Inquiry Calculus and the Issue of Negative Higher Order Informations

H. R. Noel van Erp [1,*], Ronald O. Linger [1] and Pieter H. A. J. M. van Gelder [1,2] (iD)

[1] Safety and Security Science Group, TU Delft, 2628 BX Delft, The Netherlands; r.o.linger-1@tudelft.nl (R.O.L.); p.h.a.j.m.vangelder@tudelft.nl (P.H.A.J.M.v.G.)

[2] TU Delft Safety and Security Institute, 2628 BX Delft, The Netherlands

[*] Correspondence: h.r.n.vanerp@tudelft.nl; Tel.: +31-15-278-3887

Abstract: In this paper, we will give the derivation of an inquiry calculus, or, equivalently, a Bayesian information theory. From simple ordering follow lattices, or, equivalently, algebras. Lattices admit a quantification, or, equivalently, algebras may be extended to calculi. The general rules of quantification are the sum and chain rules. Probability theory follows from a quantification on the specific lattice of statements that has an upper context. Inquiry calculus follows from a quantification on the specific lattice of questions that has a lower context. There will be given here a relevance measure and a product rule for relevances, which, taken together with the sum rule of relevances, will allow us to perform inquiry analyses in an algorithmic manner.

1. Introduction

Cox asked the question: Is it possible to construct a consistent set of mathematical rules for carrying out plausible, rather than deductive, reasoning? He found that, if we try to represent degrees of plausibility by real numbers, then the conditions of consistency can be stated by functional equations whose general solutions can be found. The results were: out of all possible monotonic functions which might in principle serve our purpose, there exists a particular scale, that is, class of functions, on which to measure degrees of plausibility, which we henceforth call "probability". The consistent rules on how to combine these probabilities take the form of Laplace's product and sum rules [1]. Cox, thus, proved that any method of inference in which we represent degrees of plausibility by real numbers, is necessarily either equivalent to Laplace's or inconsistent. The Cox formulation encompasses and generalizes the Kolmogorov formulation [2], as it derives, rather than postulates, the properties of normalization, non-negativity, and additivity, and does so in the more general context of proposition logic, rather than in the more confined context of set theory.

Knuth in his turn has encompassed and generalized the Cox formulations. By introducing probability as a bi-valuation defined on a lattice of statements Knuth quantifies the degree to which one statement implies another [3–11]. This generalization from logical implication to degrees of implication not only mirrors Cox's notion of plausibility as a degree of belief, but includes it. The main difference is that Cox's formulation is based on a set of desiderata derived from his particular notion of plausibility—whereas the symmetries of lattices in general form the basis of the theory and the meaning of the derived measure is inherited from the ordering relation, which in the case of statements is implication. Moreover, by introducing the measure of relevance as a bi-valuation defined on a lattice of questions, it is also possible to quantify the degree to which the answering of one question is relevant to the answering of another. This quantification gives rise to an extended information theory,

also called inquiry calculus, which is intimately connected with probability theory and has its own "Bayesian" product and sum rules for relevances.

In this paper, we first give, as a theoretical background, a discussion on Knuth's valuation calculus and his derivation of the probability theory as the calculus that is associated with the specific lattice of statements. We then proceed to present the derivation of the product rule of relevances and a general relevance measure for the elements of the lattice of questions. Together with the general sum rule and the mathematical definition of a question [10,11], this product rule and measure then will give us the understanding of the fundamental structure of the question space and its associated calculus, which is needed to come to the formulation of an algorithm for automated inquiry.

2. A Calculus of Valuation

Two elements of a set are ordered by comparing them according to a binary ordering relation, that is, by way of $\leq$, which may be read as "is contained by", or $\geq$, which may be read as "contains". Elements may be comparable, in which case they form a chain, or they may be incomparable, in which case they form an anti-chain. A set consisting of both inclusion and incomparability are called partially ordered sets, or posets for short [12].

The upper bound of a pair of elements in a poset is the set of elements that contain them. Given a pair of elements x and y, the least element of the upper bound is called the join, denoted $x \vee y$. The lower bound of a pair of elements is defined dually by considering all the elements that the pair of elements share. The greatest elements of the lower bound is called the meet, denoted $x \wedge y$. A lattice is a partially ordered set where each pair of elements has a unique meet and unique join. There often exist elements that are not formed from the join of any pair of elements. These elements are called join-irreducible elements. Meet-irreducible elements are defined similarly. In Figure 1, we give a general lattice isomorphic to $\mathbf{3}^2$ [12]. It can be seen in this figure that the element $x \vee y$ is located above the element x, whereas the element $x \wedge y$ is located below: i.e., $x \wedge y \leq x \leq x \vee y$.
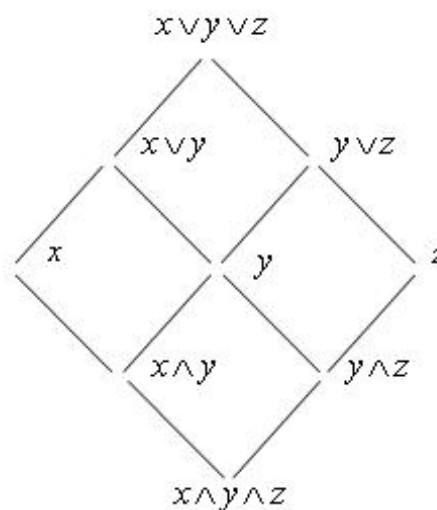
Figure 1. General $\mathbf{3}^2$ lattice.

We can choose to view and join and meet as algebraic operations that take any two lattices elements to a unique third lattice element. From this perspective, the lattice is an algebra. An algebra can be extended to a calculus by defining functions that take lattice elements to real numbers. This enables one to quantify the relationships between the lattice elements. A valuation v is a function that takes a single lattice element x to a real number $v(x)$ in a way that respects the partial order, so that, depending on

the type of algebra, either $v(x) \leq v(y)$ or $v(y) \leq v(x)$, if, in the poset, we have that $x \leq y$. This means that the lattice structure imposes constraints on the valuation assignments, which can be expressed as a set of constraint equations [10], or, equivalently, a valuation calculus [11].

In closing, given any statement about ordered sets, we may obtain a dual statement by replacing each occurrence of $\leq$ by $\geq$. In addition, given any statement about lattices, a special class of ordered sets, we may obtain a dual statement by replacing each occurrence of the join $\vee$ by the meet $\wedge$, and vice versa. This is the duality principle of order theory [12].

2.1. Unconditional Valuations

Valuations v may be assigned to the lattices that take the lattice elements x, $x \wedge y$, and $x \vee y$ to the numbers $v(x)$, $v(x \wedge y)$, and $v(x \vee y)$. The structure of the lattice constrains the valuations v and these constraints are enforced by way of constraint equations. In order for the valuations to be consistent with all the binary ordering relations within the lattice, the valuations of the contained elements in the lattice must be either smaller-equal or greater-equal than the valuations of their containing elements. This gives rise to the order-preserving fidelity constraint equations [11]:

$$x \leq y \qquad \text{implies} \qquad v(x) \leq v(y), \tag{1}$$

or, dually,

$$x \leq y \qquad \text{implies} \qquad v(y) \leq v(x). \tag{2}$$

The third constraint equation on valuations is the so-called general sum rule [3–11]:

$$v(x \vee y) = v(x) + v(y) - v(x \wedge y). \tag{3}$$

The sum rule ensures that the binary ordering relations of the valuations are consistent with those of the lattice itself, just like the fidelity constraints (1) and (2). However, the sum rule provides much more structure, as it relates the valuations of the binary ordered elements as an identity, rather than an inequality.

It is to be noted that the sum rule allows for a strictly monotonic decreasing or increasing one-to-one regrade

$$\Theta[v(x)] = v(x), \tag{4}$$

such that

$$v(x) = \Theta^{-1}[v(x)]. \tag{5}$$

Substituting (4) into (3), we obtain

$$\Theta[v(x \vee y)] = \Theta[v(x)] + \Theta[v(y)] - \Theta[v(x \wedge y)], \tag{6}$$

from which it follows that the regraded valuation (5) also admits a sum rule:

$$v(x \vee y) = \Theta^{-1}(\Theta[v(x)] + \Theta[v(y)] - \Theta[v(x \wedge y)]). \tag{7}$$

This observation is useful in that it is found that a linear rescaling, by some constant factor k, and division of valuations, i.e.,

$$\Theta(x) = kx \qquad \text{and} \qquad \Theta(x) = 1/x, \tag{8}$$

do not destroy the ordering, which is enforced by the sum rule.

2.2. Bi-Valuations

If we want to quantify the *degree* of ordering of element x relative to some context element c, then we need to go from our initial univariate valuation $v(x)$ to the conditional valuation $m(x|c)$. In order for the bivaluations to be consistent with all the binary ordering relations within the lattice, as with the valuations, (1) and (2), we must either have

$$x \leq y \qquad \text{implies} \qquad m(x|c) \leq m(y|c), \tag{9}$$

or, dually,

$$x \leq y \qquad \text{implies} \qquad m(y|c) \leq m(x|c). \tag{10}$$

In addition, the introduction of the degree of ordering does not do away of our need to maintain the order of the unquantified lattice in our valuations. It follows that for conditional valuations the general sum rule must also hold as a third constraint equation (3):

$$m(x \vee y|c) = m(x|c) + m(y|c) - m(x \wedge y|c). \tag{11}$$

Moreover, by introducing the concept of a context, there is introduced a commensurate concept of a change of context. From changes of context, there then follows the chain rule as a fourth constraint equation [10,11]:

$$m(x|c) = m(x|y)\,m(y|c), \tag{12}$$

for chained lattice elements $x \leq y \leq c$ and $c \leq y \leq x$. The chain rule ensures that the binary ordering relations of the conditional valuations are consistent with those of the lattice itself as we go from one context to the other.

