

Non-standard theories of uncertainty in knowledge representation and reasoning⁽¹⁾

DIDIER DUBOIS and HENRI PRADE

Institut de Recherche en Informatique de Toulouse (IRIT)—CNRS Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse Cedex, France

Abstract

This paper provides a survey of the state of the art in plausible reasoning, that is exception tolerant reasoning under incomplete information. Three requirements are necessary for a formalism in order to cope with this problem: (i) making a clear distinction between factual information and generic knowledge; (ii) having a correct representation of partial ignorance; (iii) providing a nonmonotonic inference mechanism. Classical logic fails on requirements (i) and (iii), whilst the Bayesian approach does not fulfil (ii) in an unbiased way. In this perspective, various uncertainty modelling frameworks are reviewed: MYCIN-like fully compositional calculi, belief functions, upper and lower probability systems, and possibility theory. Possibility theory enables *classical* logic to be extended to layered sets of formulae, where layers express certainty levels. Finally, it is explained how generic knowledge can be expressed by constraints on possibility measures, and how possibilistic inferences can encode nonmonotonic reasoning in agreement with the Lehmann et al. postulates.

1 Introduction

The last 15 years have witnessed a noticeable research effort towards a rational theory of exception-tolerant reasoning. However, the literature on this research is spread through a wide range of publications, partially due to the diversity of scientific backgrounds of those active in the field. While probability theory has recently blossomed in this area with the emergence of Bayesian networks, there is still a prominent role to be played by logical and symbolic representations. In addition, the monopoly of probability theory has recently been challenged by a number of alternative numerical approaches, such as belief functions and possibility theory. Currently, work is focusing on the specification of a knowledge representation framework which combines the apparently incompatible merits of classical logic and Bayesian probability.

This paper provides a perspective view of uncertainty models in the context of exception-tolerant, plausible reasoning. This is achieved by stressing some of the important ideas and problems which have emerged as common to all approaches.

2 Exception-tolerant plausible inference

The problem considered in this paper is at the core of the whole knowledge-based systems enterprise, namely how to handle the presence of (possibly hidden) exceptions in the rule-base of an expert system. The kind of plausible reasoning that is involved here can be summarized as follows: *How can we automatically derive plausible conclusions about an incompletely specified situation, given a generic knowledge-base which represents what is “normally” the case?*

For example, in a medical expert system generic knowledge may encode relationships between symptoms and diseases. The situation at hand would be a given patient on which some test results are available. Given this generic knowledge and specific situation, some form of plausible inference

⁽¹⁾This paper is the revised version of an invited plenary talk at the 4th Conference on the Principles of Knowledge Representation and Reasoning, Bonn, April, 1994.

will then be required to perform, say, a diagnosis task. This kind of problem can be cast in more general terms as a form of taxonomic reasoning; generic knowledge describes links between classes and subclasses, and some factual evidence provides an incomplete description of an instance to be classified. The essence of the problem is that the generic knowledge encoded as a set of rules is pervaded with uncertainty due to the presence of exceptions to those rules.

To solve this problem in a satisfactory way, three requirements must be met:

1. *The necessity of a clear distinction between factual evidence and generic knowledge*

This distinction is fundamental and has been explicitly acknowledged in the expert systems literature at the implementation level (facts versus rules). The generic rules encode background knowledge which is used to “jump” to conclusions which would not be strictly inferable given only the available factual evidence. Clearly, assimilating a new piece of evidence does not produce the same effect as the arrival of a new rule or the modification of a rule. The arrival of new evidence does not affect the generic knowledge. Rather, it modifies the reference class of the case under study. In contrast, assimilating a new rule will cause a revision of the generic knowledge.

2. *The need for representing partial ignorance in an unbiased way*

There are three extreme epistemic attitudes with regard to a proposition p . On the basis of current evidence and background knowledge, one could be: sure that p is true; sure that p is false; or in a state of ignorance as to the truth-value of p . The third situation corresponds to partial ignorance. Its representation should not depend upon a simple count of situations in which p is true, since the count can depend upon how these situations are described. That is, the count is language-dependent.

3. *The inference at work cannot be monotonic*

As has been indicated above, a plausible reasoning system will not be cautious; it must go beyond the conclusions strictly entailed by incomplete evidence. This is achieved by making the assumption that the particular situation under study is as normal as possible. It is then possible to jump to adventurous, yet plausible, conclusions. The price paid for such deductive efficiency is that such conclusions may need to be withdrawn, should new evidence arrive which indicates that the current situation is not as normal as had been supposed. This is an obvious contradiction with the monotonicity property of classical logic, which prevents the retraction of existing conclusions as further axioms are added.

The classical solutions proposed by the expert systems literature were either based on the propagation of certainty coefficients (as in MYCIN and PROSPECTOR), or based on an explicit handling of the reasons for uncertainty at the control level. However, these solutions were partially *ad hoc*, motivating the development of further, better founded streams of work; namely Bayesian networks and nonmonotonic reasoning. While the first of these approaches was safely founded on a strong probabilistic tradition, the second line of research proved to be more adventurous, although fruitful. The next two sections will identify the limitations of classical logic and of Bayesian networks which restrict their applicability to exception-tolerant reasoning. However, it turns out that many of the lessons that have been learnt from the Bayesian network literature can be used to solve the exception-tolerant inference problem, whilst remaining within the tradition of logic.

3 Limitations of classical logic

The problems of exception handling in classical logic are well-known (e.g. Léa Sombé, 1990). The essential dilemma is that although we may wish to express generic knowledge as simple rules which are valid in an implicit context, everything must be explicitly encoded in classical logic. Unless exceptions are explicitly encoded, encountering exceptions will lead to inconsistencies. The traditional example of this arising is when we encounter a penguin; given that penguins typically do not fly, but penguins are birds and birds typically do fly. However, if exceptions *are* explicitly encoded, then the rules will no longer be triggered in the face of incomplete information.

Continuing the example, if Tweety is only known to be a bird, we can conclude nothing about its flying capabilities using a rule that says that birds that are not penguins fly. What is more, the list of exceptions is typically an ever open one, and the addition of an exception will normally require a non-trivial modification to the knowledge base.

Of the three requirements of section 2, it can easily be seen that classical logic only satisfies the second. If we stick to propositional logic there will be no distinction between factual evidence and generic knowledge; both will be propositional formulae. If we use first order logic, factual evidence will be encoded as grounded formulae, and generic knowledge will be encoded as universally quantified formulae. But the latter, by definition, rule out the possibility of exceptions. The alternative of using existentially quantified formulae simply does not have the expressive power to describe the normal course of things. So the first requirement is not met.

The second requirement, namely modelling partial ignorance, is basically fulfilled. If K is a set of formulae (encoding factual evidence and generic knowledge), a conclusion p is certainly true in the context of K if and only if $K \vdash p$ (p is a logic consequence of K). p is certainly false if $K \vdash \neg p$. When neither $K \vdash p$ nor $K \vdash \neg p$ hold, the truth of p is unknown, and this is precisely the definition of total ignorance *about* p . We can define *partial* ignorance as the epistemic state described by a set of formulae K such that neither $K \vdash p$ nor $K \vdash \neg p$ hold, for some p . This phenomenon occurs each time K is not a complete characterization of a situation; in semantic terms, it has more than one model.

The third requirement, nonmonotonic inference, is clearly not met, since it is a fundamental property of classical logic that

$$K \vdash p \text{ implies } \forall q, K \cup \{q\} \vdash p \quad (1)$$

Monotonicity has strong similarities to conditional independence in probability theory. Let A, B, C be events that represent the set of models of K, q and p , respectively. C is conditionally independent of B in the context A if and only if

$$P(C | A \cap B) = P(C | A). \quad (2)$$

If we interpret $K \vdash p$ as $P(C | A) = 1$, then clearly (1) is a particular case of (2) when $P(C | A \cap B) = P(C | A) = 1$. The fact that classical logic is monotonic can be interpreted as systematic conditional independence. However, when the inference symbol \vdash no longer means deduction in the usual sense, but plausible inference that can be viewed as high probability inference, the general validity of (1) looks dubious. In classical logic the only available notion of independence is logical independence; more expressive notions of independence and dependence cannot be expressed as supplementary information.

