

Entropy in classical and quantum physics

John F. Cyranski

Theoretical and Physical Chemistry Institute, The National Hellenic Research Foundation, 48, Vassileos Constantinou Avenue, Athens, Greece

(Received 13 September 1978; accepted for publication 28 April 1980)

The enigma of "entropy" is reconsidered from the viewpoint of generalized information theory on a lattice generated from measurements that define the system. A small (incomplete) set of natural axioms for a global information measure is developed sufficiently to deduce as a special case a generalization of Segal's entropy on a W^* -algebra (classical and quantum). A simple relationship between monotonicity of entropy and a semigroup on $[0, \infty]$ (representing composibility of information) is presented. Various extensions of information-theoretic results are incidentally proven, including relations between regular composable informations (on an orthocomplemented complete lattice) and measures (on σ -ideals of the lattice).

PACS numbers: 05.70.Ce, 05.50.+q

I. INTRODUCTION

Although dating from the grimy dawn of the industrial age, entropy has remained one of the most controversial, esoteric, and enigmatic concepts of physics. Recently Wehrl¹ has reviewed the properties of various entropy expressions useful in physics. In addition to the numerous "entropies" he has collected, one stumbles across other formulas which are also dubbed "entropy."²⁻⁵ This proliferation of "entropies" suggests the need for a comprehensive approach to the basic issue, namely: "What *is* entropy?"

Rather than dismiss all expressions that do not conform to " $-\sum p_i \ln p_i$," " $-\text{Tr} \hat{\rho} \ln \hat{\rho}$," or even Boltzmann's " $-k \ln W$ "—as some reviewers would happily have us do—we prefer to seek a fundamental meaning and axiomatic formulation of a quantity which, when suitably constrained, reduces to such special formulas. Indeed, the need for such an approach is evident from the difficulty of reconciling the second law of thermodynamics with Hamiltonian motion. This problem has led to various approaches to nonequilibrium statistical physics, including different forms of "coarse-graining"¹ (which effectively reject the second law for the *true* system "state," and only work for infinite systems in general¹), "master equations" (e.g., those based on projections of "relevant" data⁶), semigroups,⁷ and modifications of entropy itself.⁸

Of course, all concepts of "entropy" are ultimately related to probabilities (or "states"). This relationship in fact unites the mystery of entropy to the enigma of "probability." Good⁹ has distinguished at least five different concepts of "probability." More recently, Fine¹⁰ has compared the various concepts and ultimately concluded with the doubt that "probability" is even necessary for scientific endeavors. Most physicists are familiar only with the so-called "objective" probabilities—relative frequencies over infinite ensembles which, though by definition unmeasurable, are nevertheless considered to represent a physical property ("state"). Rayski¹¹ has shown how quantum mechanics paradoxes vanish if one relinquishes this interpretation. Elsewhere¹² we have argued how certain arguments against theories of "hidden variables" depend implicitly on the interpretation of probability, and thus cannot be convincing.

An extremely successful step in explaining both enigmas—"entropy" and "probability"—has been made by Jaynes.¹³ In essence, Jaynes interpreted "entropy" as the "expected self-information" of a class of mutually exclusive and exhaustive "events," or propositions. Although the association of "entropy" with "information" (negative uncertainty) is neither original with Jaynes¹⁴ nor without opposition,¹⁵ it nevertheless provides an explanation for this hitherto misunderstood quantity. Simultaneously, Jaynes interpreted "probability" to mean the rational "degree of belief"¹⁶ we assign to events (on a zero to one scale) based on available evidence, generally of the form of expectation values of observables. The choice of data used effectively defines the total evidence assumed relevant to the properties being estimated with the degree of belief. This condition is quantified by taking as degree of belief that measure which maximizes the "entropy", i.e., the average uncertainty, subject to the chosen evidence. In a single blow, Jaynes not only explained, simplified (no ensembles!), and generalized statistical thermodynamics, but also provided scientists and engineers with a new, (more or less) consistent technique of *inference* which has been successfully employed in many areas.¹⁷

Despite these successes, and in part because of them, we again find a need for an axiomatic explication of "entropy." For one thing, Jaynes's program is strictly constrained by the choice of entropy formula¹⁸ most generally, the Gibbs formula classically, and the von Neumann expression in quantum mechanics). Although Shore and Johnson¹⁹ have recently given an axiomatic derivation of the classical entropy in terms of inductive inference, there is no reason to assert that this specialized "entropy" is appropriate or even correct in general. Indeed, as we will elaborate below, most "real-life" situations do not admit the requisite idealizations for Borel algebras, and thus probability-related information measures. It thus may be not only desirable but essential to eliminate probability altogether. Because of this and certain other difficulties,²⁰ the MEF (maximum entropy formalism) has not been exploited as thoroughly as is possible.

In order to arrive at a fundamental understanding of the significance of "entropy" in physics, we believe it necessary

to reexamine the essence of physics as distilled in its lattice-theoretic formulation. It is convenient to visualize a physical theory as a data-processing algorithm: Empirical data are digested by the theory and predictions result. This is consistent with the view that theories model reality: They are not “explanatory” in the sense of being in one-to-one correspondence with aspects of reality.²¹ The model is constructed by first postulating an “operational definition” of the objects under study in terms of the set of all “relevant” empirical relations that can be defined on the objects. This definition is then naturally imbedded in a complete lattice—the “theory,” which is the mathematical context for calculations. We review these ideas in Sec. II, extending our earlier work.^{22,23} Also, we present certain new results concerning composable informations.

Our approach to “entropy” is based on the idea of “local” informations on a lattice (of “propositions”) first studied by Sallantin.²⁴ In Sec. III we define “entropy” to be a *global* measure of information associated with possibly several “local” informations on the “theory.” We do not attempt to characterize our “entropies,”²⁵ but rather consider certain special cases that ultimately reduce, as desired, to the usual “entropies.” In Sec. IV we provide an interpretation of certain qualitatively novel features of our “entropies,” which appears to have bearing on irreversible dynamics. Finally, we conclude in Sec. V with a general commentary and some open problems.

II. ABSTRACT THEORIES AND INFORMATIONS

The essence of physics is generally considered to be one or the other of various logico-algebraic models²⁶ based on abstract lattice theory.²⁷ While physicists are concerned specifically with deriving quantum and classical physics, we have found a broader viewpoint to be more enlightening. In Refs. 22 and 23 we have proposed a somewhat unusual approach which we describe and extend below. Our approach is reminiscent of the constructive descriptions of logic and language by Watanabe²⁸ and Sallantin²⁹; however, we do not attempt to derive specific theories, but rather describe universal properties. We assume the reader to be familiar with standard terminology of lattice theory.

Our starting point is the obvious one: How do we *define* an object about which we wish to “talk” (i.e., discuss in a literal sense, or scientifically study)? This question is by no means trivial or metaphysical, since practical problems of computer recognition are involved. It is natural to define an object (actually, the class A of all such objects) by a complete list of its properties. Although for real-life objects such a list may not be possible, one can effectively define varying degrees of approximation to “real-life” by a judicious choice of the relevant properties. To account for the intrinsic “fuzziness” of “real-life” descriptions^{30,31} we assert that the properties consist of relations $R: A^{n_R} \rightarrow Q$, where Q is a quasiordered set of “truth values.” (This generalizes slightly the work of Goguen³² and DeLuca and Termini.³³ Ali and Prugovecki³⁴ have applied “fuzzy” set theory directly to quantum physics, but not in the “logical” context.) To be operationally consistent we must insist that the order n_R of the

relations be finite. The set of all “relevant” relations is denoted by $F: \langle A, Q, F \rangle$ constitutes an empirical relational system. Note that the relations are essentially generalized characteristic functions.

The relations in F are naturally ordered by an “implication,” \Rightarrow , defined in terms of the quasi-ordering of the truth set Q . This binary (classical) relation induces a quasi-order in F ; we denote by P the partially ordered set obtained from F by associating “ \Rightarrow -equivalent” relations. Thus, P represents a minimal set of relations defining an object class A , with a built-in partial order. See Ref. 23 for details.

At this point we wish to clarify the above “abstract definition” by a further condition on P . Essentially, if R is a relation relevant to an object, it has a quasi-ordered range $\mathcal{R}(R) = \{R(a) \in A^{n_R}\} \subseteq Q$. If we reverse the order of $\mathcal{R}(R)$ —which corresponds in the simplest case to interchanging binary (yes-no) responses on an elementary experimental apparatus³⁵—we have a new relation, R' , whose range is the dual of $\mathcal{R}(R)$. It is natural to allow for each $R \in P$ that $R' \in P$ as well, since the same apparatus or operation that defines R also defines R' . The map $R \rightarrow R'$ is an involution by construction:

$$R'' = R; \quad R \leq S \Rightarrow S' \leq R'. \quad (1)$$

Notice, however, that there exists no meaningful relation in P which can in any sense “imply” *both* R and R' . That is

$$\{S \in P: S \leq R \text{ and } S \leq R'\} = \phi. \quad (2)$$

In this axiom, we diverge from the custom of introducing an “absurd” relation $\emptyset \in P$ such that $R \wedge R' = \emptyset$. Thus the “logic” of measurements is not assumed at this level.

So far the objects have been abstractly defined by the poset P with involution; this structure, constructively obtained, corresponds to the rational numbers. It is essential for further progress to introduce idealized elements—analogs to real numbers. We do this in the corresponding manner; we imbed P in a complete lattice $L(P)$ by the “cut procedure.”³⁶ (This method is not alien to quantum mechanics³⁷; however, the construction and properties of P used by previous authors are less general than ours.)

Theorem 1: If P is a poset with involution satisfying (2), then $L(P)$ is a complete orthocomplemented lattice.

Proof: Recall²⁷ that $\bar{E} = \{R \in P: R \leq S \text{ for all } S \geq T \text{ for all } T \in E\}$ defines the closure of $E \subseteq P$, and that $L(P)$ is just the set of closed subsets of P , a complete lattice under set inclusion (\subseteq). For $E \in L(P)$ define

$$E^\perp = \{R \in P: R \leq S' \text{ for all } S \in E\}. \quad (3)$$

Then, $(E^\perp)^\perp = \{R \in P: R \leq S' \text{ for all } S' \geq T \text{ for all } T \in E\} = E$; $E \subseteq F$ implies $F^\perp \subseteq E^\perp$ since $R \in F^\perp$ implies $R \leq S'$ for all $S \in F$ implies $R \leq S'$ for all $S \in E$, and thus $R \in E^\perp$. Finally, (3) defines an involution since $E^\perp \in L(P)$: $\overline{E^\perp} = \{R \in P: R \leq S \text{ for all } S \geq T \text{ for all } T \in E^\perp\} = \{R \in P: R \leq S \text{ for all } S' \leq T', \text{ for all } T' \geq W' \text{ for all } W \in E\} = \{R \in P: R \leq S \text{ for all } S' \in \bar{E} = E\} = E^\perp$. Now $E \cap E^\perp = \{R \in P: R \leq S \text{ and } R \leq S', \text{ for all } S \in E\} = \phi$ by (2). Thus, $E \rightarrow E^\perp$ is an orthocomplement.³⁸

Q.E.D.

The importance of the universality of orthocomplemented complete lattices ($L(P)$) as “theories” lies in the ability we obtain to construct partitions. If $A \in L(P)$, a (finite) *par-*

tion of A is the set

$$\pi(A) = \{E_n \in L(P) : A = \bigvee_{n=1}^N E_n, E_n \perp E_m\} \quad (E \perp F \text{ iff } E \subset F^\perp). \quad (4)$$

We denote by Π the set of all (finite) partitions of P , and $\pi(P) = \pi$ for short. The set Π is partially ordered by the relation

$$\pi_1 < \pi_2 \text{ iff for each } E \in \pi_1, \quad \pi(E) \subseteq \pi_2. \quad (5)$$

If $\pi_1 < \pi_2$, then π_2 is a “finer” partition; note that $\pi_0 = \{\phi, P\}$ is the universal lower bound, and that unless $L(P)$ is finite, Π has no upper bound in general.