We may apply (12) to the chained elements $x \leq x \leq c$ and $c \leq x \leq x$. This gives for both chains the identity

$$m(x|c) = m(x|x)\,m(x|c),$$

which is only consistent for

$$m(x|x) = 1. \tag{13}$$

It follows that the degree of ordering of any element in the lattice relative to itself is constrained by the chain rule to be one [11].

In the sum and chain rules, (11) and (12), we have the general calculus of valuation for platonic lattices, like the one in Figure 1, which have lattice elements whose meaning are not yet specified. If we specify the nature of the elements and the join and meet operators in our lattice, then we go from the one universal lattice, the elements of which carry no specific meaning, to all the specific lattices we might conceive of.

We will now apply to the general calculus of valuation to the specific lattices of propositions and questions. It will be found that the specifics of these lattices introduce additional constraints that are unique to these lattices.

3. Valuations on the Specific Lattice of Propositions

If we define the elements in the platonic lattice of Figure 1 to be propositions while letting the join $\vee$ and the meet $\wedge$ be the OR- and AND-operators of Boolean logic, then we have the specific (Boolean) lattice of propositions. The ordering relation of the lattice of propositions naturally encodes logical implication, such that a given proposition implies all the propositions above it. Logical deduction is straightforward in this framework since every proposition in the lattice implies (i.e., is included by) all the proposition above it with certainty—for example, x implies x, $x \vee y$, $x \vee y \vee z$, etc. The lattice of propositions is in this sense an algebra of deduction [3–11].

Logical induction, however, works backwards. In induction, we quantify the degree to which one's current state of knowledge implies a proposition of lower certainty below it. Thus, in order to go from deduction to induction, we need to generalize the algebra of deduction to a calculus of induction, by way of a bi-valuation on the lattice of propositions. In what follows, we derive the constraints on a bi-valuation measure, called probability, that quantifies the degree to which one proposition implies another [3–11].

Since we let the ordering relation be degree of implication, we may interpret the constraint (13) to signify that the proposition x implies itself absolutely. Moreover, by way of the lattice of propositions' natural encoding of logical implication, we have that any proposition above x is implied with absolute certainty:

$$x \leq c \quad \text{implies} \quad m(c \mid x) = 1. \tag{14}$$

By way of this additional constraint, which is introduced by the specific meaning, we have assigned to the lattice elements (i.e., propositions) and lattice join and meet (i.e., OR- and AND-operators), together with the sum rule, we may further constrain the chain rule into a product rule that is specific for upper contexts.

If, for the small diamond in Figure 1, which is defined by x, $x \vee y$, y, and $x \wedge y$, we consider the context to be x. Then, the sum rule for this diamond may be written down as, (11),

$$m(x \vee y \mid x) + m(x \wedge y \mid x) = m(x \mid x) + m(y \mid x). \tag{15}$$

Since $x \leq x$ and $x \leq x \vee y$, we have that the statement x implies both statements x and $x \vee y$ with absolute certainty, that is, (14),

$$m(x \mid x) = m(x \vee y \mid x) = 1. \tag{16}$$

Substituting (16) into sum rule (15), we obtain the further constraint:

$$m(x \wedge y \mid x) = m(y \mid x), \tag{17}$$

which holds for arbitrary elements x and y in the lattice of propositions that are closed under the join and meet, and which is expressed by the equivalence of the arrows in Figure 2.
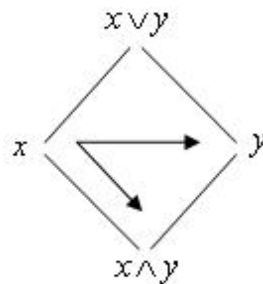
Figure 2. 2^2 lattice of propositions x, $x \vee y$, y, and $x \wedge y$.

Consider the chain where the bi-valuation $m(x \wedge y \wedge z \mid x)$ with context x is decomposed into two parts, by introducing the intermediate context $x \wedge y$. The chain rule (12) gives

$$m(x \wedge y \wedge z \mid x) = m(x \wedge y \wedge z \mid x \wedge y) \, m(x \wedge y \mid x) \tag{18}$$

and the constraint (17) gives

$$m\left(x \wedge y \wedge z \mid x\right) = m\left(y \wedge z \mid x\right),$$
$$m\left(x \wedge y \wedge z \mid x \wedge y\right) = m\left(z \mid x \wedge y\right), \qquad (19)$$
$$m\left(x \wedge y \mid x\right) = m\left(y \mid x\right).$$

If we substitute the simplifications (19) into (18), we obtain the specific product rule for upper contexts:

$$m\left(y \wedge z \mid x\right) = m\left(z \mid x \wedge y\right) m\left(y \mid x\right). \qquad (20)$$

For the lattice of propositions the meet $\wedge$ is the Boolean AND-operator. Since this operator is commutative (i.e., $y \wedge z = z \wedge y$), the constraint (20) relaxes to [10,11]

$$m\left(y \mid x \wedge z\right) m\left(z \mid x\right) = m\left(y \wedge z \mid x\right) = m\left(z \mid x \wedge y\right) m\left(y \mid x\right). \qquad (21)$$

Therefore, by assigning meaning to the lattice, we obtain the additional constraints (14). This then translates to the constraint (17), which then refines the chain rule of the general bi-valuation calculus into the product rule for upper contexts.

The product rule for upper contexts allows us determine the valuation for impossibility. If proposition y implies impossibility under context x, then so must the logical conjunction of the propositions y and z: i.e.,

$$y = y \wedge z = \text{impossibility}. \qquad (22)$$

Relabeling the propositions in (20) by way of (22), it is found that impossibility must either translate to a valuation of 0 or ∞, in order for this identity to hold:

$$m\left(\text{impossibility} \mid x\right) = m\left(z \mid x \wedge y\right) m\left(\text{impossibility} \mid x\right).$$

We then have, from (14) and the sum rule (11), for a context $x \vee y$ where the propositions x and y are mutually exclusive and, as a consequence, $x \wedge y$ is impossible, the following identity:

$$1 = m\left(x \mid x \vee y\right) + m\left(y \mid x \vee y\right) - m\left(\text{impossibility} \mid x \vee y\right),$$

from which it follows that the valuation of the impossibility $x \wedge y$ necessarily equals zero:

$$m\left(\text{impossibility} \mid c\right) = 0. \qquad (23)$$

By way of (14) and (23), it then follows that for the sum rule (11) the fidelity constraint (9) must hold, rather than (10). The removal of this degree of freedom allows us to determine the unknown bivariate function F that quantifies the degree of ordering of element x relative to some higher context element y in the lattice with an upper context.

Because of (17), we have that

$$m\left(x \mid y\right) = m\left(x \wedge y \mid y\right).$$

Thus, we are looking for the bivariate bi-valuation function

$$F\left[v\left(x \wedge y\right), v\left(y\right)\right] = m\left(x \mid y\right). \qquad (24)$$

Since the sum rule allows for linear rescaling of the valuations, (8), we want our bi-valuation to be invariant for such a rescaling. This then puts the following constraint on the unknown function F:

$$F\left[v\left(x \wedge y\right), v\left(y\right)\right] = F\left[k\, v\left(x \wedge y\right), k\, v\left(y\right)\right]. \qquad (25)$$

The general solution of this homogenous function of degree zero is [13]

$$F\left[v\left(x \wedge y\right), v\left(y\right)\right] = f\left[\frac{v\left(x \wedge y\right)}{v\left(y\right)}\right], \tag{26}$$

where f is some unknown function. Because of (9) and (14), it then follows that f is the identity function:

$$f\left(x\right) = x. \tag{27}$$

Substituting (26) and (27) into (24), we obtain the bi-valuation for the lattices with an upper context:

$$m\left(x\middle|y\right) = \frac{v\left(x \wedge y\right)}{v\left(y\right)}. \tag{28}$$

By relabeling the measure m to p, the context symbol c to I, and the operator symbols $\vee$ and $\wedge$ to the corresponding Boolean OR- and AND-operators, "+" and (optional) "·", we may recognize in the constraints (11) and (21) the sum and product rules of probability theory:

$$p\left(x + y\middle|I\right) = p\left(x\middle|I\right) + p\left(y\middle|I\right) - p\left(x \cdot y\middle|I\right) \tag{29}$$

and

$$p\left(y\middle|I\right)p\left(x\middle|y\right) = p\left(x \cdot y\middle|I\right) = p\left(x\middle|I\right)p\left(y\middle|x\right), \tag{30}$$

where use has been made of the fact that for $x, y \leq I$

$$x \cdot I = x \quad \text{and} \quad y \cdot I = y.$$

The probability measure has a range, (14) and (23),

$$0 \leq p\left(x\middle|I\right) \leq 1 \tag{31}$$

and is defined as a ratio of valuations, (28):

$$p\left(x\middle|y\right) = \frac{v\left(x \wedge y\right)}{v\left(y\right)}, \tag{32}$$

where, the first option in (9),

$$x \wedge y \leq x \quad \text{implies} \quad v\left(x \wedge y\middle|I\right) \leq v\left(y\middle|I\right). \tag{33}$$

By introducing probability as a bi-valuation defined on a lattice of propositions, we can quantify the degree to which one proposition is implied by another. The symmetries of lattices in general form the basis of the theory and the meaning of the derived measure is inherited from the ordering relation, which in the case of propositions is implication. Because of the concept of context, we have that probability is necessarily conditional, and a Bayes' theorem for probability calculus follows as a direct result of the chain rule in terms of a change in context.

4. What Is a Question?

Before we can proceed to construct a specific lattice of question, we first need to have a clear notion of the nature of questions. In his last scientific publication, Cox explored the logic of inquiry. In this paper, he defined a question as the set of all possible statements that answer it [14]. Thus, Cox proposed that a question is a set of specific statements (i.e., answers).

