Invited paper
Proceedings of the 2006 IEEE World Congress on Computational Intelligence (IEEE WCCI 2006), Vancouver,BC,Canada, July 2006.

Valuations on Lattices and their Application to Information Theory

Kevin H. Knuth

***Abstract*— Valuations, or more specifically, bi-valuations, are functions that map two lattice elements to a real number. The zeta function $\zeta(\cdot,\cdot)$ is a very important function since it encodes inclusion on the lattice by setting $\zeta(x,y)=1$ if a lattice element *y* includes *x*, and 0 otherwise. This function can be generalized to a degree of inclusion, which induces a measure on the partial order. For distributive lattices, these degrees of inclusion follow a sum rule, a product rule and a Bayes' theorem analog. Applied to a lattice of logical statements, we recover Bayesian Probability Theory. The lattice of questions is dual to the statement lattice (in the sense of Birkhoff's Representation Theorem). Bi-valuations on the lattice of questions give rise to entropy, mutual information, and other higher-order informations suggested by several previous researchers. I show that this order-theoretic framework results in a natural generalization of information theory.**

I. Introduction

THE utility of Lattice Theory, also known as order theory, has become widely recognized as a basic mathematical field that introduces a new perspective on the concept of an algebra. While an algebra considers a set of elements along with a set of operations that map elements of the algebra on to one another, a lattice takes the perspective of a set of elements and an ordering relation, called a partial order. This partial order, denoted by $x \le y$ and read "*y* includes *x*", describes the situation where one element of the set "includes" a second element as defined by the ordering relation. A set of elements and an ordering relation give rise to what is called a partially ordered set, or *poset*. It is in the special case where each pair of elements in a poset possesses a unique least upper bound, denoted $x \vee y$, and a unique greatest lower bound, denoted $x \wedge y$, that the poset is called a *lattice*. For more information on posets and lattices, we refer the reader to the introductory text by Davey and Priestley [1].

This notion of inclusion, which is central to the partially ordered set, is encoded neatly by a function called the zeta function $\zeta(\cdot,\cdot)$, defined as

Manuscript received January 30, 2006. This work was supported in part by the College of Arts and Sciences and the College of Computing and Information of the University at Albany, and the NASA SISM IS Program.

K. H. Knuth is with the Department of Physics in the College of Arts and Sciences and the College for Computing and Information at the University at Albany, Albany, NY 12222 USA (phone: 518-442-4653; fax: 518-442-5260; e-mail: kknuth@albany.edu).

$$\zeta(x,y) = \begin{cases} 1 & \text{if} \quad x \le y \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

This function belongs to an important class of real-valued functions of two variables defined on the poset called the *incidence algebra* [2]. These functions $f(x,y)$ are non-zero only when $x \le y$, and can be multiplied by performing a convolution over the interval of elements *z* in the poset where $x \le z \le y$

$$h(x,y) = \sum_{x \le z \le y} f(x,z) g(z,y). \tag{2}$$

The inversion of the zeta function relies on the Möbius function, $\mu(x,y)$, which is the inverse of the zeta function [2,3,4] so that

$$\delta(x,y) = \sum_{x \le z \le y} \zeta(x,z) \mu(z,y), \tag{2}$$

where $\delta(x,y)$ is the Kronecker delta function. As one might expect, these functions are indeed related to the more familiar Riemann zeta function of number theory, which originates from the partially ordered set of integers ordered by the relation "divides". These functions play an important role in order theory, and will play an even greater role in the generalizations that we will discuss here.

In this paper we will consider a particular class of lattices called distributive lattices, which include Boolean lattices.

II. VALUATIONS AND DISTRIBUTIVE LATTICES

A. *Valuations and Measures*

A *measure m* typically refers to a function on a Boolean lattice *B* that takes a lattice element to a real number so that given $x \in B$, $m : x \to \mathbb{R}$. More generally, we define valuations as being functions *v* that take a lattice element to a real number such that given $x \in L$, $v : x \to \mathbb{R}$. The concept of valuations can be extended to include multiple lattice elements as arguments, such as the bi-valuation, which takes two lattice elements to a real number. The zeta function and Möbius functions above are two such examples. In this paper we will call all such functions valuations.

B. Valuations and the Lattice Structure

It is clear from the definition of the zeta function that the values that this function takes depend on the structure of the lattice. In fact, all valuations must obey the requirement that the mapping of lattice elements to real numbers must be consistent with the lattice structure. All lattices obey the following algebraic properties common to all posets

P1. For all a, $a \le a$ Reflexivity
P2. If $a \le b$, $b \le a$, then $a = b$ Antisymmetry
P3. If $a \le b$, $b \le c$, then $a \le c$ Transitivity

as well as the following properties common to all lattices

L1. $x \vee x = x$, $x \wedge x = x$ Idempotency
L2. $x \vee y = y \vee x$, $x \wedge y = y \wedge x$ Commutativity
L3. $x \vee (y \vee z) = (x \vee y) \vee z$ Associativity
$x \wedge (y \wedge z) = (x \wedge y) \wedge z$
L4. $x \vee (x \wedge y) = x \wedge (x \vee y) = x$ Absorption.