An important additional remark is that the problem of exception-handling in classical logic is closely related to two other problems, namely inconsistency management and belief revision (Gärdenfors, 1988). Indeed, encountering exceptions leads to inconsistencies, as has been discussed, and resolving this inconsistency means revising the set of formulae that are derivable. Makinson and Gärdenfors (1991) have pointed out that belief revision and nonmonotonic reasoning are two sides of the same coin. Indeed, the claim that p is a plausible conclusion of K in the context where q is also true comes down to accepting that p lies in the deductively closed belief set which is the revision of K by input q (that is, the deductive closure of K is modified in such a way that p is a member of the closure, consistency is maintained and the minimal changes made to include p and restore consistency conform to certain ‘‘rationality’’ criteria). This equivalence is important, especially at the computational level. However, it is conceptually limited, as in this approach K is encoded in propositional logic. That is, there is no clear distinction between factual evidence and generic knowledge.

4 Limitations of Bayesian networks

Classical logic represents beliefs by means of a set K of formulae that implicitly identify a subset of possible states of the world (the set of models of K), one of which is the actual one. Bayesian

networks, on the other hand, encode belief in a weighted acyclic graph that represents a single probability distribution on the set of states of the world. In this paper, only Bayesian networks that carry binary variables will be considered.

Bayesian networks are currently the most popular and most widely implemented model of reasoning under uncertainty. They are particularly successful from a computational point of view, being a good example of an approach where efficient local propagation algorithms succeed in properly handling complex probabilistic knowledge in a rigorous way. In this sense, they make certainty factor-based expert systems obsolete. The problem with Bayesian networks is neither in their mathematical, nor in their algorithmic, side. The main difficulty is to grasp precisely which kind of reasoning tasks they can address.

In the theory of Bayesian nets, as presented by Pearl (1988), it is not always obvious whether the probabilities involved are subjective or objective ones. Bayesian networks certainly look like a very powerful tool for the efficient encoding of any complex multivariate probability distribution, starting from statistical data. The problem is then how to extract the simplest acyclic graph that explicates as many independence relationships as possible. The strength of the approach relies on the fact that *any* probability distribution can be encoded as an (acyclic) Bayesian network. In other words, you do not need to be a subjective Bayesian probabilist to enjoy Bayesian nets.

However, if one does adopt a subjectivist Bayesian point of view, it looks neither convincing nor feasible to assess directly a complex joint probability distribution. In this situation, the Bayesian network methodology basically runs as follows: first draw a directed acyclic graph where links express direct causation, and indicate dependencies; assess conditional probabilities on the links, and prior probabilities on the roots of the graph; these data then uniquely determine a probability distribution underlying the graph, using appropriate conventions for the graphical representation of independence.

Objections to this approach are as follows:

1. The results of the approach rely heavily on the independence assumptions encoded in the topology of the graph, as well as on the numerical values put in the network. These values must be precise numbers which in practice, and when no statistical knowledge is available, may be out of reach (for instance $\text{Prob}(\text{other symptom}|\text{other disease})$ when exhaustive lists of observable symptoms and diseases are not available). Invariably there is some *a priori* probability to be supplied. It is not clear that experts are always capable of supplying all the numbers.
2. The network building method never produces inconsistencies. The “expert” is asked exactly the amount of data required for ensuring the unicity of the underlying distribution. Hence this distribution depends upon the set of questions you ask.
3. The assumption of an acyclic network, if innocuous with statistical data, becomes very questionable with a subjectively defined Bayesian net. One of the often used justifications is that arrows in the graph express causation. However, a generic rule such as “most students are young” does not mean that the youth of individuals is caused by their being registered as student. It is not clear how to encode the set of two generic rules {“most students are young”, “few young people are students”} under the form of an acyclic graph.

The first objection can be tackled by using sensitivity analysis techniques and, more generally, by allowing that knowledge can be represented by a set of conditional probability bounds that constrain a set of admissible probability distributions. For instance, we might only know that $P(B|A) \in [0.7, 0.8]$, or that “most A’s are B”, where the linguistic quantifier “most” is modelled by an interval restricting the possible values of the proportion $|A \cap B|/|A|$ (where $||$ denotes cardinality). This goes against the Bayesian credo that a unique probability is necessary and sufficient to encode subjective beliefs. However, of the alternative Bayesian proposals for handling the lack of knowledge about probabilities, many essentially reduce to introducing either higher-order probabilities or extra variables. These proposals basically make the picture even more complex. It is not at all clear that people can quantify their uncertainty about probabilities. As for the use of extra variables, this looks to be very efficient for the purpose of learning from cases.

However, in this paper we take the knowledge for granted and do not consider the learning problem.

The second objection is a significant problem. When representing knowledge in logic, inconsistency is almost unavoidable at the acquisition level. Tools for the detection of such inconsistencies must then be used to improve and refine the knowledge base. Inconsistency handling is apparently ignored by the Bayesian network approach. This question is clearly related to the acyclicity assumption: if the network has cycles, then the risk for an inconsistent assignment of probabilities is high.

It is clear that classical logic possesses none of the above drawbacks. Now as it turns out, Bayesian nets do fulfil the two requirements of plausible reasoning that classical logic violates. First, the distinction between generic knowledge and factual evidence is carefully made by the Bayesians. Generic knowledge is encoded by probabilities, especially conditional probabilities. Factual evidence is modelled as in propositional logic by allocating values to some variables. Plausible inference is then made by focusing the generic knowledge on the factual evidence E . This is achieved by conditioning, i.e. computing $P(C | E)$ for the events C of interest. Note that finding the most plausible values of (binary) variables in the network, in the presence of evidence E , is often called “revision” by some authors. We believe that the term “revision” is more convincingly applied to the modification of the network itself.

The nonmonotonicity requirement is also met by the Bayesian approach. Namely, it is possible to have a high value for $P(C | A)$ (hence C is a plausible conclusion when only A is known) and a very low value for $P(C | A \cap B)$ when B is learned. Interestingly, the Bayesian network topology is also tailored to encode those cases when nonmonotonicity does not occur, i.e. when $P(C | B) = P(C | B \cap A)$ holds.

Bayesian probability fails on requirement (ii) of plausible reasoning. That is, in contrast to classical logic, it cannot encode incomplete knowledge. This is due to the assumption that a Bayesian network encodes a single probability distribution. By way of illustration, let K be a propositional knowledge base on a finite language L . Now let $M(K)$ be the set of models of K and assume that $M(K)$ contains just three models. We may define the set $U(K)$ of unknown propositions in the context K by $U(K) = \{p \mid \text{neither } K \vdash p \text{ nor } K \vdash \neg p \text{ hold}\}$. All these propositions are equally unknown in the sense that there is no reason to consider any one to be more plausible than any other. However, assume now that you are a subjective Bayesian, and that you allocate positive probabilities a_1, a_2, a_3 respectively to each model in $M(K)$, such that $a_1 + a_2 + a_3 = 1$. Now, by definition each member of $U(K)$ must be true in some models in $M(K)$ and false in others. It then immediately follows that there is *no* probability assignment (a_1, a_2, a_3) to $M(K)$ that ensures that for all $p \in U(K)$ and $q \in U(K)$, $P(p) = P(q)$. That is, some equally unknown propositions must have distinct probability values. This gives a precise characterization of the fact that a unique probability distribution cannot model partial ignorance.