For example, say we have three mutually exclusive and exhaustive weather states

$$a \equiv \text{raining}, \quad b \equiv \text{snowing}, \quad c \equiv \text{sunny}, \tag{34}$$

then these three weather states map to a Boolean lattice of logical statements having elements [11]

$$\{a+b+c,$$
$$a+b, \quad a+c, \quad b+c,$$
$$a, \quad b, \quad c,$$
$$\perp\},$$

where + is the OR-operator of Boolean logic and where the lattice elements are ordered by implication. For example, a implies a, $a+b$, $a+c$ and $a+b+c$, whereas c implies c, $a+c$, $b+c$ and $a+b+c$. The top element $a+b+c$ is the so-called truism, which, for an exhaustive enumeration of weather states (34), is always true.

The question whether it is raining or not is answered by the statements a and $b+c$, respectively, the statements "it is raining" and "it is either snowing or sunny", the latter statement being equivalent to the statement "it is not raining". However, the statements b and c, respectively, the statements "it is snowing" and "it is sunny" also answer that same question, seeing that they imply the statement "it is not raining". Moreover, absurdity, designated as $\perp$, also constitutes an answer should the actual weather state be other than either a, b or c. Thus, the question whether it is raining or not is answered by the set of statements $\{b+c, a, b, c, \perp\}$.

Now, the down-sets $\downarrow$ of a the (non-absurdity) elements of the Boolean lattice with generating atomic elements a, b and c are given as [12]:

$$
\begin{aligned}
A &\equiv \downarrow a = \{a, \perp\}, \\
B &\equiv \downarrow b = \{b, \perp\}, \\
C &\equiv \downarrow c = \{c, \perp\}, \\
AB &\equiv \downarrow (a+b) = \{a+b, a, b, \perp\}, \\
AC &\equiv \downarrow (a+c) = \{a+c, a, c, \perp\}, \\
BC &\equiv \downarrow (b+c) = \{b+c, b, c, \perp\}, \\
ABC &\equiv \downarrow (a+b+c) = \{a+b+c, a+b, a+c, b+c, a, b, c, \perp\}.
\end{aligned}
\tag{35}
$$

The down-sets (35) are called *ideal questions*. Ideal questions are mathematical abstractions since they do not entertain all the possible mutually exclusive statements (34) as answers, with the exception of the ideal question ABC that is answered by every statement. *Real questions* are constructed by taking the set unions $\cup$ of the ideal questions such that all the possible mutually exclusive statements (34) are entertained as answers [4–10].

For example, we may construct the question whether it is raining or not as a union of ideal questions, or, equivalently, down-sets, A and BC, (35):

$$A \cup BC = \downarrow a \cup \downarrow (b+c) = \{a, \perp\} \cup \{b+c, b, c, \perp\} = \{b+c, a, b, c, \perp\}. \tag{36}$$

Thus, the set of all real questions is given as [4–10]

$$\{ABC,$$
$$AB \cup AC \cup BC,$$
$$AC \cup BC, \quad AB \cup BC, \quad AB \cup AC,$$
$$A \cup BC, \quad B \cup AC, \quad C \cup AB,$$
$$A \cup B \cup C\},$$

where the real questions have been ordered by set inclusion $\subseteq$ with the least concise question, that is, the question with the largest set of possible answers (i.e., ABC) at the top, and the most concise question, that is, the question with the smallest set of possible answers (i.e., $A \cup B \cup C$) at the bottom.

This most concise question is called the *central issue*, as it is the question which, when answered, answers all real questions that lie above it in the lattice of real questions [4–10].

Since a question is the set of statements that can be given as an answer to that question, each question represents a set of answers. It follows that related questions may be ordered by set inclusion. This ordering relation of set inclusion implements the concept of answering [4–10]. Thus, if question Q_1 is a subset of question Q_2, then $Q_1 \subseteq Q_2$, and by answering question Q_1, we will have necessarily answered question Q_2.

For example, the question "is it raining, snowing or sunny?",

$$A \cup B \cup C = \{a, b, c, \perp\}, \tag{37}$$

also answers the question "is it raining or not?",

$$A \cup BC = \{b + c, a, b, c, \perp\}. \tag{38}$$

Comparing the sets corresponding with questions (37) and (38), we may easily check that the former is indeed included by the latter

$$\{a, b, c, \perp\} = A \cup B \cup C \subseteq C \cup AB = \{a + b, a, b, c, \perp\}. \tag{39}$$

Since questions are just sets of all the possible statements that answer that question, we have that the logical meet $\cap$ and join $\cup$ of set theory may be applied to questions. The meet of the questions $AB \cup C$, "is it sunny or not?", and $A \cup BC$, "is it raining or not?", gives the question "is it raining, snowing, or sunny?":

$$\begin{aligned} (C \cup AB) \cap (A \cup BC) &= \{a + b, a, b, c, \perp\} \cap \{b + c, a, b, c, \perp\} \\ &= \{a, b, c, \perp\} \\ &= A \cup B \cup C. \end{aligned} \tag{40}$$

This may be seen as follows. If we first ask if it is sunny, then we will either know that it is sunny or not. If it is not sunny, then we may inquire further, and ask whether it is raining or not, after which we will know exactly what kind of weather it is. We would have gotten the same result had we asked directly whether it was raining, snowing, or sunny. Thus, the meet of two questions tends to give us a question that is more informative, when answered, than either question alone.

The join of the questions "is it sunny or not?", $AB \cup C$, and "is it raining or not?", $A \cup BC$, gives the question "is it not raining or is it not sunny?":

$$\begin{aligned} (C \cup AB) \cup (A \cup BC) &= \{a + b, a, b, c, \perp\} \cup \{b + c, a, b, c, \perp\} \\ &= \{a + b, b + c, a, b, c, \perp\} \\ &= AB \cup BC. \end{aligned} \tag{41}$$

We can see that a join of two given questions tends to give us a less informative question than either of the questions alone, or, for that matter, the meet of those same questions. This observation will prove to be crucial in the derivation of the product rule of the lattice of questions.

The definition of real questions as a set union of down-sets allows us to come to an intrinsic order of real questions by way of set-inclusion, with the most informative real question being a subset of the least informative real question. However, the set of formally admissible answers for a given answer is a subset of the total set of statements that make up a real question.

For, example the question "is it sunny or not?",

$$C \cup AB = \{a + b, a, b, c, \perp\}, \tag{42}$$

has as its set of formally admissible answers the subset $\{a + b, c\}$. The question $C \cup AB$ is called a *partition question*, since the set of formally admissible answers of this question neatly partitions the set of answers [8]. Partition questions are concrete questions in the sense that for each weather state there can only be one correct answer from the subset of formally admissible answers. That is, if either of the weather states a or b holds then the correct answer to the question whether it is sunny or not is the statement that it is not sunny, $a + b$, and, under weather state c, the only correct answer is that it is sunny, c. Alternatively, the question "is it not sunny, is it not snowing, or is it not raining?",

$$AB \cup AC \cup BC = \{a + b, a + c, b + c, a, b, c, \perp\},\tag{43}$$

has as its set of formally admissible answers the subset $\{a + b, a + c, b + c\}$. The question $AB \cup AC \cup BC$ is not a partition question. Questions that are not partition questions are ambiguous in the sense that, for a given weather state, two answers may be given. For example, for an actual weather state of raining, a, one may answer with either "it is not sunny", $a + b$, or "it is not snowing", $a + c$.

Thus, the formally admissible answers of the real questions map to the following statements:

$$
\begin{aligned}
ABC &\mapsto \{a + b + c\}, \\
AB \cup AC \cup BC &\mapsto \{a + b, a + c, b + c\}, \\
AC \cup BC &\mapsto \{a + c, b + c\}, \\
AB \cup BC &\mapsto \{a + b, b + c\}, \\
AB \cup AC &\mapsto \{a + b, a + c\}, \\
C \cup AB &\mapsto \{a + b, c\}, \\
B \cup AC &\mapsto \{a + c, b\}, \\
A \cup BC &\mapsto \{b + c, a\}, \\
A \cup B \cup C &\mapsto \{a, b, c\}.
\end{aligned}
\tag{44}
$$

Furthermore, it can be checked that this mapping can be accomplished by taking the set unions of the ideals in (35), but with the down-set operators $\downarrow$ dropped in these ideals.

5. Valuations on the Specific Lattice of Questions

If define the elements in the platonic lattice of Figure 1 to be questions, the join $\vee$ to be the union-operator of set theory, and the meet $\wedge$ to be the intersection-operator, then we have the specific lattice of questions. The ordering relation of the lattice of questions naturally encodes relevance, such that a given question answers all the questions above it. Relevance assignment is straightforward in this framework since a question in the lattice is absolutely relevant for (i.e., is included by) every question above it. For example, the questions x, $x \wedge y$, $x \wedge y \wedge z$, etc., are all absolutely relevant for the answering of question x. The lattice of questions is in this sense an algebra of relevance, just like the lattice of statements is an algebra of deduction [3–10].

Now, if want to quantify the degree to which a given question is relevant for some other question that is not located directly above it in the lattice of questions, then this will require a generalization of the algebra of questions to a calculus of questions. In what follows, we derive the constraints on a bi-valuation measure, called relevance, that quantifies the degree to which the answering of one question will contribute to the answering of another question [3–10].

Since we let the ordering relation be degree of relevance, we may interpret the constraint (13) to signify that the answering of question x is absolutely relevant to the answering of itself. Moreover, by way of of the lattice of questions' natural encoding of relevance, we have that any question below x will be absolutely relevant for its answering:

$$c \leq x \quad \text{implies} \quad m(c \mid x) = 1.\tag{45}$$

By way of this additional constraint, which is introduced by the specific meaning we have assigned to the lattice elements and lattice join and meet, together with the sum rule, we may now constrain the chain rule into a product rule that is specific for lower contexts.