We digress briefly to interpret $L(P)$ and Π . Since $R \in P$ is a relation, $\theta(R) = \overline{\{R\}} = \{S \in P : S \leq R\}$ (closure of $\{R\}$) is the “proposition” or idealization consisting of all (nonequivalent) relations in P that “imply” R . Since $\bar{E} = \bigvee_{R \in E} \theta(R)$ (see Ref. 22), \bar{E} is a set of relations plus their antecedents. The elements of $L(P)$ thus effectively summarize sets of properties. This ability is desirable in a theory—i.e., to combine consistently large numbers of experimental features. A partition $\pi \in \Pi$ is a subdivision of $L(P)$ into a finite set of “orthogonal” sets. It is natural to consider such a partition as an experiment whose outcomes are manifestations of the property. Each manifestation is a summary of more “elementary” measurements that “imply” relations comprising the manifestation. Such an interpretation does not disagree with typical physics usage, especially in regard to observables in physics (countable partitions!).³⁵ One may look upon the elements of P as “outcomes” and the elements of $L(P)$ as “events” in the language of probability theory.

Recall²⁷ that $\mathcal{T} \subseteq L(P)$ is a σ -ideal if (a) $E \in \mathcal{T}, X \in L(P)$ and $X \subset E$ implies $X \in \mathcal{T}$; and (b) $E_n \in \mathcal{T}, n = 1, 2, \dots$, implies $\bigvee_{n=1}^\infty E_n \in \mathcal{T}$. Note that if $E, F \in \mathcal{T}$, $E \cap F^\perp \in \mathcal{T}$ as well, which suggests that any σ -ideal is analogous to a σ -ideal in Boolean set theory.³⁹ We thus are led to define a *generalized measure* as a mapping $\mu : \mathcal{T} \rightarrow \mathbb{R}^+ = [0, \infty]$ such that \mathcal{T} is a σ -ideal, $\mu(\phi) = 0$, $\mu(E) = \mu(F) = 0$ implies $\mu(E \vee F) = 0$ and

$$\mu(\bigvee_{n=1}^\infty E_n) = \sum_{n=1}^\infty \mu(E_n), \quad E_n \perp E_m. \quad (6)$$

This definition is a natural extension of Piron’s³⁵ definition 4.38, the distinction being that Piron defines μ on a “tribe”: \mathcal{T} such that $E \in \mathcal{T}$ implies $E^\perp \in \mathcal{T}$ and \mathcal{T} countably closed. In special cases of interest, the σ -ideals reduce to tribes, and the concepts coincide.

Viewing the lattice $L(P)$ as an ordered set of propositions,⁴⁰ one naturally discovers certain minimal requirements for a local information measure⁴¹ on $L(P)$. First, it is natural to demand that no proposition have negative information:

$$J : L(P) \rightarrow \mathbb{R}^+ = [0, \infty]. \quad (7a)$$

Next, if $E \subset F$, then proposition E is more specific than F in that it contains fewer relations (from P). Thus, as one expects, the more specific proposition have the more informative value:

$$E \subset F \text{ implies } J(E) \geq J(F). \quad (7b)$$

Finally, as P is maximally unspecific, it should have the

minimal information value, while ϕ (being absurdly specific) should have undefined value:

$$J(\phi) = \infty; \quad J(P) = 0. \quad (7c)$$

We have studied properties of such information—defined by (7)—in Ref. 22. In this paper we will restrict attention to the subclass of informations which are (weakly) σ -composable. An information is said to be *composable*⁴¹ (weakly) if there exists a real-valued function F such that

$$J(A \vee B) = F[J(A), J(B)], \quad \text{when } A \perp B. \quad (8)$$

We are motivated largely by the consequences to consider informations which are composed by ROC’s:

*Definition 1*⁴²:

A *regular operation of composition* (ROC) is a function F satisfying:

$$F : \mathbb{R}^{+2} \rightarrow \mathbb{R}^+ \text{ is continuous,} \quad (9a)$$

$$F(x, y) = F(y, x), \quad (9b)$$

$$F[x, F(y, z)] = F[F(x, y), z], \quad (9c)$$

$$F(x, \infty) = x, \quad (9d)$$

$$x < y \text{ implies } F(x, z) \leq F(y, z). \quad (9e)$$

In fact, any composition map must satisfy (9b), (9c), (9d), and (9e) on the set of values defined by

$$\Gamma_n(J) = \{(x_1, x_2, \dots, x_n) : A_1, A_2, \dots, A_n \in L(P), A_i \perp A_j, x_j = J(A_j)\} \quad (10)$$

for the appropriate $n \geq 1$. Condition (9a) extends the domain of F from $\Gamma_2(J)$ to all \mathbb{R}^{+2} , and furthermore makes F a topological semigroup on \mathbb{R}^+ . Since $J(\bigvee_{n=1}^\infty E_n) = \lim_{N \rightarrow \infty} F_N [J(E_1), J(E_2), \dots, J(E_N)]$ exists, (where $F_1(x) = x$ and $F_N(x_1, \dots, x_N) = F[F_{N-1}(x_1, \dots, x_{N-1}), x_N]$), J is σ -composable under F . Incidentally, (9) implies that

$$F(x, y) \leq \inf\{x, y\}. \quad (11)$$

Among the remarkable properties of ROC’s is the following characterization theorem⁴³:

Theorem 2: Let A be a closed subset of \mathbb{R}^+ so

$\mathbb{R}^+ - A = \bigcup_{i \in I} (a_i, b_i) - I = \phi$, is finite, or countable. Let $\bar{\mu}_i \in [0, \infty]$, $\theta_i : [0, \bar{\mu}_i] \rightarrow [a_i, b_i]$ be strictly decreasing, continuous with $\theta_i(0) = b_i$, $\theta_i(\bar{\mu}_i) = a_i$. Then

$$F(x, y) = \begin{cases} \inf\{x, y\}, & (x, y) \in \mathbb{R}^{+2} - \bigcup_{i \in I} (a_i, b_i)^2 \\ \phi_i[\theta_i^{-1}(x) + \theta_i^{-1}(y)], & (x, y) \in [a_i, b_i]^2 \end{cases} \quad (12)$$

is a ROC with $A = A(F) = \{x \in \mathbb{R}^+ : F(x, x) = x\}$. Here

$$\phi_i(x) = \begin{cases} \theta_i(x), & x \in [0, \bar{\mu}_i] \\ a_i, & x > \bar{\mu}_i \end{cases}. \quad (13)$$

Conversely, any ROC is of form (12) with $A = A(F)$.

Perhaps yet more remarkable is the intimate association of ROC-composable informations with certain ideals and measures on ideals. Because $L(P)$ lacks the essential distributivity properties of a Borel algebra, the beautiful results of Ref. 42 apparently do not survive intact. For this reason we extend these results (as best we can) in the following somewhat unaesthetic theorems:

Theorem 3: Let $A \subseteq \overline{\mathbb{R}^+}$ be closed with $0, \in \in A$,

$\mathbb{R}^+ - A = \bigcup_{i \in I} (a_i, b_i)$. To each $x \in A$, associate a σ -ideal \mathcal{T}_x such that for all $x > 0$, \mathcal{T}_x is proper, $\mathcal{T}_0 = L(P)$, and $x < y$ implies $\mathcal{T}_x \supseteq \mathcal{T}_y$. On each \mathcal{T}_{a_i} , $i \in I$, assume a generalized measure $\mu_i: \mathcal{T}_{a_i} \rightarrow [0, \bar{\mu}_i]$ such that $\bar{\mu}_i \in (0, \infty]$ and $\mu_i[\mathcal{T}_{a_i} \cap \mathcal{T}_{b_i}] = 0$. Let $\theta_i: [0, \bar{\mu}_i] \rightarrow [a_i, b_i]$ be any continuous, strictly decreasing function with $\theta_i(0) = b_i$ and $\theta_i(\bar{\mu}_i) = a_i$. Let $z_1(A) = \sup\{x \in A: A \in \mathcal{T}_x\}$ and $z_2(A) = \inf\{x \in A: A \notin \mathcal{T}_x\}$.

Then

$$J(A) = \begin{cases} z_1(A), & \text{if } z_1(A) = z_2(A) \\ \phi_i(\mu_i(A)), & \text{if } z_1(A) = a_i < b_i = z_2(A) \end{cases} \quad (14)$$

is a σ -composable information on $L(P)$.

Proof: Note that $J(A) \in [z_1(A), z_2(A)]$ by construction and that $(z_1(A), z_2(A)) \cap A = \emptyset$ if $z_1(A) < z_2(A)$, since $(a_i, b_i) \cap A = \emptyset$. Clearly J is defined on all $A \in L(P)$. Since $\phi \in \mathcal{T}_x$ for all $x \in A$, $z_1(\phi) = \infty$ so $J(\phi) = \infty$; since $P \in \mathcal{T}_0$, but for all $x > 0$, $P \notin \mathcal{T}_x$, then $z_1(P) = z_2(P) = 0$ and $J(P) = 0$, or else $J(P) = \phi_0(\bar{\mu}_0) = a_0 = 0$, if $z_1(P) < z_2(P)$. Now for all $x < z_1(B)$, $B \in \mathcal{T}_x$ and $A \subset B$ implies $A \in \mathcal{T}_x$, so $z_1(B) \leq z_1(A)$. If $J(A) < J(B)$, $z_1(A) < z_1(B)$, so $a_i = z_1(A) = z_1(B) < J(A) < J(B) \leq z_2(B) = z_2(A) = b_i$. But $\mu_i(A) \leq \mu_i(B)$ implies $\phi_i(\mu_i(A)) \geq \phi_i(\mu_i(B))$, so $J(A) \geq J(B)$.

Now $A, B \in \mathcal{T}_x$ ($x \in A$) iff $A \vee B \in \mathcal{T}_x$, so that $z_1(A \vee B) = \min\{z_1(A), z_1(B)\} \leq J(A \vee B) \leq \min\{J(A), J(B)\} \leq \min\{z_2(A), z_2(B)\}$. For the sake of definiteness, assume $z_1(A) \leq z_1(B)$ ($A \perp B$) so the following cases are possible:

- (1) $z_1(A) = z_1(B) = z_2(A) = a_i < z_2(B) = b_i$,
- (2) $z_1(A) = z_1(B) = z_2(B) = a_i < z_2(A) = b_i$,
- (3) $z_1(A) = z_1(B) = a_i < z_2(A) = z_2(B) = b_i$,
- (4) $z_1(A) = a_i < z_1(B) = z_2(A) = z_2(B) = b_i$,
- (5) $z_1(A) = z_1(B) = z_2(A) = z_2(B)$,
- (6) $z_1(A) \leq z_2(A) \leq z_1(B) \leq z_2(B)$.

In Case (1), $J(A) = J(A \vee B) = a_i$, $J(B) \geq a_i$, so $J(A \vee B) = \inf\{J(A), J(B)\}$. In Case (2), $J(B) = J(A \vee B) = a_i$, $J(A) \geq a_i$, so $J(A \vee B) = \inf\{J(A), J(B)\}$. In Case (3), $J(A), J(B), J(A \vee B) \in [a_i, b_i]$; $J(A \vee B) = \phi_i[\mu_i(A) + \mu_i(B)]$ and $J(A) = \theta_i(\mu_i(A))$, $J(B) = \theta_i(\mu_i(B))$ (definition of θ_i and μ_i), so $J(A \vee B) = \phi_i[\theta_i^{-1}J(A) + \theta_i^{-1}J(B)]$. In Case (4), $J(A), J(A \vee B) \in [a_i, b_i]$, $J(B) = b_i$, so $J(A \vee B) = \phi_i[\mu_i(A) + 0] = J(A) = \inf\{J(A), J(B)\}$. In Case (5), $x = J(A \vee B) = J(A) = J(B) = \inf\{J(A), J(B)\}$, where $x \notin (a_i, b_i)$. In Case (6), $J(A), J(A \vee B) \in [z_1(A), z_2(A)]$, $J(B) \geq z_2(A)$. Thus, either $z_1(A) = J(A) = J(A \vee B) = z_2(A)$, or $J(A \vee B) = \phi_i[\mu_i(A) + 0] = J(A)$ if $a_i = z_1(A) < b_i = z_2(A) \leq J(B)$. Thus, $J(A \vee B) = F[J(A), J(B)]$ where F is defined by (12). Since by Theorem 2 F is a ROC, J is σ -composable. Q.E.D.

