The Quest for Uncertainty

Jörg Zimmermann and Armin B. Cremers

Institute of Computer Science,
University of Bonn, Germany
{jz, abc}@iai.uni-bonn.de

Abstract. The question of how to represent and process uncertainty is of fundamental importance to the scientific process, but also in everyday life. Currently there exist a lot of different calculi for managing uncertainty, each having its own advantages and disadvantages. Especially, almost all are defining the domain and structure of uncertainty values a priori, e.g., one real number, two real numbers, a finite domain, and so on, but maybe uncertainty is best measured by complex numbers, matrices or still another mathematical structure. Here we investigate the notion of uncertainty from a foundational point of view, provide an ontology and axiomatic core system for uncertainty, derive and not define the structure of uncertainty, and review the historical development of approaches to uncertainty which have led to the results presented here.

1 Introduction

The quest for a theory of inductive logic, i.e., a logic defining the relationship between observations and hypotheses, lies at the heart of the scientific process. Accordingly, there is a plethora of research aiming at the clarification of this relationship (probability theory as Bayesian theory [Jay03], possibility theory [Dub06], Dempster-Shafer theory [Sha76], revision theory [Gär92], ranking theory [Spo09], non-monotonic logic [Gin87], ...). The application of probability theory to the problem of inductive logic is known as Bayesian inference. Despite its intuitive appeal and many successful applications, it was never considered as a solution to the problem of induction because of technical and philosophical problems. In fact, the 20th century witnessed a strong rejection of probability theory as a theory for induction. Probability theory was developed to describe the randomness of observable events, not the plausibility of unobservable hypotheses. The randomness of events can be seen as an objective property of a physical system, whereas the plausibility of hypotheses is intrinsically subjective, depending on the knowledge of an "observer". A first attempt to directly axiomatize the intuition of reasoning under uncertainty was made by Richard T. Cox in 1946 [Cox46], but despite its important role as a starting point for a new branch of mathematical, subjective uncertainty theory, Cox's axiom system has drawbacks which have prevented it from becoming a generally accepted axiomatization of uncertainty measures, most notably its assumption that uncertainty values must be measured by real numbers. Addressing this issue, it is an important goal to define alternative axiom systems with a reduced number of controversial assumptions and investigate their implications. One important contribution to this endeavor is the axiom system by S. Arnborg and G. Sjödin [AS01], which has inspired the axiom system introduced in this article.

2 An Ontology of Uncertainty

In the realm of empirical knowledge, uncertainty is unavoidable. A piece of information is in general not known to be true or false, but must be annotated by shades of certainty. But what exactly is the structure of these "shades of certainty"? Are there ontologically different types of uncertainty, and, after all, how to assess, process and communicate uncertainty? One early distinction of types of uncertainty was introduced by Frank Knight in his seminal book "Risk, Uncertainty, and Profit" [Kni21] on page 19:

"Uncertainty must be taken in a sense radically distinct from the familiar notion of risk, from which it has never been properly separated.... The essential fact is that 'risk' means in some cases a quantity susceptible of measurement, while at other times it is something distinctly not of this character; and there are far-reaching and crucial differences in the bearings of the phenomena depending on which of the two is really present and operating.... It will appear that a measurable uncertainty, or 'risk' proper, as we shall use the term, is so far different from an unmeasurable one that it is not in effect an uncertainty at all."

In today's language one would describe "risk" as the uncertainty about the occurrence of events *within* a fully specified stochastic model. The "Knightian Uncertainty" is the uncertainty with regard to the correct model, what is today sometimes called model risk, especially in financial mathematics.

In the next paragraph we introduce an ontology of uncertainty, and, even more general, an ontology of indefiniteness, accompanied by a suitable terminology.

2.1 Indefiniteness

The advance of research in artificial intelligence, knowledge representation and expert systems has led to a plethora of new approaches to represent and process information: for example possibility theory, certainty factors, and non-monotonic logics. This has led to confusion about the exact differences and commonalities between the different calculi, and where they are competing approaches and where they are complementary. One striking example is fuzzy logic, which is still regarded as an alternative calculus for processing uncertain information, where in fact it is a generalization of the notion of an event. This is clearly stated by

Judea Pearl in [Pea00]: "Fuzzyness is orthogonal to probability theory - it focuses on the ambiguities in describing events, rather than the uncertainty about the occurrence or non-occurrence of events." Classical events are called crisp, in order to express that they are definitely defined: in a specific situation the event has occurred or not – there are no "degrees of occurrence". The standard approach to represent a set of crisp events is a Boolean algebra. In this sense, one can say that a crisp event is an element of a Boolean algebra.

