, so that $\wedge_{J \in Q} J(b) \leqslant \wedge_{J \in Q} J(a)$. Similarly, $J(b) \leqslant J(a) \leqslant \vee_{J' \in Q} J'(a)$ implies that $\vee_{J \in Q} J(b) \leqslant \vee_{J \in Q} J(a)$. Thus, $\wedge_{J \in Q} J$ and $\vee_{J \in Q} J$ are both in $Z(P)$. Q.E.D.

DEFINITION 2.6. A lattice $\mathscr{L}$ is *Brouwerian* if when $a, b \in \mathscr{L}$, the set $\{x \in \mathscr{L} : a \wedge x \leqslant b\}$ has a greatest element $b : a$ (called the relative pseudo-complement of a in b).

An example of a Brouwerian lattice is the complete lattice of all open sets of any topological space. If $\mathscr{L}$ is a fixed complete Brouwerian lattice and E is a non-empty set, a *fuzzy set* A of E is a function A: $E \to \mathscr{L}$. The class of all fuzzy sets of E, $\mathscr{L}(E)$, is itself a complete Brouwerian lattice (Sanchez, 1976). This lends interest to the following result.

THEOREM 2.5. *$Z(P)$ is a convex Brouwerian lattice.*

Proof. If $J_1, J_2 \in Z(P)$, then for all $t \in [0, 1]$ $tJ_1 + (1 - t) J_2 \in Z(P)$, for if $0, \mathrm{I} \in P$, $tJ_1(0) + (1 - t) J_2(0) = \infty$ and $tJ_1(\mathrm{I}) + (1 - t) J_2(\mathrm{I}) = 0$, while if $a \leqslant b$, $J_1(a) \geqslant J_1(b)$ and $J_2(a) \geqslant J_2(b)$ means that $tJ_1(a) + (1 - t) J_2(a) \geqslant tJ_1(b) + (1 - t) J_2(b)$.

Now a complete lattice is *completely distributive* when

$$\bigwedge_{g \in C} \left[\bigvee_{a \in A_g} x_{g,a} \right] = \bigvee_{\Phi} \left[\bigwedge_{C} x_{g,\phi(g)} \right] \tag{2.12a}$$

$$\bigvee_{g \in C} \left[\bigwedge_{a \in A_g} x_{g,a} \right] = \bigwedge_{\Phi} \left[\bigvee_{C} x_{g,\phi(g)} \right] \tag{2.12b}$$

where Φ is the set of all functions with domain C and $\phi(g) \in A_g$. But $[0, \infty]$ is completely distributive under the real number ordering, so (2.12) applies to $J_{g,a}(b)$ for each $b \in P$. As $Z(P)$ is a complete lattice, it follows that $Z(P)$ is a completely distributive lattice. But by Theorem 24, page 128 (Birkhoff, 1973) it follows that $Z(P)$ is Brouwerian. Q.E.D.

We pursue our study of the algebraic properties of $Z(P)$ by noting that in any lattice $I(a) = \{x \in \mathscr{L}: x \leqslant a\}$ is a *principal ideal* of $\mathscr{L}$.

LEMMA 2.2. (a) *Each principal ideal of $Z(P)$ is convex.* (b) *$J^* \in I(J)$ iff there exists $t \in [0, 1)$ such that $tJ + (1 - t)J^* \in I(J)$.* (c) *If J_1, J_2 are not comparable, then for no $s, t \in (0, 1)$ are $sJ_1 + (1 - s)J_2$ and $tJ_1 + (1 - t)J_2$ comparable.*

Proof. (a) If $J_1, J_2 \in I(J)$, then $J_1 \leqslant J$ and $J_2 \leqslant J$ implies that $tJ_1 + (1 - t)J_2 \leqslant (t + (1 - t))J = J$.

(b) If $tJ + (1 - t)J^* \leqslant J$, then $(1 - t)J^* \leqslant (1 - t)J$, or $J^* \leqslant J$ for $t \in [0, 1)$. Conversely, $J^* \leqslant J$ implies $tJ + (1 - t)J^* \in I(J)$ by (a).

(c) Suppose there exist s, $t \in (0, 1)$ such that $sJ_1 + (1 - s)J_2 \leqslant tJ_1 + (1 - t)J_2$, with $s \neq t$. Then if, eg., $(s - t) > 0$, it follows that $(s - t)J_1 \leqslant ((1 - t) - (1 - s))J_2 = (s - t)J_2$, so $J_1 \leqslant J_2$, contradicting the assumption of (c). Q.E.D.

DEFINITION 2.7. If $\mathscr{L}$ is a complete lattice and $x_a \in \mathscr{L}$, then x_a *order converges* to x^* ($x_a \to x^*$) when

$$\bigvee_b \left[\bigwedge_{a \geqslant b} x_a\right] = \bigwedge_b \left[\bigvee_{a \geqslant b} x_a\right] = x^* \tag{2.13}$$

where $a, b \in A$, a directed set of indices. A set X is *closed* if every order convergent set of elements from X order converges to an element of X. The closed sets thus established define the *order topology* on $\mathscr{L}$.

THEOREM 2.6. *Let g be a complete automorphism on $Z(P)$. Then g is a homeomorphism in the order topology.*

Proof. Any complete morphism on $Z(P)$ preserves arbitrary meets and joins. Hence, if $x_a \to x$, $h(x) = \bigwedge_b [\bigvee_{a \geqslant b} h(x_a)] = \bigvee_b [\bigwedge_{a \geqslant b} h(x_a)]$, or $h(x_a) \to h(x)$. As g is an automorphism in particular, it follows that both g and g^{-1} are thus continuous in the order topology, so that g is a homeomorphism. Q.E.D.

Note that $I(J)$ is closed in the order topology as if $J_a \to J^*$, then $J^* = \bigvee_b [\bigwedge_{a \geqslant b} J_a] \leqslant \bigvee_b J = J$, so $J^* \in I(J)$.

THEOREM 2.7. *$Z(P)$ is metrizable.*

Proof. It is straightforward to show that the following defines a metric function:

$$d(J_1, J_2) = \sup\{| J_1(a) - J_2(a)| : a \in P\}. \tag{2.14}$$

THEOREM 2.8. *Let $K_e(J) = \{J' \in Z(P) : d(J, J') < e\}$. Then, if*

$$J_1, J_2 \in \left[\bigwedge_{J' \in K_e(J)} J', \bigvee_{J' \in K_e(J)} J'\right] = \left\{J'' \in Z(P): \bigwedge_{J' \in K_e(J)} J' \leqslant J'' \leqslant \bigvee_{J' \in K_e(J)} J'\right\},$$

then $d(J_1, J_2) \leqslant 2e$, for all $e > 0$.

Proof. Now $d(J, J') < e$ for all $J' \in K_e(J)$ implies that for all such J', $J(a) - e < J'(a) < J(a) + e$, for all $a \in P$. Thus,

$$\begin{aligned} J(a) - e &\leqslant \inf\{J'(a) : J' \in K_e(J)\} \leqslant J'(a) \leqslant \sup\{J'(a) : J' \in K_e(J)\} \\ &\leqslant J(a) + e. \end{aligned} \tag{2.15}$$

Hence, $| J(a) - \bigwedge_{J' \in K_e(J)} J'(a)| \leqslant e$ and $| J(a) - \bigvee_{J' \in K_e(J)} J'(a)| \leqslant e$ for all $a \in P$. Thus, $d(J, \bigwedge_{J' \in K_e(J)} J') \leqslant e$ and $d(J, \bigvee_{J' \in K_e(J)} J') \leqslant e$. It is also apparent from (2.15) that if $J^* \in [\wedge_{J' \in K_e(J)} J', \vee_{J' \in K_e(J)} J']$, then $d(J, J^*) \leqslant e$. Hence by the triangle inequality of metrics, $d(J_1, J_2) \leqslant d(J_1, J) + d(J, J_2) \leqslant 2e$.

Q.E.D.

III. Measurements, Theory, and Information

Let us begin by imagining that we have a non-empty set of objects. These objects are *defined* by a finite set of characteristics or properties that can be measured. For example, weight, color, intelligence, spin, valence, volume, cross-section, etc. are properties. Clearly not all properties are relevant to any given set of objects. Moreover, only finite lists of properties are realistic—operational—for if an object required an infinite list of such properties for its definition, we would never be able to recognize this object with certainty.

Now each property has *manifestations*: Red, blue, turquoise are manifestations of the property color. Manifestations of spin form the set $\frac{1}{2}$, 1, $\frac{3}{2}$, 2, $\frac{5}{2}$, etc. Intelligence is manifest by the I.Q. scale of values. When we perform a measurement, we associate an object with manifestations of one or more properties. Indeed, with any set of manifestations of properties we identify a class of objects which —on measurement—we identify with this set. Thus, electrons are identified by a particular rest mass, charge, and spin. In other words, the objects are classified according to *relations* that they are observed to satisfy. Hence, musical tones have the properties pitch and loudness, which can be used to *relate* different tones. Some typical relations are: (1) tone A has higher pitch than tone B; (2) tone A is softer than tone B; (3) tone C has pitch midway between the pitch of tone A and that of tone B. Each family of such empiricially meaningful relations defines the structure of a property. Formally, then, measurement theory deals with what is called an *empirical relational system*: a set A together with a family P of relations on A. In particular, the classical discipline turns toward an examination of "scales"—essentially homomorphisms from the empirical relational system to a similarly defined numerical system. The idea here is that the scales parallel or image the relations—for example, higher pitch is assigned a greater scale value. For this approach to measurement theory see (Pfanzagl, 1971).

There are several shortcomings of this "classical" approach to measurements. For one, this theory seems unsuited to describe quantum physics. Moreover, it is based on a restricted concept of relations that excludes "fuzzy relations" as well as further generalization. It does not seem to admit very comfortably the notions of "evidence", "degrees of belief", and information. It is, of course, possible that these difficulties can be overcome in the traditional context—and perhaps to some extent this has been done. Nevertheless we believe it worthwhile to avoid the idea of "scales" and approach the problem of measurements somewhat differently.

We start with the assumption that a system of objects is *defined* (i.e., completely characterized) by some class of empirically definable relations on the family of objects. Thus, we have an empirical relational system (A, P) associated with each class of objects, P containing all relations that are (or we believe to be) relevant to the objects. In other words, P is a *defining model* for the class.

