INFORMATION AND CONTROL 15, 1–21 (1969)

Modified Concepts of Logic, Probability, and Information Based on Generalized Continuous Characteristic Function*

SATOSI WATANABE

University of Hawaii, Honolulu, Hawaii

All the basic laws of the traditional logic can be derived from the characteristic function $f(A/a)$ which is 1 or 0 according as object a satisfies predicate A or not. There is good reason to believe that it is worthwhile to extend this formalism to the case where $f(A/a)$ can take any value in the continuous domain [0, 1]. The implications of this generalization to the concepts of logic, probability and information theory are studied.

1. INTRODUCTION

The concept of information is based on the concept of probability. The concept of probability is defined on a sigma algebra or a distributive lattice. This is justified by the fact that the usual logic is isomorphic to a distributive lattice. All the laws of the Boolean logic can be derived from the characteristic function $f(A/a)$ which is 1 if object a belongs to class A and is 0 if it does not belong to class A. A class is understood as the extension of a predicate. It is assumed that there exists an empirically based, well-defined procedure by which to determine whether a belongs to A or not. This assumption is obviously not always satisfied.

The basic postulate that a and A determine the value of $f(A/a)$ which is either 0 or 1 is called here "the postulate of fixed truth set." The breakdown of this postulate was already noticed when philosophers made distinction between the primary quality and the secondary quality.

* This paper was scheduled to be presented at the abortive Fifth Prague Conference on Information Theory, Statistical Decision Functions and Random Processes, September 9–13, 1968 and to be published in the Proceedings. A simplified version of this paper was orally presented at the IEEE Systems Science and Cybernetics Conference, October 14–15, 1968 at San Francisco. The work reported in this paper was partly supported by the research grant AF-AFOSR-68-1466 from the Air Force Office of Scientific Research.

The characteristic function of a secondary quality depends not only on a and A but also on a third argument, x, which is the observer. But this kind of trouble could be avoided easily either by requiring precise conditions about x or by considering x as an index of A, i.e., considering an A with different values of x as different A's.

Another case of deviation of $f(A/a)$ from 0 or 1 that can be easily avoided happens when the specification of the object a is not sufficiently accurate so that a usual sample of a is in reality a probabilistic mixture of two kinds of objects, one of them satisfying $f(A/a) = 1$ and the other satisfying $f(A/a) = 0$. This case of a mixture will be discussed later in (41) and again in (58) below. We can show, however, that in the majority of cases, we cannot reduce the continuous value of $f(A/a)$ to a binary case.[1]

An irreducibly continuous characteristic function seems to be needed in the typically cybernetical situation, i.e., in the case where the observer and the object are in such a strong interaction that the act of observation leaves an uncontrollable disturbance on the observed object. This is obviously the case in many psychological tests and in quantum mechanical measurement. A consequence of such uncontrollable disturbance is that when we make two observations in different orders the result will not be the same in general. Such order-dependence of observations will play one of the basic roles in the following derivation. See in particular (43) and (62) below.

The purpose of this paper is to reconstruct logic, probability and information theory on the basis of irreducibly continuous characteristic functions. As stated above, if the postulate of fixed truth set is tenable, as it is the case with the binary-valued characteristic functions, everything with regard to predicates can be reduced to the set theory of objects, and it is easy to derive directly therefrom logical operations such as conjunction, disjunction, negation, etc. But, in our case, since the postulate of fixed truth set is not valid, we have to derive these logical operations of predicates independently of the set-theoretical operations applied to objects. That is what makes our approach at once difficult and challenging.[2] The leading idea is that "implication" is the most basic

[1] In Ref. [2], a predicate whose characteristic function is irreducibly nonbinary (not necessarily 0 or 1) for an object was called an improper predicate for the object.

[2] This marks a fundamental difference from Zadeh's theory of fuzzy set, which uses from the beginning the notions of set, conjunction, disjunction, etc., as if they were already known to us. His determination of the values of the membership function for conjunction and disjunction is arbitrary. See Ref. [5].

operation in logic and therefore it should be defined first and then we should try to reconstruct conjunction, disjunction, etc. from implication. Not everything follows necessarily in the derivation, and certain heuristic groping is necessary. What is important during this heuristic groping is the idea that probability is more basic and more natural to our thinking than logic, and that we should always try to uphold the laws of probability as much as possible. For more about the philosophical background of this work, see Section 7 of this paper and Refs. [*1*] and [*2*]. The present paper is an improved and streamlined version of the author's earlier works on a similar topic [Refs. 1–4].

It is to be kept in mind that any meaningful generalization of an existing valid theoretical scheme must be such that the new generalized form of the theory in conjunction with some additional restrictive conditions will result in the present form of the theory. Thus, we should aim at a new form of logical theory such that it becomes identical with the existing logic with its inseparable set-theoretic notion of truth-sets if an additional restriction is imposed. See (45) below. This is also necessary because the metalanguage we use in formulating the new theory assumes the usual logic and the usual notion of truth-set, and the last two named must be justified as a permissible special case of the more general framework of the object language which we are going to develop.

In the foregoing paragraphs and in the main part of the following sections (except Section 6), it is assumed that for each predicate and each object considered there exists a test procedure which, if applied, gives either an affirmative ("true") or negative ("false") result.[3] Even in the usual logical framework, there are groups of propositions which are not "testable" in this sense. Hypotheses of general nature cannot be empirically demonstrated to be true or false by a finite body of experimental results. It is known that some mathematical theorems cannot be proven to be true or false. It is usually taken for granted that in spite of their direct untestability they have to be either true or false. In this paper, we refrain from asserting anything definite about the directly untestable cases in the new generalized framework.

It is safer to assume that what follows refers only to the directly testable cases. In the last section (Section 6) of this paper, however, we shall briefly introduce an entirely different case, in which a single experiment directly gives the "probability" of an object a satisfying a predicate

[3] Any proposition can be reduced to a pair of object and predicate by transcribing it in the form: "The case is such that the proposition is true."

A instead of an yes-no answer. This can also be formulated in terms of a continuously-valued characteristic function $f(A/a)$, and leads to a similar type of new logic. In this case, we do not need to talk about the "uncontrollable disturbance" by observation, and we should not confuse this case with the general problem dealt with in the main body (Sections 1–5) of this paper.

2. NON-DISTRIBUTIVE LOGIC

We shall now develop more systematically the ideas explained in Section 1.