The most commonly found reason why Bayesians insist on unique probability distributions is because of the betting behaviour interpretation of subjective probability. Clearly, this setting models a decision problem in which betting rates depend on one’s degree of belief about what pertains in the actual world. However, partial ignorance differs from uncertainty about how to bet, as argued in Dubois, Prade and Smets (1993). For example, one may have a lot of knowledge about a (fair) die, and still be uncertain about the result of the next outcome; uncertainty about how to act does not imply total ignorance. Degrees of belief may govern betting rates, but it is not clear that the correspondence between them is one to one. We would argue that plausible reasoning is about entertaining beliefs, not about decision-making, and that degrees of belief can be construed as distinct from betting rates. In particular, when revising plausible conclusions upon receipt of new evidence, it is not at all clear that we should revise the prior *betting rates* (through Bayesian conditioning).

It is clear from the above discussion that the deficiencies of classical logic and of Bayesian networks with respect to the plausible reasoning endeavour are not the same. Indeed, they are strikingly complementary. The overriding ambition for knowledge representation and reasoning,

in the domain of plausible inference, is to identify a logic which combines the advantages of Bayesian networks with those of classical logic. In fact, there are several useful lessons to be remembered from the field of Bayesian networks, if classical logic is to be suitably augmented to capture the intuitive properties of exception-tolerant inference. These will be discussed in the next section.

5 Lessons from Bayesian networks

Despite their limitations regarding the representation of incomplete knowledge, the Bayesian network approach has significantly modified the notion of a knowledge base. As pointed out by Pearl (1988), an uncertain “if . . . then . . .” rule cannot be regarded as a “production rule” in the sense of expert systems, because the meaning of sentences such as “if x is A then plausibly y is B ” is not “if A is in the factual base then B should be added to it”, but rather “if all I know about x is A , then tentatively conclude that y is B ”. Hence the whole of the factual evidence should be taken into account simultaneously, to derive plausible conclusions.

Another lesson from the Bayesian approach is that material implication should not be used without caution to model the logical part of an uncertain rule. Material implications forget about the directedness of rules, as intended by the person that provides them, and they imbed monotonicity. The meaning of uncertain “if . . . then . . .” rules is more in accordance with conditional probability. A conditional probability is directed in the sense that $P(p | q)$ differs from $P(\neg q | \neg p)$, and in particular it differs from $P(\neg q \vee p)$ except for very special cases. (Contrast this with classical logic in which $q \rightarrow p$, $\neg q \vee p$ and $\neg p \rightarrow \neg q$ are all logically equivalent, where “ \rightarrow ” is material implication). The Bayesian way of using the rule base Δ is to infer from it a rule whose antecedent exactly matches the contents of the factual base E , under the form of a conditional probability $P(C | E)$. However, what can be done if we want to keep the idea of a directed conditional and drop the probability? One appealing answer is the use of a conditional object, denoted $p | q$, which is a three-valued entity. If ω is an interpretation of the language, then ω satisfies $p | q$ if and only if $\omega \models p \wedge q$ (ω is an example of $p | q$); ω falsifies $p | q$ if and only if $\omega \models \neg p \wedge q$ (ω is an exception to $p | q$); otherwise $p | q$ does not apply to ω . This case corresponds to a third truth-value.

This notion, originally proposed by De Finetti (1936; 1937), has been rediscovered several times in the literature (see Goodman et al., 1993; Dubois and Prade, 1994b), including Adams (1975) and Calabrese (1987). $p | q$ can be viewed as a set of propositions rather than a single one, namely $\{r : p \wedge q \models r \models \neg q \vee p\}$, and the probability of a conditional event $p | q$ is indeed a conditional probability $P(p | q)$ since the latter can be expressed in terms of $P(p \wedge q)$ and $P(\neg q \vee p)$ only. Conditional objects are not at the same level as formulae of propositional calculus but are constructed on top of them. Defining a body of generic knowledge Δ as a set of conditional objects Δ , while the body of evidence E is a set of propositional formulae, the first requirement of plausible reasoning as per section 1 is potentially met. Dropping the need for assigning precise probabilities opens the door to the fulfilment of requirement (ii).

A last lesson from Bayesian networks is that a knowledge base is not just a collection of unordered well-formed formulae. For a Bayesian, a knowledge base is a directed graphical structure, which is used to reveal conditional independence properties that are implicit in the knowledge. Graphical representations of knowledge bases also pervade the literature of taxonomic reasoning, without reference to independence. However a set of exception-tolerant rules Δ once represented as a graph, will significantly differ from a Bayesian net. In a Bayesian net, nodes are (logical or n -ary) variables while in a graph built from Δ will contain nodes representing literals. Moreover, nothing forbids cycles in the latter graph, while cycles are prohibited in Bayesian representations. The problem is then to develop a capability for “reading” independence assertions from a structured set of rules. This question will become an important issue in the future.

6 Graded representations of incomplete knowledge

One obvious limitation of classical logic when representing partial ignorance is its relative crudeness. In the face of partial information represented by a set of propositions K , the language

can at best be partitioned into three sets: the propositions that are certainly true, that form the set $C(K)$ of consequences of K ; the set $F(K)$ of propositions that are certainly false; and the remainder $U(K)$ whose truth or falsity is not known. Very naturally, one may wish to refine this classification by distinguishing in $U(K)$ propositions that are more certainly true (or false) than others. The introduction of certainty factors into expert systems was one attempt to achieve this. More generally, this is an aim of all quantitative theories of uncertainty. One reason why Bayesian probability theory has not been considered as the only candidate is its failure to represent partial ignorance. For example, the MYCIN certainty factors were deliberately tailored to express the difference between a non-certainly true proposition and a certainly false one. However, as with certainty factors, many of the uncertainty calculi developed for rule-based expert systems were required to be compositional. In this section, we will first indicate why compositionality is impossible in a well-founded uncertainty calculus. Then we will review a number of theories of uncertainty, and discuss their relevance in the problem of exception-tolerant knowledge-based systems.

6.1 The compositionality problem

A major simplicity of classical logic is its compositionality property. That is, the truth-value of a formula is always a function of the truth-values of the atomic propositions it involves. In the early days of expert system development, many people assumed that this property could be carried over to degrees of uncertainty. Compositional quantified extensions of Boolean logic do exist. These are called multiple-valued logics (Rescher, 1969), and were developed in the early 1930s by a number of logicians, including Lukasiewicz. Hence the temptation to found uncertainty calculi on multiple-valued logics.

Contemporaneously, fuzzy set theory (Zadeh, 1965) was construed as the set-theoretic counterpart of multiple-valued logics. This led to the often made claim that fuzzy logic is an approach to handling uncertainty in knowledge-based systems. A side-effect of this claim is that it has led to a rejection of fuzzy set theory by those people who realize the impossibility of imposing a compositional uncertainty calculus onto Boolean logic. For instance, Elkan (1993) concludes that fuzzy logic collapses on two-valued logic in such a case.

This apparent state of confusion is basically due to the lack of acknowledged distinction between a degree of truth and a degree of uncertainty. The idea of intermediary truth completely differs from the idea of not knowing if a proposition (in the classical sense) is true or false. This is a crucial distinction whose importance was pointed out as early as 1936 by De Finetti. Confounding these two situations is analogous to making the claim that an unseen bottle which is either empty or full is actually half full.

The treatment of propositions as 2-valued entities is a matter of convention, and leads to the Boolean algebraic structure of classical logic. Modifying this convention can make sense, but it affects the meaning of the word "proposition"; introducing intermediary truth-values means that we use a language with entities that differ significantly from what is usually understood by the term "proposition". This is what fuzzy logic, a logic of graded properties, not of uncertainty, does. As a consequence, the Boolean algebra structure is lost. In particular, many formulae which are tautologies in classical logic are no longer so in fuzzy logic. Collapse results such as Elkan's are due to the impossibility of reconciling a Boolean algebra structure with intermediary truth-values which remain compositional. The former must be sacrificed if we wish to allow the latter.