If for the small diamond in Figure 1 which is defined by x, $x \vee y$, y, and $x \wedge y$ we consider the context to be x, then the sum rule for this diamond may be written down as, (11),

$$m\left(x \vee y \mid x\right) + m\left(x \wedge y \mid x\right) = m\left(x \mid x\right) + m\left(y \mid x\right). \tag{46}$$

Since $x \geq x$ and $x \geq x \wedge y$, we have that the questions x and $x \wedge y$ are absolutely relevant for question x, that is, (45),

$$m\left(x \mid x\right) = m\left(x \wedge y \mid x\right) = 1. \tag{47}$$

Substituting (47) into sum rule (46), we obtain the further constraint:

$$m\left(x \vee y \mid x\right) = m\left(y \mid x\right), \tag{48}$$

which holds for arbitrary elements x and y in the lattice of propositions that are closed under the join and meet, and which is expressed by the equivalence of the arrows in Figure 3.
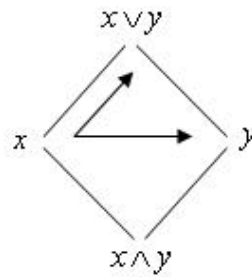
Figure 3. 2^2 lattice of questions x, $x \vee y$, y, and $x \wedge y$.

Consider the chain where the bi-valuation $m\left(x \vee y \vee z \mid x\right)$ with context x is decomposed into two parts, by introducing the intermediate context $x \vee y$. The chain rule (12) gives

$$m\left(x \vee y \vee z \mid x\right) = m\left(x \vee y \vee z \mid x \vee y\right) m\left(x \vee y \mid x\right) \tag{49}$$

and the constraint (48) gives

$$\begin{aligned} m\left(x \vee y \vee z \mid x\right) &= m\left(y \vee z \mid x\right), \\ m\left(x \vee y \vee z \mid x \vee y\right) &= m\left(z \mid x \vee y\right), \\ m\left(x \vee y \mid x\right) &= m\left(y \mid x\right). \end{aligned} \tag{50}$$

If we substitute the simplifications (50) into (49), we obtain the specific product rule for lower contexts [15]:

$$m\left(y \vee z \mid x\right) = m\left(z \mid x \vee y\right) m\left(y \mid x\right). \tag{51}$$

For the lattice of question, the union $\vee$ is the set-theoretical union-operator. Since this operator is commutative (i.e., $y \vee z = z \vee y$), the constraint (51) relaxes to

$$m\left(y \mid x \vee z\right) m\left(z \mid x\right) = m\left(y \vee z \mid x\right) = m\left(z \mid x \vee y\right) m\left(y \mid x\right). \tag{52}$$

Thus, by assigning meaning to the lattice, we obtain the additional constraints (45). This then translates to the constraint (48), which then refines the chain rule of the general bi-valuation calculus into the product rule for lower contexts. Furthermore, it is to be noted that product rules for upper and lower contexts, (21) and (52), are dual to each other.

The product rule for lower contexts allows us determine the valuation for absolute irrelevance. If question y is absolutely irrelevant under context x, then so must be the set union of the questions y and z: i.e.,

$$y = y \vee z = \text{irrelevancy}. \tag{53}$$

Relabeling the questions in (51) by way of (53), it is found that irrelevancy must either translate to a valuation of 0 or ∞, in order for this identity to hold:

$$m\left(\text{irrelevancy} \middle| x\right) = m\left(z \middle| x \vee y\right) m\left(\text{irrelevancy} \middle| x\right).$$

We then have, from (45) and the sum rule (11), for a context $x \wedge y$ where the joint question $x \vee y$ is irrelevant, the following identity:

$$1 = m\left(x \middle| x \wedge y\right) + m\left(y \middle| x \wedge y\right) - m\left(\text{irrelevancy} \middle| x \wedge y\right),$$

from which it follows that the valuation of the irrelevance $x \vee y$ necessarily equals zero:

$$m\left(\text{irrelevancy} \middle| c\right) = 0. \tag{54}$$

By way of (14) and (54), it then follows that for the sum rule (11) the fidelity constraint (10) must hold, rather than (9). The removal of this degree of freedom allows us to determine the unknown bivariate function F that quantifies the degree of ordering of element x relative to some higher context element y in the lattice with an upper context as, (28) and (48),

$$m\left(x \middle| y\right) = \frac{v\left(x \vee y\right)}{v\left(y\right)}. \tag{55}$$

By relabeling the measure m to d, the context symbol c to I, and the operator symbol $\wedge$ to the (optional) "$\cdot$" symbol, while keeping the the join symbol $\wedge$ as it is, in order to keep the $+$ symbol free for the OR-operator in the probability calculus, we now have in the constraints (11) and (52) the sum and product rules of the inquiry calculus:

$$d\left(x \vee y \middle| I\right) = d\left(x \middle| I\right) + d\left(y \middle| I\right) - d\left(x \cdot y \middle| I\right) \tag{56}$$

and

$$d\left(y \middle| I\right) d\left(x \middle| y\right) = d\left(x \vee y \middle| I\right) = d\left(x \middle| I\right) d\left(y \middle| x\right),$$

where use has been made of the fact that for $I \leq x, y$

$$x \vee I = x \quad \text{and} \quad y \vee I = y. \tag{57}$$

The relevance measure has a range, (45) and (54),

$$0 \leq d\left(x \middle| I\right) \leq 1 \tag{58}$$

and is defined as a ratio of valuations, (55):

$$d\left(x \middle| y\right) = \frac{v\left(x \vee y\right)}{v\left(y\right)}, \tag{59}$$

where, the second option in (9),

$$y \leq x \vee y \quad \text{implies} \quad v\left(x \vee y \mid I\right) \leq v\left(y \mid I\right). \tag{60}$$

By introducing relevance as a bi-valuation defined on a lattice of questions, we can quantify the degree to which one question is relevant to another. The symmetries of lattices in general form the basis of the theory and the meaning of the derived measure is inherited from the ordering relation, which in the case of questions is relevance. Because of the concept of context, we have that relevance is necessarily conditional, and a Bayes' theorem for inquiry calculus follows as a direct result of the chain rule in terms of a change in context.

6. The Three Spaces

There are three fundamental spaces of interest: the space of states, the hypothesis space, and the inquiry space [9,11]. These spaces enable us to describe a system, to describe what we know about a system, and to describe what can potentially be known about a system.

6.1. The State Space

We model the world or some interesting aspect of it as being in a particular state out of a finite set of mutually exclusive states. The enumeration of all the possible states that our system may be in gives rise to the state space. For example, if we measure our system with respect to variables A and B, and if each variable can take only two values, then the state space is given as

$$AB = \{ab_{11}, ab_{12}, ab_{21}, ab_{22}\}. \tag{61}$$

6.2. The Hypothesis Space

A given individual may not know precisely which state the system is in, but may have some information that rules out some states, but not others. Thus, the set of potential states defines what one can say about the state of the system. For this reason, we call a potential state a statement. If we let the elements in the state space AB be propositions, (61),

$$AB = \{ab_{11}, ab_{12}, ab_{21}, ab_{22}\},$$

then we may denote a set of potential states by way of the OR-operator $+$ of Boolean logic. For example, if our system can be in the set of states $\{ab_{11}, ab_{12}\}$, then we may denote this as

$$ab_{11} + ab_{12} = a_1\left(b_1 + b_2\right) = a_1. \tag{62}$$

Alternatively, if our system can be in states in the set of states $\{ab_{11}, ab_{12}, ab_{22}\}$, then we may denote this as

$$ab_{11} + ab_{12} + ab_{22} = (ab_{11} + ab_{12}) + (ab_{12} + ab_{22}) = a_1 + b_2. \tag{63}$$

A statement describes a state of knowledge about the state of the system. The set of all possible statements is called the *hypothesis space*. If we let the join $\vee$ of the lattice be the OR-operator of Boolean logic and the meet $\wedge$ be the AND-operator, then we may may construct a lattice of statements, that is, the hypothesis space may be represented by way of a lattice of statements, or, equivalently, propositions.

The lattice of statements is generated by taking the power set, which is the set of all possible unions of the elements of the set of states AB, and ordering them according to inclusion. For a system of n mutually exclusive possible states, there are

$$\sum_{i=0}^{n} \binom{n}{i} = 2^n \tag{64}$$

statements, including the null-meet. Thus, the state space AB with $n = 4$ elements has 16 possible statements, which can be ordered as, (61) and (64):

$$\{ ab_{11} + ab_{12} + ab_{21} + ab_{22},$$
$$ab_{11} + ab_{12} + ab_{21}, \quad ab_{11} + ab_{12} + ab_{22}, \quad ab_{11} + ab_{21} + ab_{22}, \quad ab_{12} + ab_{21} + ab_{22},$$
$$ab_{11} + ab_{12}, \quad ab_{11} + ab_{21}, \quad ab_{11} + ab_{22}, \quad ab_{12} + ab_{21}, \quad ab_{12} + ab_{22}, \quad ab_{21} + ab_{22},$$
$$ab_{11}, \quad ab_{12}, \quad ab_{21}, \quad ab_{22},$$
$$\perp \},$$

where the statement at the top is the truism, which represents the state of knowledge where one only knows (for this particular instance) that the system can be in one of four possible states and where the statement $\perp$ at the bottom is the null-meet that represents logical impossibility.

Degrees of implication may be inferred from the hypothesis space by assigning bi-valuations (i.e., probabilities) to the hypotheses of the state space (61).

All the elements in the hypothesis space are closed under both the join + and the meet $\cdot$, seeing that join of the null-meet $\perp$ with any element x in the hypothesis space maps to that element (i.e., $\perp + x = x$) and the meet of any two elements from the set AB is taken to the null-meet $\perp$. Thus, it follows that the sum rule may be applied to any two elements in the hypothesis space.