(1) Let $\mathfrak{O} = \{a, b, c, \cdots\}$ and $\mathfrak{R} = \{A, B, C, \cdots\}$ be, respectively a set of objects and a set of predicates such that for each object $a \in \mathfrak{O}$ the set $\mathfrak{R}$ is the set of all possible predicates applicable (affirmatively or negatively) to a, and for each predicate $A \in \mathfrak{R}$ the set $\mathfrak{O}$ is the set of all possible objects to which A is applicable (affirmatively or negatively). Note that the notion of truth-set is used in the metalanguage here.

(2) By an object is meant an object at a particular instant, and an object is in a particular (pure or mixed) "state" at a particular instant. If the state of an object at instant t_1 and that at t_2 are different, we describe them as two different objects.

(3) To each predicate $A \in \mathfrak{R}$ corresponds a well-defined experimental test procedure by which we can determine whether the predicate A is affirmed (true) or negated (false) by an object.

(4) Each pair, $a \in \mathfrak{O}$ and $A \in \mathfrak{R}$, determines a real number $f(A/a)$ in the domain [0, 1]. $f(A/a)$ is called the characteristic function of A on a.

(5) It is assumed that we can produce any number of samples of the same a. The characteristic function $f(A/a)$ is interpreted as the relative frequency of affirmative results obtained by the test of A on a collection of infinitely many samples of a.[4] The object a may be understood to mean a (random) sample of the collection of objects.

(6) If, for all predicates $A \in \mathfrak{R}$ and two given objects $a, b \in \mathfrak{O}$, we have $f(A/a) = f(A/b)$, then we say that a and b are equivalent and write $a = b$. If, for all objects $a \in \mathfrak{O}$ and two given predicates $A, B \in \mathfrak{R}$, we have $f(A/a) = f(B/a)$, then we say that A and B are equivalent and write $A = B$.

[4] Alternatively, $f(A/a)$ may be considered as the degree of expectation on our part of obtaining an affirmative result on the observation of A applied to object a.

(7) The set $\mathfrak{R}$ contains two members $\varnothing$ and $\square$ such that for all objects $a \in \mathfrak{O}$ we have $f(\varnothing/a) = 0$ and $f(\square/a) = 1$.

(8) By the product AB ($A, B \in \mathfrak{R}$) is meant a predicate which is true if and only if we obtain affirmative results both in the test of A and in the test of B, whereby the test of A follows immediately the test of B. If $A, B \in \mathfrak{R}$ then $AB \in \mathfrak{R}$.

(9) In performing the test of AB on a, the object a changes after the test of B to another object a' which is determined by a and the test result of B. If B is affirmative, we write $a' = Ba$. If $a \in \mathfrak{O}$ and $B \in \mathfrak{R}$, then $Ba \in \mathfrak{O}$. If $B = C$, then $Ba = Ca$, and if $a = b$ then $Ba = Bb$. From Definition (8), we have $f(AB/a) = f(A/Ba)\, f(B/a)$.

(10) From (9) follow for all $A, B, C \in \mathfrak{R}$:

(10a) If $B = C$, then $AB = AC$ and $BA = CA$.

(10b) $\varnothing A = A\varnothing = \varnothing$

(10c) $\square A = A\square = A$

[The proof of (10c) requires an additional assumption that $\square a = a$, for all $a \in \mathfrak{O}$, which we adopt.]

(11) If $A \in \mathfrak{R}$ and $AA = A$, we say that A is a simple predicate. The set $\mathfrak{S}$ of simple predicates is a subset of $\mathfrak{R}$. $\mathfrak{S} \subset \mathfrak{R}$. (10b) and (10c) imply, $\varnothing, \square \in \mathfrak{S}$.

(12) If $A, B \in \mathfrak{R}$ and $AB = BA$, we say A and B are compatible and write $A \sim B$. The compatibility relation is by definition reflexive and symmetric but not necessarily transitive. (10b) and (10c) imply $A \sim \varnothing$ and $A \sim \square$ for all $A \in \mathfrak{R}$. If $A, B \in \mathfrak{S}$ and $A \sim B$, then $AB \in \mathfrak{S}$.

(13) If $A \sim B$ and $AB = A$ for $A, B \in \mathfrak{S}$, then (and only then) we say that A implies B and write $A \rightarrow B$. The implication relation is by definition reflexive [note $AA = A$ for $A \in \mathfrak{S}$], and transitive, but not necessarily symmetric. [Proof of transitivity: If $A \rightarrow B$ and $B \rightarrow C$, then $AC = ABC = AB = A = BA = CBA = CA$].

(14) If $A = B$, then $A \rightarrow B$ and $B \rightarrow A$, and if $A \rightarrow B$ and $B \rightarrow A$, then $A = B$. [Note: the equivalence ($=$) was defined in (6).]

(15) For all $A \in \mathfrak{S}$, $\varnothing \rightarrow A \rightarrow \square$. See (10b) and (10c).

(16) Definition. For given $A, B \in \mathfrak{S}$, C is said to be the conjunction of A and B and denoted $C = A \cap B$ if and only if $C \rightarrow A$, $C \rightarrow B$ and $X \rightarrow C$ whenever $X \in \mathfrak{S}$ and $X \rightarrow A$ and $X \rightarrow B$.

(17) Theorem (Existence of Conjunction). *Consider the infinite product* $C = \cdots ABAB$. *Then* $C \in \mathfrak{S}$, *and* C *satisfies the definition* (16) *of* $C = A \cap B$. *If* $A \sim B$, *then* $C = AB = BA$.

[*Proof.* Start with finite products of the types, $AB \cdots AB$, $BA \cdots BA$, $AB \cdots BA$, $BA \cdots AB$, and pass to the limit where all four become C. (See Ref. [1] for the details).

(18) $\neg A$ is defined by the same test as A, but when A is true $\neg A$ is false and when A is false $\neg A$ is true. If $A \in \mathfrak{s}$, then $\neg A \in \mathfrak{s}$.

(19) Hence $f(A/a) + f(\neg A/a) = 1$ for all $a \in \mathfrak{o}$. Due to (6), this equation defines $\neg A$ up to equivalence.

(20) From (9) and (19), $f(AB/a) + f(\neg AB/a) = f(B/a)$ for all $a \in \mathfrak{o}$ and all $A, B \in \mathfrak{s}$.

(21) By putting $B = A$ in (20), we get from (11) $\neg AA = \varnothing$. Similarly, $A\neg A = \varnothing$. Hence also $A \sim \neg A$.