In contrast, degrees of uncertainty are *not* intermediary truth-values. They correspond to an epistemic attitude expressing that some individual does *not know* if a proposition (in the usual sense) is true or false. To quote De Finetti (1936): "This attitude does not correspond to a third truth-value distinct from yes or no, but to the *doubt* about yes or no." Hence the third modality expressing ignorance should not enter as a third truth-value, but should be layered on top of the two values "true" or "false". In particular, all tautologies of classical logic should be retained. For instance even if p and $\neg p$ are unknown, $p \vee \neg p$ is still ever true as $p \leftrightarrow p \wedge p$, and $p \wedge \neg p$ is still ever

false. This simple requirement means that we must sacrifice the compositionality hypothesis, if we are to allow more than two uncertainty levels (see Dubois and Prade, 1988b). But it does not impact on the validity of fuzzy logic, where $p \vee \neg p$, $p \leftrightarrow p \wedge p$ are never simultaneously acknowledged as being tautologies.

Uncertainty calculi can, however, be partially compositional. For example, probabilities are compositional with respect to negation only, whilst possibility measures (Zadeh, 1978) are compositional with respect to union, as are Spohn's (1988) disbelief functions. But Shafer's belief functions are not compositional at all.

6.2 Belief functions on uncertain rules

The ad hocery of compositional uncertainty calculi in the expert systems literature has been overcome by the development of Bayesian nets. However, the lack of capability of the latter in representing partial ignorance has prompted some researchers to model uncertainty in knowledge bases in the setting of a more flexible uncertainty calculus, namely belief functions (Shenoy and Shafer, 1990). The idea is to use the hypergraph machinery underlying the local propagation of probabilities, as described by Lauritzen and Spiegelhalter (1988), and extend it to belief functions.

A belief function (Shafer, 1976) is defined on a set Ω , which represents the universe of discourse or "frame of discernment". It is defined by a family F of non-empty subsets of Ω , called "focal elements", to which positive numbers called masses $\{m_i, A_i \in F\}$ are attached. These masses sum to unity. m_i is interpreted as the probability that the available information is correctly described by A_i . Let $m = \{m_i, A_i \in F\}$. In logical terms, we may regard Ω as representing a set of interpretations of a language. Then the pair (F, m) defined on Ω can be viewed in terms of a random propositional knowledge base K such that $A_i = M(K_i)$ and $m_i = P(K \text{ is semantically equivalent to } K_i)$; that is, the probability that A_i is the correct representation of the available evidence. The degree of belief $\text{Bel}(A)$ in a proposition p such that $A = M(p)$ is defined by

$$\text{Bel}(A) = \sum_{i: A_i \subseteq A} m_i = \sum_{i: K_i \vdash p} m_i$$

Hence $\text{Bel}(A)$ can be viewed as the probability of provability of p (Pearl, 1988).

In the belief function approach to knowledge bases, each piece of knowledge is represented as a belief function relating some variables, and is viewed as an hyperedge of the hypergraph formed by all pieces of knowledge. A global belief function is constructed using the Dempster rule of combination. This combination rule is also the one used to absorb new evidence. Question-answering is achieved by projecting the global belief function over the range of the variables of interest. When the hypergraph is an hypertree, uncertainty propagation and combination can be done locally. The generic machinery that is put at work in the hypergraph approach is capable of encompassing Bayesian networks and classical constraint propagation techniques (when the masses m_i take on value 0 or 1). Yet, this approach does not adapt well to exception-tolerant plausible reasoning for several reasons:

1. There is a failure to distinguish between factual evidence and generic knowledge. Indeed, belief functions are construed as a theory of evidence, not as a theory of generic knowledge; the Dempster rule of combination is effective for pooling *independent* pieces of uncertain evidence. In the above scheme, the hypergraph is supposed to encode both uncertain rules and (possibly uncertain) evidence. All information receives a uniform treatment, as in classical logic.
2. The question of how to encode an exception-prone rule in a belief function setting has received little consideration from the advocates of hypergraph methods. An uncertain rule is often encoded by putting some mass α on a material implication $\neg\psi \vee \varphi$, and the remainder $1 - \alpha$ on the tautology (e.g. Smets, 1988). If such pieces of information are then combined with the Dempster rule, the results obtained are often counter-intuitive. This is because the assumption

of independence between the rules in a knowledge base is seldom satisfied (Dubois and Prade, 1994a).

By way of illustration, consider again the penguin example. Let $b = \text{bird}$, $f = \text{fly}$, $p = \text{penguin}$ and suppose we have the following mass distributions for the rules in the knowledge base Δ :

$$\begin{aligned} b \rightarrow f: & m_1(\neg b \vee f) = \alpha; m_1(\mathbf{T}) = 1 - \alpha \text{ (}\mathbf{T}\text{: tautology)} \\ p \rightarrow b: & m_2(\neg p \vee b) = 1 \\ p \rightarrow \neg f: & m_3(\neg p \vee \neg f) = \alpha; m_3(\mathbf{T}) = 1 - \alpha. \end{aligned}$$

Now, suppose some evidence E is obtained as another belief function identifying Tweety is a penguin, i.e. $m_4(p) = 1$. Then it is straightforward to check that combining these pieces of information using the Dempster rule leads to the counter-intuitive result that $\text{Bel}(f) = \text{Bel}(\neg f) = (\alpha(1 - \alpha)/(1 - \alpha^2))$. This is simply a variant of the inconsistency result which one would get in pure classical logic (letting $\alpha = 1$). Indeed the presence of the normalization factor $1 - \alpha^2$ indicates that the bodies of evidence $\{m_1, m_2, m_3, m_4\}$ are deemed to be partially inconsistent. There is no way to account for the intended interaction between the sources of evidence.

6.3 Upper and lower probabilities

Another approach to relaxing the Bayesian framework is to admit that the available knowledge does not necessarily determine a single probability distribution. We retain the idea that an uncertain rule is modelled by a conditional probability, but its exact value may be ill-known.

A collection of pieces of information of this kind leads us to consider a knowledge base Δ as a set of statements of the form $P(B_i | A_i) \in [\alpha_i, \beta_i]$, that form an interval-valued network. We are interested in computing the tightest bounds on the value of some other conditional probability of interest, without using systematic independence assumptions as in Bayesian networks. A missing arrow in such a network corresponds to a quantifier or a conditional probability which is completely unknown (then represented by the interval $[0,1]$). All probabilities that we handle are (bounds of) conditional probabilities; no prior probability information is required to start the inference process in this approach. Such networks do not obey the same conventions as Bayesian nets. In the latter the network is the result of a data compression procedure that accounts for a joint probability distribution. In the present approach the data consists of incomplete statistics under the form of interval-valued conditional probabilities. A network is now only a display of the raw data expressing local constraints on an unknown joint probability distribution. This approach leaves room for inconsistent specifications (which can be detected by constraint propagation), and thus addresses the other objections to Bayesian networks. In particular, cycles are allowed.

The non-Bayesian view of a probabilistic knowledge base is thus as a collection of general statements regarding a population X of objects. These statements express, possibly in unspecific terms, the proportions of objects in various subclasses of X that belong to other subclasses. This knowledge base allows for answering queries about a given object, given a subclass E to which evidence assigns it (also called its "reference class" by Kyburg, 1974). To make an inference we just apply to this object the properties of this subclass, implicitly assuming that it is a typical element of it. That is, we compute a new rule "if E then C " and optimal bounds on $P(C | E)$, for a class C of interest. If more information becomes available for this object we just change its reference class accordingly. The computation of optimal bounds can be done by linear programming techniques (Amarger et al., 1991). But local techniques have been studied as well (Thöne et al., 1992; Dubois, Godo et al., 1993). The latter are interesting because they are more efficient and are capable of handling independence assumptions more easily, if needed. However they are sometimes suboptimal.