The greatest lower bound $x \wedge y$ can be viewed as an algebraic operation called the *meet*, and the greatest upper bound $x \vee y$ can be viewed as an operation called the *join*. The relationship between the algebraic operations meet and join, and the ordering relation of the lattice can be illustrated by what is called the *consistency relation*

$$x \le y \quad \Leftrightarrow \quad \begin{matrix} x \wedge y = x \\ x \vee y = y \end{matrix} . \qquad (2)$$

It should be noted that the meaning of the join and the meet depend on the particular ordering relation. For example, when the elements are logical statements and the ordering relation is logical implication, the join is the logical OR and the meet is the logical AND. When the elements are sets and the ordering relation is set inclusion, the join is the set union and the meet is the set intersection. Finally, when the elements are the set of positive integers and the ordering relation is divides, the join is the least common multiple and the meet is the greatest common divisor.

Lattices have a top element called the top, which is denoted by $\top$, so that $x \vee \top = \top$ for all $x \in L$, and dually an element called the bottom, denoted by $\perp$ where $x \wedge \perp = \perp$ for all $x \in L$.

Lattices can have elements that cannot be written as the join of any two elements, so that for such elements $z \in L$ there do not exist an $x \in L$ and a $y \in L$ such that $z = x \wedge y$. Such elements, called *join-irreducible elements*, play an important role in lattice theory.

Now given a valuation $v : x \to \mathbb{R}$, and the values $v(x)$ and $v(y)$ for $x \in L$ and $y \in L$, we may wish to know the value for the join of x and y, $v(x \vee y)$, or the meet of x and y, $v(x \wedge y)$. It turns out that there is a unique consistent way to do this [5]

$$v(x) + v(y) = v(x \wedge y) + v(x \vee y) , \qquad (3)$$

which, when re-written may be recognized as Gian-Carlo Rota's famous *inclusion-exclusion principle*

$$v(x \vee y) = v(x) + v(y) - v(x \wedge y) . \qquad (4)$$

This principle appears over and over in mathematics: for example, in probability theory

$$p(x \vee y \mid t) = p(x \mid t) + p(y \mid t) - p(x \wedge y \mid y), \qquad (5)$$

where $x \wedge y$ represents the logical AND of two statements, and $x \vee y$ represents the logical OR, in information theory

$$MI(X,Y) = H(X) + H(Y) - H(X,Y), \qquad (6)$$

where *MI* is the mutual information, *H* is the entropy, and even in straightforward integer relations as Rota [3] demonstrates from Pölya and Szegö's work [6]

$$\max(x, y) = x + y - \min(x, y) . \qquad (7)$$

This is a direct result of the associativity of the underlying lattice structure [7, 8, 9].

C. Distributive Lattices

In this paper we will focus on what are called *distributive lattices*, which satisfy the following distributive properties in addition to the previous properties P1-P3 and L1-L4

D1. $x \wedge (y \vee z) = (x \wedge y) \vee (x \vee z)$ Distributivity $\wedge$ over $\vee$
D1. $x \vee (y \wedge z) = (x \vee y) \wedge (x \vee z)$ Distributivity $\vee$ over $\wedge$

Any distributive lattice can be expressed as a lattice of sets ordered by set inclusion.

Boolean lattices are distributive lattices with an additional property of *complementation*, where for every element $x \in L$, there exists a unique element $\sim x \in L$, such that

C1. $x \vee {\sim}x = \top$ Complementation
C2. $x \wedge {\sim}x = \perp$

In short, Boolean lattices are complemented distributive lattices. Since the number of elements of a Boolean lattice goes as 2^N, diagrams with $N > 3$ get quite unruly. Figure 1 shows a Boolean lattice formed from three atomic elements (join-irreducible elements, which cover $\perp$). In this example the set of elements are logical statements ordered by logical implication so that *y* includes *x*, $x \le y$,