As an example, we consider the situation in quantum physics. In this context, an atomic system is defined by an appropriate Hilbert space; the one-dimensional vector subspaces of this space characterize various "pure states" of the system —in fact define equivalence relations. Empirically, one can determine to which *finite*-dimensional subspace a particular system belongs, but infinite-dimensional subspaces are clearly not operationally defined. This suggests, therefore, that P actually consists of all finite-dimensional sub-spaces of the Hilbert space. The *theoretical model* must be generated in terms of P so that P is somehow imbedded in the general theory. That is, we wish to extend the finitary definition to an infinitary definition.

In fact, the defining model P forms a poset under sub-space inclusion, and its completion by cuts, $\mathscr{L}(P)$ corresponds to the complete lattice of *all* subspaces of the Hilbert space. (Birkhoff, 1973). The approach we take below is an attempt to extend and generalize this description to measurement theory in general.

DEFINITION 3.1. An *n-ary relation* on set A (non-empty) is a mapping $R: A^n \to Q$, where Q is a poset.

This generalization of the concept of relation admits fuzzy relations (if Q is a Brouwerian lattice) and classical relations (if $Q = \{0, I\}$). In the latter case, $R(a_1, ..., a_n)$ is just the "characteristic function" $C_R(a_1, ..., a_n)$ where the "relation" is the subset R of A^n. We thus define our system class by a family of such relations, P, with fixed Q.

DEFINITION 3.2. If $R, T \in P$, define $R \to T$ iff $n_R \geqslant n_T$ and $R(x_1, ..., x_{n_R}) \leqslant T(x_{k_1}, ..., x_{k_{n_T}})$ in Q for every subsequence $(x_{k_1}, ..., x_{k_{n_T}})$ of $(x_1, ..., x_{n_R})$ and for all $(x_1, ..., x_{n_R}) \in A^{n_R}$.

The order thus defined actually corresponds to a natural extension of the idea of "implication". For example, suppose that $n_R = n_T$. Then, $R \to T$ means that $R(x_1, ..., x_{n_R}) \leqslant T(x_1, ..., x_{n_R})$ for all $(x_1, ..., x_{n_R}) \in A^{n_R}$. In particular, if $Q = \{0, I\}$, this can be restated as follows: $R \to T$ iff the subsets $R, T \subseteq A^{n_R}$ satisfy $R \subseteq T$—whenever $(x_1, ..., x_{n_R})$ satisfies relation R, it satisfies T as well. The "truth" of R "implies" the "truth" of T.

To demonstrate the more general case of $n_R \neq n_T$ (with the same classical Q), consider the following. Suppose $A = \{1, 2, 3, ...\}$ and $R_1 = \{(a, b) : a < b\}$, $R_2 = \{(a, b, c) : a < c < b\}$, and $R_3 = \{(a, b, c) : a < b < c\}$. Now if $(a, b, c) \in R_3$, then (a, b), (a, c), and (b, c) are in R_1 so that whenever R_3 is "true" for a triple (a, b, c), then R_1 is "implied" for each *pair* of elements in the triple taken in order. Note, however, that $(a, b, c) \in R_2$ does not "imply"

R_1 in this sense, as (a, b) and $(a, c) \in R_1$, but $(b, c) \notin R_1$. In this case it is not possible to say that R_2 "implies" R_1 in any sense.

We believe definition 3.2 to be the natural ordering on relations. Definition 3.2 is further justified by the following result.

THEOREM 3.1. *P is a poset under* $\rightarrow$.

Proof. Clearly $R \rightarrow R$. If $R \rightarrow R'$ and $R' \rightarrow R$, we have $n_R = n_{R'} = n$. Thus, $R(x_1, ..., x_n) \leqslant R'(x_1, ..., x_n)$ and $R'(x_1, ..., x_n) \leqslant R(x_1, ..., x_n)$ for all $(x_1, ..., x_n) \in A^n$. As Q is a poset, however, this implies $R = R'$. (Note that two relations R and T are equal iff $R(x_1, ..., x_n) = T(x_1, ..., x_n)$—they must have the same order, of course—for all $(x_1, ..., x_n) \in A^n$.)

Finally, $R \rightarrow T$ and $T \rightarrow V$ means $n_R \geqslant n_V$. Also, $R(x_1, ..., x_{n_R}) \leqslant T(x_{k_1}, ..., x_{k_{n_T}})$ for all n_T-subsequences of $(x_1, ..., x_{n_R})$, for all such n_R-tuples, and $T(x_{j_1}, ..., x_{j_{n_T}}) \leqslant V(x_{m_1}, ..., x_{m_{n_V}})$ for all n_V-subsequences of $(x_{j_1}, ..., x_{j_{n_T}})$ and for all such elements of A^{n_T}, but therefore for all such n_V-subsequences of $(x_1, ..., x_{n_R})$, for all n_R-tuples. By transitivity in Q, we thus have $R(x_1, ..., x_{n_R}) \leqslant V(x_{m_1}, ..., x_{m_{n_V}})$ for all n_V-subsequences of $(x_1, ..., x_{n_R})$, for all such n_R-tuples. Hence, $R \rightarrow V$. Q.E.D.

As noted above, we take P to be the operational definition of the system of objects, and $\mathscr{L}(P)$ to be the theoretical model for the system. Now the experimenter generally does not observe the relations of P directly. Rather, he infers (largely through consistency with a proposed definition) that he is observing a system characterized by P from a (finite) list of real numbers that represent empirical *evaluations* of certain functions defined on P, the system. That is, the result of experiment is an "estimate" or evaluation $e(X)$ of some function $X: P \rightarrow R^1$.

Notice that we consider here functions defined on the relational system poset P, and not directly on the "objects". This reflects, of course, our attitude that the "objects" can only be operationally defined via relations—"manifestations of properties". Quantum theorists have long been aware of the difficulties of defining "isolated" objects, for the act of observing (according to the usual interpretations at least (Jammer, 1974)) constitutes an interaction between object and observer. During the time of interaction—the only time when we can be sure there exists an object without further metaphysical assumptions—the object is really "part" of the observer. Thus, direct interpretation of the observed properties as those *of the object* is meaningless. It is, however, less obvious that the relations the "objects" satisfy among themselves are so closely related to the act of observation. For example, position and momentum observations on individual electrons are so interaction-dependent as to lead some physicists to the conclusion that the properties thus manifested do not simultaneously exist in "nature" (Landau and Lifshitz, 1965). On the other hand, *ensembles* of "identically prepared" electrons can be discussed in quantum theory by measuring position and momentum separately on different electrons. As these electrons all belong to the same equivalence class of electrons by nature of their preparation, it is possible to attribute position and momentum to the system or class, rather than to the individuals in that class (Belinfante, 1975). In fact, the position and momentum are not precise, but define an average position and average momentum associated with this preparation which we can understand more clearly as evaluations of position and momentum functions defined on the class of electrons (represented by a one-dimensional subspace of Hilbert space for "pure states").

We thus propose that the *objects* of measurements are relational posets P, that "theories" are their "cut completions" $\mathscr{L}(P)$, and that the *objective* of measurements is the evaluation of real-valued functions defined on P.

Let us now turn to the evaluations. Clearly an evaluation $e(X)$ depends on the function X. (We assume throughout a fixed P and $\mathscr{L}(P)$.) It is obvious from everyday usage that in order to evaluate a function we require an information input (e.g., the "preparation" or "state" of the quantum system). In particular, if an information $J \in Z(\mathscr{L}(P))$ can somehow be chosen to represent the measure of this data input, the evaluation depends on J. We shall outline below how the data can be associated with a measure of the information it provides about the "propositions" in $\mathscr{L}(P)$. Thus we assert that $e(X) = G(X, J)$. We call G the *evaluation procedure*. Among the obvious properties $G(X, J)$ must satisfy are:

$$G(X, J) = x_0, \text{ for all } J \in Z(\mathscr{L}(P)), \text{ if } X(a) = x_0, \text{ for all } a \in P \quad (3.1)$$

$$G(X, J) \leqslant G(Y, J), \text{ for all } J \in Z(\mathscr{L}(P)), \text{ if for all } a \in P$$
$$X(a) \leqslant Y(a) \quad (3.2)$$

If $X(a)$ is replaced by $X'(a) = bX(a) + c$, b, c reals, for all $a \in P$, then we expect that $e(X)$ should be replaced by $be(X) + c$:

$$G(bX + c, J) = bG(X, J) + c, \text{ for all } J \in Z(\mathscr{L}(P)). \quad (3.3)$$

These conditions will, of course, restrict the possible forms that G can assume. We have not investigated to what extent G can be determined by (3.1)–(3.3), however.

We shall digress in the next section to consider some special cases of this approach, and will continue our general discussion in Section V.

IV. Some Special Cases

Let us first suppose that $\mathscr{L}(P)$ is *orthocomplemented*: There exists a unary mapping on $\mathscr{L}(P)$ to itself that satisfies

$$E \cap E^{\perp} = \phi; \qquad E \vee E^{\perp} = P \quad (4.1a)$$
$$(E \cap F)^{\perp} = E^{\perp} \vee F^{\perp}; \qquad (E \vee F)^{\perp} = E^{\perp} \cap F^{\perp} \quad (4.1b)$$
$$(E^{\perp})^{\perp} = E \quad (4.1c)$$

for all $E, F \in \mathscr{L}(P)$. Again, the lattice of all subspaces of Hilbert space demonstrates this concept. Also, the lattice of all subsets of a set, any Borel algebra, in fact, any Boolean algebra are examples. Note, however, that a Brouwerian lattice is not orthocomplemented.

DEFINITION 4.1. An information on $\mathscr{L}(P)$ is *weakly composible* if $E \perp F$ (i.e., $E \subseteq F^{\perp}$) implies $J(E \vee F) = T(J(E), J(F))$. It is *composible* if $E \cap F = \phi$ implies $J(E \vee F) = T(J(E), J(F))$.
Clearly using induction one can define both species of composibility for any finite set of E_n's in $\mathscr{L}(P)$ that satisfy the requirements of definition 4.1.

DEFINITION 4.2. If J is weakly-composible on $\mathscr{L}(P)$, it is *weakly σ-composible* if for every sequence $\{E_n \in \mathscr{L}(P) : E_n \perp E_m \ (n \neq m)\}$

$$J\left(\bigvee_{n=1}^{\infty} E_n\right) = \operatorname*{Lim}_{N\to\infty} J\left(\bigvee_{n=1}^{N} E_n\right) \tag{4.2}$$

exists. J is *σ-composible* if the limit (4.2) holds for each sequence $\{E_n \in \mathscr{L}(P) : E_n \cap E_m = \phi \ (n \neq m)\}$.