This probabilistic model also maintains the difference between querying a knowledge base and revising it, what we call "focusing" and "revision", respectively. The querying problem is to ask for $P(C | E)$ on the basis of a case belonging to class E for which we want to know its probability of belonging to C . Revision means adding a new piece of knowledge to the database, for instance the

value of a new conditional probability that has become known. A particular case is when we learn that E is true for the whole concerned population. That means we add the constraint $P(E) = 1$ to the database. Note that the bounds on $P(C)$ given that $P(E) = 1$ usually differ from the bounds on $P(C | E)$ computed from the original database. Hence focusing generally differs from revision. However, when the probability distribution underlying the database is uniquely determined (as in the case of Bayesian nets), the two operations yield the same result: a point-valued $P(C | E)$ provided by Bayes rule. Clearly, Bayes rule serves two purposes, and this is sometimes a source of confusion. When the assumption of a unique available probability distribution is dropped, the two tasks (focusing and revision) lead to different types of conditioning rules.

The representation of ignorance with upper and lower probabilities is unbiased. Namely, it is possible to obtain that the truth of a conclusion C is unknown, in the presence of evidence E , if all that can be obtained from the knowledge base is that $P(C | E) \in [0,1]$.

Moreover, the behaviour of a set of conditional probabilities is nonmonotonic. Indeed, considering again the example of section 6.2, the set $\Delta = \{P(f | b) \geq \alpha, P(\neg f | p) \geq \alpha, P(b | p) = 1\}$ is not at all inconsistent for $\alpha \neq 1$. However, the inferential power of this probabilistic approach is low because it does not allow for systematic subclass inheritance. For instance, if the available evidence is $\{b, r\}$ where r stands for red, it is not possible to conclude from Δ anything about flying, i.e. $P(f | b \wedge r) \in [0,1]$. Here classical logic can do better since $\{b, r, \neg b \vee f, \neg f \vee \neg p, \neg p \vee b\} \vdash f$. To do so in the upper and lower probability framework, a conditional independence assumption is needed for irrelevant properties. Again we face a case where the merits of classical logic and probability are complementary.

Belief functions are special cases of lower probabilities. Yet the approach of section 6.2 is not a special case of the above approach to exception-tolerant rules. First, the rule of combination used in section 6.2 is not idempotent. If you use $b \rightarrow f$ twice (i.e. m_1 twice), it affects the resulting global belief function. This is not the case here, i.e. putting $P(f | b) \geq \alpha$ twice is innocuous. Another difference is that material implication is used in section 6.2 for rule representation. To do the same here means using $P(\neg b \vee f) \geq \alpha$ instead of $P(f | b) \geq \alpha$, as in Nilsson's (1986) probabilistic logic. This would not be satisfactory. For instance, considering the knowledge base $\{P(\neg b \vee f) \geq \alpha, P(\neg p \vee b) = 1, P(\neg p \vee \neg f) \geq \lambda\}$, it is straightforward to show that if Tweety is a penguin:

- answering queries by computing conditional probabilities leads to $P(\neg f | p) \in [0,1]$ only;
- adding $P(p) = 1$ to the knowledge base leads to a contradiction as soon as $\alpha + \lambda > 1$ (which is the regular situation since birds fly and penguins do not).

6.4 Ordinal and semi-quantitative approaches to uncertainty

Apart from fully-fledged numerical methods based on additive properties, and purely symbolic representations of uncertainty, it is natural to consider ordinal methods as well. Incomplete probabilistic databases can be questioned as to their capability to model all kinds of commonsense knowledge. In many instances, it is not possible to quantify the exceptions to a default rule. We know that "birds usually fly", but not the proportion of flying birds. One may be tempted to change interval-valued probabilities into fuzzy probabilities (Zadeh, 1985), but that makes sense for a refined sensitivity analysis only. One may also try to devise a purely symbolic approach to probabilistic inference using linguistic probability terms referring to partially unspecified probability intervals, the inferred linguistic probabilities being linguistic approximations of the intervals obtained through numerical constraint propagation (Dubois, Godo et al., 1993). Experience shows that this results in a weakening of the inferential power of the reasoning system compared to the numerical setting. Moreover, reasoning with linguistic probabilities does not solve the irrelevant property problem, and results in more ambiguous responses than in the numerical case.

Instead of quantifying uncertainty, one might prefer to compare propositions in terms of their relative certainty. Namely consider a relation \geq defined on a language, such that $p \geq q$ means that one is at least as confident in the truth of p as in the truth of q . Most quantitative approaches to

uncertainty have comparative counterparts (see Dubois and Prade, 1993, for a review). It is clear that the relation \geq should be reflexive, transitive and complete, so that the language can be partitioned into classes of propositions of equal strength, these classes being totally ordered. Certainty should go along with logical consequence in the sense that $p \vdash q \Rightarrow q \geq p$. The oldest example of such a comparative uncertainty relation was introduced by De Finetti (among others), namely comparative probability, which satisfies the additivity axiom

$$p \wedge (q \vee r) = \perp \Rightarrow (q \geq r \Leftrightarrow p \vee q \geq p \vee r). \quad (\text{A})$$

This axiom has been further generalized by Savage to the comparison of acts under the name “sure thing principle”. This is foundational to subjective expected utility theory. Unfortunately, in the finite case this property does not only characterize orderings induced by probability measures (Fine, 1973). Comparative probabilities are not simple to use because, on finite sets, the ordering among propositions cannot be recovered from the knowledge of its restriction to interpretations.

The situation is much simpler if we turn the additivity axiom into the following one (Dubois, 1986):

$$q \geq_{\Pi} r \Rightarrow p \vee q \geq_{\Pi} p \vee r. \quad (\text{II})$$

This is obtained by dropping the disjointedness condition and relaxing the equivalence in the additivity axiom. The uncertainty relations so obtained are actually called “comparative possibility relations” by Lewis (1973). They can be represented equivalently by functions Π from the language to a (finite) totally ordered set S with top 1 and bottom 0, such that

$$\Pi(p \vee q) = \max(\Pi(p), \Pi(q)).$$

These functions are called possibility measures by Zadeh (1978) in the case when S is the unit interval. $q \geq_{\Pi} r$ means that q is at least as consistent as r with the current knowledge. Dually, when $\neg r \geq_{\Pi} \neg q$, q is said to be at least as certain as r , which is denoted $q \geq_c r$. Comparative certainty relations satisfy an axiom dual to (II), namely

$$q \geq_c r \Rightarrow p \wedge q \geq_c p \wedge r. \quad (\text{N})$$

$q >_c r$ (strict preference) means that in the presence of inconsistency, if dropping either q or r would restore consistency, one would rather drop r . Axiom (N) can be derived from Gärdenfors’ (1988) postulates of belief revision (see Dubois and Prade, 1991), and is satisfied by Gärdenfors’ (1988) epistemic entrenchment. If c denotes an order reversing map on S , $\Pi(p)$ is the degree of possibility, then certainty orderings can be equivalently represented by functions N with range S , such that $N(p) = c(\Pi(\neg p))$. This expresses the condition that p is all the more certain as $\neg p$ is impossible, just as in modal logics of possibility and necessity. $N(p)$ is called the degree of necessity of p .