Comment: If $J(A) < x_0$, $z_1(A) < x_0$ implies $A \notin \mathcal{T}_{x_0}$. Thus, $A \in \mathcal{T}_{x_0}$ implies $J(A) \geq x_0$. If $A \in \mathcal{T}_{a_i}$ and $J(A) \in [a_i, b_i]$, by (14) $\mu_i(A) = \theta_i^{-1}J(A)$. Because $\mu_i(A) \leq \bar{\mu}_i$, and μ_i is σ -additive, $\sum_{J(E_n) \in [a_i, b_i]} \theta_i^{-1}J(E_n) \leq \bar{\mu}_i$ for every sequence of mutually orthogonal elements of \mathcal{T}_{a_i} .

In the following theorem, if $L(P)$ is orthomodular (15) is satisfied, while if $L(P)$ is distributive (16) is unnecessary.

Theorem 4: Let $L(P)$ be a complete orthocomplemented lattice such that

$$A, B \in L(P) \text{ implies there exists } D \subseteq A, E \subseteq B \text{ such that } A \vee B = E \vee D, E \perp D. \quad (15)$$

Let $\Gamma_x = \{A \in L(P): \text{There exists a countable partition } \pi_\infty(A) \text{ such that for all } E_n \in \pi_\infty(A), J(E_n) > x\}$, where J is an information on $L(P)$ σ -composable under ROC F . Suppose the ROC is defined by (12), where

$$\bar{\mu}_i = \infty \quad \text{or else } J(A) = a_i \text{ implies } A \notin \Gamma_{a_i}, i \in I. \quad (16)$$

Then there exists a family of σ -ideals $\mathcal{T}_x, x \in A(F)$ such that

$$\mathcal{T}_0 = L(P) \supset \dots \mathcal{T}_x \supset \dots \mathcal{T}_y \supset \dots \supset \mathcal{T}_\infty \quad (x < y) \quad (17)$$

and on each \mathcal{T}_{a_i} there exists a measure $\mu_i: \mathcal{T}_{a_i} \rightarrow \overline{\mathbb{R}^+}$. The ideals and measures determine J by (14).

Proof: If $x \in A(F)$, define $\mathcal{T}_x = \{A \in L(P): J(A) \geq x\}$. If $A \in \mathcal{T}_x$, $X \subseteq A$ ($X \in L(P)$), then $J(X) \geq J(A) \geq x$ implies $X \in \mathcal{T}_x$. By (15) and (9), if $A, B \in \mathcal{T}_x$, $J(A \vee B) = J(E \vee D) = F[J(E), J(D)] \geq F(x, x) = x$, so by induction, $E_n \in \mathcal{T}_x$ implies $J(\bigvee_{n=1}^\infty E_n) \geq x$ for all $N \geq 1$, and by σ -continuity, $\bigvee_{n=1}^\infty E_n \in \mathcal{T}_x$; so $\mathcal{T}_x, x \in A(F)$, are σ -ideals and (17) is immediate by construction.

Define on \mathcal{T}_{a_i} ($i \in I$)

$$\mu_i(A) = \begin{cases} 0, & \text{if } J(A) \geq b_i \\ \theta_i^{-1}J(A), & \text{if } A \in \Gamma_{a_i} \text{ and } J(A) \in [a_i, b_i] \\ \infty, & \text{if } A \notin \Gamma_{a_i}, J(A) \in [a_i, b_i] \end{cases} \quad (18)$$

By definition of $\mathcal{T}_x, x \in A(F)$, the upper line of (14) holds. In the lower case, $J(A) \in [a_i, b_i]$; if $A \in \Gamma_{a_i}$, by (18) $J(A) = \theta_i[\mu_i(A)] = \phi_i[\mu_i(A)]$, while if $A \notin \Gamma_{a_i}$, $\phi_i[\mu_i(A)] = a_i$. But $A \notin \Gamma_{a_i}$ means that $J(A) = a_i$, for otherwise ($J(A) > a_i$), $A = A \vee \phi \vee \phi \vee \dots$ ($\phi \perp A$ for all $A \in L(P)$) implies $A \in \Gamma_{a_i}$. Thus, (14) holds.

Now $\mu_i(\phi) = 0$, since $J(\phi) = \infty \geq b_i$, for all $i \in I$. If $E, F \in \mathcal{T}_{a_i}$ and $\mu_i(E) = \mu_i(F) = 0$, then by (18) $J(E), J(F) \geq b_i$; (15) implies $J(E \vee F) = J(A \vee B) = F[J(A), J(B)] \geq F(b_i, b_i) = b_i$ ($b_i \in A(F)$), so $\mu_i(E \vee F) = 0$ also. Let $A = \bigvee_{n=1}^\infty E_n$, $E_n \perp E_m$. In case $A \notin \Gamma_{a_i}$, some E_{n_0} must satisfy $J(E_{n_0}) = a_i$. Thus, $\mu_i(A) = \infty = \sum_{n=1}^\infty \mu_i(E_n)$.

If $A \in \Gamma_{a_i}$ and $\bar{\mu}_i = \infty$,

$$J(A) = a_i = \phi_i\left[\sum_{J(E_n) \in [a_i, b_i]} \theta_i^{-1}J(E_n)\right],$$

where the sum is $\geq \bar{\mu}_i = \infty$. But

$$\begin{aligned} \sum_{n=1}^\infty \mu_i(E_n) &= \sum_{J(E_n) \in [a_i, b_i]} \mu_i(E_n) \\ &= \sum_{E_n \in \Gamma_{a_i}, J(E_n) \in [a_i, b_i]} \theta_i^{-1}J(E_n) + \sum_{E_n \notin \Gamma_{a_i}, J(E_n) \in [a_i, b_i]} (\infty) = \infty \end{aligned}$$

so that $\mu_i(\bigvee E_n) = \sum \mu_i(E_n) = \infty$. If $A \in \Gamma_{a_i}$ and $\bar{\mu}_i < \infty$, by assumption (16) $J(A) > a_i$. Thus $\sum_{J(E_n) \in [a_i, b_i]} \theta_i^{-1}J(E_n) = \sum_{n=1}^\infty \mu_i(E_n) < \bar{\mu}_i$, and

$$\phi_i\left[\sum_{J(E_n) \in [a_i, b_i]} \theta_i^{-1}J(E_n)\right] = \theta_i\left[\sum_{n=1}^\infty \mu_i(E_n)\right] = J(A).$$

Thus, $\mu_i(A) = \theta_i^{-1}J(A) = \sum_{n=1}^\infty \mu_i(E_n)$. Hence, (18) defines

a generalized measure on \mathcal{T}_{a_i} .

Q.E.D.

These theorems indicate that there is a very deep relationship between ROC-composable informations and measures. We point out that these results may perhaps be strengthened: However, they suffice for our present purposes and we have not attempted to study further interesting properties (such as uniqueness of the measures).

III. GLOBAL INFORMATIONS

In this section we consider an explanation of “entropy” as the global information associated with $L(P)$ (complete, orthocomplemented lattice) relative to a finite set of “local” informations, each ROC-composable. Although many of the concepts we employ are well known (lattice informations,⁴¹ entropies as functionals of informations on Borel algebras,⁴⁴ measure-related entropies on lattices,²⁹ entropies relative to experiments⁵) our synthesis and generalization of these ideas appear to be unique. In order to maintain universality we seek only minimal (natural) properties of such “entropies” and do not attempt to characterize the special cases we introduce as examples. On the other hand, we sacrifice some generality by considering only ROC-composable informations on $L(P)$. We believe nevertheless that the resultant compromise is of considerable interest.

Let us denote by Z the set of all informations on $L(P)$ which are ROC-composable. We do not specify any ROC, however. We begin by considering properties of the information content of an experiment (partition) $\pi \in \Pi$ relative to a finite vector, $\mathbf{J} = (J_1, J_2, \dots, J_M)$, of informations, $J_m \in Z$. We denote by \mathbf{F} the vector of corresponding ROC's, $\mathbf{F} = (F_1, F_2, \dots, F_M)$. Our first axiom is simply that the information content of π be a real-valued functional of the local informations:

Axiom 1:

$$H_\pi^F: Z^M \rightarrow \mathbb{R}, \quad \pi \in \Pi.$$

Axiom 1 generalizes Ref. 44, which considers only Borel algebras and the case $M = 1$. Other axioms from this paper are meaningful in general. Indeed, the trivial proposition, $P \in L(P)$, satisfies $J(P) = 0$ for all $J \in Z$, so it is natural that the global information of $\pi_0 = \{\phi, P\}$ be zero:

Axiom 2:

$$H_{\pi_0}^F: Z^M \rightarrow \{0\}.$$

Our next axiom is not so widely accepted,⁴⁵ but we believe that it is in fact most natural to assume that an information on experiments not depend on the labels of the manifestation. Thus we assume

Axiom 3:

$$H_\pi^F(\mathbf{J}) = \Phi_{N_\pi}^F[\{\mathbf{J}(E_n)\}_{n=1}^{N_\pi}];$$

$$\Phi_N^F: \{\{\mathbf{x}_n: n = 1, 2, \dots, N\}: \mathbf{x}_n \in \overline{\mathbb{R}^+}^M\} \rightarrow \mathbb{R},$$

where $\mathbf{J}(E_n) = (J_1(E_n), J_2(E_n), \dots, J_M(E_n))$, $E_n \in \pi$. That is, the information of $\pi \in \Pi$ relative to \mathbf{J} depends only on the set of respective informations of the manifestations of π , and not on their order (index n). This axiom defines “symmetry”⁴⁴ of H_π^F (under permutations of $n = 1, 2, \dots, N$). We assume as well that the function Φ_N^F has as domain all sets of NM -tuples of

reals; we do this primarily for simplicity, but this hypothesis also reflects the fact that the global information is a functional of the $\mathbf{J} \in Z^M$.

Recall that Π is partially ordered under “refinement.” It is natural to suppose that the more refined experiment has greater global informative value:

Axiom 4:

$$\pi_1 < \pi_2 \text{ implies } H_{\pi_1}^F \leq H_{\pi_2}^F.$$

Again, although the order $<$ has varying definitions, the essence of Axiom 4 is well accepted.^{5,46}

Now it is not inconceivable for some $\mathbf{J} \in Z^M$ that two partitions be related as follows: $\pi_1 = \{E_1, E_2, \dots, E_N\}$, $\pi_2 = \{F_1, F_2, \dots, F_{N+1}\}$, $\mathbf{J}(E_n) = \mathbf{J}(F_n)$, $n = 1, 2, \dots, N$ and $\mathbf{J}(F_{N+1}) = (\infty, \infty, \dots, \infty)$. In such a situation, since F_{N+1} is, informationally speaking, impossible, it is natural to demand that π_1 and π_2 possess the same information value. More generally, expansibility⁴⁴ takes the form:

Axiom 5:

$$\Phi_{N+1}^F[\{\{\mathbf{x}_n\}_{n=1}^N \cup \{(\infty, \dots, \infty)\}\}] = \Phi_N^F[\{\{\mathbf{x}_n\}_{n=1}^N\}], \quad \mathbf{x}_n \in \overline{\mathbb{R}^+}^M.$$

Finally, consider two experiments defined by

$$\pi_1 = \{E_1, E_2, \dots, E_N\} \text{ and } \pi_2 = \{E_1 \vee E_2, E_3, \dots, E_N\}. \text{ Clearly } \pi_2 < \pi_1; \text{ in general}$$

$$H_{\pi_1}^F(\mathbf{J}) - H_{\pi_2}^F(\mathbf{J}) = \Delta_{N_{\pi_1}}^F[\mathbf{J}(E_1), \mathbf{J}(E_2), \mathbf{J}(E_1 \vee E_2)],$$

but the F -composability of the J_m 's means that

$$\mathbf{J}(E_1 \vee E_2) = \mathbf{F}[\mathbf{J}(E_1), \mathbf{J}(E_2)], \text{ where } \mathbf{F}(\mathbf{x}, \mathbf{y}) = (F_1(x_1, y_1), F_2(x_2, y_2), \dots, F_M(x_M, y_M)) \text{—a semigroup operation on } \overline{\mathbb{R}^+}^M.$$

Then, branching⁴⁴ means

Axiom 6:

$$\Phi_N^F[\{\{\mathbf{x}_n\}_{n=1}^N\}] - \Phi_{N-1}^F[\{\{\mathbf{F}(\mathbf{x}_1, \mathbf{x}_2)\} \cup \{\mathbf{x}_n\}_{n=3}^N\}] = \Delta_N^F[\{\mathbf{x}_1, \mathbf{x}_2\}],$$

where Δ_N^F is the information gain.⁴⁴

The properties embodied in Axioms 1–6 are certainly minimal, natural, and (excepting Axiom 3) generally accepted. Other axioms involving algebraic independence⁴⁶ do not have a clear meaning in our general context, and will be avoided. We define the *global information of $L(P)$ relative to $\mathbf{J} \in Z^M$* by

Axiom 7:

$$H^F = \sup_{\pi \in \Pi} H_\pi^F.$$

Since Π generally has no upper bound, this is the next best thing to taking a “finest partition” of $L(P)$. Even such an axiom has appeared in at least one context.⁵

The following theorem cites sufficient conditions to represent a global information—henceforth “entropy.” The conditions are probably necessary as well (in case $M = 1$ this is known⁴⁴), but we will not attempt to verify this very tedious result.