We suggest the notion "indefiniteness" for describing all sorts of non-certain, non-crisp information. This leads to the following ontology of indefiniteness:

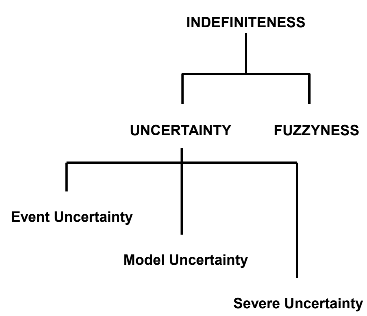

Fig. 1. Ontology of Indefiniteness

.

2.2 Types of Uncertainty

Here we propose three types of uncertainty, extending the Knightian ontology:

1. Event Uncertainty (quantitatively known unknowns)
2. Model Uncertainty (qualitatively known unknowns)
3. Severe Uncertainty (unknown unknowns)

We want to illustrate these three types of uncertainty – and their principal differences – with an example taken from Bernoulli processes:

Event Uncertainty: Consider the coin model with $p = \frac{1}{2}$. The question what will be the next outcome of an observation can be answered by a definite probability. In this case the probability is $\frac{1}{2}$, meaning that we are maximally unsure what will happen next, even under a specific, complete stochastic model, but other questions can be answered with more certainty by the coin model. For example, the probability that we will observe 450 to 550 heads out of 1000 tosses of the coin is greater than 0.998. So, for this specific question the coin model delivers an answer with near certainty.

Model Uncertainty: Here we assume that the observations are generated by a Bernoulli process, but with unknown success probability p. Without introducing a prior distribution for the model parameter, this implies that we only can infer probability intervals for events, for example the probability that we will observe between 45 and 55 successes out of 100 experiments is in the interval $[0, .71]$, regardless of the value of p.

Severe Uncertainty: This is the "black swan" case, the possibility, that the true model is not even approximately in the set of considered models. An example would be that the true process is a deterministic switch between successes and failures, leading to a probability of 1 for the above example.

The case of severe uncertainty leads to the question of how to describe all possible models. If one requires that a model has to be an algorithmic object, the answer to this question is the set of all programs, also called program space. R. Solomonoff pioneered learning in program space in the 1960s, employing a Bayesian framework for describing model uncertainty and a prior distribution on programs inspired by Occam's razor [Sol64a,Sol64b]. Unfortunately, despite the fact that all models have to be represented by programs, the learning process devised by Solomonoff for the whole program space is not computable. The question of how to essentially retain the generality of Solomonoff's approach, but render the learning process computable has spawned a research area of its own, which is today called universal induction or algorithmic probability [Hut05,Sch09].

3 Formalizing Uncertainty

First we have to discuss a subtle issue of terminology. Above we have used the notion "uncertainty values" to denote generalized truth values. Unfortunately, there is the following problem when using this term in a formalized context: no uncertainty about a proposition can be identified with sure knowledge, but maximal uncertainty about a proposition is *not* certainty with regard to the negation of the proposition. The domains of truth values we want to axiomatize contain a greatest and a least element, where the greatest element should represent certainty and the least element impossibility, i.e. certainty of the negated proposition. For this reason, we adopt the notion "confidence measure" instead of uncertainty measure in the following definitions and axioms.

3.1 The Algebra of Truth Bearers

Before delving into the structure of uncertainty, we have to define the objects and their relations which are capable to take on truth values, the *truth bearers*. In a context of crisp events, i.e., after the fact it is unambiguously decidable if the event has occurred or not, the algebra of truth bearers is normally considered to be a Boolean algebra, but when truth bearers are not crisp, then another algebra has to be used, i.e., a fuzzy algebra where the law of the excluded middle is not valid: $x \vee \neg x \neq 1$.

However, for the purpose of the present article we focus on Boolean algebras as the structure of propositions. The investigation of uncertainty measures for non-Boolean proposition algebras is open to future research.

3.2 Uncertainty: the Boolean Case

A *conditional confidence measure* for a Boolean Algebra $\mathbf{U}$ and a domain of confidence values $\mathcal{C}$ is a mapping $\Gamma : \mathbf{U} \times \mathbf{U} \setminus \{\bot\} \to \mathcal{C}$. Let $A, B \in \mathbf{U}$, then the expression $\Gamma(A|B)$ reads: "the confidence value of A given B (wrt. Γ)". The domain of confidence values is partially ordered and has a greatest ($\top\!\!\!\top$) and a least ($\bot\!\!\!\bot$) element. A *confidence space* is a triple $(\mathbf{U}, \Gamma, \mathcal{C})$. One of the following axioms (Extensibility) for confidence measures deals with relations between confidence spaces defined over different Boolean algebras. Thus it is necessary to introduce a *set of confidence spaces* all sharing the same domain of confidence values. Such a set of confidence spaces we will call a *confidence universe*, and the following axiom system is concerned with such confidence universes, and not single confidence spaces. This seemingly technical shift in perspective is essential for the formalization of natural properties like extensibility, which plays a crucial role as an intuitive axiom complementing Cox's assumptions (see section 5).