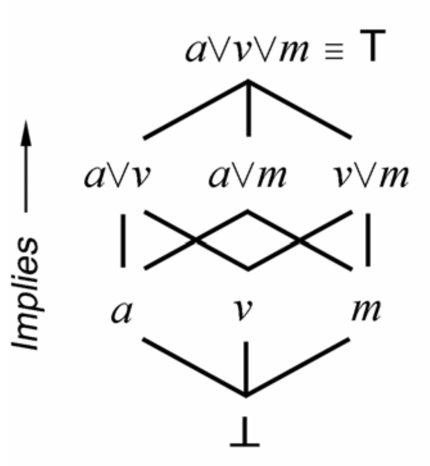

Fig. 1. A Boolean lattice of logical statements. The three atomic statements *a*, *v*, *m*, stand for animal, vegetable, and mineral. The elements are ordered by logical implication, and the top element, the truism is the statement "It is an animal, vegetable, or a mineral!" The bottom statement is the absurdity formed by a logical AND of any pair of atomic statements, such as "It is an animal and a

represents $x \to y$ with the atomic elements being the statements: [1]

a = "It is an animal!"
v = "It is a vegetable!"
m = "It is a mineral!"

The Boolean lattice consists of the power set of this set of three atomic elements, which is the set of all possible joins ordered by logical implication.

D. Generalizing the Zeta Function

We now consider a straightforward generalization of the zeta function (1). The motivation for this will become apparent.

First we define its dual [9]

$$\zeta^{\partial}(x, y) \;=\; \begin{cases} 1 & \text{if} \quad y \le x \\ 0 & \text{otherwise} \end{cases}, \tag{8}$$

which is the original zeta function, but with the inclusion condition turned around. This function simply indicates whether the lattice element *x* includes the element *y*. We now introduce the generalization by defining the function $z(\cdot,\cdot)$ expanding on the "otherwise" case

$$z(x, y) \;=\; \begin{cases} 1 & \text{if} \quad y \le x \\ 0 & \text{if} \quad x \wedge y = \perp \\ z \in (0,1) & \text{otherwise} \end{cases} \tag{9}$$

where inclusion on the lattice is now generalized to *degrees of inclusion*, which are represented by a real number in the interval $[0,1]$.[2] This function is defined so that if we are certain that *x* includes *y*, then it returns a value of 1; however, if *x* does not include *y* this function can encode the *degree* to which *x* includes *y*. This extends the incidence algebra to a calculus. The particular values that the function $z(\cdot,\cdot)$ takes in a given case will be discussed in the next section.

This bi-valuation is required to follow Rota's inclusion-exclusion principle, which is a consequence of associativity of the lattice. In general for a distributive lattice, one can show that the sum rule can be written as [8, 9]

$$\begin{aligned} & z(x_1 \vee x_2 \vee \ldots \vee x_n, t) \\ &\quad = \sum_i z(x_i, t) - \sum_{i<j} z(x_i \vee x_j, t) + \\ &\qquad \sum_{i<j<k} z(x_i \vee x_j \vee x_k, t) - \ldots \end{aligned} \tag{10}$$

Distributivity gives rise to the product rule [7, 8, 9]

$$z(x \wedge y, t) \;=\; z(x, t)\, z(y, x \wedge t)\,, \tag{11}$$

and combined with commutativity, we obtain Bayes' Theorem

$$z(y, x \wedge t) \;=\; \frac{z(y, t)\, z(x, y \wedge t)}{z(x, t)}. \tag{12}$$

III. Bayesian Inference from Valuations

The dual of the zeta function (8) indicates implication in a Boolean lattice of logical statements. The generalization to the z-function allows us to quantify degrees of implication. This generalization is quite useful. Even though the truism does not imply the statement *a* = "It is an animal!", we can use the defined bi-valuation to compute *the degree to which the truism implies* that "It is an animal!". In practical situations, this generalization is useful indeed!

The fact that the z-function quantifies degrees of implication, and is manipulated using the familiar sum rule, product rule and Bayes' Theorem, suggests that we have obtained an order-theoretic derivation of the notion of probability. With a simple change in notation, Bayesian probability theory can be recovered by defining $p(x \mid y) \;\equiv\; z(x, y)$, such that [9]

$$p(x \mid y) \;=\; \begin{cases} 1 & \text{if} \quad y \to x \\ 0 & \text{if} \quad x \wedge y = \perp \\ p \in (0,1) & \text{otherwise} \end{cases} \tag{13}$$

[1] Here we use the notation introduced by Robert Fry where a logical statement is indicated with an exclamation mark.

[2] Note that we could have defined the value of z to be over any finite real interval. We have chosen the unit interval here for convenience, but we should note that the product rule depends on the maximal value of this interval [9].

where it is understood that the area of discourse is a lattice of logical statements ordered by implication. This lattice becomes our hypothesis space, and its structure, once hidden by the algebra, is now made explicit by the perspective of lattice theory. The sum and product rules remain intact, and are now seen to be required by consistency with the structure of the hypothesis space. Any other rules for manipulating such degrees of implication would violate the properties P1-P3, L1-L4, and D1. It is important to note that this generalization is not perfect and that property D2 has been sacrificed [9, 10].