Theorem 5: For any $\mathbf{F} = (F_1, F_2, \dots, F_M)$ — F_m a ROC, $M < \infty$ —let

$$\varphi^F: \overline{\mathbb{R}^+}^M \rightarrow [0, \infty), \quad (19a)$$

$$\varphi^F(\infty, \infty, \dots, \infty) = 0, \quad (19b)$$

$$\varphi^F[\mathbf{F}(\mathbf{x}, \mathbf{y})] \leq \varphi^F(\mathbf{x}) + \varphi^F(\mathbf{y}). \quad (19c)$$

Then, if $\mathbf{J} = (J_1, J_2, \dots, J_m) \dashv J_m$ is F_m -composable on $L(P)$

$$H^F(\mathbf{J}) = -\varphi^F(0, 0, \dots, 0) + \sup_{\pi \in \Pi} \left\{ \sum_{n=1}^{N_\pi} \varphi^F[\mathbf{J}(E_n)] \right\} \quad (20)$$

is a global information of $L(P)$ relative to \mathbf{J} .

Proof: Note that (20) is the π -supremum of

$$H_\pi^F(\mathbf{J}) = -\varphi^F(0, \dots, 0) + \sum_{n=1}^{N_\pi} \varphi^F[\mathbf{J}(E_n)] \quad (21)$$

Thus we must show (21) satisfies Axioms 1–6, with φ^F defined by (19). By (19a) it is clear the (21) satisfies Axiom 1— $N_\pi < \infty$. Since $\mathbf{J}(P) = (0, \dots, 0)$, by (19b), (21) satisfies Axiom 2. Certainly (21) satisfies Axiom 3, as the sum is commutative. By (19b) Axiom 5 is immediate. Condition 6 is satisfied since

$$\begin{aligned} \Delta^F[\{\mathbf{J}(E_1), \mathbf{J}(E_2)\}] &= \{-\varphi^F(0, \dots, 0) + \sum_{n=1}^N \varphi^F[\mathbf{J}(E_n)]\} \\ &\quad - \{-\varphi^F(0, \dots, 0) + \varphi^F[\mathbf{J}(E_1 \vee E_2)] + \sum_{n=3}^N \varphi^F[\mathbf{J}(E_n)]\} \\ &= \varphi^F[\mathbf{J}(E_1)] + \varphi^F[\mathbf{J}(E_2)] - \varphi^F[\mathbf{F}(\mathbf{J}(E_1), \mathbf{J}(E_2))] \end{aligned} \quad (22)$$

That is, the difference depends only on $\mathbf{J}(E_1)$ and $\mathbf{J}(E_2)$. By (19c) we see that $\Delta^F[\{\mathbf{J}(E_1), \mathbf{J}(E_2)\}] \geq 0$. If $\pi_1 < \pi_2$, there exists a chain $\pi_1 < \pi_a < \pi_b < \dots < \pi_c < \pi_2$, with $N_{\pi_a} = N_{\pi_1} + 1$, $N_{\pi_b} = N_{\pi_a} + 1$, etc., so that employing (22) and its non-negativity successively yields Axiom 4. Q.E.D.

The next result essentially defines the effect of symmetry on the “entropy.” Here $\text{Aut}[L(P)]$ is the group of all automorphisms of $L(P)$ —these preserve orthogonality and are complete ($c -$) morphisms.

Corollary 1: Let $\mathcal{G} = \{g \in \text{Aut}[L(P)] : g(P) = P\}$. Then H^F is invariant under \mathcal{G} .

Proof: From Axiom 7 it is evident that any transformation $\Pi \rightarrow \Pi$ leaves H^F invariant. If $g \in \mathcal{G}$,

$g(\bigvee_{n=1}^{N_\pi} E_n) = \bigvee_{n=1}^{N_\pi} g(E_n)$, $g(E) \perp g(F)$ if $E \perp F$ and as $g(P) = P$, then $g(\pi) \in \Pi$ for each $\pi \in \Pi$. In addition, if $\pi \in \Pi$, for each $g \in \mathcal{G}$ there exists g^{-1} and thus $g[g^{-1}(\pi)] = \pi$, showing that g maps Π onto Π . Q.E.D.

Comment: If \mathcal{T} is a subcomplete, orthocomplemented lattice of $L(P)$ with maximum $M_{\mathcal{T}}$, then $\mathcal{G}_{\mathcal{T}} = \{g \in \text{Aut}[\mathcal{T}] : g(M_{\mathcal{T}}) = M_{\mathcal{T}}\}$ keeps invariant that portion of H^F that concerns only \mathcal{T} . This will be highly significant, as we show in Sec. IV. Note that transformations other than automorphisms may leave H^F invariant, but these revise the structure of the system and thus are eliminated.

The next result indicates that the semigroup F on \mathbb{R}^{+M} induces a potentially useful decomposition of H^F .

Corollary 2: Let $\Lambda(F) = \times_{m=1}^M \Lambda(F_m)$, and let $\mathbb{R}^{+M} - \Lambda(F) = \cup_{i \in I} [\times_{m=1}^M (a_{i_m}, b_{i_m})]$, where

$i = (i_1, i_2, \dots, i_M)$ and $I = \times_{m=1}^M I_m$, $\mathbb{R}^{+M} - \Lambda(F_m) = \cup_{i_m \in I_m} (a_{i_m}, b_{i_m})$. Define for φ^F satisfying (19)

$$T(i) = \{E \in L(P) : J_m(E) \in (a_{i_m}, b_{i_m}), m = 1, 2, \dots, M\}, \quad (23a)$$

$$S = L(P) - \cup_{i \in I} T(i), \quad (23b)$$

$$G(\pi \cap W) = \sum_{E_n \in \pi \cap W} \varphi^F[\mathbf{J}(E_n)], \quad W \subseteq 2^{L(P)}. \quad (23c)$$

Then the global information (20) can be written

$$\begin{aligned} H^F(\mathbf{J}) &= -\varphi^F(0, \dots, 0) + \sum_{i \in I} \sup_{\pi \in \Pi} G(\pi \cap T(i)) \\ &\quad + \sup_{\pi \in \Pi} G(\pi \cap S) \end{aligned} \quad (24)$$

N.B. $\{E \in L(P) : J_m(E) \geq a_{i_m}, m = 1, 2, \dots, M\} = \cap_{m=1}^M \mathcal{T}_{a_{i_m}}$ with $\mathcal{T}_{a_{i_m}} = \{E \in L(P) : J_m(E) \geq a_{i_m}\}$. These are σ -ideals.

Proof: Clearly (21) can be written as

$$H_\pi^F(\mathbf{J}) = -\varphi^F(0, \dots, 0) + \sum_{i \in I} G(\pi \cap T(i)) + G(\pi \cap S). \quad (25)$$

Thus, $H_\pi^F(\mathbf{J})$ is effectively the sum $\sum_{n \in \eta} G(\pi_n)$, where η is at most countable. The problem is therefore to show that

$$\sup_{\pi \in \Pi} \left[\sum_{n \in \eta} G(\pi_n) \right] = \sum_{n \in \eta} \left[\sup_{\pi \in \Pi} G(\pi_n) \right], \quad (26)$$

from which (24) follows. Trivially,

$$\sup_{\pi \in \Pi} \left[\sum_{n \in \eta} G(\pi_n) \right] \leq \sum_{n \in \eta} \left[\sup_{\pi \in \Pi} G(\pi_n) \right]. \quad (27)$$

For every $\epsilon > 0$ and $N \geq 1$, choose $\pi_{N, \epsilon}^n$ in Π such that

$$\epsilon/N > \sup_{\pi \in \Pi} G(\pi_n) - G(\pi_{N, \epsilon}^n). \quad (28)$$

Then, it follows by ordering η naturally that

$$\begin{aligned} \sup_{\pi \in \Pi} \sum_{n=1}^N G(\pi_n) + \epsilon &\geq \sum_{n=1}^N [G(\pi_{N, \epsilon}^n) + \epsilon/N] \\ &\geq \sum_{n=1}^N \sup_{\pi \in \Pi} G(\pi_n) = V_N, \text{ for all } N \geq 1, \epsilon > 0. \end{aligned} \quad (29)$$

Clearly V_N and $W_N = \sup_{\pi \in \Pi} \sum_{n=1}^N G(\pi_n)$ are nondecreasing in N , as $G(\pi_n) \geq 0$, so that—remember η is at most countable—

$$\begin{aligned} \lim_{N \rightarrow \|\eta\|} V_N &= \sum_{n \in \eta} \sup_{\pi \in \Pi} G(\pi_n) \leq \epsilon + \lim_{N \rightarrow \|\eta\|} W_N \\ &= \epsilon + \sup_{\pi \in \Pi} \left[\sum_{n \in \eta} G(\pi_n) \right]. \end{aligned} \quad (30)$$

As (30) holds for all $\epsilon > 0$, this proves the reverse inequality to (27) and thus (26). Q.E.D.

The next result is helpful in expressing entropy on certain atomistic lattices.³⁵

Lemma: Let $L(P)$ be atomistic such that

for each $E \in L(P)$ there exists an at most countable set of mutually \perp atoms A_n such that $E = \bigvee A_n$. (31)

Then, with Π^* the class of all countable partitions of atoms

$$\pi^* = \{A_n : A_n \perp A_m, A_n \text{ atoms}, P = \bigvee_{n=1}^\infty A_n\},$$

$$H^F(\mathbf{J}) \leq \sup_{\pi^* \in \Pi^*} H_{\pi^*}^F(\mathbf{J}). \quad (32)$$

If we also assume that

$\lim_{N \rightarrow \infty} H_{\pi_N}^F(\mathbf{J}) = H_{\pi^*}^F(\mathbf{J})$, $\pi_N \in \Pi$ such that

$$\pi_N = \{A_1, A_2, \dots, A_{N-1}, \bigvee_{n \geq N} A_n\}, \quad (33)$$

then equality holds in (32).

Proof: If $\pi \in \Pi$, there exists $\pi^* \in \Pi^*$ such that $\pi < \pi^*$ by (31) defined by $\bigcup_{n=1}^{\infty} \{A_m : m \in \mathcal{U}_n\}$, $E_n = \bigvee_{m \in \mathcal{U}_n} A_m$. Thus, by axiom 4, $H_\pi^F(\mathbf{J}) \leq H_{\pi^*}^F(\mathbf{J})$ and (32) results. For each $\pi^* \in \Pi^*$, consider the sequence $\pi_N = \{A_1, A_2, \dots, A_{N-1}, \bigvee_{n \geq N} A_n : A_n \in \pi^*\}$. Clearly $\pi_N \in \Pi$, $\pi_N < \pi_{N+1} < \pi^*$. By (33) and the fact that $H_{\pi_N}^F(\mathbf{J}) \leq H_{\pi^*}^F(\mathbf{J})$,

$$H_{\pi^*}^F(\mathbf{J}) \leq H^F(\mathbf{J}), \text{ for all } \pi^* \in \Pi^* \quad (32')$$

so that equality in (32) is apparent. Q.E.D.