We now state seven axioms, which can be grouped in three "connective axioms" and four "infrastructure axioms", where the connective axioms concern properties of the logical connectives and the infrastructure axioms deal with basic properties of the order relations, the combinability of confidence spaces and a closure property.

3.3 The Core of Uncertainty

In the following, we use $\Gamma(A)$ as an abbreviation for $\Gamma(A|\top)$.

(Not) For all $(\mathbf{U_1}, \Gamma_1, \mathcal{C})$ and $(\mathbf{U_2}, \Gamma_2, \mathcal{C})$:
If $\Gamma_1(A_1) = \Gamma_2(A_2)$, then $\Gamma_1(\bar{A}_1) = \Gamma_2(\bar{A}_2)$.

The axiom **Not** expresses that the information in the confidence value of a statement A is sufficient to determine the confidence value of $\bar{A}$. This is justified by the requirement that every piece of information which is relevant for the confidence value of A is relevant for the confidence value of $\bar{A}$ and vice versa.

(And$_1$) For all $(\mathbf{U_1}, \Gamma_1, \mathcal{C})$ and $(\mathbf{U_2}, \Gamma_2, \mathcal{C})$:
If $\Gamma_1(A_1|B_1) = \Gamma_2(A_2|B_2)$ and $\Gamma_1(B_1) = \Gamma_2(B_2)$,
then $\Gamma_1(A_1B_1) = \Gamma_2(A_2B_2)$.

The axiom **And$_1$** states that the information in the confidence values of the partial propositions determine the confidence value of the conjunction. Otherwise the confidence value of the conjunction would contain information which is not reflected in the partial propositions, although this information would be clearly relevant for at least one of them.

(And$_2$) For all $(\mathbf{U_1}, \Gamma_1, \mathcal{C})$ and $(\mathbf{U_2}, \Gamma_2, \mathcal{C})$:
If $\Gamma_1(A_1B_1) = \Gamma_2(A_2B_2)$ and $\Gamma_1(B_1) = \Gamma_2(B_2) \neq \perp\!\!\!\perp$,
then $\Gamma_1(A_1|B_1) = \Gamma_2(A_2|B_2)$.

The axiom **And$_2$** ensures that all the information contained in a conditional confidence value $\Gamma(A|B)$ will be preserved in the confidence value of the conjunction $\Gamma(AB)$ when combined with the confidence $\Gamma(B)$ (unless $\Gamma(B) = \perp\!\!\!\perp$, in which case the value of $\Gamma(A|B)$ is irrelevant). Otherwise relevant information about the partial propositions would not be contained in the confidence value of the conjunction.

(Order$_1$) For all $(\mathbf{U}, \Gamma, \mathcal{C})$ and all $A, B \in \mathbf{U}$: If $A \leq B$, then $\Gamma(A) \leq \Gamma(B)$.

(Order$_2$) For all confidence values $v, w \in \mathcal{C}$ with $v \leq w$ there is a confidence space $(\mathbf{U}, \Gamma, \mathcal{C})$ with $A, B \in \mathbf{U}$ and $A \leq B$, $\Gamma(A) = v$, $\Gamma(B) = w$.

These two axioms connect the natural ordering of the Boolean algebra ($A \leq B$ iff $A \wedge B = A$) with the ordering on the confidence domain, where **Order$_1$** specifies the forward direction and **Order$_2$** specifies the backward direction.

(Extensibility) For all $(\mathbf{U_1}, \Gamma_1, \mathcal{C})$ and $(\mathbf{U_2}, \Gamma_2, \mathcal{C})$ there is a confidence space $(\mathbf{U_3}, \Gamma_3, \mathcal{C})$, so that $\mathbf{U_3} \cong \mathbf{U_1} \otimes \mathbf{U_2}$, and for all $A_1, B_1 \in \mathbf{U}_1$, $A_2, B_2 \in \mathbf{U}_2$:

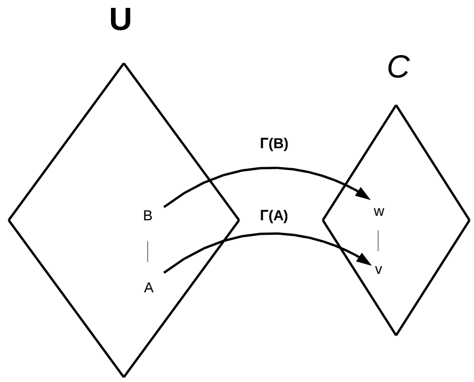

Fig. 2. Ordered confidence values v and w with corresponding propositions in a suitably chosen confidence space $(\mathbf{U}, \Gamma, \mathcal{C})$

.

$$\Gamma_3(A_1 \otimes \top_2 \,|\, B_1 \otimes B_2) = \Gamma_1(A_1|B_1) \quad \text{and} \quad \Gamma_3(\top_1 \otimes A_2 \,|\, B_1 \otimes B_2) = \Gamma_2(A_2|B_2).$$

This axiom requires the extensibility of domains of discourse, i.e., two independently defined confidence spaces shall be embeddable into one frame of reference.