DEFINITION 4.3. Let $T : [0, \infty] \otimes [0, \infty] \to [0, \infty]$ be continuous. It is a *regular operation of composition* (ROC) if it satisfies:

$$T(x, y) = T(y, x) \tag{4.3a}$$

$$T(x, T(y, z)) = T(T(x, y), z) \tag{4.3b}$$

$$T(x, \infty) = x \tag{4.3c}$$

$$x_1 < x_2 \text{ implies } T(x_1, y) \leqslant T(x_2, y) \tag{4.3d}$$

$$T(x, y) \leqslant \operatorname{Inf}(x, y). \tag{4.3e}$$

THEOREM 4.1. *Let T be a* ROC. *Define $T_1(x_1) = x_1$ and for all $n \geqslant 1$, $T_{n+1}(x_1, \dots, x_{n+1}) = T(T_n(x_1, \dots, x_n), x_{n+1})$. Then the limit of $T_N(x_1, \dots, x_N)$ as N becomes infinite exists.*

Theorem 4.1 is only part of a more comprehensive list of results in (Kampé de Fériet and Benvenuti, 1972). Note that $T_n(x_1, \dots, x_n) = J(\vee_{k=1}^{n} E_k)$ when $x_k = J(E_k)$ and the E_k are orthogonal in pairs. The real significance of ROCs resides in the following result (Kampé de Fériet, Forte, and Benvenuti, 1969):

THEOREM 4.2. *Let T be a* ROC *and $\Lambda^* = \{z \in [0, \infty] : T(z, z) = z\}$. Then $[0, \infty] - \Lambda^* = \bigcup_{n\in M} (a_n, b_n)$, M being empty, finite, or countable. The* ROC *is specified by:*

$$T(x, y) = \begin{cases} \operatorname{Inf}(x, y), & \text{if } x, y \in [0, \infty] - \bigcup\limits_{n\in M} (a_n, b_n) \\ \phi_i(\theta_i^{-1}(x) + \theta_i^{-1}(y)), & \text{if } x, y \in [a_i, b_i] \end{cases} \tag{4.4}$$

where

$$\phi_i(x) = \begin{cases} \theta_i(x), & \text{if } x \in [0, \bar{u}_i] \\ a_i, & \text{if } x \in [\bar{u}_i, \infty], \quad \bar{u}_i < \infty, \end{cases} \tag{4.5}$$

and $0 < \bar{u}_i \leqslant \infty$, *and where* θ_i: $[0, \bar{u}_i] \to [a_i, b_i]$ *is continuous and strictly decreasing with* $\theta_i(0) = b_i$ *and* $\theta_i(\bar{u}_i) = a_i$.

A σ-composible information is of *type M* if $\Lambda^* = \{0, \infty\}$ and $\bar{u} < \infty$; it is of *type M′* if $\Lambda^* = \{0, \infty\}$ and $\bar{u} = \infty$. The Shannon information $\theta(x) = -c \log x$ is of type M, while $\theta(x) = 1/x$ is of type M' (hyperbolic information). While there is obvious interest in considering more general informations on $\mathscr{L}(P)$, it is not the purpose of this paper to do so. Rather, we accept the existence of σ-composible informations of type M on the quantum lattice $\mathscr{L}(P)$ and on Borel lattices.

Let us therefore assume henceforth that $\mathscr{L}(P)$ is an ortholattice with a weakly σ-composible type M information such that $p = \theta^{-1}(J)$ defines a weakly σ-additive probability measure on $\mathscr{L}(P)$. While our discussion below parallels quantum physics in many respects, we believe that our explicit introduction of P and functions X on P is novel and brings new understanding to this well-developed theory. Moreover, it suggests how one might generally construct models starting with an empirical relational system P and imbedding this in a (continuum) theory.

Let $X: P \to R^1$ be given and consider $X^{-1}(\Delta) = \{a \in P : X(a) \in \Delta\}$ for $\Delta \in B(R^1)$, the Borel algebra on R^1. We have probability measures defined on $\mathscr{L}(P)$ but not, in general, on 2^P. Hence, it is advantageous to consider $\hat{x}(\Delta) = \overline{X^{-1}(\Delta)} \in \mathscr{L}(P)$ rather than $X^{-1}(\Delta)$. Clearly more than one X can define the same $\hat{x}$, but each $\hat{x}$ defines a unique X via $\hat{x}(\Delta) = X^{-1}(\Delta)$, for all $\Delta \in B(R^1)$. We call such functions *observables* from their origin in quantum physics (Jauch, 1968).

LEMMA 4.1. *X is an observable iff $\hat{x}$: $B(R^1) \to \mathscr{L}(P)$ is a σ-morphism.*

Proof. If X is observable and Δ_k are given, $\hat{x}(\bigcup \Delta_k) = X^{-1}(\bigcup \Delta_k) = \bigcup X^{-1}(\Delta_k) = \bigvee \hat{x}(\Delta_k)$ as $X^{-1}(\bigcup \Delta_k)$ is closed. Also, $\hat{x}(\bigcap \Delta_k) = X^{-1}(\bigcap \Delta_k) = \bigcap X^{-1}(\Delta_k) = \bigcap \hat{x}(\Delta_k)$.

Conversely, suppose $\hat{x}$ is a σ-morphism and $X^{-1}(\Delta) \subset \hat{x}(\Delta)$. Then there exists $a \notin X^{-1}(\Delta)$ such that $a \in \hat{x}(\Delta)$. But, $a \in X^{-1}(\Delta')$ and $\Delta \cap \Delta' = \phi$, so $a \in \hat{x}(\Delta) \cap \hat{x}(\Delta') = \hat{x}(\phi) = \phi$, a contradiction. Thus we conclude $\hat{x}(\Delta) = X^{-1}(\Delta)$. Q.E.D.

These observables have many strange and useful properties. For one, it is clear that the image of $\hat{x}$ in $\mathscr{L}(P)$ must be a Borel sub-algebra of $\mathscr{L}(P)$. This means that $\mathscr{L}(P)$ must be infinite and must admit Borel sub-algebras if observables are to exist. In particular, as $B(R^1)$ possesses an "orthocomplement"

—namely the set complement E^c for $E \in B(R^1)$—it is straightforward to verify[1] using (4.1) that $\hat{x}(E^c) = \hat{x}(E)^\perp$. We demonstrate some other properties:

LEMMA 4.2. *If X is observable and $a \leqslant b$, $X(a) = X(b)$. In particular, if* $\mathrm{I} \in P$, *then* $X(a) = x_0$ *for all* $a \in P$.

Proof. If $a \in E \in \mathscr{L}(P)$, $\{a\} \subseteq E$ and thus $I(a) \subseteq E$. Thus if $b \leqslant a$, then $b \in E$. Using this, suppose $a \leqslant b$ and X is observable. Then, if $z = X(b)$, $a \in X^{-1}\{z\}$, so $X(a) = z$. Clearly, if $\mathrm{I} \in P$, $X(a) = X(\mathrm{I})$ for all $a \in P$. Q.E.D.

Note that there can be no observable information J_P defined on P if 0 and I are in P, as $J_P(0) = \infty \neq 0 = J_P(\mathrm{I})$. We now consider the case where $\mathscr{L}(P)$ is an atomic lattice.

DEFINITION 4.4. An element w in a lattice $\mathscr{L}$ with universal bound 0 is an *atom* if $0 \leqslant x \leqslant w$ implies that either $x = 0$ or $x = w$; a complete lattice $\mathscr{L}$ is *atomic* if for each $a \in \mathscr{L}$ there exists an atom $w \leqslant a$ and if $a = \bigvee_{w \in a} w$.
If $\mathscr{L}(P)$ is atomic (with $\Omega = \{w \in P : I(w)$ is an atom of $\mathscr{L}(P)\}$), then observables are determined by their values on the atoms Ω:

LEMMA 4.3. *Let X be an observable and let $\tilde{X}$ be its restriction to Ω. Then X is determined by $\tilde{X}$.*

Proof. Let $X_\Omega^{-1}(\Delta) = \{w \in \Omega : w \leqslant t,\ t \in X^{-1}(\Delta)\}$. Then, for all $\Delta \in B(R^1)$, $X_\Omega^{-1}(\Delta) \subseteq \tilde{X}^{-1}(\Delta)$ using lemma 4.2. Were the inclusion strict, there would be an atom $w_0 \in \tilde{X}^{-1}(\Delta)$ such that $w_0 \nleqslant t$ for any $t \in X^{-1}(\Delta)$. Thus, $w_0 \leqslant t_0$ for $t_0 \in X^{-1}(\Delta^c)$, so $X(w_0) \in \Delta^c$, a contradiction. Thus, $X_\Omega^{-1}(\Delta) = \tilde{X}^{-1}(\Delta)$.

Now for any $\bar{E} \in \mathscr{L}(P)$, $\bar{E} = \bigvee_{t \in E} I(t) = \bigvee_{t \in E} [\bigvee_{w \in \Omega, w \leqslant t} I(w)] = \bigvee_{w \in E_\Omega} I(w)$ with $E_\Omega = \{w \in \Omega : w \leqslant t\}$. Thus, $\bar{E} = \bar{E}_\Omega$, and in particular it follows that $X^{-1}\{x\} = \overline{\tilde{X}^{-1}\{x\}}$, for all real x. Q.E.D.

Returning to the more general case (non-atomic), let B be a Borel subalgebra of $\mathscr{L}(P)$. Then, there is a set V, a σ-field $B(V)$ of subsets of V, and a σ-morphism $\chi: B(V) \to B$ that is onto ((Birkhoff, 1973), page 255, Theorem 3). For each probability p on $\mathscr{L}(P)$, let $m_p = p\chi$. This defines a σ-additive measure on $B(V)$. Note, however, that $p_1 \neq p_2$ with disagreement only outside B leads to $m_{p_1} = m_{p_2}$. The measure m_p is a probability on $B(V)$ iff $\chi(\phi) = \phi$ and $\chi(V) = P$; we assume these conditions in the sequel.

A real-valued function f is said to be a *Borel function* on V if its domain $\Omega \in B(V)$ and if for all $\Delta \in B(R^1)$, $f^{-1}(\Delta) \in B(V)$. If f is such a function, $\hat{x}(\Delta) = \chi(f^{-1}(\Delta))$ clearly defines a σ-morphism, and related observable X on P via $\hat{x}(\Delta) = X^{-1}(\Delta)$.