Let us first illustrate the above results in the case $M = 1$.

Assume $\varphi^F(\mathbf{x})$ has the form:

$$\varphi^F(x) = \begin{cases} X\psi_i(X), & X = \theta_i^{-1}(x), \quad x \in (a_i, b_i), \\ 0, & \text{otherwise,} \end{cases} \quad (34)$$

where $\psi_i : (0, \bar{\mu}_i) \rightarrow [0, \infty)$ is nonincreasing and $\psi_i(\bar{\mu}_i) = 0$. It is trivial to verify that (34) satisfies (19a) and (19b), while (19c) follows since, because ψ_i is nonincreasing,

$$(X + Y)\psi_i(X + Y) \leq X\psi_i(X) + Y\psi_i(Y), \quad X, Y \in (0, \bar{\mu}_i) \quad (35)$$

with $X + Y \in (0, \bar{\mu}_i)$.

Assume the conditions of Theorem 3, so $J(E) \in (a_i, b_i)$ implies $J(E) = \theta_i[\mu_i(E)]$ with $\mu_i : \mathcal{T}_{a_i} \rightarrow [0, \bar{\mu}_i]$ and $\mu_i(E) = 0$ for all $E \in \mathcal{T}_{a_i} \cap \mathcal{T}_{b_i}$. We write (24) as

$$H^F(J) = \sum_{i \in I} \sup_{\pi \in \Pi} \left\{ \sum_{E_n \in \pi \cap \mathcal{T}_{a_i}} \mu_i(E_n) \psi_i[\mu_i(E_n)] \right\}. \quad (36)$$

Let us further suppose that $M_i = \bigvee \{E \in \mathcal{T}_{a_i} : E \in \mathcal{T}_{a_i} \text{ and that the supremum over } \pi \cap \mathcal{T}_{a_i} \text{ may be replaced by the supremum over } \pi \in \Pi_i, \text{ where } \Pi_i \text{ consists of all (finite) partitions of } M_i \text{ with components in } \mathcal{T}_{a_i}. \text{ In particular, let each } \mathcal{T}_{a_i} \text{ be an irreducible propositional system, } \mathcal{T}_{a_i} \approx \mathcal{P}[\mathcal{H}^i] \text{ (lattice isomorphic) the closed submanifolds of a Hilbert space. Then } \mathcal{T}_{a_i} \text{ is atomistic and (31) holds; moreover, our measure coincides with the CROC-measure of Piron,}^{35} \text{ which reduces to (we choose } \bar{\mu}_i = 1, i \in I)$

$$\mu_i(E) = \text{Tr}_i \hat{W}_i \hat{\Pi}_E \quad E \in \mathcal{T}_{a_i}, \quad E \leftrightarrow \hat{\Pi}_E \in \mathcal{P}[\mathcal{H}^i], \quad (37)$$

where \hat{W}_i is a density operator on \mathcal{H}^i . Hence, (36) becomes

$$H^F(J) = \sum_{i \in I} \sup_{\pi^* \in \Pi^*} \left\{ \sum_{n=1}^{\infty} \text{Tr}_i [(\hat{W}_i \hat{\Pi}_{E_n}) \psi_i(\text{Tr}_i \hat{W}_i \hat{\Pi}_{E_n})] \right\}, \quad (38)$$

where $E_n \in \pi^*$ are atoms. Let $p_n^i(\pi^*) = \text{Tr}_i \hat{W}_i \hat{\Pi}_{E_n}$. Then, each i -term in (38) is essentially the supremum over all "measurements" of a generalized "Ingarden-Urbanik entropy." To see this, let $\psi_i(z) = -\ln z$. Then clearly $\psi_i : (0, 1] \rightarrow [0, \infty)$ is nonincreasing with $\psi_i(1) = 0$. Moreover, (33) is satisfied since in fact for all $i \in I$:

$$\begin{aligned} & \lim_{N \rightarrow \infty} \left[\sum_{n=1}^{N-1} -\text{Tr}_i (\hat{W}_i \hat{\Pi}_{E_n}) \ln (\hat{W}_i \hat{\Pi}_{E_n}) \right. \\ & \quad \left. + \left(-\text{Tr}_i \hat{W}_i \hat{\Pi}_{\bigvee_{n \geq N} E_n} \ln \text{Tr}_i \hat{W}_i \hat{\Pi}_{\bigvee_{n \geq N} E_n} \right) \right] \\ & = - \sum_{n=1}^{\infty} \text{Tr}_i \hat{W}_i \hat{\Pi}_{E_n} \ln \hat{W}_i \hat{\Pi}_{E_n} \end{aligned} \quad (33')$$

since for large enough N ,

$$\left| -\text{Tr}_i \hat{W}_i \hat{\Pi}_{\bigvee_{n \geq N} E_n} \ln (\text{Tr}_i \hat{W}_i \hat{\Pi}_{\bigvee_{n \geq N} E_n}) \right|$$

can be made arbitrarily small. Thus (38) becomes

$$\begin{aligned} H^F(J) &= \sum_{i \in I} \sup_{\pi^* \in \Pi^*} \left\{ - \sum_{n=1}^{\infty} \text{Tr}_i [\hat{W}_i \hat{\Pi}_{E_n} \ln (\hat{W}_i \hat{\Pi}_{E_n})] \right\} \\ &\geq \sum_{i \in I} (-\text{Tr}_i \hat{W}_i \ln \hat{W}_i). \end{aligned} \quad (39)$$

See Ref. 1, p. 255.

Expanding \hat{W}_i in terms of its eigenbasis (π_0^*) ,

$$\begin{aligned} p_n^i(\pi^*) &= \sum_{m=1}^{\infty} p_m^i(\pi_0^*) T_{mn}(\pi^*); \quad T_{mn}(\pi^*) \\ &= \text{Tr}_i \hat{\Pi}_{F_m} \hat{\Pi}_{E_n}, \quad F_m \in \pi_0^*, E_n \in \pi^*. \end{aligned}$$

Note that $T_{mn}(\pi^*)$ is a non-negative, doubly stochastic matrix. If we assume $\Gamma_i(z) = z\psi_i(z)$ is convex, then

$$\begin{aligned} H^F(J) &= \sum_{i \in I} \sup_{\pi^* \in \Pi^*} \sum_{n=1}^{\infty} \Gamma_i \left[\sum_{m=1}^{\infty} p_m^i(\pi_0^*) T_{mn}(\pi^*) \right] \\ &\leq \sum_{i \in I} \sup_{\pi^* \in \Pi^*} \left[\sum_n \sum_m T_{mn}(\pi^*) \Gamma_i(p_m^i(\pi_0^*)) \right] \\ &= \sum_{i \in I} \sum_m \Gamma_i(p_m^i(\pi_0^*)). \end{aligned} \quad (40)$$

But since $p_m^i(\pi_0^*)$ are just the eigenvalues of \hat{W}_i , and since $\pi_0^* \in \Pi_0^*$, we have from (40)

$$H^F(J) = \sum_{i \in I} \text{Tr}_i \Gamma_i(\hat{W}_i). \quad (41)$$

For properties of $\text{Tr} \Gamma(\hat{W})$, where Γ is concave, see Ref. 47.

Our next illustration is for the case $M = 2$. Again, without any real justification, let us assume that φ^F has the form $\varphi^F(x, \xi)$

$$\varphi^F(x, \xi) = \begin{cases} \Psi_{ij}(x, \xi), & (x, \xi) = \mathbf{x} \Delta_{ij} = (a_i, b_i) \times (a_j, b_j), (i, j) \in I_1 \times I_2 \\ 0, & \text{otherwise.} \end{cases} \quad (42)$$

If we specify that $i < j$ implies $a_i < a_j$, the conditions (19) reduce to

$$\Psi_{ij}(\mathbf{x}) \in [0, \infty) \text{ for all } \mathbf{x} \in \Delta_{ij}, \quad (i, j) \in I_1 \times I_2, \quad (43)$$

$$\Psi_{ij}(\mathbf{x}) + \Psi_{ij}(\mathbf{y}) \geq \Psi_{ij}[\mathbf{F}(\mathbf{x}, \mathbf{y})] \quad \mathbf{x}, \mathbf{y} \in \Delta_{ij}, \quad (i, j) \in I_1 \times I_2, \quad (44)$$

$$\begin{aligned} & \Psi_{ij}(\mathbf{x}) + \Psi_{ik}(\mathbf{y}) \geq \Psi_{ij}[F_1(x, y), \xi] \\ & \quad j < k; \mathbf{x} \in \Delta_{ij}, \quad \mathbf{y} \in \Delta_{ik}, \quad (F_1(x, y), \xi) \in \Delta_{ij}, \end{aligned} \quad (45a)$$

$$\begin{aligned} & \Psi_{ij}(\mathbf{x}) + \Psi_{kj}(\mathbf{y}) \geq \Psi_{ij}[x, F_2(\xi, \eta)] \\ & \quad i < k; \mathbf{x} \in \Delta_{ij}, \quad \mathbf{y} \in \Delta_{kj}, \quad (x, F_2(\xi, \eta)) \in \Delta_{ij}. \end{aligned} \quad (45b)$$

Note that $F_1(x, y) \in [a_i, b_i]$, if $x, y \in (a_i, b_i)$ in general (likewise for F_2).

We further simplify matters as follows:

$$I_2 = \{0, \infty\}; \xi = \theta_0^{(2)}(\sigma), \quad \sigma \in (0, \bar{\sigma}), \quad \bar{\sigma} \leq \infty, \quad (46)$$

$$\begin{aligned} & \Psi_{i0}(x, \xi) \\ &= \begin{cases} \sigma \psi_i(X/\sigma), & X = \theta_i^{-1}(x), \quad x \in (a_i, b_i); \quad \sigma \in (0, \bar{\sigma}) \\ 0, & \text{otherwise.} \end{cases} \end{aligned} \quad (47)$$

Note that $X \in (0, \bar{\mu}_i)$ if $x = \theta_i(X) \in (a_i, b_i)$. The constraints (43)–(45) become:

$$\psi_i : (0, \infty) \rightarrow [0, \infty), \quad i \in I_1 \quad (48)$$

$$\begin{aligned} & \sigma \psi_i(X/\sigma) + \tau \psi_j(Y/\tau) \geq (\sigma + \tau) \psi_i((X + Y)/(\sigma + \tau)), \\ & X, Y, X + Y \in (0, \bar{\mu}_i); \quad \xi, \tau, \xi + \tau \in (0, \bar{\sigma}), \end{aligned} \quad (49)$$

$$\begin{aligned} & \sigma \psi_i(X/\sigma) + \tau \psi_j(Y/\tau) \\ & \geq (\sigma + \tau) \psi_i(X/(\sigma + \tau)) \text{ if } i < j; \quad X \in (0, \bar{\mu}_i), \end{aligned}$$

$$Y \in (0, \bar{\mu}_j); \quad \sigma, \tau, \sigma + \tau \in (0, \bar{\sigma}). \quad (50)$$

A degree of justification for the arbitrariness of choices (42) and (46) can be felt from the following results.

Lemma: $\psi_i(z)$ is a convex, continuous function $\psi_i: (0, \infty) \rightarrow [0, \infty)$ and

$$\inf_{z \in (0, \infty)} \psi_j(z) \geq \psi_i(0^+) = \lim_{\epsilon \rightarrow 0^+} \psi_i(\epsilon), \quad i < j \quad (51)$$

if and only if ψ_i satisfies (48)–(50), for each $i \in I$.