(Background) For all $(\mathbf{U}, \Gamma_1, \mathcal{C})$ and all $C \in \mathbf{U}$ there is a confidence space $(\mathbf{U}, \Gamma_2, \mathcal{C})$, so that for all $A, B \in \mathbf{U}$:

$$\Gamma_1(A|BC) = \Gamma_2(A|B).$$

This closedness under conditioning assures that for every conditional confidence measure Γ_1 which is specialized by conditioning on some "background knowledge" C, there is a conditional confidence measure Γ_2 yielding the same valuations without explicitly mentioning C.

For the justification of the axioms it is important to interpret the expression $\Gamma(A|B)$ as: "*all* that can be said about the confidence of A given B (wrt. Γ)." Given this interpretation, the common justification of the connective axioms is that a violation of these axioms will necessarily lead to a loss of relevant information. Note that the axioms use only equations and inequalities between confidence values, because there are no algebraic operations defined on the domain of confidence values yet.

In order to designate this and similar axiom systems, we propose a nomenclature based on the connnective axioms. Extensionality of negation, conjunction, and disjunction is denoted as axiom N, C_1, and D_1, respectively. The reconstructibility of the confidence value of an argument of a conjunction or a disjunction, given the compositional confidence value and the confidence value of the other argument, is denoted as axiom C_2 and D_2, respectively. Using this terminology, the above introduced axiom system can be referenced as NC_{12}.

4 The Structure of Uncertainty

A first important implication of the NC_{12} axioms is stated in the following theorem.

SFG Theorem: There exist functions $S : \mathcal{C} \to \mathcal{C}$, $F : \mathcal{C}^2 \to \mathcal{C}$ and $G : \{(x, y) \in \mathcal{C}^2 | x \leq S(y)\} \to \mathcal{C}$ with:

$$\Gamma(\bar{A}) = S(\Gamma(A)), \tag{1}$$

$$\Gamma(A \wedge B) = F(\Gamma(A|B), \Gamma(B)), \tag{2}$$

$$\Gamma(A \vee B) = G(\Gamma(A), \Gamma(B)), \qquad \text{if } AB = \perp \tag{3}$$

Proof: First we prove a lemma stating that for every pair of confidence values there is a confidence measure and two independent propositions such that the confidence measure assigns the given confidence values to these propositions.

Independence Lemma: For all $v, w \in \mathcal{C}$ there is a confidence space $(\mathbf{U}, \Gamma, \mathcal{C})$ with $A, B \in \mathbf{U}$, such that:

$$\Gamma(A|B) = \Gamma(A) = v \quad \text{and} \quad \Gamma(B|A) = \Gamma(B) = w.$$

Proof: According to $\mathbf{Order}_2$, there are confidence spaces $(\mathbf{U_1}, \Gamma_1\mathcal{C})$, $(\mathbf{U_2}, \Gamma_2, \mathcal{C})$ and propositions $A \in \mathbf{U_1}$ and $B \in \mathbf{U_2}$ with $\Gamma_1(A) = v$ and $\Gamma_2(B) = w$. Then axiom **Extensibility** guarantees the existence of a confidence space $(\mathbf{U_1} \otimes \mathbf{U_2}, \Gamma_3, \mathcal{C})$ with:

$$\Gamma_3(A|B) = \Gamma_1(A) = v \quad \text{and} \quad \Gamma_3(B|A) = \Gamma_2(B) = w.$$

□

Next we show that axioms **Not** and $\mathbf{And}_1$ imply the existence of the functions S and F.

Let $v \in \mathcal{C}$ be a confidence value, so that there is a confidence space $(\mathbf{U_1}, \Gamma_1, \mathcal{C})$ and $A_1 \in \mathbf{U_1}$ with $\Gamma_1(A_1) = v$. Then define:

$$S(v) = \Gamma_1(\bar{A}_1)$$

This function is well-defined, because whenever there is another confidence space $(\mathbf{U_2}, \Gamma_2, \mathcal{C})$ having v as value of the confidence measure Γ_2, say $\Gamma_2(A_2|B_2) = v$, then axiom **Not** assures that $\Gamma_2(\bar{A}_2|B_2) = \Gamma_1(\bar{A}_1)$. That is, the value of S does not depend on the specific choice of confidence space having v as a value. Additionally, S is a total function by axiom $\mathbf{Order}_2$, which enforces that for every $v \in \mathcal{C}$ there is at least one confidence space taking v as a value.