(22) From the Definition (18) follows the law of double negation $\neg\neg A = A$.

(23) If $A \rightarrow \neg A$, then $A = \varnothing$. (Law of Self-Contradiction). [*Proof.* $A \rightarrow \neg A$ implies $A = A\neg A = \neg AA$ which according to (21) means that $A = \varnothing$.]

(24) From the Definition (18) follows: If $A \sim B$, then $A \sim \neg B$, $\neg A \sim B$, $\neg A \sim \neg B$.

(25) From (20) and (24) follows: If $A \sim B$, then $f(BA/a) + f(B\neg A/a) = f(B/a)$ for all $a \in \mathfrak{o}$.

(26) The relation $A \rightarrow B$ is equivalent to $A\neg B = \neg BA = \varnothing$.
Proof. If $A \rightarrow B$, $AB = BA = A$. But from (20) $f(A/a) = f(\neg BA/a) + f(BA/a)$ for all a. Hence $f(\neg BA/a) = 0$ for all a, i.e., $\neg BA = \varnothing$. Since $A \rightarrow B$ implies $A \sim B$, we have also $A \sim \neg B$ due to (24). Hence $f(A\neg B/a) = 0$, i.e., $A\neg B = \varnothing$ too. Conversely, $\neg BA = A\neg B = \varnothing$ implies $A \sim \neg B$, hence by (24) also $A \sim B$. We have then not only $f(A/a) = f(BA/a) + f(\neg BA/a) = f(BA/a)$ for all a due to (20) but also $f(A/a) = f(AB/a) + f(A\neg B/a) = f(AB/a)$ for all a. Hence $A = AB = BA$.]

(27) If $A \cap \neg B = \varnothing$ and $A \sim B$, then $A \rightarrow B$. If $A \rightarrow B$ then $A \cap \neg B = \varnothing$.

(28) If $A \rightarrow B$, then $\neg B \rightarrow \neg A$ (Law of Contraposition).
[*Proof.* $A \rightarrow B$, means $A\neg B = \neg BA = \varnothing$. Put $C = \neg A$ and $D = \neg B$ and we get $\neg CD = D\neg C = \varnothing$ due to the law of double negation. By (26), this means $D \rightarrow C$.]

(29) The disjunction $C = A \cup B$ is defined by (16) in which the direction of arrows is reversed. $C = A \cup B$ can be shown to be equivalent to $C = \neg(\cdots \neg A\neg B\neg A\neg B)$.

(30) We can prove the idempotent law, commutative law, associative law, absorptive law, de Morgan's law for the predicates belonging to $\mathfrak{S}$.

[*Proof.* The key is the definition of implication and the existence of the conjunction (17). Note that if the conjunction exists, it is unique due to (16) up to equivalence. (See Ref. [*1*] for the details).]

(31) The set $\mathfrak{S}$ of simple predicates is a complemented lattice (which is not necessarily distributive).

(32) If all members of $\mathfrak{S}$ are mutually compatible, then $\mathfrak{S}$ is distributive.

[*Proof.* By the use of (17), (19), (20), and (29) we can bring both $f[A \cup (B \cap C)/a]$ and $f[(A \cup B) \cap (A \cup C)/a]$ to $1 - f(\neg A \neg BC/a) - f(\neg AB \neg C/a) - f(\neg A \neg B \neg C/a)$ for all a. Under the same assumption, $f[A \cap (B \cup C)/a]$ equals $f[(A \cap B) \cup (A \cap C)/a]$ for all a.]

(33) If all members of $\mathfrak{S}$ are mutually compatible, then the characteristic function $f(A/a)$ as a function of A with a given a is a probability measure defined on $\mathfrak{S}$.

[*Proof.* By the use of (17), (19), (20), and (29), we can easily show that the axiom of probability is satisfied: $f(A \cap B/a) + f(A \cup B/a) = f(A/a) + f(B/a)$. In addition, we have (7).]

(34) When not all members of $\mathfrak{S}$ are mutually compatible, we do not have the situation described in (33). In spite of this, we require that there exist at least one particular object $g \in \mathfrak{O}$ such that $f(A/g)$ is a probability measure, i.e., $f(A \cap B/g) + f(A \cup B/g) = f(A/g) + f(B/g)$ for all $A, B \in \mathfrak{S}$ whether or not $A \sim B$.

(35) It is further required of f and g that if $A, B \in \mathfrak{S}$, $A \rightarrow B$ and $A \neq B$, then $f(A/g) < f(B/g)$. [For realizability of (34) and (35), see the next section.]

(36) (DEDEKIND'S THEOREM) [*6*]. *If a lattice is such that each element belonging to it can be assigned a weight function which satisfies the two relations mentioned in* (34) *and* (35), *then it is a modular lattice, i.e., if* A, B, *and* C *are its members and* $A \rightarrow C$, *then* $A \cup (B \cap C) = (A \cup V) \cap C$.

(37) $\mathfrak{S}$ is a modular lattice. The justification that can be given at this stage for the assumptions (34) and (35) is that they seem to represent the smallest possible generalization beyond the usual compatible case, (33), upholding the concept of probability as much as possible. The predicates belonging to $\mathfrak{S}$ obey the "modular logic."

3. MODIFIED CONCEPTS OF PROBABILITY AND TRUTH-SET

In this section we shall first see under what conditions the continuously-valued characteristic function can be interpreted in terms of a mixture of different object-types each of which having a binary-valued characteristic function. We shall show second that when this interpretation is not possible we have to assume that the act of observation alters the state of the object and a deviation from the usual logic is inevitable.

(38) If the members of $\mathfrak{s}$ are finite and all mutually compatible (i.e., if $\mathfrak{s}$ is a finite distributive lattice), there exists among the members a finite number of "atoms," $\alpha_1, \alpha_2, \cdots, \alpha_n$, such that $\alpha_i \cap \alpha_j = \varnothing (i \neq j)$, $\alpha_1 \cup \alpha_2 \cup \cdots \cup \alpha_n = \square$, and that $\varnothing \rightarrow A \rightarrow \alpha_i$ implies either $\varnothing = A$ or $A = \alpha_i$. Any member of $\mathfrak{s}$ can be expressed as a disjunction of some atoms, α_i. That is, $A = \bigcup_i^{\text{some}} \alpha_i$, where i runs over those atoms which imply A, $\alpha_i \rightarrow A$. [What is stated here is a well-known theorem about a finite distributive lattice.]