The view of probability as degrees of belief represented by real numbers goes back to 1946 with Richard Cox who derived the sum and product rules in a similar way by requiring that these degrees of belief be consistent with the rules of Boolean algebra [10, 11, 12]. While Cox relied on complementation to obtain the sum rule and associativity to obtain the product rule, the present work follows Caticha [7] and relies on associativity to obtain the sum rule and distributivity to obtain the product rule. This development is more general in that it applies to valuations on *all distributive lattices*, and is not restricted to the arena of probability theory. [3]

We are, of course, left with the dilemma of probability assignments, which we find to be outside the realm of discourse of probability theory as delineated by the sum rule, product rule and Bayes' Theorem. Those who view this as a limitation of the theory, may be comforted by the fact that there is a theorem by Rota that states

Theorem 1 *Assigning Valuations (Rota [3, Theorem 1, Corollary 2, p. 35]).* A valuation in a finite distributive lattice is uniquely determined by the values it takes on the set of join-irreducibles of *L*, and these values can be arbitrarily assigned.

In short, the constraints imposed by the lattice structure, or equivalently the algebra, do not influence the possible values that a valuation assigns to the join-irreducible elements of the lattice. With reference to our example in Fig. 1, the probability that the object considered is an animal given the fact that we know that the object must be an animal, a vegetable, or a mineral, $p(a \mid \top)$, can be assigned any value, and the calculus will suffer no contradictions. How then do we assign such values? We must take into account other consistency principles relevant to the problem, such as symmetry and constraints. Some of this groundwork has already been laid by Jaynes who introduced the *Principle of Group Invariance* [14] and the *Principle of Maximum Entropy* [15].

[3] It is worth noting that this generalized framework shows great promise in understanding quantum mechanics in that the sum and product rules for wavefunction amplitudes also derive from associativity and distributivity of experimental setups [7, 8]; although, lack of commutativity prohibits a Bayes' theorem analog.

IV. The Lattice of Questions

In his last scientific publication, Richard Cox explored the logic of inquiry. In this paper he defined a question as a set of all the possible statements that answer it [16]. At first glance, such a statement appears formidable, as most of the questions we ask can be answered by a multitude of statements. However, it is a reasonable definition, as long as one remembers that something like "2" is not an answer, but that an answer is a logical statement "Two inches of rain fell in Albany today!", and these logical statements are tied to topics of discourse, which are represented by hypothesis spaces. Cox's definition is brilliant, since given a hypothesis space of possible answers, one can construct questions, and compare their equivalence by comparing their sets of answers. For instance, the questions "Is it raining?" and "Is it not raining?" are equivalent since they are both answered by the same set of statements.

It is clear from Cox's definition that a question is a set, and as such, they form a distributive lattice when ordered by set inclusion. In the language of lattice theory, this is the definition of a down-set [1]

Definition 2 *Down-set*
A *down-set* is a subset *J* of ordered set *L* where if $a \in L$, $x \in L$, $x \le a$, then $x \in J$. Given an arbitrary subset *K* of *L*, we write the down-set formed from *K* as $J \;=\; \downarrow K \;=\; \{y \in L \mid \exists\, x \in K \text{ where } y \le x\}$.

Surprisingly, the ordering relation of set inclusion, denoted $\subseteq$ and read "is a subset of", implements the concept of answering. That is, if question *A* is a subset of question *B*, $A \subseteq B$, then by answering question *A* we will have necessarily answered question *B*. This gives us a mathematically attractive definition for a question

Definition 3 *Question*
A question *K* is a down-set of a lattice of logical statements, and the lattice of all possible questions can be constructed from the set of all possible down-sets of statements ordered by the usual set inclusion.

That is, given a lattice of logical statements *A*, such as the example presented in Fig. 1, one can construct the lattice of questions as the set of all possible down-sets of *A* ordered by the usual set inclusion. Lattices constructed from the ordered sets of down-sets of a Boolean lattice are called *Free Distributive Lattices* [1, 17], and specifically the lattice B3 gives rise to FD3 shown in Fig. 2 [18]. Because the ordering relation is that of set inclusion, the join of two questions (called the disjunction) is the set union, and the meet (called the conjunction) of two questions is the set intersection.