The analog will be proved for conjunction by introducing a binary function F. Note that a proposition B and a *conditional* proposition $A|B$ are related by F.

According to the independence lemma, for all $v, w \in \mathcal{C}$ there is a confidence space $(\mathbf{U_1}, \Gamma_1)$ with $\Gamma_1(A_1|B_1) = \Gamma_1(A_1) = v$ and $\Gamma_1(B_1) = w$. Define F as follows:

$$F(v, w) = \Gamma_1(A_1 B_1)$$

The well-definedness is implied by axiom $\mathbf{And}_1$, the value of $F(v, w)$ does not depend on the confidence space and events having v and w as confidence values. The totality of F is assured by the independence lemma, too, which is valid for all $v, w \in \mathcal{C}$.

Lemma: F is associative.

Proof: Let $x, y, z \in \mathcal{C}$ and $(\mathbf{U}, \Gamma, \mathcal{C})$ a confidence space with $A, B, C \in \mathbf{U}$ and $\Gamma(A|BC) = \Gamma(A) = x$, $\Gamma(B|C) = \Gamma(B) = y$ and $\Gamma(C) = z$. Such a confidence space always exists according to the independence lemma. Then it follows:

$$F(F(x, y), z) = F(F(\Gamma(A), \Gamma(B)), \Gamma(C)) = F(F(\Gamma(A|BC), \Gamma(B|C)), \Gamma(C)) =$$

$$= F(\Gamma(AB|C), \Gamma(C)) = \Gamma(ABC) = F(\Gamma(A|BC), \Gamma(BC)) =$$

$$= F(\Gamma(A), F(\Gamma(B|C), \Gamma(C))) = F(\Gamma(A), F(\Gamma(B), \Gamma(C))) = F(x, F(y, z)).$$

□

In the same way, by using the independence lemma to construct the appropriate confidence spaces, one can show that F is commutative and has $\top\!\!\!\top$ as a neutral element. Next we derive the cancellation property for F on $\mathcal{C}^+ = \mathcal{C} \setminus \{\perp\!\!\!\perp\}$.

Lemma: F is cancellative on $\mathcal{C}^+$, i.e., $F(x, z) = F(y, z)$ implies $x = y$.

Proof: Let $x, y, z \in \mathcal{C}^+$ and $(\mathbf{U}, \Gamma, \mathcal{C})$ be a confidence space with $A, B, C, D \in \mathbf{U}$ and $\Gamma(A|C) = \Gamma(A) = x$, $\Gamma(B|D) = \Gamma(B) = y$, and $\Gamma(C) = \Gamma(D) = z$, again using the independence lemma to show the existence of such a confidence space. Then we have $F(x, z) = \Gamma(AC)$ and $F(y, z) = \Gamma(BD)$. Thus, $F(x, z) = F(y, z)$ implies $\Gamma(AC) = \Gamma(BD)$. Invoking And_2 (without worrying about the case $z = \perp\!\!\!\perp$, because we are talking about $\mathcal{C}^+$), we get $\Gamma(A|C) = \Gamma(B|D)$, i.e., $x = y$.

□

The next step is the extension of the monoid $(\mathcal{C}^+, F)$ to a group. This can be done analogously to the classical algebraic construction of $\mathbf{Z}$ from $\mathbf{N}$, a construction which works for all cancellative commutative monoids.

Using S, F, and F^{-1}, the partial function G on $\{(x,y) \in \mathcal{C}^2 | x \leq S(y)$ is defined as follows:

$$G(x,y) = S(F(S(F(x, F^{-1}(S(y)))), S(y))), \qquad \text{if } y \neq \top\!\!\top$$
$$G(\perp\!\!\perp, \top\!\!\top) = \top\!\!\top, \qquad \text{else}$$

In order to illustrate this definition, we note that G can be seen as a solution of the problem to represent addition with the functions $x * y$, $1 - x$, and $1/x$. Using these functions, G translates into:

$$1 - (1 - \tfrac{x}{1-y})(1 - y)$$

which reduces to addition.

First we have to show that this is a well-defined function. For this, we have to establish that on the domain of G the expression $F(x, F^{-1}(S(y)))$ is in $\mathcal{C}$, because the S-function is still only defined on $\mathcal{C}$, and not on the group extension.

Lemma: $\forall x, y \in \mathcal{C},\ y \neq \top\!\!\top : x \leq S(y) \ \Rightarrow\ F(x, F^{-1}(S(y))) \in \mathcal{C}$.