Proof: Clearly (48) means $\psi_i: (0, \infty) \rightarrow [0, \infty)$ —bounded everywhere. Condition (49) is equivalent to convexity of ψ_i , which because ψ_i is bounded, means continuity as well in $(0, \infty)$. (Ref. 48, p. 91.) First, for $z_1 > z_2$, $z_i \in (0, \infty)$, choose $X \in (0, \bar{\mu}_i/2)$ and $\sigma \in (0, \bar{\sigma}/2)$ such that $z_1 = X/\sigma$ —as X/σ ranges over all $(0, \infty)$ this is always possible. Let $Y = \sigma z_2 = (z_2/z_1)X \in (0, \bar{\mu}_i/2)$ and take $\tau = \sigma$ in (49). Thus

$$\psi_i(z_1) + \psi_i(z_2) \geq 2\psi_i[(z_1 + z_2)/2] \quad \text{for all } z_1, z_2 \in [0, \infty) \quad (49')$$

so that by Ref. 48 (page 70) ψ_i is convex.

Conversely, if ψ_i is convex, then for $z_1, z_2 \in (0, \infty), \sigma, \tau \in (0, \infty)$,

$$\frac{\sigma}{\sigma + \tau} \psi_i(z_1) + \frac{\tau}{\sigma + \tau} \psi_i(z_2) \geq \psi_i \left[\frac{\sigma}{\sigma + \tau} z_1 + \frac{\tau}{\sigma + \tau} z_2 \right]$$

so taking $z_1 = X/\sigma$ and $z_2 = Y/\tau$ we obtain (49) for $X, Y, X + Y \in (0, \bar{\mu}_i)$ and $\sigma, \tau, \sigma + \tau \in (0, \bar{\sigma})$.

We show that given ψ_i are convex and continuous on $(0, \infty)$, (51) is equivalent to (50). In fact, we show that

$$\tau \psi_i(0^+) = \sup \{ (\sigma + \tau) \psi_i(X/(\sigma + \tau)) - \sigma \psi_i(X/\sigma); \\ X \in (0, \bar{\mu}_i), \sigma \in (0, \bar{\sigma}) \text{ such that } \sigma + \tau \in (0, \bar{\sigma}) \} \quad (52)$$

Let $X \in (0, \bar{\mu}_i)$, $Y \in (0, \bar{\mu}_j)$, $\sigma, \tau, \sigma + \tau \in (0, \bar{\sigma})$. Let $X = X' + \delta$ where $X', \delta \in (0, \bar{\mu}_i)$. Then by convexity

$$(\sigma + \tau) \psi_i(X/(\sigma + \tau)) - \sigma \psi_i(X/\sigma) \leq \sigma \psi_i(X'/\sigma) \\ + \tau \psi_i(\delta/\tau) - \sigma \psi_i(X/\sigma) \\ = \sigma [\psi_i((X - \delta)/\sigma) - \psi_i(X/\sigma)] + \tau \psi_i(\delta/\tau). \quad (53)$$

Now $|\psi_i((X - \delta)/\sigma) - \psi_i(X/\sigma)| \xrightarrow{\delta \rightarrow 0} 0$ by continuity, so for each $\epsilon > 0$, there exists $\delta \in (0, \bar{\mu}_i)$ such that

$$-\epsilon \sigma \leq \sigma [\psi_i((X - \delta)/\sigma) - \psi_i(X/\sigma)] \leq \epsilon \sigma, \\ -\epsilon' \tau \leq \tau [\psi_i(0^+) - \psi_i(\delta/\tau)] \leq \epsilon' \tau.$$

Take $\epsilon = \min(\epsilon, \epsilon')$ for δ .

$$(\sigma + \tau) \psi_i(X/(\sigma + \tau)) - \sigma \psi_i(X/\sigma) \leq \epsilon(\sigma + \tau) + \tau \psi_i(0^+), \\ \text{so}$$

$$(\sigma + \tau) \psi_i(X/(\sigma + \tau)) - \sigma \psi_i(X/\sigma) \leq \tau \psi_i(0^+). \quad (54)$$

Since for $X = 0^+$ equality holds in (54) we obtain (52). Now (50) is equivalent to

$$\tau \psi_j(Y/\tau) \geq (\sigma + \tau) \psi_i(X/(\sigma + \tau)) - \sigma \psi_i(X/\sigma); \\ \forall X \in (0, \bar{\mu}_i), Y \in (0, \bar{\mu}_j), \sigma, \tau, \sigma + \tau \in (0, \bar{\sigma})$$

if and only if

$$\tau \psi_j(Y/\tau) \geq \tau \psi_i(0^+), \quad \tau \in (0, \bar{\sigma}), \quad Y \in (0, \bar{\mu}_j),$$

if and only if

$$\psi_j(Y/\tau) \geq \psi_i(0^+), \quad \tau \in (0, \bar{\sigma}), \quad Y \in (0, \bar{\mu}_j),$$

if and only if

$$\psi_j(z) \geq \psi_i(0^+), \quad \forall z \in (0, \infty),$$

if and only if

$$\inf \{ \psi_j(z) : z \in (0, \infty) \} \geq \psi_i(0^+). \quad \text{Q.E.D.}$$

The following function ψ_i satisfies the Lemma's conditions:

$$\psi_i(z) = z \ln z + (1 + i)/e, \quad \text{for all } z \in (0, \infty). \quad (55)$$

Note that $\psi_i(0^+) = (1 + i)/e$ ($0 \ln 0 = 0$ convention).

Assume for J_1, J_2 the conditions of Theorem 3, and that $M_i = \bigvee \{ E \in \mathcal{T}_{a_i} : E \in \mathcal{T}_{a_i} \}$ and let $\sigma_i = \sigma|_{\mathcal{T}_{a_i}}$ (restriction of σ to \mathcal{T}_{a_i}). Then, using (24) with Π_i the set of partitions of M_i in \mathcal{T}_{a_i} again replacing $\{ \pi \cap \mathcal{T}_{a_i} \}$

$$H^F(J_1, J_2) = \sum_{i \in I} h_i, \quad (56a)$$

$$h_i = \sup_{\pi \in \Pi_i} \left\{ \sum_{\substack{\sigma_i(E_n) \in (0, \bar{\sigma}) \\ \mu_i(E_n) \in (0, \bar{\mu}_i)}} \sigma_i(E_n) \psi_i \left[\frac{\mu_i(E_n)}{\sigma_i(E_n)} \right] \right\}. \quad (56b)$$

In order to discuss classical and quantum physics, we further assume that each \mathcal{T}_{a_i} is a CROC for which only discrete superselection rules are present. We assume also that μ_i is a state on \mathcal{T}_{a_i} ($\mu_i(M_i) = \bar{\mu}_i = 1$) and $\mu_i \equiv \sigma_i$. Each CROC is associated with a W^* -algebra \mathcal{A}_i of observables on a Hilbert space \mathcal{H}_i and “state” N_i on \mathcal{A}_i . Indeed, there exists a homomorphism $\gamma_i: \mathcal{T}_{a_i} \xrightarrow{\text{onto}} \mathcal{L}_i = \mathcal{A}_i \cap \mathcal{P}(\mathcal{H}_i)$, where $\mathcal{P}(\mathcal{H}_i)$ is

the lattice of projectors on \mathcal{H}_i , so $N_i[\gamma_i(E)] = \mu_i(E)$ relates the two types of “states.” If γ_i happens to be an isomorphism, then there is a complete algebraic representation of \mathcal{T}_{a_i} . We henceforth assume this holds.⁴⁹

We also assume that the measure σ_i corresponds to a faithful, normal, semifinite trace, m_i , on \mathcal{A}_i . [See Ref. 1, p. 258, and Ref. 50.] That is, $m_i[\gamma_i(E)] = \sigma_i(E)$ for all $E \in \mathcal{T}_{a_i}$. The main characteristic of such a trace (aside from certain convergence properties) is its unitary invariance:

$m_i(\hat{U}^\dagger \hat{H} \hat{U}) = m_i(\hat{H})$ for all unitary $\hat{U} \in \mathcal{A}_i$ and all $\hat{H} \in \mathcal{L}_i$. Of the state N_i we demand that there exist $\hat{W}_i > 0$ in \mathcal{A}_i such that $N_i(\hat{X}) = m_i(\hat{W}_i \hat{X})$ for all $\hat{X} \in \mathcal{A}_i$. We can now describe h_i in terms of partitions of projectors in \mathcal{L}_i :

$$h_i = \sup_{\pi_i} \left\{ \sum_{m(\hat{P}_n) \neq 0, \infty} m_i(\hat{P}_n) \psi_i[m_i(\hat{W}_i \hat{P}_n)/m_i(\hat{P}_n)] \right\}. \quad (57)$$

[Note that from $\sigma_i \equiv \mu_i$, $\sigma_i(E) = 0$ iff $\mu_i(E) = 0$, but

$\mu_i(E) = \bar{\mu}_i < \infty$ iff $\mu_i(E) = 0$ iff $\sigma_i(E^\perp) = 0$

iff $\sigma_i(E) = \bar{\sigma}_i < \infty$. Thus we can eliminate the conditions on μ_i in (56b).] Finally, let us suppose that $\psi_i(\hat{W}_i)$ is “ m_i -integrable,” i.e., $m_i[\psi_i(\hat{W}_i)] < \infty$. Under these many conditions we have that $h_i = m_i[\psi_i(\hat{W}_i)]$. We prove this in the following Lemma (dropping the index i for convenience.)

Lemma: Let m be a faithful, normal, semifinite trace on a W^* -algebra \mathcal{A} with projection lattice \mathcal{L} (a logic). Let n be a normal state on \mathcal{A} such that there exists self-adjoint $\hat{W} \in \mathcal{A}$, $\hat{W} > 0$, such that $n(\hat{X}) = m(\hat{W} \hat{X})$ for all $\hat{X} \in \mathcal{A}$. We assume $\psi: (0, \infty) \rightarrow (0, \infty)$ is convex and continuous, and $m[\psi(\hat{W})] < \infty$. Then

$$m[\psi(\hat{W})] = \sup_{\pi} \sum_{m(\hat{P}_n) \neq 0, \infty} m(\hat{P}_n) \psi[m(\hat{W}\hat{P}_n)/m(\hat{P}_n)]. \quad (58)$$

Proof: \hat{W} being positive and self-adjoint,

$$\hat{W} = \int_0^\infty d\hat{P}(\omega) \omega \quad (59)$$

is its spectral resolution; we define $\psi(\hat{W})$ by

$$\psi(\hat{W}) = \int_0^\infty d\hat{P}(\omega) \psi(\omega) \quad (60)$$

so that with $m_{\hat{W}}(\Delta) = \int_{\Delta} dm \hat{P}(\omega)$, $\Delta \in B(R^+)$:

$$m[\psi(\hat{W})] = \int_0^\infty dm [\hat{P}(\omega)] \psi(\omega) = \int_0^\infty dm_{\hat{W}} \psi(\omega). \quad (61)$$

Since ψ is convex and continuous, its domain decomposes into two disjoint parts: D_1 on which ψ is monotonic increasing, and D_2 on which ψ is monotonic decreasing. Thus, using (61) we obtain

$$\begin{aligned} m[\psi(\hat{W})] &= \sum_{j=1}^2 \int_{\psi^{-1}(D_j)} dm_{\hat{W}} \psi(\omega) \\ &= \sum_{j=1}^2 \sup_{\{\Delta_n \cap \psi^{-1}(D_j)\}} \left[\sum_{m_{\hat{W}}[\Delta_n \cap \psi^{-1}(D_j)] \neq 0, \infty} m_{\hat{W}}[\Delta_n \cap \psi^{-1}(D_j)] \right. \\ &\quad \left. \times \inf\{\psi(\omega); \omega \in \Delta_n \cap \psi^{-1}(D_j)\} \right]. \end{aligned} \quad (62)$$