In addition, the sum rule (29) may be generalized to

$$p\left(\sum_i x_i \middle| I\right) = \sum_i p\left(x_i \middle| I\right) - \sum_{j>i} p\left(x_i \cdot x_j \middle| I\right) + \sum_{k>j>i} p\left(x_i \cdot x_j \cdot x_k \middle| I\right) - \cdots . \tag{65}$$

Thus, if the propositions x_i are mutually exclusive, then the meet of these propositions will map to the null-meet, which signifies impossibility. Since we must assign a valuation of zero to logical impossibility, it follows that for mutually exclusive propositions (65) will simplify to

$$p\left(\sum_i x_i \middle| I\right) = \sum_i p\left(x_i \middle| I\right). \tag{66}$$

Furthermore, if we assign probabilities to the exhaustive and mutually exclusive elements of the state space, then we may generate consistent probabilities for all the non-atomic compound statements in the hypothesis space by substituting these probabilities in the right-hand of (66); e.g.,

$$p\left(ab_{11} + ab_{12} + ab_{22} \middle| I\right) = p\left(ab_{11} \middle| I\right) + p\left(ab_{12} \middle| I\right) + p\left(ab_{22} \middle| I\right).$$

Now, we are free to relabel the statements of the hypothesis space in a meaningful manner, (62) and (63):

$$\{ I,$$
$$a_1 + b_1, \quad a_1 + b_2, \quad a_2 + b_1, \quad a_2 + b_2,$$
$$a_1, \quad b_1, \quad ab_{11} + ab_{22}, \quad ab_{12} + ab_{21}, \quad b_2, \quad a_2,$$
$$ab_{11}, \quad ab_{12}, \quad ab_{21}, \quad ab_{22},$$
$$\perp \},$$

where we denoted the top truism as I. In addition, it is to be noted that the hypotheses $ab_{11} + ab_{22}$ and $ab_{12} + ab_{21}$, which are legitimate elements of the lattice of propositions, will not be entertained that often. Since in most probability analyses concerning product spaces like AB, there only will be interest for the probabilities

$$p\left(A \middle| I\right), \quad p\left(B \middle| I\right), \quad p\left(AB \middle| I\right), \quad p\left(A + B \middle| I\right), \quad p\left(A \middle| B\right), \quad p\left(B \middle| A\right), \tag{67}$$

all of which do not pertain to the hypotheses $ab_{11} + ab_{22}$ and $ab_{12} + ab_{21}$.

This simple relabeling together with the ratio-structure of the probability measure also makes insightful the mechanism by which the product rule of probability theory maintains the identity

$$p\left(a_1 \mid I\right)\left(b_1 \mid a_1\right) = p\left(a_1 \mid I\right) \frac{p\left(ab_{11} \mid I\right)}{p\left(a_1 \mid I\right)} = p\left(ab_{11} \mid I\right),$$

for we have that, (30) and (32),

$$p\left(a_1 \mid I\right) p\left(b_1 \mid a_1\right) = \frac{v\left(a_1 \cdot I\right)}{v\left(I\right)} \frac{v\left(a_1 \cdot b_1\right)}{v\left(a_1\right)} = \frac{v\left(a_1\right)}{v\left(I\right)} \frac{v\left(ab_{11}\right)}{v\left(a_1\right)} = \frac{v\left(ab_{11}\right)}{v\left(I\right)} = \frac{v\left(ab_{11} \cdot I\right)}{v\left(I\right)} = p\left(ab_{11} \mid I\right),$$

where use has been made of the fact that for $x, y \leq I$

$$x \cdot I = x \qquad \text{and} \qquad y \cdot I = y.$$

6.3. The Inquiry Space

The state space is an enumeration of all the possible states that the system under consideration may be in. Now, if we measure our system only with respect to the variables A and B, then the most obvious relevances in an inquiry-analysis will be

$$d\left(A \mid I\right), \quad d\left(B \mid I\right), \quad d\left(AB \mid I\right), \quad d\left(A + B \mid I\right), \quad d\left(A \mid B\right), \quad d\left(B \mid A\right), \tag{68}$$

much like the probabilities (67) are the most obvious in a data-analysis. These relevances pertain to the four questions A, B, AB, and $A + B$ in different contexts, just like the probabilities in (67) pertain to the statements a_i, b_j, ab_{ij}, and $a_i + b_j$ in different contexts.

Thus, we wish to determine the relevances of the questions A, B, AB, and $A + B$ relative to the central issue $I = AB$ with which these questions form a chain, as well as the relevances of A and B relative to the issues of interest B and A, respectively, with which these questions form an anti-chain.

6.3.1. Normalized Entropies

Introspection suggests that certainty about the answer to some question will make that question irrelevant. Thus, we take it as our basic assumption that the relevance of a question ought to depend on the probability of its answers. If we do so, then additivity, subadditivity, symmetry with respect to disjunction, and expansibility, which describe how the lattice collapses when a statement is found to be false, under the condition that small probabilities tend to small relevances, dictate that relevances relative to the central issue be in the form of a Shannon entropy [7]. Stated differently, if we wish the relevance of a question to depend on the probability of its answers, then the use of the Shannon entropy becomes a forced choice.

Let us have some system with possible states x having probabilities $p\left(x \mid I\right)$. Then, the quantity

$$h\left(x\right) = \log \frac{1}{p\left(x \mid I\right)} \tag{69}$$

is called the surprise, since it is large for improbable events and small for probables ones. Averaging this quantity over all of the possible states of the system x gives the Shannon entropy [16]

$$H\left(X\right) = \sum_x p\left(x \mid I\right) \log \frac{1}{p\left(x \mid I\right)}. \tag{70}$$

The Shannon entropy can be thought of as a measure of the amount of uncertainty, or, equivalently, surprise, we possess about a system. In addition, let us have some system with possible states $x \cdot y$ having probabilities $p(x \cdot y|I)$. Then, the mean conditional entropy is given as (70)

$$H(Y|X) = \sum_x p(x|I) \sum_y p(y|x) \log \frac{1}{p(y|x)} = \sum_{x,y} p(x \cdot y|I) \log \frac{p(x|I)}{p(x \cdot y|I)}, \tag{71}$$

or, equivalently,

$$H(Y|X) = \sum_x p(x|I) H(Y|x) = H(XY) - H(X). \tag{72}$$

The mean conditional entropy is the amount of uncertainty regarding the states y that remains after having obtained knowledge about the states x. In what follows, we take the mean conditional entropy as our departure point for the construction of a relevance measure. This departure point will bring us right back to the Shannon entropy proposal of the relevancy measure [7], while also making explicit the link with the mutual information measure.

The question AB, when asked, immediately reveals the state of the system and has as its elements

$$ab_{11}, \quad ab_{12}, \quad ab_{21}, \quad ab_{22}.$$

If we ask question AB, then the remaining uncertainty regarding the state of our system can be quantified by way of the mean conditional entropy, (71),

$$H(AB|AB) = \sum_{ij} p(ab_{ij}|I) \log \frac{1}{p(ab_{ij}|ab_{ij})} = \sum_{ij} p(ab_{ij}|I) \log \frac{p(ab_{ij}|I)}{p(ab_{ij}|I)},$$

or, equivalently, (72),

$$H(AB|AB) = H(AB) - H(AB) = 0, \tag{73}$$

which is intuitive enough.

The question A, when asked, tells us if our system is in state a_1 or in state a_2, and is the down-set that has as its top elements

$$a_1 = ab_{11} + ab_{12} \quad \text{and} \quad a_2 = ab_{21} + ab_{22}.$$

If we ask question A, then the remaining uncertainty regarding the state of our system can be quantified by way of the mean conditional entropy, (71),

$$H(AB|A) = \sum_{ij} p(ab_{ij}|I) \log \frac{1}{p(ab_{ij}|a_i)} = \sum_{ij} p(ab_{ij}|I) \log \frac{p(a_i|I)}{p(ab_{ij}|I)},$$

or, equivalently, (72),

$$H(AB|A) = H(AB) - H(A). \tag{74}$$

The question $A + B$ when asked, tells us that our system is not in one of either states, ab_{11}, ab_{12}, ab_{21}, or ab_{12}, and is the down set that has as its top elements

$$a_1 + b_1, \quad a_1 + b_2, \quad a_2 + b_1, \quad a_2 + b_2.$$

Questions A, B, and AB are concrete in that, for a given system state, they only admit one correct answer. The question $A + B$, however, is ambiguous in that, for a given system state, multiple correct answers are admitted. For example, for a given system state of ab_{11}, the answers $a_1 + b_1$, $a_1 + b_2$ and