(39) The probabilities assigned [according to (33)] to all members of a finite distributive lattice can be derived from the probability of each atom. If $A = \bigcup_i^{\text{some}} \alpha_i$, then $f(A/a) = \sum_i^{\text{some}} f(\alpha_i/a)$, where the index i runs over the same range in the two expressions, i.e., over the subset of atoms such that $\alpha_i \rightarrow A$. [This follows from the probability axiom and the definition of atoms.]

(40) Consider a special object a_i such that $f(\alpha_i/a_i) = 1$ and $f(\alpha_j/a_i) = 0$, for $i \neq j$. This means that for any A we have $f(A/\alpha_i) = 1$, if $\alpha_i \rightarrow A$ and $f(\mathrm{A}/\alpha_i) = 0$ if not $\alpha_i \rightarrow A$. Suppose we now mix these object-types a_i, $i = 1, 2, \cdots, n$, in the ratio of $\rho_1 : \rho_2 : \cdots : \rho_n$, where $\rho_i \geqq 0$ and $\sum_{i=1}^{n} \rho_i = 1$. Then the probability of obtaining an affirmative result in the observation of α_i will be precisely ρ_i and the probability of obtaining an affirmative result in the observation of A is $\sum_i^{\text{some}} \rho_i$ where the index i runs over those atoms α_i such that $\alpha_i \rightarrow A$. The truth-set of α_i is a_i, and the truth-set of A is a mixture of a_i such that $\alpha_i \rightarrow A$.

(41) Combining (39) and (40), we conclude that in the case of a finite lattice consisting of mutually compatible predicates, we can interpret the nonbinary characteristic function of an object a as the result of a mixture of atomic object-types a_i in the ratio of $f(\alpha_i/a)$, whereby each object a_i [defined in (40)] is in a pure state, i.e., it has a binary characteristic function for any member of the lattice.

(42) Consider the effect of a compound observation AB (8) on a mixture defined in (41) in the case of a distributive lattice. The state right after the observation (with an affirmative result) of B, the state

of a passes to $a' = Ba$ according to (8). If the interpretation (41) is correct, among the atomic object-types a_i existing in a only those a_i such that $a_i \rightarrow B$ will be surviving after the affirmative observation of B. In the same way, after AB (with affirmative results for A and B), only those atomic object-types a_i will be surviving such that $a_i \rightarrow A$ and $a_i \rightarrow B$. This characterization of the state after AB is symmetrical with respect to A and B, hence it must be also true for the state after BA. This is in agreement with the relation $AB = BA$ which is true in a distributive lattice. In a word, observation in a distributive lattice has the effect of simple filtering, eliminating some existing components and retaining some others without changing them.

(43) In a non-distributive lattice, in general $AB \neq BA$. This means that the interpretation in terms of mixture and filtering does not work, implying that the idea of truth-set is untenable. This is in agreement with the fact that the distributive law breaks down in this case.

(44) The untenability of the idea of mixture implies also the untenability of the idea of filtering. As the effect of an affirmative observation of A the state changes from a to $a' = Aa$ as before, but in the non-distributive case this cannot be interpreted as an elimination and retention of some existing components. This point may become more convincing if we consider the value of $f(AB/a)$ when it is known that $f(A/a) = 0$. In the commutative case, $f(AB/a) = f(BA/a) = f(B/Aa)f(A/a) = 0$. But, in the non-commutative case, there is no reason why $f(AB/a)$ should become zero on the ground that $f(A/a) = 0$. The relation $f(AB/a) \neq 0$ implies $f(A/Ba) \neq 0$ because $f(AB/a) = f(A/Ba) \cdot f(B/a)$. In this case we have both $f(A/a) = 0$ and $f(A/Ba) \neq 0$. The state which negates A entirely is changed, as a result of the observation of B, to a state which does not negate A. This fact does not agree with our notion of elimination, retention or filtering but has to be interpreted as a change of the essential feature of the state. See (62) for more about this point.

(45) Within a modular lattice, there exist distributive sublattices. Hence by limiting our language to predicates of a single distributive sublattice, we can use the usual logic and the usual notion of probability. If we use $f(A/a)$ for the entire modular lattice, this function has to be considered as a generalization of the usual notion of probability. In the distributive case, $f(A/a) = 1$ means that a belongs to the truth-set of A, $f(A/a) = 0$ means that a belongs to the truth-set of $\neg A$, and $0 \neq f(A/a) \neq 1$ mean that the specification of a is so loose that a has to be

interpreted as a mixture of a member of the truth-set of A and a member of truth-set of $\neg A$. In the non-distributive case, $0 \neq f(A/a) \neq 1$ cannot be reduced to a mixture of two cases $f(A/a) = 0$ and $f(A/a) = 1$. In the last case, the concept of truth-set cannot be used and the logic has to be non-distributive.

4. GEOMETRICAL MODEL

The task of this section is to give a mathematical expression to the characteristic function $f(A/a)$ to substantiate that all that we have stated about predicates, objects and probability in the foregoing sections is free from internal contradiction.

(46) The clue is that the subspaces of any vector space constitute a modular lattice. We assume a real space for simplicity but we can easily generalize the formalism to a complex space. Hence, each member of $\mathfrak{s}$ will be represented by a subspace, the isomorphism being established by interpreting the conjunction as the largest common subspace contained by two subspaces and the disjunction as the smallest subspace which contain both subspaces. The convenient mathematical expression of a subspace is the projection operator corresponding to it. Hence, the mathematical expression of each member of $\mathfrak{s}$ will be a projection operator.

(47) Let $\mathfrak{M}(A)$ be the subspace corresponding to predicate $A \in \mathfrak{s}$, and $\mathcal{P}[\mathfrak{M}(A)]$ denote the projection operator corresponding to $\mathfrak{M}(A)$. If ξ_i, $i = 1, 2, \cdots, r$ are r orthogonal, normalized vectors which subtend the subspace $\mathfrak{M}(A)$, then $\mathcal{P}[\mathfrak{M}(A)]$ can be expressed by

$$\mathcal{P}[\mathfrak{M}(A)] = \sum_{i=1}^{r} \xi_i \xi_i^T$$

The dimension of $\mathfrak{M}(A)$ which is r can be computed by trace $\mathcal{P}[\mathfrak{M}(A)]$. We abbreviate sometimes $\mathcal{P}[\mathfrak{M}(A)]$ as $\mathcal{P}[A]$ and trace $\mathcal{P}[\mathfrak{M}(A)]$ as $D[A]$. Corresponding to $\square$ and $\varnothing$ which are members or $\mathfrak{s}$, we shall have $\mathcal{P}[\square] = \mathbb{1}$, $\mathcal{P}[\varnothing] = 0$ and $D[\square] = n$ and $D[\varnothing] = 0$, where n is the number of dimensions of the entire vector space.