[1] Note that $\hat{x}(\phi) = \overline{X^{-1}(\phi)} = \phi$ and $\hat{x}(P) = \overline{X^{-1}(P)} = P$. For an arbitrary σ-morphism $\chi: B(R^1) \to \mathscr{L}(P)$ not of the form $\overline{X^{-1}(\Delta)}$ this fails, and $\chi(E^c) \neq \chi(E)^\perp$ as (4.1a) fails.

Suppose instead that we begin with an observable X on P. As χ: $B(V) \to B$ is onto, for each $\Delta \in B(R^1)$ there exists $E(\Delta) \in B(V)$ defined by $X^{-1}(\Delta) = \chi(E(\Delta))$—remember, the range of $\hat{x}$ is a Borel subalgebra of $\mathscr{L}(P)$. Define $f^{-1}(\Delta) = E(\Delta)$. This defines a Borel function almost everywhere. That is, in order that f be a function on $V, f^{-1}\{y\} \cap f^{-1}\{z\} = \phi$ when $y \neq z$. But although $\{y\} \cap \{z\} = \phi$, we can only conclude that $f^{-1}\{y\} \cap f^{-1}\{z\} \in \operatorname{Ker} \chi$ (the kernel of χ is the set of elements of $B(V)$ which are mapped by χ onto ϕ). But in this case, $m_p(f^{-1}\{y\} \cap f^{-1}\{z\}) = 0$, so the set of points in V on which f is not a function has measure zero. Call this set N. As $N \in B(V)$, $\Omega = V - N \in B(V)$ and $f_x(v) = f(v)$ thus defines a Borel function *with domain* Ω. Thus, any observable on P is associated with a Borel function on Ω.

Note that if f_1 and f_2 are two Borel functions such that $\hat{x}(\Delta) = \chi(f_1^{-1}(\Delta)) = \chi(f_2^{-1}(\Delta))$, then in particular $\hat{x}\{x\} = \chi(f_1^{-1}\{x\} \cap f_2^{-1}\{x\})$. However, it is possible that $f_2^{-1}\{x\} = f_1^{-1}\{x\} \cap N(x)$, where $N(x) \in \operatorname{Ker} \chi$. Thus f_1 and f_2 are only equal up to sets in $\operatorname{Ker} \chi$. Clearly, however, $f_1 = f_2$ almost everywhere in the sense that the measure m_p of the set of points on which they are not equal is zero.

An especially interesting observable is generated by the characteristic function C_E: $V \to \{0, 1\}$ defined by[2]

$$\hat{K}_E(\Delta) = \chi(C_E^{-1}(\Delta)), \qquad E \in B(V). \tag{4.6}$$

Clearly $\hat{K}_E\{1\} = \chi(E)$, $\hat{K}_E\{0\} = \chi(E^c) = \hat{K}_E\{1\}^\perp$. This defines the observable

$$K_E(a) = \begin{cases} 1, & \text{if } \ a \in \chi(E) \\ 0, & \text{if } \ a \in \chi(E^c) = \chi(E)^\perp. \end{cases} \tag{4.7}$$

Note that $K_E^{-1}\{0, 1\} = K_E^{-1}\{0\} \cup K_E^{-1}\{1\} = P$, so that $P = \chi(E) \cup \chi(E)^\perp$. Thus, (4.7) can be considered an observable characteristic function on P.

As is well-known (Munroe, 1953), any Borel function on V is expressed as the limit of a sequence of "simple functions":

$$f(v) = \operatorname*{Lim}_{N\to\infty} \left\{ \sum_{k=1}^{N2^N} \left[\frac{k-1}{2^N}\right] C_{f^{-1}[(k-1)/2^N, k/2^N)}(v) + N C_{f^{-1}[N,\infty]}(v) \right\} \tag{4.8}$$

Thus, letting $f_N(v)$ represent the Nth such simple function and with the obvious shorthand notation:

$$f_N(v) = \sum a_n C_{f^{-1}(E_n)}(v) \tag{4.9}$$

we find

$$f_N^{-1}(\Delta) = \bigcup_{n\,\mathrm{s.t.}\, a_n \in \Delta} f^{-1}(E_n) = f^{-1}\left(\bigcup_{(a_n \in \Delta)} E_n\right). \tag{4.10}$$

[2] Again, this only works if $\chi(\phi) = \phi$ and $\chi(V) = P$.

Thus, for every N,

$$\chi(f_N^{-1}(\Delta)) = \bigvee_{(a_n \in \Delta)} \chi(f^{-1}(E_n)) = \hat{f}_N(\Delta) = F_N^{-1}(\Delta), \tag{4.11}$$

where the observable[3] $F_N = \sum a_n K_{f^{-1}(E_n)}$. As χ is a σ-morphism, this holds for infinite N, so we propose that every Borel function on V defines an observable on P of the form

$$X(a) = \operatorname*{Lim}_{N\to\infty} \sum a_n K_{f_X^{-1}(E_n)}(a). \tag{4.12}$$

But *any* observable X on P defines a Borel function f_X and thus an observable F_X expressed by (4.12). As $X^{-1}(\Delta) = \chi(f_X^{-1}(\Delta)) = F_X^{-1}(\Delta)$ holds for all $\Delta \in B(R^1)$, we conclude that $X = F_X$, i.e., that every observable on P has representation (4.12). (The non-uniqueness of f_X is irrelevant.)

Now K_E is an observable, provided $E \in B(V)$ and $\chi(\phi) = \phi$, $\chi(V) = P$. We might as well have denoted K_E by $K_{\chi(E)}$. More generally, we ask when is K_A a "characteristic observable" for $A \in \mathscr{L}(P)$. Certainly if A and $A^\perp$ are contained in *some* Borel sub-algebra B of $\mathscr{L}(P)$, we can associate A with $\chi(E)$ and the desired properties follow. I.e., $A \cap A^\perp = \phi$, $A \vee A^\perp = P \in B$ and χ being onto assures us that us that $\chi(\phi) = \phi$ and that $\chi(V) = P$ for the σ-morphism χ: $B(V) \to B$. In this case every K_A defined by

$$K_A(a) = \begin{cases} 1, & \text{if } a \in A \\ 0, & \text{if } a \in A^\perp \end{cases} \tag{4.13}$$

defines an observable characteristic function on P. This leads us to study the evaluation $e(K_A)$, $A \in \mathscr{L}(P)$.

Let us now assume that the valuation $e(K_A) = F(J(A))$, where $A \in \mathscr{L}(P)$ and J is weakly σ-composible. In other words, as K_A is a characteristic function associated with set A, it is reasonable to anticipate that our evaluation of this function depend solely on the information content of A. In addition, if $J(A)$ is changed to $J'(A)$ where $|J'(A) - J(A)|$ is "small", we do not expect our evaluation to change by much. (We ignore possibilities related to questions about extraterrestrial life, for example,where small changes in our information can lead to drastic changes in functions related to our beliefs.) Thus, we suppose $F(x)$ is continuous on $[0, \infty]$. If $x \neq y$, we assume $F(x) \neq F(y)$, so that we forbid the case that two informations disagree on A yet yield the same evaluation.

LEMMA 4.4. *$F(x)$ is strictly decreasing on $[0, \infty] \to [0, 1]$ with $F(0) = 1$ and $F(\infty) = 0$.*

[3] Certainly $F_N^{-1}(\Delta) = \bigcup_{(a_n \in \Delta)} \chi(f^{-1}(E_n)) \subseteq \chi(f_N^{-1}(\Delta))$. Were the inclusion strict, there would be an element $a \in P$ such that $a \in F_N^{-1}(\Delta^c) \cap \chi(f_N^{-1}(\Delta))$. But, $F_N^{-1}(\Delta^c) \subseteq \chi(f_N^{-1}(\Delta^c))$ precludes this, so (4.11) follows.

Proof. Since $e(K_P) = 1 = F(J(P))$, $F(0) = 1$; as $e(K_\phi) = 0 = F(J(\phi))$, $F(\infty) = 0$. (Note that $K_P(a) = 1$, for all $a \in P$ and $K_\phi(a) = 0$, for all $a \in P$.) As F is $1 - 1$ and continuous on $[0, \infty]$, it must be strictly monotonic, and hence strictly decreasing. Let $F = \Theta^{-1}$. Q.E.D.

From lemma 4.4 and the weak σ-composibility of J, if J is of type M, then $e(K_A) = \Theta^{-1}(J(A)) = p(A)$ is a weak σ-additive probability measure on $\mathscr{L}(P)$. Note that if $A = \hat{x}(\varDelta)$, $\varDelta \in B(R^1)$, $\hat{x}$ a σ-morphism, then $p\hat{x}$ is a probability measure on the Borel sets of the reals.

We now propose the following property of the evaluation procedure:

$$e\left(\sum_{n=1}^{N} a_n K_{A_n}\right) = \sum_{n=1}^{N} a_n e(K_{A_n}) \tag{4.14}$$

for all finite N, where a_n are reals and $\{A_n \in \mathscr{L}(P) : A_n \perp A_m \ (n \neq m)\}$ is contained in some Borel subalgebra of $\mathscr{L}(P)$. Clearly $\{A_n = \hat{x}(E_n)\}$ for the E_n defined by (4.8) satisfies these conditions. Using (4.12) and (4.14) we *define*

$$e(X) = \operatorname*{Lim}_{N\to\infty} \sum a_n p\hat{x}(E_n) = \operatorname*{Lim}_{N\to\infty} \sum a_n m_p(f_X^{-1}(E_n)), \tag{4.15}$$

where X is given by (4.12) with $K_{f_X^{-1}(E_n)} = K_{\chi(f_X^{-1}(E_n))} = K_{\hat{x}(E_n)}$ as explained above. Following (Munroe, 1953) we conclude that if X is an observable on P,

$$e(X) = G(X, J) = \int_{-\infty}^{\infty} t \, dp \, \hat{x} = \int_V f_X \, dm_p \tag{4.16}$$

where J is weakly σ-composible of type M on $\mathscr{L}(P)$, $p = F(J)$, and $m_p = p\chi$.

Note that if X and Y are observables which do not relate to a common space V, then $e(X + Y)$ is meaningless. If the two observables do refer to a common space—i.e., $\hat{x}$ and $\hat{y}$ have ranges in a common Borel subalgebra of $\mathscr{L}(P)$—, then $f_X + f_Y$ is a well-defined Borel function on the space V, and one can define $e(X + Y) = \int_V (f_X + f_Y) \, dm_p = e(X) + e(Y)$. In this case we say that X and Y are *compatible* observables, again following quantum theory (Jauch, 1968). Clearly if $\mathscr{L}(P)$ is a Borel algebra, all observables on P are compatible.