Proof: With Order_2 and $x \leq S(y)$ it follows that there is a confidence space $(\mathbf{U}, \Gamma, \mathcal{C})$ with $A, B \in \mathbf{U}$, $A \leq B$, $\Gamma(A) = x$, and $\Gamma(B) = S(y)$. Now, because of $A \leq B$, it holds that $\Gamma(AB) = \Gamma(A) = x$. Let $\Gamma(A|B) = z$, which is uniqely determined according to And_2. z satisfies the equation $x = F(z, S(y))$, which is equivalent to $z = F(x, F^{-1}(S(y)))$. Because z is, by definition, in the range of a confidence measure, we have $z \in \mathcal{C}$ and hence $F(x, F^{-1}(S(y)) \in \mathcal{C}$.

□

Having the well-definedness of G established, it is easy to show that G has the desired property. Let $\Gamma(B) \neq \top\!\!\top$ and $AB = \perp$:

$$\Gamma(A \vee B) = S(\Gamma(\bar{A} \wedge \bar{B})) = S(F(S(\Gamma(A|\bar{B})), S(\Gamma(B))))) =$$
$$= S(F(S(F(\Gamma(A \wedge \bar{B}), F^{-1}(S(\Gamma(B))))), S(\Gamma(B))))$$

Now $A \wedge \bar{B}$ is equal to A, because we assumed $AB = \perp$. Hence G has the stated property for $\Gamma(B) \neq \top\!\!\top$. In the case of $\Gamma(B) = \top\!\!\top$, we invoke Order_1 to show that $\Gamma(A \vee B) = \top\!\!\top$, too. Furthermore, because $AB = \perp$ implies $A \leq \bar{B}$, we have, again by Order_1, $\Gamma(A) \leq \Gamma(\bar{B}) = S(\Gamma(B)) = \perp\!\!\perp$. Hence $\Gamma(A) = \perp\!\!\perp$, and we can apply the second part of the definition of G, which yields $\Gamma(A \vee B) = G(\Gamma(A), \Gamma(B)) = G(\perp\!\!\perp, \top\!\!\top) = \top\!\!\top$. This finishes the proof of the SFG theorem.

□

Arnborg and Sjödin proved in [AS01] a theorem clarifying the algebraic structure of $\mathcal{C}$ resulting from their axioms: it is the $[0, 1]$-interval of a totally ordered field.

Analyzing their proof, we find that the construction of a field from a ring will fail if one does not assume a total order on $\mathcal{C}$. In lemma 13 of [AS01] they state that the ring they have constructed is a totally ordered integral domain, i.e. a ring without zero divisors. Then they use a theorem from S. MacLane and G. Birkhoff in [MB67] which states that every totally ordered integral domain can be embedded in a totally ordered field. But this will not work in the case of partial order because without the total order assumption one cannot prove that the constructed ring will not contain zero divisors. So, lemma 13 of [AS01] cannot be transferred to the partial order case, which blocks the application of the MacLane-Birkhoff theorem. This is an interesting example of the interplay between order properties and algebraic properties: a total order assumption has strong algebraic implications, while partial order has not. Accordingly, order properties and algebraic properties cannot, as one might have hoped, treated separately. Based on these observations we formulate the following conjecture:

Ring Conjecture: The domain of confidence values $\mathcal{C}$ of a confidence universe satisfying the axiom system NC_{12} can be embedded in a partially ordered ring $(\hat{\mathcal{C}}, 0, 1, \oplus, \odot, \leq)$. Let $\hat{\ } : \mathcal{C} \to \hat{\mathcal{C}}$ be the embedding map, then the following holds:

$$\hat{\perp\!\!\!\perp} = 0, \quad \hat{\top\!\!\!\top} = 1, \quad \forall v, w \in \mathcal{C} : \; v \leq w \;\Leftrightarrow\; \hat{v} \leq \hat{w} \,.$$

Furthermore, all confidence measures Γ of the confidence universe satisfy:

$$\hat{\Gamma}(\top) = 1 \tag{4}$$
$$\hat{\Gamma}(A \vee B) = \hat{\Gamma}(A) \oplus \hat{\Gamma}(B), \qquad \text{if } AB = \bot \tag{5}$$
$$\hat{\Gamma}(A \wedge B) = \hat{\Gamma}(A|B) \odot \hat{\Gamma}(B) \tag{6}$$

We state this as a conjecture, because a full proof is beyond the scope of this article. A proof outline can be found in [Zim10], which documents work in progress. If it can be confirmed, it will yield an algebraic characterization of uncertainty based on NC_{12}:

Uncertainty can be represented by elements of the $[0, 1]$*-interval of partially ordered rings.*

Furthermore, with regard to the ring operations, the uncertainty measures satisfy the same axioms as probability measures satisfy with regard to the real numbers. But in contrast to the real numbers, a ring may be only partially ordered or may contain infinitesimal elements, like the hyperreal numbers (see section 6).

5 The Lineage of NC_{12}

Our approach of axiomatizing uncertainty measures extends a line of thinking started by R. T. Cox in 1946. In [Cox46], based on axioms which should hold for all uncertainty measures, Cox derived a theorem stating that uncertainty measures are essentially probability measures, although his axioms are very different from the axioms of probability theory. A recent exposition of his result can be found in [Jay03].