In this figure, I employ a simplifying notation where

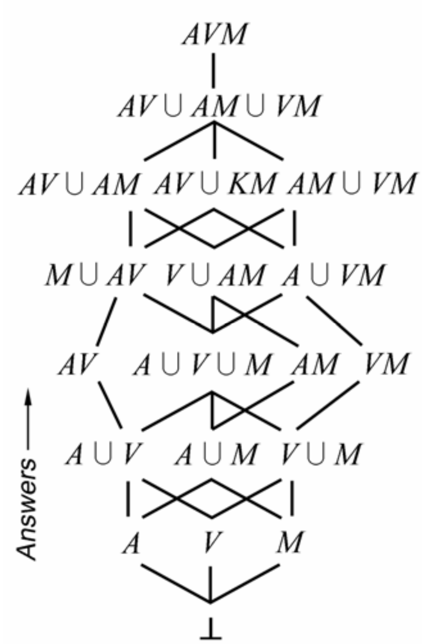

Fig. 2. The lattice of all possible questions formed from the Boolean lattice of logical statements in Fig. 1. Questions are ordered by set inclusion, which implements the notion of answering. Answering questions lower in the lattice necessarily answers the questions above them. Of special note is the central issue $A \cup V \cup M$, which resolves all the real questions above it.

$$A = \{a, \bot\}$$
$$AV = \{a \vee v, a, v, \bot\}$$
$$AVM = \{a \vee v \vee m, a \vee v, a \vee m, v \vee m, a, v, m, \bot\}$$

and elements like

$$A \vee M = \{a, m, \bot\}$$
$$AV \vee M = \{a \vee v, a, v, m, \bot\} \equiv \{a \vee v, a, v, \bot\} \cup \{m, \bot\}$$

represent joins (set unions) of questions. Note that $A \vee M \neq AM$. It is important to be aware that these are mathematical constructs, and that not all questions defined mathematically correspond with natural questions that one might ask, just as all integers do not correspond to the amount of money in your pocket. However, above we see one familiar example, the question $AV \vee M$, which can be read as "Is it or is it not a mineral?" Such a question can be answered by the set of all statements that it contains, where one might read $a \vee v$ as "It is not a mineral!" rather than "It is an animal or a vegetable!" Also, note that all questions can be answered by the absurdity. This feature is inherited from the fact that in Boolean logic the absurdity (or a false statement) implies everything.

The join-irreducible elements of the question lattice, such as A, AM, and AVM form a sub-lattice that is isomorphic to the Boolean lattices of assertions. These are special down-sets called *ideals*, and for this reason I have called them *ideal questions* [18, 19, 9]. They are ideal from a mathematical perspective, since writing down all possible joins of the ideals allows you to construct the full lattice. However, from a practical perspective, these questions do not allow a wide array of answers and are not useful in discourse. With the exception of the top question, which is answered by every statement, the other ideal questions, which do not consider all the possible mutually exclusive atoms of the statement lattice as answers, may not be answerable by a true statement.

Questions which include all atomic statements as potential answers are assured to be answerable by a true statement. Cox called such questions *real questions* [16]. There are of course some very special questions. One such question is the *central issue* [9], denoted by *I*, which is the minimal real question. In our example it is the question

$$I = A \vee V \vee M = \{a, v, m, \bot\},$$

which asks "Is it an animal, a vegetable or a mineral?", but does not accept an ambiguous result such as "It is either an animal or a vegetable!" Given our Boolean statement lattice, this is the issue we want to resolve. Only one of the three possible atomic statements can be true, and this question cuts straight to the heart of the matter. Answering this question necessarily answers all other real questions (above it) in the lattice.

Unfortunately, we are not always allowed to ask such straightforward questions, and this has led to human activities such as games and scientific inquiry, which we will discuss later. Often in games, we are only allowed to ask yes-or-no questions, which I call *binary questions*. We explicitly mentioned one earlier, the question $AV \vee M$, which can be read as "Is it or is it not a mineral?"

Last, it is useful to identify *partition questions* as questions that neatly partition the set of answers. In our example there are five partition questions AVM, $A \vee VM$, $V \vee AM$, $M \vee AV$, $A \vee V \vee M$, which together form a lattice isomorphic to the partition lattice or three elements.

There is much that can be said about questions, and I would direct the interested reader to the following papers [16, 18, 19, 20, 9] for more details.

V. Valuations on Questions and Information Theory

It is now a straightforward matter to generalize the order-theoretic concept of answering, which is implemented by set inclusion, to a more flexible measure of the degree to which one question answers another. Such a generalization is extremely useful since even though the binary question $AV \vee M$ does not answer the central issue $A \vee V \vee M$ (since $A \vee V \vee M \subseteq AV \vee M$), we could compute the *degree* to which it resolves the central issue. For example, $AV \vee M$ may be a better question to ask than either $V \vee AM$ or $M \vee AV$.