The restriction to elements of the partitions of R^+ with $m_{\hat{W}}[\Delta_n \cap \psi^{-1}(D_j)] \neq 0, \infty$ follows from the integrability of $\int_{\psi^{-1}(D_j)} dm_{\hat{W}} \psi(\omega)$ while the supremum representation of the integral is standard. (See, for example, Ref. 39, p. 115.) On the other hand, for all $\Delta \in B(R^+)$ we have

$$\begin{aligned} m_{\hat{W}}[\Delta \cap \psi^{-1}(D_j)] \inf\{\omega \in \Delta \cap \psi^{-1}(D_j)\} &\leq \int_{\Delta \cap \psi^{-1}(D_j)} dm_{\hat{W}} \omega \\ &\leq m_{\hat{W}}[\Delta \cap \psi^{-1}(D_j)] \sup\{\omega \in \Delta \cap \psi^{-1}(D_j)\}, \end{aligned} \quad (63)$$

where $m_{\hat{W}}[\Delta \cap \psi^{-1}(D_j)] \neq 0, \infty$. Exploiting the monotonicity of ψ on D_j yields

$$\begin{aligned} \inf\{\psi(\omega); \omega \in \Delta \cap \psi^{-1}(D_j)\} &\leq \psi \left[\int_{\Delta \cap \psi^{-1}(D_j)} dm_{\hat{W}} \omega / m_{\hat{W}}(\Delta \cap \psi^{-1}(D_j)) \right]. \end{aligned} \quad (64)$$

Thus, by (64) in (62) we get

$$\begin{aligned} m[\psi(\hat{W})] &\leq \sum_{j=1}^2 \sup_{\{\Delta_n \cap \psi^{-1}(D_j)\}} K[\{\Delta_n \cap \psi^{-1}(D_j)\}] \\ &= \sup_{\{\Delta_n\}} \sum_{j=1}^2 K[\{\Delta_n \cap \psi^{-1}(D_j)\}], \end{aligned} \quad (65)$$

where

$$\begin{aligned} K[\{\Delta_n \cap \psi^{-1}(D_j)\}] &= \sum_{m_{\hat{W}}[\Delta_n \cap \psi^{-1}(D_j)] \neq 0, \infty} m_{\hat{W}}[\Delta_n \cap \psi^{-1}(D_j)] \\ &\quad \times \psi \left[\int_{\Delta_n \cap \psi^{-1}(D_j)} dm_{\hat{W}} \omega / m_{\hat{W}}(\Delta_n \cap \psi^{-1}(D_j)) \right] \end{aligned}$$

and an argument as in Corollary 2 to Theorem 5 applies. But, $\{\Delta_n \cap \psi^{-1}(D_j)\}$ is a finer partition of R^+ than $\{\Delta_n\}$, and $\{\hat{P}_{\Delta_n \cap \psi^{-1}(D_j)} = \int_{\Delta_n \cap \psi^{-1}(D_j)} d\hat{P}(\omega)\}$ is among all projector partitions $\{\hat{P}_n; n = 1, 2, \dots, N_\pi\} \in \Pi$, so that

$$m[\psi(\hat{W})] \leq \sup_{\pi} \sum_{m(\hat{P}_n) \neq 0, \infty} m(\hat{P}_n) \psi[m(\hat{P}_n \hat{W})/m(\hat{P}_n)]. \quad (66)$$

On the other hand, for any projector \hat{P} such that $m(\hat{P}) \neq 0, \infty$

$$\begin{aligned} \psi \left[\frac{m(\hat{P}\hat{W})}{m(\hat{P})} \right] &= \psi \left[\int dm [\hat{P}\hat{P}(\omega)] \omega / m(\hat{P}) \right] \\ &\leq \int dm [\hat{P}\hat{P}(\omega)] \psi(\omega) / m(\hat{P}) = m[\hat{P}\psi(\hat{W})] / m(\hat{P}), \end{aligned} \quad (67)$$

using Jensen's inequality. Employing (67) with (66) yields

$$\begin{aligned} m[\psi(\hat{W})] &\leq \sup_{\pi \in \Pi} \sum_{m(\hat{P}_n) \neq 0, \infty} m(\hat{P}_n) \psi[m(\hat{P}_n \hat{W})/m(\hat{P}_n)] \\ &\leq \sup_{\pi} \sum_{m(\hat{P}_n) \neq 0, \infty} m(\hat{P}_n \psi(\hat{W})) \leq m[\psi(\hat{W})] \end{aligned} \quad (68)$$

since $\sum_{m(\hat{P}_n) \neq 0, \infty} \hat{P}_n \leq \hat{I}$. Thus, we have (58). Q.E.D.

The above result is based on the proof of Theorem 1.1 of Ref. 51, a classical version of the Lemma (without the restrictions $\psi \geq 0$ and integrable). Our "entropy" is thus given by

$$H^F(J_1, J_2) = \sum_{i \in I_1} m_i[\psi_i(\hat{W}_i)]. \quad (69)$$

In case $\sigma_i(M_i) < \infty$ and $\psi_i(z)$ is given by (55), this becomes

$$\begin{aligned} H^F(J_1, J_2) &= \sum_{i \in I_1} [m_i(\hat{W}_i \ln \hat{W}_i) + (1+i)m_i(\hat{I}_i)/e] \\ &= \sum_{i \in I_1} [-S_i(\hat{W}_i) + (1+i)m_i(\hat{I}_i)/e], \end{aligned} \quad (70)$$

where $S_i(\hat{W}_i)$ is the "Segal entropy," which includes both the (classical) Gibbs-Boltzmann-Shannon entropy and the (quantum) von Neumann entropies as cases.⁵⁰ If, however, $\sigma_i(M_i) = \infty$ —our result fails since $m_i[\psi_i(\hat{W}_i)] = \infty \geq h_i$. In words, the "global information" is essentially the "neg-entropy" plus the prior measure of the space.

Since our purpose in introducing these illustrations was simply to indicate the relationship of the "global information" concept to typical "entropies," we will not discuss further properties or problems associated with expressions like (70). Rather, we postpone such discussion until the axiom set is completed by an appropriate description of the information of compound systems.

IV. A TENTATIVE INTERPRETATION

In the preceding sections we have established minimal universal criteria for a "global measure of information on a 'theory' $L(P)$ relative to M (generalized) local information measures"—a quantity which by the examples we considered, apparently relates to a generalization of the neg-entropies of physics. We have purposely avoided questions of characterization, as the universal structure of $L(P)$ is not yet fixed. Nevertheless, a truly surprising result appears already, namely, the introduction of a natural semigroup F (on R^{+M}) and its natural segmentation of the "entropy." In keeping with our policy of specialization, let us consider the case $M = 1$. (The arguments apply as well to the $M = 2$ case considered in Sec. III.)

With $a_0 = 0 < b_0 \leq a_1 < b_1 \leq a_2 \dots$ by the construction of Theorem 3, we may write (24) as

$$H^F(J) = \lim_{I \rightarrow \infty} H^F_I(J), \quad (71)$$

where we assume for convenience that $\varphi^F(0) = 0$ and define for all $t > 0$:

$$H_t^F(J) = \sum_{i \in I, t > a_i} \sup_{\pi \in \Pi} \left[\sum_{J(E_n) \in [a_i, b_i]} \varphi^F[J(E_n)] \right] + \sum_{i \in I, t > b_i} \sup_{\pi \in \Pi} \left[\sum_{J(E_n) \in [b_i, \min(t, a_{i+1})]} \varphi^F[J(E_n)] \right]. \quad (72)$$

Note that $H_t^F(J)$ is constant for all $t \in (a_i, b_i)$ and is generally increasing otherwise. The connection between this monotonic nondecreasing “entropy” and dynamics—the “Second Law”—comes about as follows.

Let us model *time* by the positive real line $\overline{R^+}$. This corresponds to the usual “initial value” problem of physics, but can also represent an operational model of a general relativistic cosmos because of the finite age of the universe. A dynamical object is described³⁵ by

$$P(T) = \bigvee_{z \in [0, T]} P_z; \quad \overline{P} = \lim_{T \rightarrow \infty} P(T), \quad (73)$$

where \bigvee is the direct product of posets (the cartesian set product ordered in the obvious way²⁷). According to our model, each $P_z, z \in \overline{R^+}$, represents the abstract definition—i.e., the ordered skeleton of defining relations—at the “instant” z . The structure may endure over finite intervals, i.e. $P_z = P$ for all $z \in (a, b)$, whereas it may continuously change over other intervals. Since $L(\bigvee P_z) = \bigvee L(P_z)$ ²² these comments extend directly to the abstract “theory” of the “objects.”

Note that the “object” thus described is a “dynamical object.” In fact, this description is analogous to the world-line models in relativity. Let us define

$$\mathcal{T}_T = [\bigvee_{z \in [0, T]} \phi_z] \vee [\bigvee_{z \geq T} L(P_z)] \subseteq L(\overline{P}). \quad (74)$$

Clearly each \mathcal{T}_T is a complete sublattice of $L(\overline{P})$ (and thus a σ -ideal). Moreover, if $T_1 < T_2$, $\mathcal{T}_{T_1} \supset \mathcal{T}_{T_2}$ is evident. If we impose on $\overline{R^+}$ the semigroup F with idempotents of $A(F)$, the ideals $\mathcal{T}_z, z \in A(F)$ correspond to *distinct* components in $L(\overline{P})$, while $\bigvee_{z \in (a_i, b_i)} \mathcal{T}_z$ corresponds to $\bigvee_{z \in (a_i, b_i)} L(P_z)$, where $L(P_z)$ is the *same* structure for all z in (a_i, b_i) . [Compare with the usual physical model of Ref. 35 where the $L(P_z)$ are related to the mutually isomorphic Hilbert spaces.] Thus, intuitively each choice of F (a real-line semigroup!) induces a world-line model that consists of (open) segments with constant entropy (and reversible unitary dynamics) and of (closed) segments of constantly increasing entropy in which dynamics is irreversible in that the physical structure of the object changes. The “entropy” $H_t^F(J)$ defined by (72) is the total information about the system up to “time t ”—i.e., the entropy of the “world-line” segment defined by $[0, t]$. As more data about the “world-line” is obtained, the information increases towards the maximum represented by (71).

Each “dynamical object” can therefore be associated with a semigroup $(\overline{R^+}, F)$ —not the usual physical choice²—and F — σ -composable informations on $L(\overline{P})$ which are related to unitarily evolving measures on \mathcal{T}_{a_i} (that vanish on $\mathcal{T}_{a_i} \cap \mathcal{T}_{b_i}$). In order to see more clearly the significance of this model, consider the example of an atomic sys-

tem which decays at time T . This system can be described by $L(P_A)$ —the lattice of projections of a Hilbert space defining the atomic system—and $L(P_D)$ —the lattice of projections of the Hilbert space defining the decay products. Then

$$L(\overline{P}) = [\bigvee_{t \in [0, T]} L(P_A)] \vee [\bigvee_{t \geq T} L(P_D)] \quad (75)$$

is the “dynamical object.” Note the dependence on T :

$A(F) = \{0, T, \infty\}$ defines the corresponding information. We have $\mathcal{T}_0 = L(\overline{P})$ and $\mathcal{T}_T \approx \bigvee_{t \geq T} L(P_D)$. A measure μ_0 on \mathcal{T}_0 acts, in fact, solely on $\bigvee_{t \in [0, T]} L(P_A)$ since $\mu_0: \mathcal{T}_0 \cap \mathcal{T}_T \rightarrow 0$. Note that $\mu_0(\overline{\mu}_i = 1)$ is a *joint measure* on $\bigvee_{t \in [0, T]} L(P_A)$ whose marginals on $L(P_A)_t$ define the dynamically evolving “states” $\overline{W}_A(t)$ on \mathcal{H}_A . The same remarks apply to μ_T on \mathcal{T}_T . It should be appreciated that the present description in fact extends the ideas of Piron [see Ref. 35, pp. 117–119] and explains the lattice theoretic model for general irreversible systems in terms of information theory.

The potentiality of the present viewpoint revolves about the interpretation of (71) as an information measure. If one knew precisely J, F , and $L(\overline{P})$, (71) would be fixed and there would be complete knowledge of the “object.” In reality, we can only *infer* J, F , and $L(\overline{P})$ from generally inadequate evidence (or hypotheses). Thus, the role of information theory in the inductive process, described in the Introduction, can be exploited in the present context to seek estimates of J and $A(F)$, e.g., if we fix θ_i and the distinct $L(P_z)$ ’s. We shall consider these applications elsewhere.