$a_2 + b_1$ are all legitimate answers to the question $A + B$. The probability of the system state ab_{11} and either one of these legitimate answers τ_k, for $k = 1, 2, 3$, equals

$$p\left(ab_{11} \cdot \tau_k \middle| I\right) = p\left(ab_{11} \middle| I\right) p\left(\tau_k \middle| I\right) = p\left(ab_{11} \middle| I\right) \frac{1}{3},$$

seeing that we do not have any information as to which of the three possible answers τ_k to expect. Moreover, each system state ab_{ij} and "answering state" τ_k correspond together with exactly one state of surprise (69). Thus, if we ask question $A + B$, then the mean conditional entropy may be computed unambiguously as the average of the conditional surprises over the compound system-answering states $\left(ab_{ij} \cdot \tau_k\right)$:

$$
\begin{aligned}
H\left(AB \middle| A+B\right) = {} & p\left(ab_{11} \middle| I\right) \frac{1}{3} \left[\log \frac{1}{p\left(a_{11} \middle| a_1 + b_1\right)} + \log \frac{1}{p\left(a_{11} \middle| a_1 + b_2\right)} + \log \frac{1}{p\left(a_{11} \middle| a_2 + b_1\right)}\right] \\
& + p\left(ab_{12} \middle| I\right) \frac{1}{3} \left[\log \frac{1}{p\left(a_{12} \middle| a_1 + b_1\right)} + \log \frac{1}{p\left(a_{12} \middle| a_1 + b_2\right)} + \log \frac{1}{p\left(a_{12} \middle| a_2 + b_2\right)}\right] \\
& + p\left(ab_{21} \middle| I\right) \frac{1}{3} \left[\log \frac{1}{p\left(a_{21} \middle| a_1 + b_1\right)} + \log \frac{1}{p\left(a_{21} \middle| a_2 + b_1\right)} + \log \frac{1}{p\left(a_{21} \middle| a_2 + b_2\right)}\right] \\
& + p\left(ab_{22} \middle| I\right) \frac{1}{3} \left[\log \frac{1}{p\left(a_{22} \middle| a_1 + b_2\right)} + \log \frac{1}{p\left(a_{22} \middle| a_2 + b_1\right)} + \log \frac{1}{p\left(a_{22} \middle| a_2 + b_2\right)}\right],
\end{aligned}
$$

or, equivalently,

$$
\begin{aligned}
H\left(AB \middle| A+B\right) = {} & p\left(ab_{11} \middle| I\right) \frac{1}{3} \left[\log \frac{p\left(a_1 + b_1 \middle| I\right)}{p\left(a_{11} \middle| I\right)} + \log \frac{p\left(a_1 + b_2 \middle| I\right)}{p\left(a_{11} \middle| I\right)} + \log \frac{p\left(a_2 + b_1 \middle| I\right)}{p\left(a_{11} \middle| I\right)}\right] \\
& + p\left(ab_{12} \middle| I\right) \frac{1}{3} \left[\log \frac{p\left(a_1 + b_1 \middle| I\right)}{p\left(a_{12} \middle| I\right)} + \log \frac{p\left(a_1 + b_2 \middle| I\right)}{p\left(a_{12} \middle| I\right)} + \log \frac{p\left(a_2 + b_2 \middle| I\right)}{p\left(a_{12} \middle| I\right)}\right] \\
& + p\left(ab_{21} \middle| I\right) \frac{1}{3} \left[\log \frac{p\left(a_1 + b_1 \middle| I\right)}{p\left(a_{21} \middle| I\right)} + \log \frac{p\left(a_2 + b_1 \middle| I\right)}{p\left(a_{21} \middle| I\right)} + \log \frac{p\left(a_2 + b_2 \middle| I\right)}{p\left(a_{21} \middle| I\right)}\right] \\
& + p\left(ab_{22} \middle| I\right) \frac{1}{3} \left[\log \frac{p\left(a_1 + b_2 \middle| I\right)}{p\left(a_{22} \middle| I\right)} + \log \frac{p\left(a_2 + b_1 \middle| I\right)}{p\left(a_{22} \middle| I\right)} + \log \frac{p\left(a_2 + b_2 \middle| I\right)}{p\left(a_{22} \middle| I\right)}\right],
\end{aligned}
$$

or, more succinctly, as we gather all the terms,

$$H\left(AB \middle| A+B\right) = \sum_{ij} p\left(ab_{ij} \middle| I\right) \log \frac{1}{p\left(ab_{ij} \middle| I\right)} - \frac{1}{3} \sum_{ij} p\left(a_i + b_j \middle| I\right) \log \frac{1}{p\left(a_i + b_j \middle| I\right)}. \tag{75}$$

We now define for the ambiguous question whose answers, for a given system state ab_{ij}, have probabilities $p\left(a_i + b_j \middle| I\right)$ and where the variables a_i and b_j can take on n and m values, respectively, the Shannon entropy to be

$$H\left(A+B\right) = \frac{1}{n+m-1} \sum_{ij} p\left(a_i + b_j \middle| I\right) \log \frac{1}{p\left(a_i + b_j \middle| I\right)}, \tag{76}$$

where $n + m - 1$ is number of legitimate answers allowed for each system state as well as the sum

$$\sum_{ij} p\left(a_i + b_j \middle| I\right) = \sum_{ij} p\left(a_i \middle| I\right) + \sum_{ij} p\left(b_j \middle| I\right) - \sum_{ij} p\left(ab_{ij} \middle| I\right) = n + m - 1.$$

Then, the amount of uncertainty that will remain after asking question $A + B$, can be written down, (70), (75) and (76),

$$H\left(AB \middle| A+B\right) = H\left(AB\right) - H\left(A+B\right). \tag{77}$$

Now, seeing that our total uncertainty in regards to the state of our system is quantified by the Shannon entropy $H(AB)$, we may take as the relevance valuation v for a given question the total uncertainty in the system minus the uncertainty that remains after having asked that question, (73), (74) and (77):

$$
\begin{aligned}
v(AB) &= H(AB) - H(AB|AB) = H(AB), \\
v(A) &= H(AB) - H(AB|A) = H(A), \\
v(A+B) &= H(AB) - H(AB|A+B) = H(A+B).
\end{aligned}
\tag{78}
$$

Since the questions include the central issue $I = AB$,

$$
AB, A, (A+B) \geq I,
$$

we may use the fact that, (57),

$$
x \geq I \qquad \text{implies} \qquad x = x \vee I,
$$

to come to the needed valuations v:

$$
\begin{aligned}
v(AB \vee I) &= v(AB), \\
v(A \vee I) &= v(A), \\
v[(A+B) \vee I] &= v(A+B),
\end{aligned}
\tag{79}
$$

where we let

$$
v(I) = v(AB).
\tag{80}
$$

Substituting (78)–(80) into (59)

$$
d(x|I) = \frac{v(x \vee I)}{v(I)},
$$

we obtain the relevances to the questions AB, A, and $A + B$, relative to the central issue $I = AB$:

$$
d(AB|I) = \frac{H(AB)}{H(AB)}, \quad d(A|I) = \frac{H(A)}{H(AB)}, \quad d(A+B|I) = \frac{H(A+B)}{H(AB)},
\tag{81}
$$

where the Shannon entropy H for ambiguous (i.e., disjunctive) questions like $A + B$ is defined as (76):

$$
H(A+B) = \frac{1}{n+m-1} \sum_{ij} p(a_i + b_j | I) \log \frac{1}{p(a_i + b_j | I)},
\tag{82}
$$

where $n + m - 1$ is the sum

$$
\sum_{ij} p(a_i + b_j | I) = \sum_{ij} p(a_i | I) + \sum_{ij} p(b_j | I) - \sum_{ij} p(ab_{ij} | I) = n + m - 1.
$$

It is to be noted that the assigning of Shannon entropies to ambiguous questions solves the problem of negative higher order informations [7,17], as it allows us to forgo the use of these higher order informations in our relevance assignments.

In addition, the Shannon entropies are order-preserving [12], or, equivalently, adhere to the fidelity constraint (2) on the lattice of real questions, in the sense that for the questions AB, A, B and $A + B$ on the lattice of real questions, which are ordered as

$$
AB \leq A, B \leq A + B,
$$

we have

$$H\left(A+B\right) \leq H\left(A\right), H\left(B\right) \leq H\left(AB\right).$$

In Appendix A.2, these order-preserving inequalities are demonstrated for an inquiry space that is generated by the state space (61).

6.3.2. Normalized Mutual Informations

If now take as our system B, rather than AB, and ask question A, then the remaining uncertainty regarding the state of our system B can be quantified by way of the mean conditional surprise, (71),

$$H\left(B|A\right) = \sum_{ij} p\left(a_i|I\right) p\left(b_j|a_i\right) \log \frac{1}{p\left(b_j|a_i\right)} = \sum_{ij} p\left(ab_{ij}|I\right) \log \frac{p\left(a_i|I\right)}{p\left(ab_{ij}|I\right)},$$

or, equivalently, (72),

$$H\left(B|A\right) = H\left(AB\right) - H\left(A\right), \tag{83}$$

where it is to be noted that, (74) and (83),

$$H\left(B|A\right) = H\left(AB|A\right),$$

which is not that surprising, seeing that, by way of the product rule of probability theory (30),

$$p\left(b_j|a_i\right) = p\left(ab_{ij}|a_i\right).$$

Again, as our total uncertainty in regards to the state of our system is quantified by the Shannon entropy $H\left(B\right)$, we may take as our relevance valuation v for a given question the total uncertainty in the system minus the uncertainty that remains after having asked that question, (78),

$$v\left(A \vee B\right) = H\left(B\right) - H\left(B|A\right). \tag{84}$$

Now, the mutual information adheres to the sum rule (3) and is defined as the following function of Shannon entropies:

$$MI\left(X,Y\right) = H\left(X\right) + H\left(Y\right) - H\left(XY\right). \tag{85}$$

The mutual information may be interpreted as the amount of entropy (i.e., uncertainty) that remains in Y after we have subtracted the conditional entropy that is "explained" by X, (72):

$$MI\left(X,Y\right) = H\left(Y\right) - H\left(Y|X\right). \tag{86}$$

Thus, it follows that, (83) through (71),

$$v\left(A \vee B\right) = H\left(A\right) + H\left(B\right) - H\left(AB\right) = MI\left(A,B\right). \tag{87}$$

Let (79)

$$v\left(B \vee I\right) = v\left(B\right) = H\left(B\right). \tag{88}$$

Then, substituting (87) and (88) into (59), we obtain the relevance we are looking for:

$$d\left(A|B\right) = \frac{H\left(A\right) + H\left(B\right) - H\left(AB\right)}{H\left(B\right)} = \frac{MI\left(A,B\right)}{H\left(B\right)}, \tag{89}$$

which is also known as the η_H statistic [18].