We begin by restricting ourselves to the lattice of real questions (Fig. 3), which are the questions that include the central issue, $I \subseteq Q$. We also discard the join-irreducible question at the top of the lattice. In this example, the lattice is Boolean, however, for larger lattices this is not the case

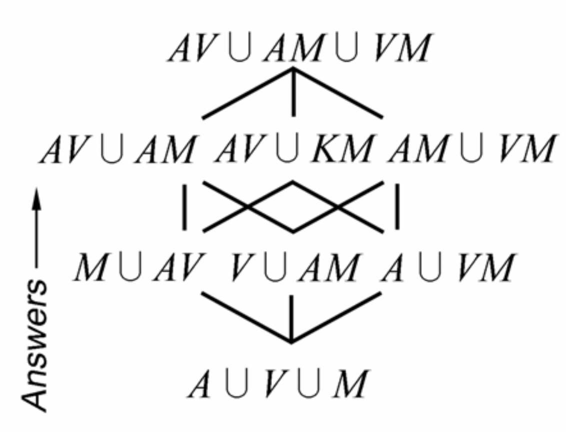

Fig. 3. The lattice of real questions. The bottom of the lattice is the central issue *I*; answering the central issue answers all other real questions. The relevance allows one to compute the degree to which a question in this lattice resolves the central issue, or any other question. The join-irreducible elements are the partition questions, and again, by assigning the relevance to these questions, we can use the calculus to compute any other relevance.

[18]. We again generalize the zeta function for this lattice by defining a bi-valuation called the relevance [9],

$$d(I \mid Q) \;=\; \begin{cases} 1 & \text{if} \quad Q \subseteq I \\ 0 & \text{if} \quad Q \wedge I = \perp \\ d \in (0,1) & \text{otherwise} \end{cases} \tag{14}$$

which describes the degree to which question *Q* resolves issue *I*. This quantity, originally suggested by Richard Cox [16] and further studied by Robert Fry [21] under the name of bearing, is presented here in a very different context as a generalized zeta function for the question lattice. The calculus assures us that computations performed using the sum rule, the product rule, and Bayes' Theorem analog will be consistent with the underlying distributive algebra. However, as always the real trick is to assign the valuations on the join-irreducible elements in a manner that is consistent with the problem of interest.

In real question lattice, the join-irreducible elements are the partition questions, which in this example are the five questions AVM , $A \vee VM$, $V \vee AM$, $M \vee AV$, $A \vee V \vee M$. Note that they cannot be formed as the join of any other two questions. The statements that answer these questions are neatly partitioned, and the set of greatest statements that answer them can be read directly from the notation. For example, $A \vee VM$ is answered by the two statements a , $v \vee m$, and all the statements that imply them. These two statements form an *irreducible set*, and the down-set of this irreducible set defines the question.

Since the relevance of a question ought to depend on the probability of its answers, we make a simple assumption. The degree to which a partition question resolves the central issue is a function of the probabilities of the statements comprising its irreducible set, so that

$$d(I \mid P) = K_n\big(p(x_1 \mid \top),\, p(x_2 \mid \top),\, \ldots,\, p(x_n \mid \top)\big)\,, \tag{15}$$

where $H_n(\cdot)$ are functions to be determined.[4] Due to the structure of the distributive lattice, the relevance satisfies four important properties: *additivity*, and *subadditivity*, which are assured by the sum rule, *symmetry* with respect to disjunction, and *expansibility*, which describes how the lattice collapses when a statement is found to be false. An important result from Aczél et al. [22] states that the unique form of H_n is a linear combination of the Shannon and Hartley entropies

$$K_n(p_1, p_2, \ldots, p_n) \;=\; a\, H_n(p_1, p_2, \ldots, p_n) \;+\; b\; {}_oH_n(p_1, p_2, \ldots, p_n) \tag{16}$$

where a and b are arbitrary non-negative constants, and the Shannon entropy [23] is defined as

$$H_n(p_1, p_2, \ldots, p_n) \;=\; -\sum_{i=1}^{n} p_i \log_2 p_i \tag{17}$$

and the Hartley entropy [24] is defined as

$${}_oH_n(p_1, p_2, \ldots, p_n) \;=\; \log_2 N(P) \tag{18}$$

where $N(P)$ are the number of non-zero arguments p_i. An additional result found by Aczél [22] states that the Shannon entropy is the unique solution if the result is to be small for small probabilities so that the relevance varies continuously with the probability. According to our definition, the relevance must be normalized. To accomplish this, we set the non-negative constant a in (16) equal to the entropy of the central issue, so that

$$a \;=\; -\sum_{i=1}^{N_A} p_i \log_2 p_i \;, \tag{19}$$

where N_A is the number of atomic statements in the statement lattice, and the p_i's are their respective probabilities. The degree to which a partition question resolves the central issue is then

$$d(I \mid P) \;=\; a\, H_n(p_1, p_2, \ldots, p_n) \tag{20}$$

where n is the number of statements in the partition question's irreducible set, and the p_i's are their respective probabilities. It follows simply that

$$d(I \mid I) \;=\; 1 \tag{21}$$

as required.