The approach used by Cox was one of the first attempts to justify the use of probabilities as a representation of uncertainty by directly axiomatizing the intuition on uncertainty measures and then *deriving* that uncertainty measures have the same mathematical structure as probability measures. This was a surprising result, given the fact that Cox's axioms look totally different from the Kolmogorov axioms of probability theory. But despite its new and far reaching conclusions, Cox's theorem was not widely acknowledged. This can be attributed to at least two factors: first, it became clear that Cox's derivation of his theorem was not complete. The assumptions he made were not sufficient to reach the conclusion in its full generality. This was noted by several authors, and J. Halpern showed in detail where Cox's proof failed by constructing a counterexample in [Hal99]. It was not before 1994 that J.B. Paris completed Cox's proof by introducing a new axiom [Par94]. This axiom closes the loopholes in Cox's proof, but is very technical in nature. Thus it is not acceptable as an axiom which should hold for all reasonable uncertainty measures. This leads to the second factor contributing to the slow adoption of Cox's result: there is at least one axiom which is too strong to be considered as a general property of uncertainty measures, yet is inherently necessary for the proof approach adopted by Cox. This axiom is the assumption that uncertainty can be measured by one real number. This is a strong structural assumption, implying that the uncertainty values are totally ordered. This prevents, for example, the applicability of Cox's theorem to calculi like Dempster-Shafer theory, which uses two real numbers for the representation of uncertainty.

The remaining question after the result of J. B. Paris is the following: are there extensions or modifications of the Cox axioms, which are justifiable as general properties of uncertainty measures and which imply a result essentially similar to Cox's theorem? One important step in this direction was taken by S. Arnborg and G. Sjödin. They replaced the axiom introduced by J.B. Paris by a more intuitive statement which they called "Refinability axiom". Furthermore, they dropped the requirement that uncertainty values are real numbers. By this step, they transformed the Cox approach to a genuine algebraic approach, constructing the structure of the domain of uncertainty values and not assuming it. But in order to get the result they want, they introduced a total of 16 axioms (when one counts every discernible requirement they formulate as a separate axiom, as we do for our core system), with different degrees of foundational justifiability. Additionally, at a crucial step in their proof they introduce a total order assumption for the domain of uncertainty values, thus restricting the range of their result in a fundamental way.

This was the situation where we entered the development, seeing that Arnborg and Sjödin made a crucial step in the amelioration of the original Cox's approach, but still leaving some major issues open, which have blocked the general applicability of their result. Accordingly, our goal was the following: to devise an axiom system as minimal as possible, with as weak and as general properties as possible, especially to drop the total order assumption, but still be able to derive a Cox-style result.

6 Relations to existing Uncertainty Calculi

Today, there exist many approaches for dealing with uncertainty, for example lower probabilities, which have only partially ordered uncertainty values or non-monotonic logic, which can be interpreted as using infinitesimal probabilities. In the following, we try to analyse these calculi in the light of our results.

6.1 Lower Probabilities

The problem of dealing with "imprecise" probabilities has led to the development of calculi known under the common name "lower probabilities". The main distinction from the probability calculus is that the uncertainty of a proposition is judged by *two* numbers instead of one. Accordingly, there are two functions mapping the elements of a proposition algebra to $[0, 1]$, the *lower probability* P_* and the *upper probability* P^*. The most general notion of a lower probability is defined wrt. a set of probability distributions $\mathcal{P}$ (see, for example, [Hal03]):

$$P^*(A) = \sup_{P \in \mathcal{P}} P(A) \quad \text{and} \quad P_*(A) = \inf_{P \in \mathcal{P}} P(A).$$

One can show that lower and upper probabilities satisfy the following inequalities if A and B are disjoint:

$$P_*(A \cup B) \geq P_*(A) + P_*(B) \quad \text{and} \quad P^*(A \cup B) \leq P^*(A) + P^*(B).$$

These properties are called super-additivity and sub-additivity, respectively. Furthermore, lower and upper probability are connected via the following relations:

$$P_*(A) \leq P^*(A) \quad \text{and} \quad P^*(A) = 1 - P_*(\bar{A}).$$

The inequality says that lower and upper probabilities can be seen as defining an interval, thus making lower and upper probabilities an uncertainty calculus having a partially ordered domain of uncertainty values. The equation implies that from both uncertainty values, upper and lower probability, of a proposition one can derive the upper and lower probabilities of its negation. Hence lower and upper probabilities together satisfy axiom Not.