A possibility (or a certainty) ordering on a finite set can be characterized by a complete preordering of the set Ω of interpretations, encoded by an assignment $\pi(\omega) \in S$ to each interpretation ω . π is called a possibility distribution, and encodes a preference relation describing the respective levels of plausibility of ω as being the actual world. $\pi(\omega) = 0$ means that ω is impossible, while $\pi(\omega) = 1$ means that ω is a most plausible world (say a most normal one). Then we have

$$\begin{aligned} \Pi(p) &= \max\{\pi(\omega), \omega \vDash p\} \\ N(p) &= \min\{c(\pi(\omega)), \omega \vDash \neg p\}. \end{aligned}$$

This kind of uncertainty measure can be viewed as completely symbolic (S is a totally ordered set) or numerical (S is the unit interval). When numerical, $N(p)$ is a special kind of belief function, called a consonant belief function (Shafer, 1976), where the focal elements are nested. Spohn (1988) has considered a very similar kind of uncertainty function k such that $k(p \vee q) = \min(k(p), k(q))$, where $k(p)$ is a non-negative integer. $k(p) = n$ is viewed as expressing that the probability of

p is of the form ε^n , where ε is a very small number. It has been pointed out that the set function Π_k such that $\Pi_k(p) = a^{-k(p)}$ for $a > 1$ is also a possibility measure, valued on a subset of $[0,1]$.

Rather than demonstrate clear failings with respect to the requirements of exception-tolerant plausible reasoning, the above approaches show a great deal of promise. In the remainder of the paper, we will expand on this line of argument and show how close we can get to satisfying the requirements of section 2.

7. Possibilistic logic

Possibility theory enables classical logic to be extended to layered sets of formulae, where layers express certainty levels. This is possibilistic logic (Dubois et al., 1991). The encoding of the layers is simply achieved by assigning to each formula φ of a knowledge base K an element $\alpha \in S$, where S is a totally ordered set as in the previous section. We then have a pair (φ, α) , expressing the constraint that $N(\varphi) \geq \alpha$.

Inference in possibilistic logic can be achieved by means of a straightforward extension of the resolution principle from automated theorem proving. Let $\text{Res}(\varphi, \psi)$ be the resolvent of the two clauses φ and ψ (for example, if $\varphi = \neg p \vee q$ and $\psi = p \vee r$, then $\text{Res}(\varphi, \psi) = q \vee r$). Then, for a possibilistic database

$$(\varphi, \alpha), (\psi, \beta) \vdash (\text{Res}(\varphi, \psi), \min(\alpha, \beta)). \quad (\text{R})$$

The use of this rule presupposes that a knowledge base K can be transformed into clausal form, but this is always possible.

Resolution is refutation complete. To prove a formula is a logical consequence of a consistent database, we add its negation to that database and attempt to draw a contradiction. To prove (φ, β) from a possibilistic knowledge base K , denoted $K \vdash (\varphi, \beta)$, the procedure can be modified as follows:

1. Add $(\neg\varphi, 1)$ to K (or the clauses corresponding to $\neg\varphi$) and let $K' = K \cup \{(\neg\varphi, 1)\}$.
2. Try to derive the contradiction \perp from K' , using a sequence of applications of (R), with a sufficient level α of certainty. That is, $K' \vdash (\perp, \beta)$ with $\beta \geq \alpha$. Then, it can be proved that $N(\varphi) \geq \beta$.

Unlike classical logic and Bayesian probability, possibilistic logic can accommodate a degree of inconsistency. A knowledge base K is said to be totally consistent if it is *not* true that $K \vdash (\perp, \alpha)$ for $\alpha > 0$, by successive applications of (R). However, more generally,

$$\text{Inc}(K) = \sup\{\alpha, K \vdash (\perp, \alpha)\}$$

is called the “degree of inconsistency” of K .

The main advantage of graded inconsistency is to avoid the drawback of classical logic, whereby the presence of contradiction trivializes the consequence relation (everything follows from an inconsistent knowledge base). In contrast, non-trivial deductions can be performed from an inconsistent possibilistic knowledge base. A formula φ is called a non-trivial deduction from K , denoted $K \vdash_{\pi} \varphi$, if and only if

$$K \vdash (\varphi, \alpha) \quad \text{with } \alpha > \text{Inc}(K).$$

$K \vdash_{\pi} \varphi$ means that only a consistent subpart of K has been used to derive φ ; that is, a sub-base containing formulae of certainty higher than the inconsistency degree. Unsurprisingly, the consequence relationship \vdash_{π} does not obey the postulates usually expected of consequence relationships. Because of the possibility of different sub-bases being mutually inconsistent, it is not transitive. It is reflexive up to the contradiction and it is not monotonic. For example, if we reconsider the Tweety example and let

$$K = \{(\neg b \vee \neg f, \alpha), (\neg p \vee f, \beta), (\neg p \vee b, \beta)\} \quad \text{with } \beta > \alpha.$$

Then $K \cup \{b\} \vdash_{\pi} f$ while $K \cup \{b, p\} \vdash_{\pi} \neg f$.

To clarify why this is so, one must turn to the semantics of possibilistic logic. Essentially, the ordering over formulae in K induces an ordering over interpretations. The latter can be obtained by first representing each uncertain assertion (φ_i, α_i) in K by means of a possibility distribution π_i over the set of interpretations ω . This is defined as

$$\begin{aligned}\pi(\omega) &= 1 \text{ if } \omega \models \varphi_i \\ &= 1 - \alpha_i \text{ if } \omega \models \neg\varphi_i.\end{aligned}$$

That is, $1 - \alpha_i$ is interpreted as the degree of possibility that φ_i is false. The ordering over the models $M(K)$ is then defined by π_K , where

$$\pi_K = \min_{i=1,n} \pi_i.$$

The construction of π_K from K obeys the principle of minimal specificity, namely each interpretation is assigned the highest possibility degree so as not to violate any constraint of the form $N(\varphi_i) \geq \alpha_i$. The possibility distribution π_K gives a grading over models of K , and so defines a fuzzy set of preferred models. The maximal possibility degree $\sup \pi$ turns out to be equal to $1 - \text{Inc}(K)$. A notion of semantic entailment is then defined following Zadeh. Namely, if $\pi_{(\varphi,\alpha)}$ is the possibility distribution deriving from (φ, α) , $K \models (\varphi, \alpha)$ means $\pi_K \leq \pi_{(\varphi,\alpha)}$. Possibilistic logic is sound and complete in the sense that $K \vdash (\varphi, \alpha)$ if and only if $K \models (\varphi, \alpha)$ (see Dubois et al., 1991). The non-trivial inference \vdash_π can now be expressed semantically as follows: $K \vdash_\pi \varphi$ if and only if all preferred interpretations of K (in the sense of π_K) are models of φ . This is clearly a kind of preferential inference, which Shoham (1988) has proved to be at work in many non-monotonic logics. The absence of monotony of inference \vdash_π should not be surprising.

8 Properties of exception-tolerant inference

Now, how well does possibilistic logic match the requirements of section 2? From the preceding, it should be clear that possibilistic logic obeys requirements 2 and 3 of exception-tolerant reasoning. It models ignorance in an unbiased way; if p is ignored, $N(p) = N(\neg p) = 0$ and the proposition p is simply omitted. Moreover, inference is indeed nonmonotonic, as we have just seen. However, as it stands possibilistic logic does not allow for a clear separation of generic knowledge from factual evidence. This is clear from the treatment of the Tweety example in section 7. In particular, modelling “birds fly” as $(\neg b \vee f, \alpha)$ is hard to justify in any generality. This is because the clean and intuitive treatment of the example relies entirely on the proper choice of the weight α ; α must be less than β in $(\neg p \vee \neg f, \beta)$. If this is not the case, the expected inference will not occur. So, the question is: where do the weights come from? The answer is: from a proper modelling of the generic knowledge.