(48) Implication $A \rightarrow B$ is translated as meaning that if $x \in \mathfrak{M}(A)$, then $x \in \mathfrak{M}(B)$, where x is a vector. The conjunction is interpreted as meaning $\mathfrak{M}(A \cap B) = \{x \mid x \in \mathfrak{M}(A) \text{ and } \mathfrak{M}(B)\}$. The disjunction is interpreted as meaning $\mathfrak{M}(A \cup B) = \{x \mid x = ay + bz,\ y \in \mathfrak{M}(A),\ z \in \mathfrak{M}(B),\ a, b: \text{real numbers}\}$. The negation is interpreted as meaning $\mathfrak{M}(\neg A) = \{x \mid x \perp y \text{ for all } y \in \mathfrak{M}(A)\}$, where $\perp$ means "perpendicular to." The compatibility $A \sim B$ means that there exists a set of com-

plete orthogonal coordinates, e_1, e_2, $\cdots$ e_n such that $\mathfrak{M}(A)$ as well as $\mathfrak{M}(B)$ is a subspace subtended by some of the e's.

(49) An object (in a particular state) is essentially characterized by the probabilities with which the test results will turn out. Hence, it is not surprising that it can be expressed also by some quantities definable in reference to the vector space we have introduced above. We cannot describe all the heuristic considerations which lead to the following results, but it is easy to see that if we introduce the mathematical expression of an object (in a particular state) in the following way, we can satisfy the required properties of the characteristic function. Let $\{e_1, e_2, \cdots, e_n\}$ be a set of complete, orthogonal unit vectors constituting a coordinate system, and let w_i, $i = 2, \cdots, n$ be a weight (probability) distribution, so that $w_i \geqq 0$ and $\sum_{i=1}^{n} w_i = 1$. Then the state of an object a is expressed by

$$Z(a) = \sum_{i=1}^{n} w_i \mathcal{P}[e_i]$$

The state of an object is determined by the coordinate system $\{e_i\}$ and the probability distribution $\{w_i\}$. Z is thus a nonnegative, symmetric matrix with trace one.

(50) The state of an object which consists of one term in the above formula is called a pure state and characterized by the fact

$$[Z(a)]^2 = Z(a)$$

(51) The characteristic function $f(A/a)$ can now be defined as

$$f(A/a) = \text{trace } (\mathcal{P}[A] \cdot Z[a])$$

(52) We can prove all the properties of $f(A/a)$ mentioned in the foregoing by the formula given by (51).

(53) It can be shown that, for a given a, the characteristic function thus defined $f(A/a)$ behaves like a probability, i.e., $f(A \cap B/a) + f(A \cup B/a) = f(A/a) + f(B/a)$, and $f(\square/a) = 1 - f(\varnothing/a) = 1$, insofar as we limit the predicates to a family of predicates which are mutually compatible.

(54) Whether or not the predicates considered are mutually compatible, there exists a special object g for which $f(A/g)$ behaves like a probability for all $A \in \mathfrak{s}$. See (34) above. This special object g is described by

$$Z(g) = \sum_{i=1}^{n} \cdot \mathcal{P}[e_i]/n = \mathbb{1}/n$$

(55) In the compatible case considered under (33) and (45), all A can be expressed in the form: $\mathcal{P}[A] = \sum_i^{\text{some}} \xi_i \xi_i^T$, where ξ_i belongs to one and the same coordinate system $\{\xi_i\}$.

(56) If the predicates under consideration are all mutually compatible, so that the lattice $\mathfrak{s}$ is distributive, the most precise description by a combination of available predicates corresponds to the atoms α_i of the distributive lattice $\mathfrak{s}$. See (38). Such an atomic predicate can be expressed by a single term in the summation appearing in (47). $\mathcal{P}[\alpha_i] = \xi_i \xi_i^T$.

(57) In the distributive case the atomic object a_i corresponding to atomic predicate α_i is characterized by $f(\alpha_i/a_j) = \delta_{ij}$. See (40). This object a_i is according to (51) expressible by the formula of (49) consisting of one term, $Z(a_i) = \mathcal{P}[\alpha_i] = \mathcal{P}[\xi_i]$. This is, according to (50), a pure state.

(58) In the distributive case, an object a which has the values of characteristic function $f(\alpha_i/a) = \rho_i$ for the atomic predicates can be expressed, according to (49) and (51), as $Z(a) = \sum_{i=1}^n \rho_i \mathcal{P}[\xi_i]$. This can be interpreted as a mixture of atomic objects a_i in the mixing ratio, $\rho_1 : \rho_2 : \cdots : \rho_n$, each atomic object a_i satisfying $f(A/a_i) = 0$ or 1 for all $A \in \mathfrak{s}$.

(59) This shows that in the distributive case, everything can be expressed by a single coordinate system $\{\xi_i\}$. This is not the case in the general non-distributive case.

(60) When the lattice $\mathfrak{s}$ is distributive and an object is described by the ρ_i, there is an alternative expression for Z, which is—insofar as observation is limited to $\mathfrak{s}$—equivalent to the one given in (58). This second alternative is $Z(a) = \mathcal{P}[e]$, where the vector e is defined by trace $(\mathcal{P}[e] \cdot \mathcal{P}[\xi_i]) = \rho_i$. If we allow ourselves to use observation of predicates outside the $\mathfrak{s}$, these two expressions of $Z(a)$ behave differently. The $Z(a)$ of this section is a pure state, while $Z(a)$ of (58) is not, in general.

(61) The state a of an object becomes after the affirmative test of A, $a' = Aa$ as defined in (9). All that has been stated about a' can be satisfied if we give the following mathematical expression to $a' : Z(a') = \mathcal{P}[A] Z[a] \mathcal{P}[A] / \text{trace}\{\mathcal{P}[A] \cdot Z(a)\}$. This formula is valid also in the non-distributive case when $\mathcal{P}[A]$ and $Z(a)$ have to be expressed in terms of different coordinate systems.