[4] Note that the ⊤ in $p(x|\top)$ refers to the top statement, the truism.

Bayes' Theorem can be employed to move questions from one side of the solidus to the other by relying on the top question of the lattice, which I will call T, since it is not the original top

$$\begin{aligned} d(A\,|\,B) &= d(A\,|\,B \wedge T) \\ &= d(B\,|\,A \wedge T)\,\frac{d(A\,|\,T)}{d(B\,|\,T)} \\ &= d(B\,|\,A)\,\frac{d(A\,|\,T)}{d(B\,|\,T)} \end{aligned} \tag{22}$$

In the case where A is the central issue, this simplifies since $d(B\,|\,I)=1$

$$\begin{aligned} d(I\,|\,B) &= d(I\,|\,B \wedge T) \\ &= d(B\,|\,I \wedge T)\,\frac{d(I\,|\,T)}{d(B\,|\,T)} \\ &= d(B\,|\,I)\,\frac{d(I\,|\,T)}{d(B\,|\,T)} \\ &= \frac{d(I\,|\,T)}{d(B\,|\,T)} \end{aligned} \tag{23}$$

The sum rule can be used to generate the relevance for more complex questions

$$d(A \vee B\,|\,C) = d(A\,|\,C) + d(A\,|\,C) - d(A \wedge B\,|\,C), \tag{24}$$

which was noted by Robert Fry to be related to (6), which is the equation for the mutual information [21, 9]. The result that the relevance is the Shannon entropy combined with the generalized sum rule verifies the form of Cox's *generalized entropy* [11, 16]. It should also be noted that it is straightforward to reproduce a number of earlier results such as McGill's *multi-information* [25], Han's *entropy space* [26], and Bell's *co-information lattice* [27]. Last, it is interesting that consistency with the structure of the distributive lattice rules out non-logarithmic entropies, such as Renyi and Tsallis entropies for use in this particular domain.

VI. Extending Information Theory

A. The Game of Twenty Questions

Almost everyone is familiar with the Game of Twenty Questions, which is often played on rainy days or long car rides. The idea of the game is that one person thinks of something and the other person has to guess what they are thinking by asking at most twenty questions. It would be no fun if we were allowed to ask the *central issue* "What are you thinking?" Instead, we are restricted to asking only binary questions: "Is it or is it not colored red?", "Is it or is it not larger than an elephant?", "Is it or is it not a mineral?" ($M \vee AV$). Each of these questions has the potential to provide information relevant to the central issue, with some questions being more relevant than others.

If we restrict ourselves to a finite hypothesis space, we can easily perform these calculations without explicitly describing either lattice. We simply compute the degree to which a considered binary question B resolves the central issue I. This is straight-forward since all binary questions are partition questions

$$d(I\,|\,B) = a\,H_n(p_1, p_2, \ldots, p_n) \tag{25}$$

This answer is proportional to the entropy of the binary question, and specifically for the question $M \vee AV$ we have

$$d(I\,|\,M \vee AV) = -a\left(p_m \log_2 p_m + p_{a\vee v} \log_2 p_{a\vee v}\right)$$

Again, the relevance depends on the probabilities we assign, and the fact that your friend is a geologist may mean that the question, $M \vee AV$, "Is it or is it not a mineral?" may be much more relevant than if you were playing with a botanist.

B. Experimental Design

Games such as The Game of Twenty Questions are fun because we impose constraints on the questions we are allowed to ask. However, the world around us imposes natural constraints. We perceive the world through our senses, and each perception begins as a question. Any physiologist will tell you that at the very root the questions are extremely basic: "Was that rhodopson molecule excited by a photon?", or "Did a particular hair cell in my inner ear bend sufficiently to distort a particular protein in the cell membrane?"

To overcome our human limitations, we build scientific instruments, but these too ask very limited questions. More often than not, these questions do not neatly partition the possible answers. For example a spectrometer may measure the quantity of green light at 530 nm emitted by the object. What does this tell us about whether the object is an animal, a vegetable, or a mineral? We are currently working to further develop this formalism to perform such computations.

Acknowledgments

The author would like to thank John Skilling, Philip Goyal, Steve Gull, Ariel Caticha, Carlos Rodríguez, and Janos Aczél for inspiring discussions, invaluable remarks, and comments, and much encouragement.