One way out of this difficulty is to consider a generic rule as a constraint on the result of a plausible inference between two facts. This is the path followed by Lehmann and colleagues (Kraus et al., 1990; Lehmann and Magidor, 1992). Their idea is that the rule $b \rightarrow f$ constrains the pair (b, f) to belong to a consequence relationship denoted \vdash , i.e. $b \vdash f$. The problem is then to prescribe what the properties of this consequence relationship should be. They take the form of the following postulates:

Left logical equivalence:

$$\text{from } p \leftrightarrow p' = T \text{ and } p \vdash q \text{ deduce } p' \vdash q \quad (\text{LLE})$$

Right weakening:

$$\text{from } q \models q' \text{ and } p \vdash q \text{ deduce } p \vdash q' \quad (\text{RW})$$

Reflexivity: $p \vdash p$

Left OR:

$$\text{from } p \vdash r \text{ and } q \vdash r \text{ deduce } p \vee q \vdash r \quad (\text{LOR})$$

Cautious monotony:

$$\text{from } p \vdash q \text{ and } p \vdash r \text{ deduce } p \wedge q \vdash r \quad (\text{CM})$$

Weak transitivity:

$$\text{from } p \wedge q \vdash r \text{ and } p \vdash q \text{ deduce } p \vdash r \quad (\text{cut})$$

The three last rules had already appeared in Gabbay (1985). Kraus, Lehmann and Magidor (1990) call a non-monotonic consequence relation \vdash satisfying the above postulates “preferential”, and name the corresponding logic P . They prove that the following rules of inference can be derived from the above set of postulates

$$\text{from } p \vdash q \text{ and } p \vdash r \text{ deduces } p \vdash q \wedge r \quad (\text{RAND})$$

$$\text{from } p \wedge q \vdash r \text{ deduce } p \vdash q \rightarrow r \quad (\text{S})$$

It is interesting to check the relevance of the above inference rules to plausible reasoning. CM restricts the use of monotonicity to the case when q has already been plausibly inferred from p . That is, property q is not exceptional for models of p . The cut rule restricts transitivity to the case when models of p are normal models of q with respect to property r . LLE is merely a consistency condition with respect to classical logic. The rules RAND and RW ensure that the set of plausible consequences of p is deductively closed. Property (S) is one half of the classical deduction theorem $p \wedge q \vdash r \Leftrightarrow p \vdash p \rightarrow r$.

Based on these axioms, Kraus et al. (1990) define a syntactic deduction operation, which we will denote by \vdash in the following. This deduction operation acts on a set Δ of conditional assertions of the form $p_i \vdash q_i$. $\Delta \vdash p \vdash q$ if and only if $p \vdash q$ can be derived from Δ using $p \vdash p$ as an axiom schema and the inference rules LLE, RW, LOR, CM and cut of logic P . $\text{CONSP}(\Delta)$ is then the set of conditional assertions deduced from Δ in P .

There are a number of alternative proposals for giving a semantics to a preferential consequence relation \vdash . Kraus et al. (1990) have proposed a semantics for $p \vdash q$ that is based on a two level structure involving a set X of states, a mapping f from X to the set Ω of interpretations of the language and a preference relation among states which is symmetric and transitive. A state x satisfies p iff $f(x) \vDash p$. Then $p \vdash q$ means that all the preferred states satisfying p satisfy q . Another semantics of $p \vdash q$ is that $P(q | p) \geq 1 - \varepsilon$, i.e. the conditional probability $P(q | p)$ should be very close to 1. This type of conditional has been studied by Adams (1975) and revived by Pearl (1988). Probabilistic inference with infinitesimal lower probabilities does satisfy the rules of system P , hence showing the consistency of the above logical approach with the one of section 6.3, based on upper and lower probabilities. Dubois and Prade (1994b) give a simpler semantics of conditional assertions based on the three-valued conditional events of De Finetti.

Finally, Fariñas del Cerro et al. (1992) have proposed another semantics of conditional assertions based on possibility theory, whereby $p \vdash q$ is viewed as a constraint restricting a set of possibility distributions on Ω that verify the condition $\Pi(q \wedge p) >_{\Pi} \Pi(\neg q \wedge p)$. This is saying that in the framework where p is true, q is more plausible than $\neg q$. It can also be interpreted as saying that all the most plausible worlds satisfying p satisfy q . That is, $p \vDash_{\pi} q$ in the sense of possibilistic logic. Then the inference of $\varphi \vdash \psi$ under the form $\Pi(\varphi \wedge \psi) > \Pi(\neg \psi \wedge \varphi)$ from a set of constraints of the form $\Pi(p_i \wedge q_i) > \Pi(p_i \wedge \neg q_i)$ satisfies all the properties of system P . This semantics can be interpreted in terms of a qualitative notion of conditioning in possibility theory, namely $N(q | p) > 0$ where $N(q | p) = 1 - \Pi(\neg q | p)$ and $\Pi(q | p)$ is the greatest solution of the Bayesian-like equation (Dubois and Prade, 1988a).

$$\Pi(p \wedge q) = \min(\Pi(q | p), \Pi(p)).$$

Whatever the chosen semantics, the logic P satisfies all three of the requirements for an exception-tolerant logic. Exception-handling is imbedded in the inference process. Instead of enriching the factual base by triggering rules in order to derive conclusions, new rules $p \vdash q$ are produced from the knowledge base until one rule is produced that fits the evidence as a whole. Namely, if E contains the available (propositional) evidence and Δ the generic knowledge in the form of conditional assertions, then C is a plausible conclusion from (E, Δ) if and only if $E \vdash C$ can be derived from Δ .

9 Possibilistic encoding of rational inference

There is, however, still a weakness: the above logic of conditional assertions is very cautious. Like the approach based on upper and lower conditional probability, it does not solve the problem of irrelevant properties. For instance, using the same notation as the example in section 6.3, if $\Delta = \{b \vdash f, p \vdash \neg f, p \vdash b\}$ it does not follow in P that red birds fly, i.e. $b \wedge r \vdash f$. This is a by-product of the absence of the monotony property. The cautious monotony property is too weak, because in this example, we do not have that $b \vdash r$ and so are unable to make the step from $b \vdash f$ to $b \wedge r \vdash f$.

This difficulty has been solved by Lehmann by making the inference more monotonic. Namely, although we do not have $b \vdash r$, we do not have that $b \vdash \neg r$ either. This should enable us to jump to the conclusion that $b \wedge r \vdash f$, and that $b \wedge \neg r \vdash f$ as well; that is, the redness of the bird is irrelevant to its flying. The basic idea is to augment the preferential closure $\text{Cons}_P(\Delta)$ with additional conditional assertions, and construct the so-called Rational Closure, $\text{Cons}_R(\Delta)$, of Δ . This latter satisfies the following property, called rational monotony:

$$\text{if } p \vdash q \text{ and } \neg(p \vdash \neg r) \text{ then } p \wedge r \vdash q. \quad (\text{RM})$$

Unfortunately, $\text{Cons}_P(\Delta)$ cannot be defined using (RM) as an inference rule like those of section 8. In general, there will be several supersets of $\text{Cons}_P(\Delta)$ that are closed under (RM), and the intersection of two rational extensions will not necessarily be rational (Lehmann and Magidor, 1992). However, it is possible to define rational closure appropriately. It can be proved (Lehmann and Magidor, 1992; Gärdenfors and Makinson, 1994; Benferhat et al., 1992) that any rational extension corresponds to a unique complete partial ordering of interpretations. In terms of possibilistic logic, this result reads: for any rational consequence relationship \vdash , there exists a possibility distribution π on the set of interpretations such that \vdash coincides with \vdash_π ; conversely \vdash_π is rational. The problem is then to find the “best” ordering of interpretation that defines the rational closure of $\text{Cons}_P(\Delta)$.