(62) In the distributive case, if we use $Z(a)$ of (58), the difference between $Z(a)$ and $\mathcal{P}[A] Z(a) \mathcal{P}[A]$ is only that some of the terms in the former is missing (filtered out) in the latter. In the non-distributive

case $Z(a)$ and $Z(a')$ have to use different coordinate systems if we express them in the form given in (49). This shows that $Z(a')$ is not just a portion of $Z(a)$, but entirely different in nature. Observation changes the object.

5. MODIFIED CONCEPT OF INFORMATION

In this section, we mention some of the elementary features of the generalized non-distributive information theory.

(63) The simplest information quantity S that can be defined in the non-distributive case is

$$S = -\text{trace}\,[Z(a)\,\log Z(a)]$$

This is equal to

$$S = -\sum_{i=1}^{n} w_i \log w_i$$

where w_i is given in (49), but if we write in this last form, we cannot see which coordinates the w's refer to.

(64) Most of the significant results in information theory can be mathematically derived from Gibbs' Theorem. Its generalization to our case is

$$-\text{trace}\,[Z(a)\,\log Z(a)] \leqq -\text{trace}\,[Z(a)\,\log Z(b)]$$

where equality holds if and only if

$$Z(a) = Z(b)$$

This last equation means not only equality of the w's but also equality of the coordinate systems in which $Z(a)$ and $Z(b)$ are diagonal. The proof of this generalized Gibbs' Theorem requires two lemmas

(65) LEMMA 1. (GIBBS' THEOREM) *If* $p_i \geqq 0$, $\sum_i p_i = 1$, $q_i \geqq 0$, $\sum_i q_i = 1$, for $i = 1, 2, \cdots, n$, *we have*

$$-\sum_i p_i \log q_i \geqq -\sum_i p_i \log p_i$$

where the equality holds if and only if $p_i = q_i$ *for all* i.

(66) LEMMA 2. (SIMPLE H-THEOREM) *Let* $\| A_{ji} \|$ *be a matrix such that* $A_{ji} \geqq 0$, $\sum_j A_{ji} = 1$, *and* $\sum_i A_{ji} = 1$. *If the probabilities* p_i *and* q_i *are so related that* $q_j = \sum_i A_{ji} p_i$, *we have*

$$-\sum_i q_i \log q_i \geqq -\sum_i p_i \log p_i$$

The equality holds if and only if those p_i's that belong to the same "terminally connected family" have the same value.

(67) DEFINITION. Two indices i and l are said to be terminally connected if there exists a sequence of indices $(i, \cdots, m, n, \cdots, l)$ in which each pair of consecutive indices, say, m and n, are such that there is at least one index j that satisfies both $A_{jm} \neq 0$ and $A_{jn} \neq 0$. A family of terminally connected indices is such that any two members of the family are terminally connected and a member and a non-member of the family are not terminally connected.

(68) *Proof of (64).* Let three states $Z(a), Z(b), Z(c)$ be defined as

$$Z(a) = \sum_i u_i \mathcal{P}[e_i]$$
$$Z(b) = \sum_j v_j \mathcal{P}[\xi_j]$$
$$Z(c) = \sum_j w_j \mathcal{P}[\xi_j]$$

where $w_j = \sum_i A_{ji} u_i$, $A_{ji} = (e_i \cdot \xi_j)^2 = (T_{ji})^2$. The expression $T_{ji} = (e_i \cdot \xi_j)$ means the scalar product of two vectors e_j and ξ_j. The fact that both coordinate-systems $\{e_i\}$ and $\{\xi_j\}$ are orthogonal coordinate-systems entails $T^T = T^{-1}$, which in turn entails the "double stochasticity": $\sum_j A_{ji} = \sum_i A_{ji} = 1$. From Lemma 1 follows (using the coordinate system of $\{\xi_i\}$)

$$-\text{trace } [Z(a) \log Z(b)] \geqq S[Z(c)].$$

From Lemma 2 further follows

$$S[Z(c)] \geqq S[Z(a)]$$

Combining these two results, we obtain the first part of (64). A careful study of the condition that equality should hold in both of these results leads to the conclusion that this happens if and only if $Z(a) = Z(b) = Z(c)$. See Ref. [1] for details.

(69) In the distributive case (i.e., in the usual information theory), the interdependence analysis in terms of entropy functions is based on two basic theorems. Suppose the object to be described consists of two parts of which part 1 can be described by index $i (i = 1, 2, \cdots, m)$ and part 2 can be described by index $j (j = 1, 2, \cdots, n)$ so that the total system can be described by a double index i, j. The probability p_i of i and the probability p_j of j can be derived from the probability p_{ij} of i, j, according to $p_i = \sum_j p_{ij}$ and $p_j = \sum_i p_{ij}$. Then we have two

theorems

$$\text{(I)} \qquad S(1) + S(2) \geqq S(1, 2)$$

and

$$\text{(II)} \qquad S(1, 2) \geqq S(1), \qquad S(1, 2) \geqq S(2),$$

where $S(1, 2)$, $S(1)$ and $S(2)$ are the entropies defined respectively by p_{ij}, p_i and p_j. Theorem I allows us to introduce a non-negative quantity

$$J(1, 2) = S(1) + S(2) - S(1, 2) \geqq 0$$

as a measure of the interdependence between part 1 and part 2. The equality holds if and only if part 1 and part 2 are probabilistically independent. Theorem II further allowed us to conclude $J(1, 2) \leqq S(1)$ and $J(1, 2) \leqq S(2)$ where the equality holds respectively when part 1 depends entirely on part 2 and when part 2 depends entirely on part 1. In the non-distributive case, we have an equivalent of Theorem I but not of Theorem II.

(70) First, a remark about the notation. The matrix elements of $Z(a)$ of (68) can be written in the coordinate-system $\{e_i\}$ as $[Z(a)]_{ii'} = u_i\delta_{ii'}$. This we write $(i \mid Z(a) \mid i') = u_i\delta_{ii'}$. The same matrix $Z(a)$ expressed in the coordinate system $\{\xi_j\}$ will have the elements: $(j \mid Z(a) \mid j') = \sum_i T_{ji}u_iT_{ij}^{-1}$. We consider the case where the index of the total system can be expressed by a double index i, l, where i refers to part 1 and l refers to part 2. The quantity in question will then be expressed as $(i, l \mid Z \mid i', l')$.