References

[1] B.A. Davey, H.A. Priestley, *Introduction to Lattices and Order*, 2nd Edition, Cambridge:Cambridge University Press, 2002.

[2] G.-C. Rota, "On the foundations of combinatorial theory I. Theory of Möbius functions." *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 2, 1964, pp. 340–368.

[3] G.-C. Rota, "On the combinatorics of the Euler characteristic," in *Studies in Pure Mathematics (Presented to Richard Rado)*, London:Academic Press, 1971, pp. 221–233.

[4] M. Barnabei and E. Pezzoli "Gian-Carlo Rota on combinatorics," in *Möbius functions*, J.P.S. Kung, Ed.., Boston:Birkhauser, 1995, pp. 83–104.

[5] D.A. Klain and G.-C. Rota, *Introduction to Geometric Probability*, Cambridge:Cambridge University Press, 1997.

[6] G. Pölya and G. Szegö, *Aufgaben und Lehrsätze aus der Analysis*, vol. 2, Berlin:Springer, 1964, p. 121.

[7] A. Caticha, "Consistency, amplitudes and probabilities in quantum theory," *Phys. Rev*., A57, 1998, pp. 1572-1582.

[8] K.H. Knuth, "Deriving laws from ordering relations," in *Bayesian Inference and Maximum Entropy Methods in Science and Engineering, Jackson Hole WY, USA*, G. J. Erickson, Ed. Melville NY:AIP, 2003, pp. 204-235.

[9] K.H. Knuth, "Lattice duality: The origin of probability and entropy. *Neurocomputing*, vol. 67C, 2005, pp. 245-274.

[10] R.T. Cox, "Probability, frequency, and reasonable expectation", vol. 14, 1946, pp. 1-13.

[11] R.T. Cox, *The Algebra of Probable Inference*, Baltimore:Johns Hopkins Press, 1961.

[12] E.T. Jaynes, *Probability Theory: The Logic of Science*, Cambridge:Cambridge University Press, 2003.

[13] K.H. Knuth, "Explorations in hypothesis space," unpublished.

[14] E.T. Jaynes, "Prior probabilities," *IEEE Trans. System Sci. Cyb.*, vol. SSC-4, 1968, p. 227.

[15] E.T. Jaynes, "Where do we stand on maximum entropy?," in: *The Maximum Entropy Formalism*, R.D. Levine, M. Tribus, Eds., Cambridge:MIT Press, 1979, pp. 15–118.

[16] R.T. Cox, "Of inference and inquiry," in: *The Maximum Entropy Formalism*, R.D. Levine, M. Tribus, Eds., Cambridge:MIT Press, 1979, pp. 119–167.

[17] G.D. Birkhoff, *Lattice Theory*, Providence:American Mathematical Society, 1967.

[18] K.H. Knuth, "What is a question?" in *Bayesian Inference and Maximum Entropy Methods in Science and Engineering, Moscow ID 2002, AIP Conference Proceedings,* vol. 659, C. Williams, Ed., Melville NY:AIP, 2002, pp. 227-242.

[19] K.H. Knuth "Intelligent machines in the 21st century: Automating the processes of inference and inquiry," *Phil. Trans. Roy. Soc. Lond. A, Triennial Issue*, vol. 361(1813), 2003, pp. 2859-2873.

[20] K.H. Knuth Toward question-asking machines: The logic of questions and the inquiry calculus. *Tenth International Workshop on Artificial Intelligence and Statistics (AISTATS 2005), Barbados*, 2005.

[21] R.L. Fry, "The engineering of cybernetic systems," in Bayesian Inference and Maximum Entropy Methods in Science and Engineering, Baltimore MD, USA, August 2001, R. L. Fry, Ed., Melville NY:AIP, 2002, pp. 497–528.

[22] J. Aczél, B. Forte, C.T. Ng, "Why the Shannon and Hartley entropies are 'natural'", *Adv. Appl. Probab.*, vol. 6, 1974, pp. 131–146.

[23] C.E. Shannon, W. Weaver, *A Mathematical Theory of Communication*, Chicago:University of Illinois Press, 1949.

[24] R.V. Hartley, "Transmission of information", *Bell System Tech. J.*, vol. 7, 1928, pp. 535–563.

[25] W.J. McGill, "Multivariate information transmission", *IEEE Trans. Inform. Theory* vol. 4, 1955, pp. 93–111.

[26] T.S. Han, "Linear dependence structure of the entropy space", *Inform. Contr.* Vol. 29, 1975, pp. 337–368.

[27] A.J. Bell, "The co-information lattice", in *Proceedings of the Fifth International Workshop on Independent Component Analysis and Blind Signal Separation: ICA 2003*, S. Amari, A. Cichocki, S. Makino, N. Murata, Eds., 2003.