In closing, mutual informations $MI\left(x,y\right)$ are used to find the relevance of x relative to the anti-chained y, or, vice versa, the relevance of y for x, for arbitrary anti-chained elements x and y of the lattice of questions. Furthermore, since relevance assignments can be decomposed in a

question to which we want to assign a relevance and a question to which that relevance is relative, it follows that only mutual information is needed to assign relevances to a context that is anti-chained. Stated differently, for anti-chained contexts, the higher order information with its potentially negative values are not needed to assign relevances.

6.3.3. A General Relevance Measure

If we compare (78) with (86), then it can be observed that the Shannon entropy of a question is equivalent to a mutual information between that question and the central issue that answers all questions. Thus, it follows that all relevances, basically, are normalized informations. That is, let x be a question for which we wish to determine the relevance in relation to the issue of interest y. Then, the relevance of x relative to y is defined as the ratio of valuations, (59):

$$d\left(x|y\right) = \frac{v\left(x \vee y\right)}{v\left(y\right)} = \frac{MI\left(x,y\right)}{H\left(y\right)}, \tag{90}$$

where H is the Shannon entropy and (57), (81) and (89),

$$MI\left(x,y\right) = H\left(x\right) + H\left(y\right) - H\left(x \cdot y\right) = \begin{cases} H\left(y\right), & x \leq y, \\ H\left(x\right), & y \leq x, \\ H\left(x\right) + H\left(y\right) - H\left(x \cdot y\right), & x \parallel y, \end{cases} \tag{91}$$

where $\parallel$ is the symbol for incomparability.

Thus, it follows that the inquiry calculus has the following algorithmic structure. Assign Shannon entropies H as a mono-valuation v to all the elements of the lattice of (concrete and ambiguous) real (i.e., answerable) questions, [4–9]. This valuated lattice can then be used to generate all the needed relevances by taking for a specified issue of interest y all the remaining elements of the lattice of questions, by way of the sum rule, i.e., (91), to all possible mutual informations (i.e., joins). These mutual informations are then normalized, by way of the product rule, i.e., (90), to the desired relevances.

For comparison, the probability calculus has the following algorithmic structure. Assign probabilities to the join-irreducible atomic elements of the state space. Use these probabilities to generate, by way of the sum rule, the valuations on the total lattice of statements. This valuated lattice can then be used to generate all the needed probabilities by taking for a specified conditional variable y all the valuations of the lattice of statements, by way of the sum rule, to the valuation of all the possible Boolean conjunctions (i.e., meets). These valuated Boolean conjunctions are then normalized to the desired probabilities, by way of the product rule, as we divide these valuated conjunctions by the valuation of the conditional variable y.

7. Discussion

By introducing relevance as a bi-valuation defined on a lattice of questions, we can quantify the degree to which one question is relevant to another. The symmetries of lattices in general form the basis of the theory and the meaning of the derived measure is inherited from the ordering relation, which in the case of questions is relevance. Because of the concept of context, we have that relevance is necessarily conditional, and a Bayes' Theorem for inquiry theory follows as a direct result of the chain rule in terms of a change in context.

By way of a quantification on the lattice of statements, it can be proven that the product rule of probability calculus gives the degree of implication of the logical meet of two statements relative to some upper context [10,11]. Now, in the lattice of statements, the join of statements $x \vee y$ is absolutely implied by the lower contexts x and y, whereas, because of the definition of a question, in the lattice of questions it is the meet of questions $x \wedge y$ that is absolutely relevant for the upper contexts x and y.

Thus, if, for the alternative lattice of questions, we follow the derivations for the product rule in [10,11] with this simple observation in mind, then there effortlessly flows forth a dual product rule of inquiry calculus, which gives the degree of relevance of the logical join of two questions relative to some lower context [15]. It follows that the derivation of the product rule of inquiry calculus is more an uncovering of what is already there, hiding in plain sight, beneath the theoretical scaffolding laid down in [10,11]. Furthermore, it is in this specific sense that the uncovering of the information theoretical product rule is nothing more than a technical and conceptual refinement of Knuth's inquiry calculus.

Further conceptual refinements have now been presented in this paper in that we have extended the Shannon entropy from concrete to ambiguous questions, and we have come to the realization that the mutual information is the only information needed. That is, since all relevancy assignments for anti-chained contexts can be decomposed in a question x and a context y, say, where both x and y can be join or a meet of multiple lattice elements, there is no need to introduce a third lattice element z into the relevancy equation. This then solves the problem of negative higher order information [7,17], as we simply have no need for this information.

An unexpected consequence of these new insights is that it follows that all elements of the lattice of real questions are generators for the canvas of inquiry, just like the atomic propositions in the lattice of statements are generators for the canvas of inquiry [19]. In addition, any pair of elements of the valuated lattice of real questions is closed under both the join and the meet, where the meet always will map into the valuated lattice of real questions itself, while the join maps outside of this valuated lattice. To be more specific, for a given issue of interest y, the join will map the remaining elements of lattice of real questions that do not involve the context question y, be it in conjunction or disjunction, to a valuated sub-lattice of the lattice of real questions. Finally, the Shannon entropies that are used for the uni-valuations adhere to the fidelity constraint on the lattice of real questions and, as a consequence, also all bi-valuations on the sub-lattices of real questions for some issue of interest y.

Author Contributions: H.R. Noel van Erp and Ronald. O. Linger came with the proposed solution to the problem of the potential negative higher order informations in inquiry calculus; Pieter H.A.J.M. van Gelder provided feedback and supervision. All authors have read and approved the final manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Miscellanea

We give here some further examples of the the assignment of Shannon entropies to ambiguous question, in order to give the reader a better sense of how to assign such entropies, as well as the definition of the transfer entropy in terms of relevance.

Appendix A.1. Some Further Examples of Ambiguous Questions

If we have the product space ABC, with measurements on the variables A, B, and C, where the variables can take on n_A, n_B, and n_C different values, then for the ambiguous question $A + B + C$ we have that

$$H(A + B + C) = \frac{1}{C} \sum p(a_i + b_j + c_k \mid I) \log \frac{1}{p(a_i + b_j + c_k \mid I)},$$

where, (65),

$$p\left(a_i + b_j + c_k \mid I\right) = p\left(a_i \mid I\right) + p\left(b_j \mid I\right) + p\left(c_k \mid I\right)$$
$$- p\left(a_i \cdot b_j \mid I\right) - p\left(a_i \cdot c_k \mid I\right) - p\left(b_j \cdot c_k \mid I\right)$$
$$+ p\left(a_i \cdot b_j \cdot c_k \mid I\right),$$

and

$$C = \sum_{i,j,k} p\left(a_i + b_j + c_k \mid I\right) = n_B n_C + n_A n_C + n_A n_B - n_C - n_B - n_A + 1,$$

which, for a given system state abc_{ijk}, is the number of permitted answers to the ambiguous question $A + B + C$. Alternatively, for the ambiguous question $AB + C$, we have

$$H\left(AB + C\right) = \frac{1}{C} \sum p\left(ab_{ij} + c_k \mid I\right) \log \frac{1}{p\left(ab_{ij} + c_k \mid I\right)},$$

where, (65),

$$p\left(ab_{ij} + c_k \mid I\right) = p\left(ab_{ij} \mid I\right) + p\left(c_k \mid I\right) - p\left(ab_{ij} \cdot c_k \mid I\right)$$

and

$$C = \sum_{i,j,k} p\left(ab_{ij} + c_k \mid I\right) = n_C + n_A n_B - 1,$$

which, for a given system state abc_{ijk}, is the number of permitted answers to the ambiguous question $AB + C$. Finally, as a last example, for the ambiguous question $AB + AC + BC$, we have

$$H\left(AB + AC + BC\right) = \frac{1}{C} \sum p\left(ab_{ij} + ac_{ik} + bc_{jk} \mid I\right) \log \frac{1}{p\left(ab_{ij} + ac_{ik} + bc_{jk} \mid I\right)},$$

where, (65),

$$p\left(ab_{ij} + ac_{ik} + bc_{jk} \mid I\right) = p\left(ab_{ij} \mid I\right) + p\left(ac_{ik} \mid I\right) + p\left(bc_{jk} \mid I\right) - 2\, p\left(abc_{ijk} \mid I\right)$$

and

$$C = \sum_{i,j,k} p\left(ab_{ij} + ac_{ik} + bc_{jk} \mid I\right) = n_C + n_B + n_A - 2,$$

which, for a given system state abc_{ijk}, is the number of permitted answers to the ambiguous question $AB + AC + BC$.