There is one last case of special interest, namely that where P (and thus $\mathscr{L}(P)$) is finite. If one considers $B(R^1)$ to be the *Boolean* algebra of subsets of R^1 generated by its open sets, then $\hat{x}(\varDelta) = X^{-1}(\varDelta)$ defines what we call a *finitary observable* (*f*-observable). Lemma 4.1 remains valid with $\hat{x}$ a morphism. (Lemmas 4.2 and 4.3 are independent of the cardinality of P, so are unaffected.) Suppose that $\mathscr{L}(P)$ is a Boolean algebra. Then it turns out that $\mathscr{L}(P)$ is isomorphic to the power set of its atoms (see Theorem 17, p. 119 of (Birkhoff, 1973)) and that the isomorphism associates each $E \in 2^P$ with its completion $\bar{E}$ $(\chi(E) = \bar{E} = \bigvee_{w\in E} I(w))$. Now any function $f \colon \Omega \to R^1$ determines a unique *f*-observable F on P (lemma 4.3) as $f_X^{-1}(\varDelta)$ is in $1 - 1$ correspondence with $\overline{f_X^{-1}(\varDelta)} = F_X^{-1}(\varDelta)$

for all $\Delta \in B(R^1)$. Thus, every observable is associated with a unique function on Ω. Obviously, $f(w) = \sum_{w_k \in \Omega} f(w_k)\, C_{\{w_k\}}(w)$ and $\chi(f^{-1}(\Delta)) = F^{-1}(\Delta) = \bigvee_{(f(w_k) \in \Delta)} \chi(C^{-1}_{\{w_k\}}(\Delta))$ where, as before, $K^{-1}_{\{w_k\}}(\Delta) = \chi(C^{-1}_{\{w_k\}}(\Delta))$ is an observable characteristic function on P, and $F(a) = \sum f(w_k)\, K_{\{w_k\}}(a)$. Finally,

$$e(X) = \sum f(w_k)\, p(I(w_k)) = \sum f(w_k)\, m_p(w_k) \tag{4.17}$$

with $m_p = p\chi$. In particular, $\Theta^{-1} m_p$ defines an observable information J_P on P and

$$e(J_P) = \sum (\Theta^{-1} m_p(w_k))\, m_p(w_k). \tag{4.18}$$

We have already stated that if P is the lattice of finite-dimensional Hilbert sub-spaces, $\mathscr{L}(P)$ is the complete lattice of all subspaces of the Hilbert space. Suppose P is a poset *all of whose elements are incomparable* in the order. Then $P = \Omega$ defines the atoms of $\mathscr{L}(P)$. If we demand that $\mathscr{L}(P)$ be a completely distributive complete Boolean algebra (this includes the finite case), then $\mathscr{L}(P)$ is isomorphic to 2^P. In particular, every function on P is a function on a set Ω of points, while the probabilities on $\mathscr{L}(P)$ are actually (via the isomorphism) probabilities on 2^Ω (trivially a Borel algebra). Note that this is why in the above example $F(a) = \sum f(w_k)\, K_{\{w_k\}}(a) = f(w_k)$, for $a = w_k$, and $=0$ if a is not an atom is an observable: The only elements of P *are* the atoms, so in particular there is no $a \in P$, $a > w_k$ to conflict with the requirements of Lemma 4.2.

V. MEASUREMENT, THEORY, AND INFORMATION (Continued)

In Section III we argued that an empirical system is operationally defined by the poset P of its (relevant) empirical relations ordered by "implication". Functions on P are to be evaluated by associating with experimental data an information $J \in Z(\mathscr{L}(P))$ in an unambiguous manner. In this section we will consider this problem.

Once the system P is defined, we know $\mathscr{L}(P)$ and the complete lattice of all informations definable on $\mathscr{L}(P)$. In general, the laboratory provides us with evaluations of some appropriate functions X on P. These serve to limit $Z(\mathscr{L}(P))$ to some subset Q of informations all consistent with the given data. For example, $Q = \{J \in Z(\mathscr{L}(P)) : G(X, J) = x_0\}$. That this situation is typical is evident from probability theory with $e(X)$ the mathematical expectation: $e(X) = x_0$ does not determine probability uniquely. Clearly $Z(\mathscr{L}(P))$ may contain informations not of type M (not related to probabilities). Thus, one must a priori define a suitable class $K = \{Q \in Z(\mathscr{L}(P))\}$ such that each $Q \in K$ can be associated uniquely with some relevant laboratory datum. That is, the result of each measurement should correspond to some $Q \in K$, and there should be no $Q \in K$ that cannot be related to any measurement.

Beyond this correspondence, it is also necessary that to each $Q \in K$ we can assign a *unique* $J_Q \in Q$ that serves to characterize or "label" Q. The reason, of course, is that from this measurement we want to be able to unambiguously evaluate any function X on P—at least any of a class of functions (eg. "observables") that the overall consistency of the model permits. Thus, $e(X//Q) = G(X, J_Q)$ is the required procedure.

Hence, we require a priori P (system definition), G (evaluation procedure), K (class of information sets corresponding to experimental outcomes), a class Γ of allowed functions X on P, and an algorithm to choose exactly one $J_Q \in Q$ for each $Q \in K$. If either $\wedge_{J \in Q} J$ or $\vee_{J \in Q} J$ exist in Q, either of these is unique and serves the purpose. Another approach is given by:

Definition 5.1. A family $K = \{Q \in Z(\mathscr{L}(P))\}$ is *G-admissible* if there exists $\chi_G: K \to [0, \infty]$ such that for each $Q \in K$ the set $\{J \in Q: G(J_P, J) = \chi_G(Q)\}$ is the singleton $\{J_Q\}$.

We call $G(J_P, J)$ the *entropy* and $H(Q) = G((J_Q)_P, J_Q)$ the *entropy of* Q. The motivation for definition 5.1 is that K is thus associated intimately with the evaluation procedure G and the information measures. No doubt other algorithms can be devised. The criterion of G-admissibility is of special interest, however, as it leads to the generalization of the MEP below.

As an illustration of a "metatheory" that synthesizes these ideas, consider the following. Let $v_t(a, X)$ be a rule or procedure that associates with an object $a \in A$ some real number *valid* at time t which represents the value of X for a. We do not claim that the object a actually *has* this value at t, or at any other time, for that matter. As experiments are of finite duration, t is hardly well-defined. Hence, we simply *agree* that at t it is appropriate—valid—to say $v_t(a, X) = x$. An "experiment" or "measurement" consists in the simultaneous determination of some finite set of values for functions in a class:

$$E_t^N(a_1, \dots, a_N; X_1, \dots, X_N) = \{v_t(a_k, X_k) : a_k \in A, X_k \in \Gamma, k = 1, \dots, N\} \quad (5.1)$$

for N finite. The class of all possible measurements is

$$E = \{E_t^N(a_1, \dots, a_N; X_1, \dots, X_N) : t, N \text{ finite}, a_k \in A, X_k \in \Gamma\}. \quad (5.2)$$

Note that while the measurements of E can be considered in terms of empirical relations, we assume here that these relations are *not* part of the family of empirical relations that *define* the system class A. As noted above, one generally *assumes* a model P (defining relations for the system are hypothesized) and "measures" functions X defined on P. Overall consistency of observations, model, and predictions provides verification for the model.

Let $\Phi: E \to K$ uniquely associate with each experiment a subset $Q \in K$, where K is G-admissible, and Φ is onto. At this point Γ is restricted by requiring that $\Phi(E_t^N(a_1, \dots, a_N; X_1, \dots, X_N)) \neq \phi$, for all $a_k \in A$ and $X_k \in \Gamma$. Thus, only functions whose values correspond to nonempty subsets of K are allowed. Such a model is still not very specific, but reflects the ideas we are proposing. We trust that the reader will see also the reflection of a reasonable skeleton for a measurement-theory-information model of science.

In fact, one may consider the "scientific method" in terms of a sequence $T_k = \{P_k, G_k, K_k, \chi_{G_k}, \Gamma_k\}$ where the definition of the system, method of evaluation, concept of data input, allowed function class evolve. Certainly the transition from classical to quantum physics is a dramatic example of this evolution.

The reader should beware of the obvious temptation to call the elements of the class K "evidences". This word—"evidence"—is much abused and rarely defined by writers on measurement theory and "classical" probability. (Indeed, in a recent book entitled "A Mathematical Theory of Evidence" (Shafer, 1976), one never finds any attempt to define or even characterize "evidence".) (Carnap, 1950) at least stipulates that "evidence" is a proposition, defined by being the second argument in the "degree of belief" $p(h/e)$, where the first argument (h) is the hypothesis, another proposition. In the case where our informations are of type M, our model might be considered as a means of defining "degrees of belief" $p(h//Q)$ for "propositions" $h \in \mathscr{L}(P)$ given "evidence" $Q \in K$. We will consider this possibility in the next section.

We now turn to a generalization of Jaynes' MEP.

THEOREM 5.1. *Let g: $Z' \to W$ be a $1 - 1$ correspondence such that $Z' \subseteq Z(\mathscr{L}(P))$ and W is a topological space in which $S_{ab} = \{w \in W : w = tw_a + (1-t)w_b, t \in [0, 1]\}$ is an arc. Moreover, let $g(Q) = R$ be convex and compact for all $Q \in K$. Finally, let $\bar{H}(w) = \bar{H}(g(J)) = G(J_P, J)$ be continuous on R* (for all R) *and strictly convex down. Then, the only choice of χ_G for which K is generally G-admissible is*

$$\chi_G(Q) = \sup\{\bar{H}(w) : w \in g(Q)\}. \tag{5.3}$$

Proof. (A) Proof that (5.3) makes K G-admissible.

For K to be G-admissible, for all $Q \in K$ $\{J \in Q : G(J_P, J) = \chi_G(Q)\}$ must be a singleton. Let $f_G(R) = \chi_G(Q)$ where $R = g(Q)$. Then,

$$\begin{aligned}\{w \in R : f_G(R) = \bar{H}(w)\} &= \{g(J) : J \in Q, \chi_G(Q) = G(J_P, J)\}\\ &= g\{J \in Q : \chi_G(Q) = G(J_P, J)\}.\end{aligned}$$

Thus, K is G-admissible iff for all $Q \in K$ $\{w \in R = g(Q) : f_G(R) = \bar{H}(w)\}$ is a singleton (g is a $1 - 1$ correspondence).