V. CONCLUSION

In this paper we have attempted to provide a universal, intuitively reasonable axiomatic description of “entropy.” Our approach takes seriously the information-theoretic interpretation of “entropy,” and explores the universal explication of this concept in the context of what might be called “algebraic measurement theory.” Although we have not characterized our “entropy” species, we have found that with sufficient effort one can (more or less) recover the more common entropy formulas as special cases of our definition. Besides this encouraging result, we have discovered a surprising relationship between semigroups on $\overline{R^+}$ and Piron’s model for irreversibility. This, in conjunction with information-theoretic inference procedures, suggests a new (and relatively simple) way to devise models, etc. for irreversible processes.

While this paper represents a first toddle in hopefully the right direction, it clearly leaves much work to be done. Most urgent, perhaps, is the incorporation of an appropriate “tensor product” in the scheme to describe compound objects whose components preserve a certain identity: The “entropy” of such a compound system will evidently suffer restrictions similar to the usual (sub-)additivity that favors the logarithmic function in a central entropic role. Beyond this, one can establish a “communication theory” between such abstract “languages” which (among interesting social potentials) could perhaps form the basis of a quantitative “meta-theory” to compare distinct mathematical theories or physical models. We have already spoken at length of the more and more respectable techniques of information-theoretic

inference which provide the means to use the full “data-processing” potential of a theory to predict based on observations—as well as defining useful criteria for the relevance of parameters within a theory. We anticipate that the “communication theory” envisioned will play a pivotal role in establishing the universal validity of a generalized MEF¹⁷ and in removing the various difficulties it now suffers.

In closing, we note that our axioms already disqualify certain candidates from the entropy zoo. In particular, we have pointed out that the classical and quantal “entropies” correspond to *infinite* global information ($h_i \leq \infty$ actually) when the “prior” information corresponds to an infinite measure. In fact, this illustrates the obvious: That in “real-life” priors based on invariance over infinite spaces (Lebesgue measure classically, infinite Hilbert spaces quantally) require an infinite amount of information to specify. In reality, one should account for the finite limits of the *observable* cosmos (operational view) by appropriately delimiting the priors. Otherwise, one must apparently seek other formulas for “entropy.”⁵²

Noted added in Proof: We have discovered an omission in conditions (43)–(45): Further consequences of Axioms 2 and 4 compel $\psi_i(z)$ in (47) to vanish. Deletion of these axioms permits the representation (20) without $-\phi^F(0, \dots, 0)$, where ϕ^F has real range and (19c) is not in force. In this case $\psi_i(z) = z \ln z$ is admissible in (47) and, since non-negativity of ψ_i is not essential in deriving (58) (see Ref. 51), we obtain $H^F(J_1, J_2) = -\sum_{i \in I} S(\hat{W}_i)$. If $\hat{W}_i^2 \leq \hat{W}_i$, $S(\hat{W}_i) \geq 0$ [M. B. Ruskai, Ann. Inst. H. Poincaré **19**, 357 (1973)] so that $H^F(J_1, J_2)$ is monotonic (non-increasing) in “time” (Sec. IV). The error noted here does not effect any other result of this paper.

ACKNOWLEDGMENTS

This work has been funded in part by a research grant from the N.H.R.F. and in part by the Theoretical and Physical Chemistry Institute of the N.H.R.F. I wish to acknowledge the hospitality of the members of this Institute, and useful discussions with Dr. I. Komninos and Prof. C. A. Nicolaides.

¹A. Wehrl, Rev. Mod. Phys. **50**, 221 (1978).

²J.-P. Marchand and W. Wyss, J. Stat. Phys. **16**, 349 (1977).

³W. Band and J. L. Park, Found. Phys. **6**, 249 (1976).

⁴M. D. Srinivas, J. Math. Phys. **19**, 1952 (1978).

⁵M. Mastrangelo and V. Mastrangelo, Ann. Inst. Henri Poincaré **A 30**, 295 (1979).

⁶B. Robertson, in *The Maximum Entropy Formalism*, edited by R. D. Levine and M. Tribus, (MIT Press, Cambridge, 1979), p. 289.

⁷J. Mehra and E. C. G. Sudarshan, N. Cim. **B 11**, 215 (1972).

⁸I. Prigogine, C. George, F. Henin, and L. Rosenfeld, Chem. Scripta **4**, 5 (1973).

⁹I. J. Good, Science **129**, 443 (1959).

¹⁰T. Fine, *Theories of Probability* (Academic, New York, 1973).

¹¹J. Rayski, Found. Phys. **3**, 89 (1973); **7**, 151 (1977); **9**, 217 (1979).

¹²J. F. Cyranski, Found. Phys. **8**, 493 (1978).

¹³E. T. Jaynes, Phys. Rev. **106**, 620 (1957); **108**, 171 (1957).

¹⁴L. Brillouin, *Science and Information Theory* (Academic, New York, 1956).

¹⁵J. B. Barón and J. M. Jauch, Helv. Phys. Acta **45**, 220 (1972).

¹⁶R. Carnap, *Logical Foundations of Probability* (University of Chicago, Chicago, 1950).

¹⁷(a) B. J. Berne, P. Pechukas, and G. D. Harp, J. Chem. Phys. **49**, 3125 (1968). (b) H. Heiner, Phys. Stat. Sol. (b) **91**, 87 (1979). (c) S. Gadre and S. B. Sears, J. Chem. Phys. **71**, 4321 (1979). (d) C. Y. She and H. Heffner, Phys. Rev. **152**, 1103 (1966). (e) B. Robertson, Phys. Rev. **144**, 151 (1966); **160**, 175 (1967). (f) A. Corbet and H. Morowitz, Phys. Rev. A **6**, 2298 (1972). (g) J. D. Bekenstein, Phys. Rev. **D 12**, 3077 (1975). (h) A. G. Wilson, *Entropy in Urban and Regional Modelling* (Pion, London, 1970). (i) See also articles and Refs. in: *The Maximum Entropy Formalism*, edited by R. D. Levine and M. Tribus (MIT, Cambridge, 1979).

¹⁸W. Ochs, Rep. Math. Phys. **8**, 109 (1975); **9**, 135 (1976); and **9**, 331 (1976).

¹⁹J. E. Shore and R. W. Johnson, I.E.E.E. Trans. Inf. Theory **IT-26**, 26 (1980).

²⁰The MEF suffers from the restrictive form the evidences must take and the problem of choosing the correct “prior” measure. We plan to offer at least a partial solution (classical case) to these problems elsewhere.

²¹J. F. Cyranski, Found. Phys. **8**, 805 (1978).

²²J. F. Cyranski, Inform. Contr. **41**, 275 (1979).

²³J. F. Cyranski, Found. Phys. **9**, 641 (1979).

²⁴J. Sallantin, C. R. Acad. Sci. Paris **A 274**, 986 (1972).

²⁵There exist numerous characterization theorems for various “entropies.” See Ref. 1 for examples, and Refs. 18.

²⁶For a summary of these models, see *The Logico-Algebraic Approach to Quantum Mechanics*, vols. I and II, edited by C. A. Hooker (Reidel, Dordrecht, Holland, 1975; 1979).

²⁷G. Birkhoff, *Lattice Theory*, 3rd ed. (Am. Math. Soc., Providence, 1967).

²⁸S. Watanabe, Inform. Contr. **15**, 1 (1969).

²⁹J. Sallantin, *Structures de L'Information* No. 8 (C.N.R.S., Paris 1979).

³⁰A. Kaufmann, *Introduction to the Theory of Fuzzy Subsets*, Vol. I (Academic, New York, 1975).

³¹M. M. Gupta, G. N. Saridis, and B. R. Gaines, editors, *Fuzzy Automata and Decision Processes* (North-Holland, Amsterdam, 1977).

³²J. A. Goguen, J. Math. Anal. Appl. **18**, 145 (1967).

³³A. De Luca and S. Termini, J. Math. Anal. Appl. **40**, 373 (1972).

³⁴S. T. Ali and E. Prugovecki, J. Math. Phys. **18**, 219 (1977).

³⁵C. Piron, *Foundations of Quantum Mechanics* (Benjamin, New York, 1976).

³⁶Ref. 27; for other completion alternatives, see W. R. Tunnicliffe, Proc. Lond. Math. Soc. **3**, 13 (1974).

³⁷M. D. MacLaren, Pac. J. Math. **14**, 597 (1964); K. Bugajska and S. Bugajski, Ann. Inst. Henri Poincaré **A 19**, 333 (1973).

³⁸This proof extends the relevant part of Lemma (29.11) of F. Maeda and S. Maeda, *Theory of Symmetric Lattices* (Springer, Berlin, 1970).

³⁹P. R. Halmos, *Measure Theory* (van Nostrand, Princeton, N.J., 1950).

⁴⁰Strictly speaking, $L(P)$ is not a “logic.” However, recent work suggests how an operation of implication (conditional) may be introduced in general (at least, orthomodular) lattices: G. M. Hardegree, Z. Naturforsch. **30A**, 1347 (1975). Moreover, quantifiers can be introduced as mappings (Boolean context), which suggests a possible avenue to extend quantification to $L(P)$: P. Halmos, *Algebraic Logic* (Chelsea, New York, 1962). If these extensions are accomplished, one could define a theory of “semantic information” on a good deal of everyday language. J. F. Cyranski, in *Advances in Communications*, edited by D. G. Lainiotis and N. S. Tzannes (Reidel, Dordrecht, 1980).

⁴¹G. Comyn and J. Losfeld, C. R. Acad. Sci. Paris **A 278**, 633 (1974); see also Ref. 24.

⁴²J. Kampé de Fériet and P. Benvenuti, C. R. Acad. Sci. Paris **A 274**, 655 (1972).

⁴³J. Kampé de Fériet, B. Forte, and P. Benvenuti, C. R. Acad. Sci. Paris **A 269**, 529 (1969).

⁴⁴B. Forte and C. T. Ng, Ann. Mat. Pura. Appl. **101**, 355 (1974).

⁴⁵B. Forte, in *Théorie de L'Information: Développements Récents et Applications*, edited by C. Picard (Edits. du C.N.R.S., Paris, 1978). Forte argues that in statistical physics neither the degrees of freedom nor the phase space (subset of R^{2n}) are known, so that the “entropy” cannot depend solely on the distribution $f_{2n}(x)$ over R^{2n} . Our model has $L(P)$ fixed. See Sec. 4, however.

⁴⁶B. Forte, R.I.R.O. **3**, R-2, 63 (1969).

⁴⁷A. Wehrl, Found. Phys. **9**, 939 (1979).

⁴⁸G. H. Hardy, J. E. Littlewood, and G. Polya, *Inequalities*, 2nd ed. (Cambridge U. P., London, 1973).

⁴⁹For details see R. Cirelli and F. Gallone, Ann. Inst. H. Poincaré **A 19**, 7

(1973). If one prefers the algebraic method, one can utilize the fact that the projectors of a W^* -algebra form a "quantum logic" and use *this* "logic" as \mathcal{T}_{α} . See V. S. Varadarajan, *Geometry of Quantum Theory*, Vol. 1 (van Nostrand, Princeton, N.J., 1968).

⁵⁰W. Ochs and H. Spohn, *Rep. Math. Phys.* **14**, 75 (1978).

⁵¹S. G. Ghurye, *Ann. Math. Stat.* **38**, 2056 (1968).

⁵²See Ref. 21 and E. T. Jaynes, *IEEE Trans. Syst. Sci. Cyber.* **SSC-4**, 227 (1968). If $\nu(\Omega) < \infty$ ($M = 2$ context) in general unitary invariance will be lost. Thus, alteration of the "prior" from a "trace" to a weight (Ref. 50, fn. 4) will necessarily modify the form of the "entropy," since $n(\hat{X}) = m(T\hat{X}T)$, $\hat{T} > 0$, is the new "Radon-Nikodym" Theorem. See S. Sakai, *Bull. Am. Math. Soc.* **71**, 149 (1951).