(71) Define

$$(i \mid Z_1 \mid i') = \sum_l (i, l \mid Z \mid i', l)$$

$$(l \mid Z_2 \mid l') = \sum_i (i, l \mid Z \mid i, l')$$

$$(i, l \mid Z' \mid i', l') = (i \mid Z_1 \mid i')(l \mid Z_2 \mid l')$$

Then by the use of the generalized Gibbs' Theorem, we get the next theorem.

$$\text{(72)} \qquad S(Z') = S(Z_1) + S(Z_2) \geqq S(Z),$$

The equality holding if and only if $Z = Z'$. This is the generalization of Theorem I above.

(73) If Z_1 is diagonal in the i-coordinate system for part 1 and if

there exists a l-coordinate system for part 2 such that the total Z becomes diagonal in the i-l-coordinate system, then $S(Z) - S(Z_1)$ is non-negative, corresponding to Theorem II. (See Ref. [1] for proof). But this is rather an unusual exception, and in general there is no theorem equivalent to Theorem II in the non-distributive case.

(74) This shows that part of the usual information theory can be recovered in the non-distributive, modular case but the information theory as a whole needs a reformulation. This reformulation is of course such that the usual theory can be rediscovered as a special case.

6. PATTERN RECOGNITION AND NON-DISTRIBUTIVE LOGIC

What follows is not the product of an intentional effort on the part of the author to force some domain of experience into the framework of a non-distributive logic, but it has been so to speak forced upon him during his experimentation in the domain of pattern recognition. He had to resort to a non-distributive logic in interpreting some results in pattern recognition. What distinguishes this case from the cases discussed in the foregoing sections is that we do not need to assume that the operation of observation has in general an effect of not only filtering but altering the state of the object. This situation might encourage those "realistic" interpreters of quantum mechanics who do not want to recognize the contraction of wave-packet by observation as a real effect. But, such an analogy between pattern recognition and quantum mechanics is impossible because in pattern recognition the "probability" $f(A/a)$ itself is directly measurable while in quantum mechanics each observation gives only a yes-or-no answer.

In pattern recognition, each object is usually expressed by an n-dimensional vector $\{x_1, x_2, \cdots, x_n\}$ which for simplicity we assume to be normalized $\sum x_i^2 = 1$. The suffix i here is not the label of a vector but the index of a component. In the standard type of problems, we are first given a certain number (ν_A) of samples (paradigms) of each class A, i.e., objects satisfying predicate A-ness. There could be two or more classes, $(A, B, C, \cdots)$. After the training period during which the paradigm vectors are shown, we are given in the application period a new vector whose class-belonging is not known, and we are required to place this newcomer into one of the classes which have been illustrated by the paradigms. The usual pattern recognition techniques assume that we can assign a zone (an n-dimensional volume) to each class in accordance with paradigms, so that the newcomer can be at-

tributed to the class in whose zone its vector falls. In other words, pattern recognition can be considered as a search for the extension of a predicate, when the name of the predicate (such as A-ness, B-ness, etc.) and some members of the extension are given as samples.

In many experimental cases, the present author was led to discover that this zone picture of a class is a very poor one and that the paradigm vectors in reality lie mostly in a subspace (usually passing the origin) of dimensions m which is considerably smaller than the total dimension as well as the number of paradigms: $m \ll n$, $m \ll \nu_A$. In other words, if $\{x_i^\alpha\}$, $\alpha = 1, 2, \cdots, \nu_A$ are paradigm vectors of class A, there exists a subspace whose projection operator is $\mathcal{P}[A]$ such that

$$\frac{\sum_{\alpha=1}^{\nu_A} (\mathcal{P}[A]x^\alpha)^2}{\nu_A} = \sigma$$

where σ can be made very close to unity (say, 0.95), with a very small number of dimensions: $m = \text{trace}\,(\mathcal{P}[A]) \ll n$ and $m \ll \nu_A$. The σ is the fraction of the total vector components that lie in the subspace, hence may be considered as a measure of the fidelity of this subspace picture of a class. When a newcomer $\{y_i\}$ without class-label arrives, we should compare $(\mathcal{P}[A]y)^2$ for different A's and assign the newcomer to the class which maximizes this quantity (which is the square of the cosine between the vector y and the subspace A). This subspace approach (called CLAFIC method) has been tested again and again and has proved to be quite successful [*9*, *10*].

Suppose we interpret the situation we envisage here as follows: A class of objects corresponds to a subspace in an n-dimensional representation space and any deviation therefrom is due to some kind of noise which inevitably occurs during processing. Then it is inevitable to conclude that $A \rightarrow B$ (A-ness implies B-ness) means that the subspace of A is a sub-subspace of the subspace of B, i.e., $\mathcal{P}[A]\mathcal{P}[B] = \mathcal{P}[B]\mathcal{P}[A] = \mathcal{P}[A]$. Assigning $\varnothing$ and $\square$ to the subspace of dimension zero and the entire representation space, this definition of implication is sufficient to conclude that the set of classes $A, B, C, \cdots$ constitute a modular lattice. This lattice is non-distributive insofar as there are subspaces involved which do not belong to a single family of orthogonal subspaces (subspaces spanned by some axes of one and the same orthogonal coordinate system). The negation of A can be represented by the largest subspace that is perpendicular to the subspace of A.

This is enough to let us conclude that the logic governing the predicates, A-ness, B-ness, $\cdots$, etc., is in general a modular logic which is not necessarily distributive. Since there is no restriction on the kind of predicates we can deal with, except that the objects satisfying these predicates can be submitted to measurements (or binary tests) of some sort, this conclusion has a very large domain of validity.

We can identify $(\wp[A]y)^2$ as $f(A/y)$, i.e.,

$$(\wp[A]y)^2 = f(A/y)$$

and interpret it as the probability of object y belonging to class A. This is mathematically identical with the formula given in (51), when the object a is in a pure state. This interpretation of $f(A/y)$ as a probability is perfectly acceptable with the same limitations that were explained in Section 3. The only difference is that in the present case the measurement of y gives directly the value of $f(A/y)$ provided the subspace of A is already known. The mathematical quantities of the type $f(AB/y)$ when $A \sim B$ is not the case have no clear interpretation here although $f(A \cap B/y)$ has.