By strict convexity and continuity of $\bar{H}(w)$ over a compact, convex set R, if $f_G(R) = \sup\{\bar{H}(w) : w \in R\}$, it is clear that $\{w \in R : f_G(R) = \bar{H}(w)\}$ is a singleton, as H attains it supremum over R at exactly one point of R. Thus, (5.3) suffices for K to be G-admissible.

(B) Proof that only (5.3) admits K to be G-admissible in all cases.

A priori there are three cases: (i) R is a single point of W; (ii) R is an arc (contains two distinct points, and thus the connecting line); (iii) R contains at least three distinct points, not all on the same arc.

Consider case (iii): Let w_1, w_2, w_3 be three points with $w_1 \notin S_{23}$, $w_2 \notin S_{31}$, $w_3 \notin S_{21}$. In particular, consider that $\bar{H}(w_1) = \sup\{\bar{H}(w) : w \in R\}$. Suppose $f_G(R) \neq \bar{H}(w_1)$. Then, if $f_G(R) > \bar{H}(w_1)$, there is clearly no $w \in R$ such that $f_G(R) = \bar{H}(w)$. Such a choice cannot make K G-admissible. Thus, assume $f_G(R) < \bar{H}(w_1)$. Assuming $f_G(R) = \bar{H}(w_2)$, for K to be G-admissible, $\bar{H}(w_3) \neq f_G(R)$. Suppose, therefore, that $\bar{H}(w_3) < f_G(R) < \bar{H}(w_1)$. Continuity of $\bar{H}$ on S_{31} means there exists t_0 such that $\bar{H}(t_0 w_3 + (1 - t_0)w_1) = f_G(R) = \bar{H}(w_2)$; as $w_2 \notin S_{31}$, K is clearly not G-admissible. Suppose we take $f_G(R) < \bar{H}(w_3)$. Either $f_G(R) = \inf\{\bar{H}(w) : w \in R\}$ or not. If not, continuity on the line connecting w_3 with any point at which $\bar{H}$ attains its inf leads to a contradiction similar to the above. Finally, if $f_G(R) = \inf\{\bar{H}(w) : w \in R\}$, in general more than one element in R satisfies $\bar{H}(w) = f_G(R)$. Thus, in general, for case (iii), (5.3) alone admits G-admissibility of K.

Thus we conclude that (as (5.3) suffices in cases (i) and (ii) to make K G-admissible), that the only choice for χ_G that makes K G-admissible in *all* cases is given by (5.3). Q.E.D.

We now illustrate this theorem by formulating the usual version of the MEP. Let $P = \Omega$ be a finite set of non-comparable elements and $\mathscr{L}(P)$ a Boolean algebra; $\chi: 2^{\Omega} \to \mathscr{L}(P)$ is an isomorphism. If $J_P(w_k) = \Theta(q_k)$, where $q_k = m_p\{w_k\}$, then J_P is an observable information. Let $W = \{\mathbf{q}\} : k = 1, 2, \ldots \| P \|$, $q_k \geqslant 0$, $\sum q_k = 1\}$, and let $Z' = \{J \in Z(\mathscr{L}(P)) : J = \Theta p$, p a probability on $\mathscr{L}(P)\}$. Then, $g: W \to Z'$ is a $1 - 1$ correspondence. For, each $\{\mathbf{q}\} \in W$ determines an unique J as $J(\bar{E}) = \Theta p(\bar{E}) = \Theta(\sum_{w_k \in E} q_k)$ and Θ^{-1} is a $1 - 1$ correspondence. Also, every $J = \Theta p$ is defined by some $\{\mathbf{q}\} \in W$, since $J(I(w_k)) = \Theta(q_k)$ defines $\{\mathbf{q}\} \in W$ and $J(\bar{E}) = \Theta(\sum_{w_k \in E} q_k)$ fixes J on $\mathscr{L}(P)$ Finally, g is $1 - 1$ since $\{\mathbf{q}\} \neq \{\mathbf{q}'\}$ implies for some $w_k \in P$, $J(I(w_k)) = \Theta(q_k) \neq \Theta(q_k') = J'(I(w_k))$.

Now W is closed and bounded in $\| P \|$-dimensional Euclidean space. It is thus compact (Theorem 33, page 89, (Bushaw, 1963)). We define $R(x_1, \ldots, x_N) = \{\{\mathbf{q}\} \in W : x_n = \sum_{w_k \in P} X_n(w_k) q_k, n = 1, \ldots, N\}$. This is, of course, the closed, bounded subset of W (and hence compact) of distributions consistent with $e(X_n) = x_n$, $n = 1, \ldots, N$. Note that $R(x_1, \ldots, x_N)$ is obviously convex, and the line segment $S_{ab} = \{\mathbf{q}(t) : q_k(t) = t q_k{}^a + (1 - t) q_k{}^b,\ t \in [0, 1]\}$ is an arc (problem 154, page 101, (Bushaw, 1963)). Now $\bar{H}(\mathbf{q}) = \sum_{w_k \in P} q_k \Theta(q_k)$ is certainly continuous, so it remains to pose conditions on $\Theta(x)$ such that $\bar{H}$ is strictly convex down. The usual $\Theta(x) = -c \log x$ certainly satisfies these demands. Thus, the MEP algorithm for choosing $\{\mathbf{q}\} \in W$ emerges as the only choice which makes the class $K = \{g(R(x_1, \ldots, x_N))\}$ G-admissible for the usual $G(X, J) = \sum_{w_k \in P} X(w_k)\, \Theta^{-1}(J(I(w_k)))$.

To illustrate that we need not restrict the "MEP" to probability-related informations, we now consider a rather interesting example of topological space W.

THEOREM 5.2. *Let W be any complete convex lattice. The* interval topology τ *is defined by having as subbasis of closed sets the family* $[w_a, w_b] = \{w \in W : w_a \leqslant w \leqslant w_b\}$. *Let K_W be a family of closed, convex subsets of W, and let $\bar{H}(w)$ be continuous and strictly convex down on $R \in K_W$. Then, $\{w \in R : f(R) = \bar{H}(w)\}$ is a singleton in general iff* $f(R) = \sup\{\bar{H}(w) : w \in R\}$.

Proof. According to (Birkhoff, 1973), page 250, W is a compact topological space. Thus, the closed sets in K_W are compact—Theorem 21, page 77 (Bushaw, 1963). We show below in Lemmas 5.1 and 5.2 that S_{ab} is an arc. Thus, according to the proof to theorem 5.1, the conclusion follows. Q.E.D.

LEMMA 5.1. *If τ^{ab} is the topology induced on S_{ab} by τ, then the subbasis of closed sets for τ^{ab} is given by*

$$S_{ab} \cap [w_l, w_u] = \{w(t) : t \in [t', t'']\} \tag{5.4}$$

where $w(t) = tw_a + (1 - t)w_b$ and $[t', t''] \subseteq [0, 1]$.

Proof. Let $T^* = \{t \in [0, 1] : w_l \leqslant w(t) \leqslant w_u\}$. Then the result follows once we have shown that $T^* = [\bigwedge_{t^* \in T^*} t^*, \bigvee_{j^* \in T^*} t^*]$.

Now if $t_1 \leqslant t_2$, $t_i \in T^*$, then $[t_1, t_2] \subseteq T^*$: Let s_{12} be the segment $uw_1 + (1 - u)w_2$, $u \in [0, 1]$, where $w_l \leqslant w_i = t_i w_a + (1 - t_i)w_b \leqslant w_u$. Then $w_l \leqslant w_1 \wedge w_2 \leqslant s_{12} \leqslant w_1 \vee w_2 \leqslant w_u$. Also, it is clear that as u varies over $[0, 1]$, s_{12} represents a variation of t over $[t_1, t_2]$.

Next, let $\{t_1(a)\}$ be any non-increasing sequence in T^* such that $\bigwedge t_1(a) = \bigwedge t^*$ and let $\{t_2(b)\}$ be any non-decreasing sequence in T^* such that $\bigvee t_2(b) = \bigvee t^*$. Finally, choose $t_1(a_0) \leqslant t_2(b_0)$. Define $I(a, b) = [t_1(a), t_2(b)]$. By the first part, $I(a, b) \subseteq T^*$, for all a, b. Thus, as $I(a, b)$ under inclusion is a non-decreasing chain, $\sup\{I(a, b)\} = [\bigwedge t_1(a), \bigvee t_2(b)] \subseteq T^*$. Obviously, $T^* \subseteq [\bigwedge_{j^* \in T^*} t^*, \bigvee_{j^* \in T^*} t^*]$, so we are done. Q.E.D.

LEMMA 5.2. *S_{ab} is an arc.*

Proof. The map w: $[0, 1] \to S_{ab}$ defined by $w(t) = tw_a + (1 - t)w_b$ is clearly $1 - 1$, onto. Every closed set on S_{ab} is of the form $K = \bigcap_{a \in A} (\bigcup_{n_a=1}^{N_a} K_{n_a})$, where $K_{n_a} = \{w(t) : t \in [t'_{n_a}, t''_{n_a}]\}$ is the subbasis element $S_{ab} \cap [w'_{n_a}, w''_{n_b}]$—(Bushaw, 1963), chapter 4. Thus, the pre-image of any closed set of S_{ab} under w is the set $\bigcap_{a \in A} (\bigcup_{n_a=1}^{N_a} T^*_{n_a})$ which is closed on $[0, 1]$ as $T^*_{n_a} = [t'_{n_a}, t''_{n_a}]$ (by lemma 5.1) is in the subbasis of closed sets on $[0, 1]$. Thus, w is continuous.

Now $w[t_1, t_2] = [w(t_1) \wedge w(t_2), w(t_1) \vee w(t_2)] \cap S_{ab}$, so it follows that the pre-image of any closed set on $[0, 1]$ under w^{-1} is a closed set in S_{ab}, i.e. that w^{-1} is continuous.

We therefore conclude that w is a homeomorphism $[0, 1] \to S_{ab}$ and that S_{ab} is an arc—page 100, (Bushaw, 1963). Q.E.D.

Theorem 5.2 is of special interest because $Z(\mathscr{L}(P))$ is itself a complete convex lattice. Any family of closed sets in the interval topology will thus define a G-admissible class of informations provided one can construct $\bar{H}(w) = G(J_P, J)$ continuous and strictly convex down. Once one such class is found, one can ask about the effects of the group of automorphisms on $Z(\mathscr{L}(P))$ on this class. We will pursue this study elsewhere.