Appendix A.2. The Order-Preserving Property of the Shannon Entropy

The inequalities $H\left(A\right) \leq H\left(AB\right)$ and $H\left(B\right) \leq H\left(AB\right)$ are a simple consequence of the fact that, when there are more many possibilities, there will be a higher entropy than when there are a few [20]. For example, let $p_1, \ldots, p_n$ be n probabilities of a some probability distribution and, reverting to an alternative notation, let H_n be the Shannon entropy

$$H_n\left(p_1, \cdots, p_n\right) = \sum_{i=1}^{n} p_i \log \frac{1}{p_i}.$$

Then, we have, by construction [20], that

$$H_3\left(p_1, p_2, p_3\right) = H_2\left(p_1, p_2 + p_3\right) + \left(p_2 + p_3\right) H_2\left(\frac{p_2}{p_2 + p_3}, \frac{p_3}{p_2 + p_3}\right).$$

Thus, it follows that

$$H_3\left(p_1, p_2, p_3\right) \le H_2\left(p_1, p_2 + p_3\right).$$

In order to demonstrate the inequalities $H\left(A + B\right) \le H\left(A\right)$ and $H\left(A + B\right) \le H\left(B\right)$, we rewrite the identity

$$
\begin{aligned}
H\left(A + B\right) = {}& p\left(ab_{11}\vert I\right) \frac{1}{3}\left[\log \frac{1}{p\left(a_1 + b_1\vert I\right)} + \log \frac{1}{p\left(a_1 + b_2\vert I\right)} + \log \frac{1}{p\left(a_2 + b_1\vert I\right)}\right] \\
&+ p\left(ab_{12}\vert I\right) \frac{1}{3}\left[\log \frac{1}{p\left(a_1 + b_1\vert I\right)} + \log \frac{1}{p\left(a_1 + b_2\vert I\right)} + \log \frac{1}{p\left(a_2 + b_2\vert I\right)}\right] \\
&+ p\left(ab_{21}\vert I\right) \frac{1}{3}\left[\log \frac{1}{p\left(a_1 + b_1\vert I\right)} + \log \frac{1}{p\left(a_2 + b_1\vert I\right)} + \log \frac{1}{p\left(a_2 + b_2\vert I\right)}\right] \\
&+ p\left(ab_{22}\vert I\right) \frac{1}{3}\left[\log \frac{1}{p\left(a_1 + b_2\vert I\right)} + \log \frac{1}{p\left(a_2 + b_1\vert I\right)} + \log \frac{1}{p\left(a_2 + b_2\vert I\right)}\right],
\end{aligned}
$$

into the inequality

$$
\begin{aligned}
H\left(A + B\right) \le {}& p\left(ab_{11}\vert I\right) \frac{1}{3}\left[\log \frac{1}{p\left(a_1\vert I\right)} + \log \frac{1}{p\left(a_1\vert I\right)} + \log \frac{1}{p\left(b_1\vert I\right)}\right] \\
&+ p\left(ab_{12}\vert I\right) \frac{1}{3}\left[\log \frac{1}{p\left(a_1\vert I\right)} + \log \frac{1}{p\left(a_1\vert I\right)} + \log \frac{1}{p\left(b_2\vert I\right)}\right] \\
&+ p\left(ab_{21}\vert I\right) \frac{1}{3}\left[\log \frac{1}{p\left(b_1\vert I\right)} + \log \frac{1}{p\left(a_2\vert I\right)} + \log \frac{1}{p\left(a_2\vert I\right)}\right] \\
&+ p\left(ab_{22}\vert I\right) \frac{1}{3}\left[\log \frac{1}{p\left(b_2\vert I\right)} + \log \frac{1}{p\left(a_2\vert I\right)} + \log \frac{1}{p\left(a_2\vert I\right)}\right],
\end{aligned}
$$

from which follows, by gathering the terms, a first inequality, (70):

$$H\left(A + B\right) \le \frac{2}{3}H\left(A\right) + \frac{1}{3}H\left(B\right). \tag{A1}$$

Alternatively, we may also obtain the inequality

$$
\begin{aligned}
H\left(A + B\right) \le {}& p\left(ab_{11}\vert I\right) \frac{1}{3}\left[\log \frac{1}{p\left(b_1\vert I\right)} + \log \frac{1}{p\left(a_1\vert I\right)} + \log \frac{1}{p\left(b_1\vert I\right)}\right] \\
&+ p\left(ab_{12}\vert I\right) \frac{1}{3}\left[\log \frac{1}{p\left(a_1\vert I\right)} + \log \frac{1}{p\left(b_2\vert I\right)} + \log \frac{1}{p\left(b_2\vert I\right)}\right] \\
&+ p\left(ab_{21}\vert I\right) \frac{1}{3}\left[\log \frac{1}{p\left(b_1\vert I\right)} + \log \frac{1}{p\left(b_1\vert I\right)} + \log \frac{1}{p\left(a_2\vert I\right)}\right] \\
&+ p\left(ab_{22}\vert I\right) \frac{1}{3}\left[\log \frac{1}{p\left(b_2\vert I\right)} + \log \frac{1}{p\left(a_2\vert I\right)} + \log \frac{1}{p\left(b_2\vert I\right)}\right],
\end{aligned}
$$

from which follows, by gathering the terms, a second equality, (70):

$$H\left(A + B\right) \le \frac{1}{3}H\left(A\right) + \frac{2}{3}H\left(B\right). \tag{A2}$$

Multiplying (A2) times a half and subtracting this product from (A1), we obtain the inequality

$$\frac{1}{2}H\left(A + B\right) \le \frac{1}{2}H\left(A\right), \tag{A3}$$

or, equivalently,

$$H\left(A + B\right) \le H\left(A\right). \tag{A4}$$

Likewise, multiplying (A1) times a half and subtracting this product from (A2), we obtain the other sought after inequality,

$$H(A+B) \leq H(B). \tag{A5}$$

Appendix A.3. Transfer Entropy

If we have the product space ABC, with measurements on the variables A, B, and C, then the tranfer entropy of A, relative to B, in the prediction of C is proportional to the difference in relevances, (89) and [21]:

$$
\begin{aligned}
d(AB|C) - d(B|C) &= \frac{MI(AB,C) - MI(B,C)}{H(C)} \\
&= \frac{H(AB) + H(C) - H(ABC) - [H(B) + H(C) - H(BC)]}{H(C)} \\
&= \frac{H(AB) - H(ABC) - H(B) + H(BC)}{H(C)} \\
&\propto T(A|BC),
\end{aligned}
$$

which is intuitive enough.

References

1. Cox, R.T. Probability, frequency and reasonable expectation. *Am. J. Phys.* **1946**, *14*, 1–13.
2. Kolgomorov, A.N. *Foundations of the Theory of Probability*, 2nd ed.; Chelsea: New York, NY, USA, 1956.
3. Knuth, K.H. Inductive logic: From data analysis to experimental design. In *Bayesian Inference and Maximum Entropy Methods in Science and Engineering, Proceedings of the AIP Conference, Baltimore, ML, USA, 4–9 August 2001*; Fry, R.L., Ed.; American Institute of Physics: New York, NY, USA, 2002; Volume 617, pp. 392–404.
4. Knuth, K.H. What is a question? In *Bayesian Inference and Maximum Entropy Methods in Science and Engineering, Proceedings of the AIP Conference, Moscow, ID, USA, 3–7 August 2003*; Williams, C., Ed.; American Institute of Physics: New York, NY, USA, 2003; Volume 659, pp. 227–242.
5. Knuth, K.H. Intelligent machines in the twenty-first century: Foundations of inference and inquiry. *Philos. Trans. R. Soc. Lond.* **2003**, *361*, 2859–2873.
6. Knuth, K.H. Deriving laws from ordering relations. In *Bayesian Inference and Maximum Entropy Methods in Science and Engineering, Proceedings of the AIP Conference, Jackson Hole, WY, USA, 3–8 August 2003*; Erickson, G.J., Zhai, Y., Eds.; American Institute of Physics: New York, NY, USA, 2004; Volume 707, pp. 204–235.
7. Knuth, K.H. Lattice Duality: The origin of probability and entropy. *Neurocomputing* **2004**, *67*, 245–274.
8. Knuth, K.H. Valuations on lattices and their application to information theory. In Proceedings of the 2006 IEEE World Congress on Computational Intelligence, (IEEE WCCI 2006), Vancouver, BC, Canada, 24–29 July 2016.
9. Knuth, K.H. The origin of probability and entropy. In *Bayesian Inference and Maximum Entropy Methods in Science and Engineering, Proceedings of the AIP Conference, Sao Paulo, Brazil, 6–11 July 2008*; Lauretto, M.S., Pereira, C.A.B., Eds.; American Institute of Physics: New York, NY, USA, 2008; pp. 35–48.
10. Knuth, K.H. Measuring on lattices. In *Bayesian Inference and Maximum Entropy Methods in Science and Engineering, Proceedings of the AIP Conference, Oxford, MS, USA, 5–10 July 2009*; Goggans, P., Chan, C.Y., Eds.; American Institute of Physics: New York, NY, USA, 2009; Volume 1193, pp. 132–144.
11. Knuth, K.H.; Skilling J. Foundations of inference. *Axioms* **2012**, *1*, 38–73, doi:10.3390/axioms1010038.
12. Davey, B.A.; Priestley, H.A. *Introduction to Lattices and Order*; Cambridge University Press: Cambridge, UK, 2002.
13. Aczel, J. *Lectures on Functional Equations and Their Applications*; Academic Press: New York, NY, USA, 1966.
14. Cox, R.T. Of inference and inquiry, an essay in inductive Logic. In *The Maximum Entropy Formalism*; Levine R.D., Tribus, M., Eds.; The MIT Press: Cambridge, MA, USA, 1979; pp. 119–167.
15. Van Erp, H.R.N. Uncovering the specific product rule for the lattice of questions. *arXiv* **2013**, arXiv:1308.6303.

16. Shannon, C.E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–623.

17. Center, J.L. Inquiry calculus and information theory. In *Bayesian Inference and Maximum Entropy Methods in Science and Engineering, Proceedings of the AIP Conference, Oxford, MS, USA, 5–10 July 2009;* Goggans, P., Chun, C.Y., Eds.; American Institute of Physics: New York, NY, USA, 2010; pp. 69–78.

18. Wickens, T.D. *Multiway Contingency Tables Analysis for the Social Sciences*; Lawrence Erlbaum Associates, Inc.: Mahwah, NJ, USA, 1989.

19. Skilling, J. The Canvas of Rationality. In *Bayesian Inference and Maximum Entropy Methods in Science and Engineering, Proceedings of the AIP Conference, Sao Paulo, Brazil, 6–11 July 2008*; Lauretto, M.S., Pereira, C.A.B., Eds.; American Institute of Physics: New York, NY, USA, 2008; pp. 67–79.

20. Jaynes, E.T. *Probability Theory: The Logic of Science*; Cambridge University Press: Cambridge, UK, 2003.

21. Schreiber, T. Measuring Information Transfer. *Phys. Rev. Lett.* **2000**, *85*, 461–464.