The conclusion reached in this section is quite startling, because it would extend the domain that necessitates a non-distributive logic to the domain of ordinary propositions (such as this letter is A, that sound is "ah") of ordinary life. But we have to recognize that the whole argument in this section is based on two simplifying assumptions. First, if we make exactly $\sigma = 1$, the number of dimensions of each subspace becomes close to n, if not exactly n. Second, the vectors belonging to a class are found in a good approximation in a subspace, but it is not entirely certain that any arbitrary vector in this subspace conversely corresponds in reality to the class. In fact, it is confirmed that certain directions within a subspace are heavily represented by actual objects and some are not. According to our subspace picture, if we make the disjunction of A and B, the subspace corresponding to $A \cup B$ becomes the smallest subspace which includes both the subspace of A and the subspace of B. Hence, the subspace of $A \cup B$ contains vectors which belong neither to the subspace of A and the subspace of B. This is formally admissible due to the probability interpretation of the characteristic function but it is questionable whether all the vectors of the subspace of $A \cup B$ are inhabited by actual objects. In spite of these limitations, the subspace picture of concepts is perfectly tenable as a good approximation to

actuality, and a deviation from the distributive logic even on the level of ordinary language may not be dismissed as impossibility.[5]

7. BACKGROUND AND RELATED PROBLEMS

When Birkoff and von Neumann published in 1936 their paper [7] on what they called quantum logic, they depended strongly on the theoretical structure of quantum mechanics, that had been formulated (partly by von Neumann himself) in terms of the Hilbert space. It was K. Husimi [*11*] who tried as early as in 1937 to found the non-distributive modular logic directly on the experimental basis of atomic phenomena without relying on the Hilbert space used in quantum mechanics. The present paper can be said to be in the stream of thought started by Husimi. It was the present author's belief from the beginning that the modular logic would have a wide application outside atomic physics. It was in that spirit that he wrote in 1948 his article on the non-Boolean logic in a Japanese handbook of philosophy [*12*].

In 1956 he gave at the IBM Research Laboratory a series of lectures on quantum logic along this line, which was repeated at Yale Graduate School of Physics in 1959. The lecture notes were soon thereafter circulated quite widely, but they appeared in a printed form only in 1966 [*3*]. In these lectures, he introduced already the function $f(A/a)$ which is discussed more succinctly in the present paper. The central idea was that in order to reconstruct a logic free from the concept of truth-set we should rely on Peirce's principle ("there is one primary and fundamental logical relation, that is illation [implication]" [*13*]) and derive implication from a function of the type, $f(A/a)$. In the meantime, in 1959 and 1960, the author suggested that the modular logic may be useful in resolving some of the riddles of mind-body problem [4, 14]. In a paper presented at a conference on the philosophical problems in psychology in 1968, he mentioned six reasons for which a continuously-valued characteristic function of the type $f(A/a)$ should be taken as the starting point of a general discussion of logical problems. It may be of some interest to the readers of the present paper to repeat briefly three of them here. (1) According to the present day theory of probability, we have, as mentioned already

[5] In fact, most of the subspaces we obtained in experiments in pattern recognition belong approximately to a family of orthogonal subspaces, allowing application of the usual logic but some of them definitively do not, suggesting the existence of limitations to the usual logic.

in Section 1, to have first a Boolean lattice of propositions (or a sigma algebra) and then define probability (or measure) on it. But, it seems more natural and truthful to our human experience to derive logic from a probability-like function such as $f(A/a)$. Probability precedes logic. (2) About 20 years ago, an erroneous view was widespread that the neurons process information on a binary basis as in Boolean logic. But, now it is clear that information in the nervous system is encoded and processed in terms of continuous variables (such as the interval between two pulses). (3) In reductionistic efforts, it is usually assumed that the state on a higher level can be determined by the state on a lower level. Thus the psychological state is determined by the neurophysiological state, the biological state is determined by the physical state, etc. But, this is not true even in the simplest reductionist effort of explaining thermodynamics by atomic physics. Strictly speaking, the temperature T of a system is not completely determined by the energy of the microscopic state a. This means that $f(T/a)$ is not binary.

As we have shown in this paper, starting from a probability-like continuously-valued characteristic function, we can reconstruct a logic, but this logic is in general non-distributive, and perhaps modular in many cases, and it becomes distributive only in a rather exceptional case where compatibility of all predicates is guaranteed.

The brand of information theory the present author introduced 30 years ago [*15*] (i.e., 10 years before the ordinary information theory) was the general theory sketched above in Section 5, of which the usual information theory is a special case.

After conclusion of the present work, an interesting paper by P. A. Heelan [*16*] came to the present author's attention, in which he raises a legitimate objection against the usual version of "quantum logic." The notion of improper predicate and the notion of object as used in the present paper seem to be the right way to circumvent his objection.

RECEIVED: November 7, 1968; revised: May 19, 1969

REFERENCES

1. WATANABE, S., "Knowing and Guessing," Wiley, New York, 1969.
2. WATANABE, S., "Logic of the Empirical World," in *the Proceedings of the International Conference on Philosophical Problems in Psychology, Honolulu, Hawaii, 1968*, in preparation.
3. WATANABE, S., *Prog. Theoret. Phys. Suppl. Nos. 37-38*, 1966, 350. The original version of this paper was in the unpublished lecture notes of my course

"Physical Information Theory," given at the Yale Graduate School of Physics, Academic year 1959–60.

4. WATANABE, S., *Synthese*, **XIII**, 261 (1961).
5. ZADEH, L. A., *Inform. Control*, **8**, 338 (1965).
6. G. BIRKOFF, "Lattice Theory," Am. Math. Soc. Colloquium Publication, (1961).
7. BIRKOFF, G., AND VON NEUMANN, J., *Ann. Math. Ser. 2*, **37**, 823 (1936).
8. WATANABE, S., *in* (Bernays and Dockx, Eds.) "Information and Prediction in Science", Academic Press, New York, 1965.
9. WATANABE, S., *et al. in* (J. Tou, Ed.), "Computer and Information Science II", Academic Press, New York, 1966.
10. KAMINUMA, T., TAKEKAWA, T., AND WATANABE, S., *J. Pattern Recognition* (to appear).
11. HUSIMI, K., *Proc. Phys. Math. Soc. Japan*, **19**, 766 (1937).
12. WATANABE, S., Logic of Physics, *in* "Handbook of Philosophy," Kawade Shobo, Tokyo (1950).
13. PEIRCE, C. S., "Collected Papers," Vol. 4, Harvard Univ. Press, Cambridge, Mass., 1960.
14. WATANABE, S., *in* (S. Hook Ed.), "Dimensions of Mind," New York Univ. Press (1960).
15. WATANABE, S., *Z. Physik*, **113**, 482 (1939).
16. HEELAN, P. A., Quantum Logic and Classical Logic: Their Respective Roles, *Synthese*, (to appear shortly).