VI. Bayes' Rule

In the Introduction we raised a question concerning the use of Bayes' rule in relating "evidences" and "hypotheses". The question is whether evidence is a proposition in the same "language" as is the hypothesis, at least when "evidence" is of the form required by the MEP. We established in the preceding section that the MEP is concerned with G-admissible classes K of subsets of informations in $Z(\mathscr{L}(P))$, and warned that these sets Q are not "evidences" in the traditional sense of the word. It is clear, in fact, that Q (together with G and χ_G) serves to fix the "degree of belief" $p(h//Q)$ for all $h \in \mathscr{L}(P)$, while $p(h, e)$ is not "fixed" given h and e (problem of "priors") in classical Bayesian theory. Also, it is not clear that Q can be associated with any language $\hat{\mathscr{L}}$, although is apparent that Q cannot be associated with propositions in $\mathscr{L}(P)$. For example, $\| P \| < \infty$ in the MEP case of the preceeding section, but $\| K \| = \mathscr{C}$, as there is a continuum of different allowed sets $R(x_1, \ldots, x_N)$ corresponding to the continuum range of the probabilities. Thus, the "evidences" $Q \in K$ do not serve as part of the *definition* of the system, i.e. $\mathscr{L}(P)$. More generally, it is hardly reasonable to expect that the choice of G (evaluation procedure) should have bearing on the "attributes" in $\mathscr{L}(P)$. We further illustrate this point by noting that a list of mutually exclusive and exhaustive states of an atom defines the atom, while the expected value of its hamiltonian (energy) does not correspond to any particular state or subset of states, but serves only to provide a family of density matrices consistent with the expected energy.

Let us then *suppose* that there exists some "language" $\hat{\mathscr{L}}$ whose propositions $\hat{d}_Q$ correspond in a $1-1$, onto way with $Q \in K$. In other words, we consider some lattice ordering of K as the "language" $\hat{\mathscr{L}}$. An immediate problem presents itself: Just how is $\hat{\mathscr{L}}$ to be defined? We suggest three alternatives for the ordering:

(i) $\hat{d}_Q \leqslant \hat{d}_{Q'}$ iff $Q \supseteq Q'$ in $Z(\mathscr{L}(P))$;

(ii) $\hat{d}_Q \leqslant \hat{d}_{Q'}$ iff $J_Q \leqslant J_{Q'}$ in $Z(\mathscr{L}(P))$;

(iii) $\hat{d}_Q \leqslant \hat{d}_{Q'}$ iff $\chi_G(Q) \geqslant \chi_G(Q')$.

The first ordering represents a "quality" grading of evidence—the more restrictive the evidence (smaller allowed set in $Z(\mathscr{L}(P))$) the better the "grade". Ordering (ii) essentially imbeds $\hat{\mathscr{L}}$ in $Z(\mathscr{L}(P))$. The last possibility follows Shannon's intuitive idea that "entropy" is "uncertainty", and grades evidence inversely according to its uncertainty level.

Each of these orderings—and one can probably conceive of others—has intuitive appeal and seems natural, so unless in certain circumstances they are equivalent, there is no unique $\hat{\mathscr{L}}$ available. Still, in some situations a suitable choice for $\hat{\mathscr{L}}$ may present itself, despite the general ambiguity. Let us therefore suppose we have such a well-defined $\hat{\mathscr{L}}$ with universal bounds $\hat{d}_\phi$ and $\hat{d}_K$.

Let us consider $\mathscr{L}_1 = \mathscr{L}(P)$ and $\mathscr{L}_2 = \hat{\mathscr{L}}$, and define the lattice product $\mathscr{L}_1 \otimes \mathscr{L}_2 = \{(a_1, a_2) : a_1 \in \mathscr{L}_1, a_2 \in \mathscr{L}_2\}$ ordered by

$$(a_1, a_2) \leqslant_{12} (b_1, b_2) \quad \text{iff} \quad a_1 \leqslant_1 b_1 \quad \text{and} \quad a_2 \leqslant_2 b_2. \tag{6.1}$$

Then $\mathscr{L}_1 \otimes \mathscr{L}_2$ is a lattice with universal bounds $(0_1, 0_2)$ and (I_1, I_2). Let $J_{12} \in Z(\mathscr{L}_1 \otimes \mathscr{L}_2)$ satisfy

$$J_{12}(0_1, I_2) = J_{12}(I_1, 0_2) = \infty. \tag{6.2}$$

Then we may easily verify that $J_1 \in Z(\mathscr{L}_1)$ and $J_2 \in Z(\mathscr{L}_2)$ with

$$J_1(a_1) = J_{12}(a_1, I_2), \qquad \text{for all} \quad a_1 \in \mathscr{L}_1 \tag{6.3a}$$

$$J_2(a_2) = J_{12}(I_1, a_2), \qquad \text{for all} \quad a_2 \in \mathscr{L}_2. \tag{6.3b}$$

Consider the relation

$$J_{12}(a_1, a_2) = J_1(a_1) + S(a_2/a_1) = J_2(a_2) + T(a_1/a_2) \tag{6.4}$$

defining S and T for all $(a_1, a_2) \in \mathscr{L}_1 \otimes \mathscr{L}_2$. Provided J_{12} satisfies

$$J_{12}(a_1, 0_2) = J_{12}(0_1, a_2) = \infty, \qquad \text{for all} \quad (a_1, a_2) \in \mathscr{L}_1 \otimes \mathscr{L}_2 \tag{6.5}$$

we can verify that $S(a_2/a_1) \in Z(\mathscr{L}_2)$ for all $a_1 \neq 0_1$ in $\mathscr{L}_1$, and that $T(a_1/a_2) \in Z(\mathscr{L}_1)$ for all $a_2 \neq 0_2$ in $\mathscr{L}_2$.

We claim that (6.4) defines the natural generalization of Bayes' rule. *If* $\mathscr{L}_1 = \mathscr{L}_2 = \mathscr{L}$, then any $J \in Z(\mathscr{L})$ defines $J_{12} \in Z(\mathscr{L} \otimes \mathscr{L})$ that satisfies (6.5) and is also symmetric in its arguments. One simply takes

$$J_{12}(a, b) = J(a \wedge b), \qquad \text{for all} \quad a, b \in \mathscr{L}. \tag{6.6}$$

If $J = -c \log p$, $c > 0$, then (6.4) reduces to

$$p(a \wedge b) = p(a)\, p(b/a) = p(b)\, p(a/b) \tag{6.7}$$

where $S(b/a) = -c \log p(b/a)$ and $T(a/b) = -c \log p(a/b)$. Thus, (6.4) indeed generalizes Bayes' rule.

Returning to the more general case $\mathscr{L}_1 = \mathscr{L}(P)$, $\mathscr{L}_2 = \hat{\mathscr{L}}$, let us suppose that there exists $J_{12} \in Z(\mathscr{L}_1 \otimes \mathscr{L}_2)$ satisfying (6.5). We must now face the problem of choosing J_{12}, assuming there are more than one. The obvious way we might try to specify the assignment is by setting

$$T(A/\hat{d}_Q) = J_Q(A), \qquad \text{for all} \quad A \in \mathscr{L}(P). \tag{6.8}$$

That is, the information specified by the evidence $\hat{d}_Q(J_Q \in Q)$ is the "conditional information", $J_{12}(A, \hat{d}_Q) - J_2(\hat{d}_Q)$. This leads to

$$J(A, \hat{d}_Q) = J_1(A) + S(\hat{d}_Q/A) = J_2(\hat{d}_Q) + J_Q(A) \tag{6.9}$$

where once $J \in Z(\mathscr{L}(P) \otimes \hat{\mathscr{L}})$ is known, all informations in (6.9) are determined. Clearly, if we know $J_Q(A)$ *and* $J_2(\hat{d}_Q)$, we can find $J(A, \hat{d}_Q)$.

Expressed in terms of probabilities (assuming Shannon informations), (6.9) reduces to

$$p(A, \hat{d}_Q) = p_1(A)\, p_S(\hat{d}_Q/A) = p_2(\hat{d}_Q)\, p_Q(A). \tag{6.10}$$

Note that $p_1(A) = p(A, \hat{d}_K)$ and $p_2(\hat{d}_Q) = p(P, \hat{d}_Q)$. From (6.10) it is clear that the proposition $(A, \hat{d}_Q)$ behaves strikingly like the expression "A & $\hat{d}_Q$" cited in the Introduction, but is *not* the same object. Indeed, (6.10) is—we believe—the correct way to formalize Bayes' rule in terms of "degrees of belief".[4] The "priors" p_1 and p_2 are based on "background evidence" $\hat{B}$ which asserts the existence and structure of P, $\hat{\mathscr{L}}$, G, χ_G, and Γ; also, in order to fix p we must include additional "prior evidence". Thus, the infamous "problem of priors" returns in a new formulation.

[4] The correct statements of some typical Bayesian results follow: Equation (1), page 27, of (Jeffreys, 1961) becomes

$$p((A \wedge B, \hat{d}_Q)) = p_Q(A/B)\, p_Q(B)\, p(\hat{d}_Q).$$

The "principle of inverse probability"—equation (4), page 28 of the same reference—becomes

$$p(\hat{d}_Q/(A, H)) \propto p(A/(\hat{d}_Q, H))\, p(\hat{d}_Q/H)$$

where $H \in \hat{\mathscr{L}}_2$ is "prior evidence" and $\hat{d}_Q \in \hat{\mathscr{L}}_1$ is additional evidence on $\mathscr{L}(P)$. There appears to be no mathematical loss in considering $\mathscr{L}(P) \otimes \hat{\mathscr{L}}$ rather than assuming that "A & $\hat{d}_Q$" is a legitimate proposition per se. However, $(A, \hat{d}_Q) \neq (\hat{d}_Q, A)$, so that it is not a priori clear that results based on the "symmetry" of "A & $\hat{d}_Q$" will remain valid.