An application of our results to the analysis of lower probabilities is now the following: even if the domain of uncertainty values is only partially ordered, which is possible according to NC_{12}, there exists a function G which relates the uncertainty value of a disjunction of disjoint propositions and the uncertainty values of the single propositions by an equation, and not only by an inequality. If no such function G exists for an uncertainty calculus, it must violate at least one of the axioms Not, And_1, or And_2 (we assume that the infrastructure axioms are satisfied). Now, because lower probabilities satisfy axiom Not, they must violate And_1 or And_2. This implies that there cannot be any definition of conditioning for lower probabilities which satisfies And_1 and And_2. Seeing And_1 and And_2 as essential conditions for not loosing relevant information, this may explain why the definiton of conditioning for lower probabilities has turned out to be such a hard problem, which is still the topic of ongoing research.

This conclusion is also valid for Dempster-Shafer theory, which can be seen as lower and upper probabilities satisfying additional constraints. Accordingly, there are several proposals for conditioning in DS-theory, each having its own advantages and disadvantages. By the above analysis, this is not a transitory state until the "right" conditioning rule has been found, but a fundamental obstacle which cannot be resolved within the frame of DS-theory.

6.2 Non-monotonic Logic

A non-monotonic logic extends classical logic with a framework of "belief revision", i.e. conclusions derived at one point can be retracted at a later point. Non-monotonic logic can be seen as defining a hierarchy of "default assumptions", which are assumed valid until observed evidence directly contradicts them. If this happens, a revision process is executed, which incorporates the new evidence and eliminates contradictions while trying to preserve as much as possible from the old knowledge state. Now, as for example Lehman and Magidor have observed in [LM92], one can formalize default expressions of the type "if A then typically B" as "the probability of B given A is very high", where "very high" is equated to $1 - \epsilon$, for *infinitesimal* ϵ. This can be modeled by a generalized probability algebra using the $[0, 1]$-interval of *hyperreal* numbers as a domain of uncertainty values.

7 Conclusion

Despite many attempts, there is still no consensus on basic questions concerning uncertainty and the foundations of inductive logic. In [AS01], Arnborg and Sjödin note that reaching a consensus is not only a foundational issue but is also important outside the ivory tower: designers of complex systems struggle with difficult compatibility problems when they plan to integrate system components which happen to use different ways to describe uncertainty.

In this article, we have tried to contribute to the debate on uncertainty by discerning ontologically different types of uncertainty and introducing an axiomatic core system for uncertainty measures with the explicit aim not to prejudice structural properties of the domain of uncertainty values, but to derive them from basic assumptions.

References

[AS01] S. Arnborg and G. Sjödin. What is the plausibility of probability? *Preprint, Nada, KTH*, 2001.

[Cox46] R. T. Cox. Probability, frequency, and reasonable expectation. *Am. Jour. Phys.*, 14:1–13, 1946.

[Dub06] D. Dubois. Possibility theory and statistical reasoning. *Computational Statistics & Data Analysis*, 51(1):47–69, 2006.

[Gär92] P. Gärdenfors, editor. *Belief Revision*. Cambridge University Press, 1992.

[Gin87] M. Ginsberg, editor. *Readings in Nonmonotonic Reasoning*. Morgan Kauffman, Los Altos, CA, 1987.

[Hal99] J. Halpern. A counterexample to theorems of Cox and Fine. *Journal of A.I. Research*, 10:76–85, 1999.

[Hal03] J. Halpern. *Reasoning about Uncertainty*. MIT Press, 2003.

[Hut05] M. Hutter. *Universal Artificial Intelligence*. Springer, 2005.

[Jay03] E. T. Jaynes. *Probability Theory: The Logic of Science*. Cambridge University Press, 2003.

[Kni21] F. Knight. *Risk, Uncertainty, and Profit*. Houghton Mifflin, 1921.

[LM92] D. Lehman and M. Magidor. What does a conditional knowledge base entail? *Artificial Intelligence*, 55(1):1–60, 1992.

[MB67] S. MacLane and G. Birkhoff. *Algebra*. The MacMillan Company, 1967.

[Par94] J. B. Paris. *The Uncertain Reasoner's Companion*. Cambridge University Press, 1994.

[Pea00] J. Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 2000.

[Sch09] J. Schmidhuber. Ultimate cognition à la Gödel. *Cognitive Computation*, 1(2):177–193, 2009.

[Sha76] Glenn Shafer. *Mathematical Theory of Evidence*. Princeton University Press, 1976.

[Sol64a] R. Solomonoff. A formal theory of inductive inference, part I. *Information and Control*, 7(1):1–22, 1964.

[Sol64b] R. Solomonoff. A formal theory of inductive inference, part II. *Information and Control*, 7(2):224–254, 1964.

[Spo09] W. Spohn. A survey of ranking theory. In F. Huber and C. Schmidt-Petri, editors, *Degrees of Belief*. Springer, 2009.

[Zim10] J. Zimmermann. A proof outline for the ring conjecture. *http://www.iai.uni-bonn.de/˜jz/ring_conjecture.pdf*, 2010.