Of course, in the procedure we presented earlier, there is no need to delve into the problems associated with (6.10) for p_Q is all that we really need to evaluate functions for a given theory. Can our procedure be extended to solve the "problem of priors"? In order to calculate a unique $J \in Z(\mathscr{L}(P) \otimes \hat{\mathscr{L}})$ we need some consistent criterion, which in turn is related to the nature of such informations. It is not sensible to propose an "evidence class" analogous to K, as this leads to an infinite regress problem—we would then have a lattice $\hat{\mathscr{L}}'$ of evidences about $\mathscr{L}(P) \otimes \hat{\mathscr{L}}$, and thus require an information on $\mathscr{L}(P) \otimes \hat{\mathscr{L}} \otimes \hat{\mathscr{L}}'$; *ad infinitum*. We thus tend to believe that our method cannot solve this problem.[5]

A similar difficulty arises if one considers the indended application of Bayes' rule. For is "evidence" that defines the *structure* $\mathscr{L}(P) \otimes \hat{\mathscr{L}}$ legitimately a proposition in even *this* "language"? Clearly not, so it is necessary to consider a third "language", $\hat{\mathscr{L}}'$, that contains "background evidence" defining the models $\mathscr{L}(P) \otimes \hat{\mathscr{L}}$. Each element of this "meta-language" is thus a "scientific theory", in the sense we have described. But, evidence in favor of one theory over another cannot belong to $\hat{\mathscr{L}}'$, so we must consider a "language" of evidences for the theory. And so on...

These considerations suggest a need for a careful review of the foundations of inductive logic, based on the distinction between "attributes" and "evidences". While Bayes' rule is correct as a mathematical condition its interpretation and application are not correct in Bayesian theories.

VII. Conclusions

In this paper we have presented an answer to the following question: How are measurement, theory, and information inter-related? We have tried to formulate the answer in a very general fashion, the result being a mathematical model of interest in itself.

Among the results we feel of special importance are the following: that the empirical relations serving to define a class of systems naturally form a poset; that functions defined on such a poset can be related to "observables" and a

[5] It is not certain, however, that there is a "prior" problem for the constraints expressed by (6.10) are powerful. Thus, for example, $p_1(A) = p_K(A)$ provided $\hat{d}_K$ corresponds to some "maximal" set in K. In the MEP example, $\hat{d}_K$ refers to W, in which case $p_K(A)$ is the usual "prior probability." In fact, (6.10) gives the ratio $p_S(\hat{d}_Q/A)\colon p_2(\hat{d}_Q)$ for all $(A, \hat{d}_Q)$ in terms of the *known* ratio $p_Q(A)\colon p_K(A)$. There is hope that additivity properties of the probability measures on $\hat{\mathscr{L}}$ can be used to determine the remaining unknowns. Indeed, this is the essence of the attempt in (Friedman and Shimony, 1973) to fault the MEP. The choice of ordering of evidences in this paper is rather peculiar: If $\hat{d}_z$ is defined by $e(X//Q) = z$, the order is assumed to be $\hat{d}_x \leqslant \hat{d}_y$ iff $x \leqslant y$. Moreover, p_2 is understood to be a probability measure on the Borel subsets $\{\hat{d}_x\}$. (See Appendix of (Cyranski, 1978).) Together with a technical assumption, these conditions imply that $p_2(\{\hat{d}_x\}) = 1$, if $x = E(X//W)$ and $=0$ otherwise.

probability theory based on its "cut completion"; that the MEP can be divorced from "randomness" concepts and greatly extended; that "attributes" and "evidences" do not coexist as propositions in a common "language" but can be treated as "pairs" (h, e) in a lattice product, and that Bayesian "degrees of belief" should be based on such an approach.

As suggested in the introduction, much of the material we have presented is not original, but our synthesis of it in a consistent viewpoint is the essential contribution. Nevertheless, we have in several cases extended existing theories and, perhaps more important, our approach has raised questions for further work. For example, the study of automorphisms on the complete lattice of informations—especially with regard to G-admissible classes—has been noted. (Present work indicates some surprising results concerning "dynamics" based on information theory.) We have not fully understood the nature of "entropy": In the "classical" continuum case in particular, our approach should help the characterization of "entropies"—at least those defined essentially as expectations of observable informations. The brief study of a lattice product of "languages" indicates a way towards a deeper understanding of communication theory, following the methods of rate distortion theory (Berger, 1971) and the "Mutual Information Principle" (Tzannes and Noonan, 1973)—a method similar to the MEP.

Further applications include the definition of "entropy" on computer programs—see a statistical attempt in (Hehner, 1977). We also mentionned a similar possibility for relativistic spacetime (Carter, 1971). Naturally, the ultimate goal of the measurement theory is successful application. The model we have presented is—we hope—general enough to admit non-statistical models in social sciences, for example, where standard mathematics is often incongruous. We hope this paper serves to stimulate both further research into the basic questions raised and useful application of the philosophy and methods.

Acknowledgments

We wish to thank Professor A. Shimony for posing to us the general problem of the consistency of Bayes' rule and the MEP and Professor D. Hestenes for his explanation of the traditional approach to the MEP. In addition, we acknowledge the partial support provided by a grant from the National Hellenic Research Foundation, and the hospitality of the Theoretical Chemistry Institute of the N.H.R.F.

Received: September 16, 1977; revised: June 26, 1978

Note added in proof. Since submission of this manuscript we have modified certain ideas presented here. For example, Definition 3.2 has been extended to permit Q to be a quoset and to relax the constraint $n_R \geqslant n_T$. Also, scales $X: A \to R^1$ lead to relations $\Delta X \in P$ where Δ is a "fuzzy" subset of R^1 ($\Delta: R^1 \to Q' \subseteq P$). These results will appear in Cyranski, J. (1979), Measurement theory for physics, *Found. Phys.*, in press.

References

Bar-Hillel, Y. (1964), "Language and Information," Addison–Wesley, Reading, Mass.

Belinfante, F. J. (1975), "Measurements and Time Reversal in Objective Quantum Theory," Pergamon, Long Island City, N. Y.

Berger, T. (1971), "Rate Distortion Theory, A Mathematical Basis for Data Comprssion," Prentice–Hall, Englewood Cliffs, N. J.

Birkhoff, G. (1973), "Lattice Theory," 3rd ed., Amer. Math. Society, Providence, R. I.

Bushaw, D. (1963), "Elements of General Topology," Wiley, New York.

Carnap, R. (1950), "Logical Foundations of Probability," Univ. of Chicago Press.

Carter, B. (1971), Causal Structure in Space-Time, *Gen. Rel. Gravit.* **1**, No. 4, 349–391.

Černý, J., and Brunovský, P. (1974), A note on information without probability, *Inform. Contr.* **25**, 134–144.

Cox, R. (1946), Probability, frequency, and reasonable expectation, *Amer. J. Phys.* **14**, 1–13.

Cyranski, J. (1978), Analysis of the MEP "Debate," *Found. Phys.* **8**, 493–506.

Fine, T. L. (1973), "Theories of Probability," Academic Press, New York.

Forte, B., and Kampé de Fériet, J. (1967), Information et Probabilité, *C. R. Acad. Sci. Paris Sér. A* **265**, 110–114, 142–146, 350–353.

Friedman, K., and Shimony, A. (1971), Jaynes's maximum entropy prescription and probability theory, *J. Statist. Phys.* **3**, No. 4, 381–384.

Hehner, E. (1977), Information content of programs and operation encoding, *J. A. C. M.* **24**, No. 2, 290–297.

Hestenes, D., and Gage, D. (1973), Comment on the paper "Jaynes's maximum entropy prescription and probability theory," *J. Statist. Phys.* **7**, No. 1, 89–90.

Ingarden, R., and Urbanik, K. (1962), Information without probability, *Colloq. Math.* **9**, 131–149.

Jammer, M. (1974), "The Philosophy of Quantum Mechanics," Wiley–Interscience, New York.

Jauch, J. (1968), "Foundations of Quantum Mechanics," Addison–Wesley, Reading, Mass.

Jaynes, R. (1957), Information theory and statistical mechanics, I, *Phys. Rev.* **106**, No. 4, 620–630; II, *Phys. Rev.* **108**, No. 2, 171–190.

Jaynes, E. (1958), "Probability Theory in Science and Engineering," Field Research Laboratory Socony Mobil Oil Company, Inc. Colloquium Lectures in Pure and Applied Science, No. 4, Feb. 1958.

Jeffreys, H. (1961), "Theory of Probability," Oxford Univ. Press, London/New York.

Kampé de Fériet, J., and Benvenuti, P. (1972), Opération de composition régulière et ensembles de valeurs d'une information, *C. R. Acad. Sci. Paris Sér. A* **274**, 655–659.

Kampé de Fériet, J., Forte, B., and Benvenuti, P. (1969), Forme générale de l'opération de composition continue d'une information, *C. R. Acad. Sci. Paris Sér. A* **269**, 529–534.

Kampé de Fériet, J., and Picard, C. F. (Eds.) (1974), "Théories de l'Information," Lecture Notes in Mathematics, No. 398, Springer–Verlag, New York/Berlin.

Landau, L. D., and Lifshitz, E. M. (1965), "Quantum Mechanics, Non-Relativistic Theory," Pergamon, Long Island City, N. Y.

MacQueen, J., and Marschak, J. (1975), Partial knowledge, entropy, and estimation, *Proc. Nat. Acad. Sci. USA* **72**, No. 10, 3819–3824.

Munroe, M. (1953), "Introduction to Measure and Integration," Addison-Wesley, Reading, Mass.

Pfanzagl, J. (1971), "Theory of Measurement," Physica-Verlag, Würzburg.

Picard, C. F. (1975) (directeur du Groupe de Recherche, Structures de l'Information), Journées Lyonnaises des Questionnaires, Structures de l'Information publications (Univ. Pierre et Marie Curie, Paris).

PICARD, C. F. (1976), Information Questionnaires et Reconnaissance, Structures de l'Information publications, Paris.

PICARD, C. F. (1977), Séminaires sur les questionnaires, Structures de l'Information publications, Paris.

PICARD, C. F. (1978), "Théorie de l'Information, Développements Récents et Applications," Éditions du Centre National de la Recherche Scientifique, Paris, France.

PICARD, C. F., AND SALLANTIN, J. (1977), Data seizing and information processing, *J. Geophys.* **43**, 215–225.

RENYI, A. (1970), "Probability Theory," North-Holland, Amsterdam.

SALLANTIN, J. (1972), Systèmes de propositions et information, *C. R. Acad. Sci. Paris Sér. A* **274**, 986–988.

SANCHEZ, E. (1976), Resolution of composite fuzzy relation equations, *Inform. Contr.* **30**, 38–48.

SHAFER, G. (1976), "A Mathematical Theory of Evidence," Princeton Univ. Press, Princeton, N. J.

SHANNON, C. (1948), A mathematical theory of communication, *Bell Syst. Tech. J.* **27**, 379–523.

SHIMONY, A. (1955), Coherence and the axioms of confirmation, *J. Symb. Logic* **20**, 1.

TZANNES, N. S., AND NOONAN, J. P. (1973), The mutual information principle and applications, *Inform. Contr.* **22**, 1–12.