This problem has been solved by Lehmann and Magidor (1992) and Pearl (1990). The idea is to find the ranking of interpretations that agrees with Δ , where each interpretation is as normal as possible. This construction can be given a completely equivalent interpretation in terms of possibility theory. Namely, let $PI(\Delta)$ be the set of possibility distributions π such that $\forall p_i \vdash q_i$ in Δ , $\Pi(p_i \wedge q_i) > \Pi(p_i \wedge \neg q_i)$. A possibility distribution π is said to be less specific than another possibility distribution π' if and only if $\pi > \pi'$, that is for any interpretation ω , $\pi(\omega) > \pi'(\omega)$. Clearly, each interpretation is at least as normal in the sense of π as in the sense of π' . The ordering that determines the rational closure is then defined by the least specific possibility distribution π^* in $PI(\Delta)$, that is the largest possibility distribution associated with a possibility measure Π^* obeying the set of constraints $\Pi(p_i \wedge q_i) > \Pi(p_i \wedge \neg q_i)$.

Equivalently, as proposed by Pearl (1990), rational inference can be captured through an ordering of the conditional knowledge base Δ itself. This ordering is based on a concept of *tolerance* of a default rule by the knowledge base. If $p \vdash q$ is a conditional assertion, let $\neg p \vee q$ be its material implication counterpart. Let K be the set of material implication counterparts of the conditional assertions in Δ . Then we have the following definition:

$$\text{A rule } p \vdash q \text{ is tolerated by } \Delta, \text{ if and only if } p \wedge q \text{ is consistent with } K.$$

To get a flavour for the reason behind this definition, consider again the penguin example. The

examples to the rule $p \vdash \neg f$ are exceptions to the two other rules $b \vdash f$ and $p \vdash b$. That is, $p \vdash \neg f$ is not tolerated by them. In contrast, $b \vdash f$ is tolerated by the other rules in Δ ($= \{b \vdash f, p \vdash \neg f, p \vdash b\}$) since this rule has examples that are not exceptions to the others (birds which are not penguins). Hence using rule $b \vdash f$ does not cause any conflict in Δ . But applying the other rules does conflict with $b \vdash f$. Hence, to obtain the most specific conclusions, non-tolerated rules should be given priority over tolerated ones.

An algorithm for obtaining a priority ranking of rules has been proposed by Pearl (1990). It can be described as follows. If $\Delta = \{q_i \vdash p_i, i = 1, n\}$, then partition Δ into $\Delta_0 \cup \Delta_1 \cup \dots \cup \Delta_k$ where Δ_0 has the lowest priority and Δ_k has the highest. Δ_0 consists of those conditional objects $p \vdash q$ which are tolerated by Δ , i.e. such that $p \wedge q$ is consistent with $\{\neg p_i \vee q_i, i = 1, n\}$. Δ_1 is then made of those conditional objects which are tolerated by $\Delta - \Delta_0$, and so on. This ordering respects the priority ordering on the basis of the most specific rules (Poole, 1985).

The ordering so obtained is exactly the same as the one derived by changing each conditional assertion $p_i \vdash q_i$ in Δ into its material counterpart $\neg p_i \vee q_i$ and attaching to it the weight $N^*(\neg p_i \vee q_i)$ computed using the least specific possibility distribution π^* from $PI(\Delta)$. Let K^* be the possibilistic knowledge base so constructed from Δ . It can be proven (Benferhat et al., 1992) that $p \vdash q$ belongs to the rational closure of Δ if and only if $K^* \cup \{p\} \vdash_{\pi^*} q$. Hence, the plausible conclusions derivable from evidence p and generic knowledge Δ can be computed using the possibilistic logic proof methods of section 7. In particular, the complexity of system P and rational inference is roughly similar to that of inference in propositional logic (e.g. polynomial for Horn clause-like conditional assertions).

This approach partially solves the problem of irrelevant properties. In the Tweety example the possibilistic knowledge base obtained by the ranking procedure is the one of section 7 ($b \vdash f$ is the most tolerated of rules, and hence transformed into the material counterpart $\neg b \vee f$ in K^* with the lowest weight α), and it holds that $\{b, r\} \cup K^* \vdash \neg f$.

Unfortunately, the rational closure still has problems when a class is exceptional for a superclass with respect to some attribute. Then the least specific ranking does not allow us to conclude anything about whether this class is normal with respect to the other attributes. For instance, if we add $b \vdash \ell$ (birds have legs) to the knowledge base of the Tweety example, then $p \vdash \ell$ is not in the rational closure of $\Delta \cup \{b \vdash \ell\}$. This problem has been solved in various ways by several authors. Geffner (1990) defines a partial ordering that comes down to checking for maximal consistent sub-bases of $K^* \cup E$. Goldszmidt et al. (1990) exploit the infinitesimal probability setting and apply maximum entropy methods to characterize another ranking of worlds. More recently, Benferhat et al. (1993) and Lehmann (1993) have suggested an approach based on a lexicographic ordering whereby each interpretation should violate as few conditional assertions as possible. All these methods solve the problem of blocking of property inheritance in the Tweety example, but their motivations are unclear, and counter-examples where these techniques give counterintuitive results can be found. Moreover, all of them are syntax-dependent, and their complexity is higher than the rational closure.

10 Conclusion

Numerical and symbolic approaches to uncertainty should not be considered as competing models. It is far more interesting and fruitful to display their underlying coherence, rather than to argue in favour of one or against the other. The notion of conditional assertions, or equivalently of conditional events, as discussed here captures the basic features of plausible reasoning. What is more, it does this in a way which is in accordance not only with non-monotonic logics, but also with probability theory and possibility theory. Possibilistic logic, which can encode conditional assertions as ordered sets of default rules, can be viewed as a logic of accepted beliefs, in full agreement with a calculus of extreme probabilities. It is a natural framework for encoding Lehmann's rational closure.

However, further work is still needed. An important issue that has received little attention so far

is the notion of independence. To what extent does the probabilistic definition of independence have counterparts in the other uncertainty models? The limitations of conditional event logic for dealing with property inheritance are directly linked to this question. The RM axiom of section 9 does express a default independence assumption, but is not expressive enough to address all the difficulties.

Note that in probability theory, independence assumptions are always made explicit, while in logic distinct propositional symbols are implicitly independent. The language of logic encodes *dependencies*, but there is no *independence* symbol. Some insights into qualitative independence can be found in Pearl (1988); these results could be usefully related to the conditional event logic. It is clear that the next step in exception-tolerant reasoning is to search for a new concept of independence that is less permissive than logical independence, but more flexible than probabilistic independence, in the context of ordinal uncertainty approaches like possibility theory. Our suggestion is that rational monotonicity should be augmented by conditional independence assertions whose statement is dictated by the structure of the knowledge base Δ . This is the final lesson from Bayesian networks to be taken advantage of. Preliminary steps along the line can be found in Goldszmidt and Pearl (1992), Fonck (1993) and Benferhat et al. (1994).

Acknowledgements

The authors wish to particularly thank Paul Krause for his valuable help when preparing a revised version of the paper. This work has been partially supported by the ESPRIT Basic Action 6156 entitled "Defeasible Reasoning and Uncertainty Management Systems" (DRUMS-II).

References

- Adams, EW, 1975, *The Logic of Conditionals*. Reidel.
- Amarger, S, Dubois, D and Prade, H, 1991, "Constraint propagation with imprecise conditional probabilities." In: *Proc. 7th Conf. on Uncertainty in AI*, pp 26–34.
- Benferhat, S, Cayrol, C, Dubois, D, Lang, J and Prade, H, 1993, "Inconsistency management and prioritized syntax-based entailment." In: *Proc. IJCAI'93*, pp 640–645.
- Benferhat, S, Dubois, D and Prade, H, 1992, "Representing default rules in possibilistic logic." In: *Proc. KR'92*, pp 673–684.
- Benferhat, S, Dubois, D and Prade, H, 1994, "Expressing independence in a possibilistic framework and its application to default reasoning." In: *Proc. ECAI'94*, pp 150–154.
- Calabrese, P, 1987, "An algebraic synthesis of the foundation of logic and probability." *Information Sciences* 42 187–237.
- De Finetti, B, 1936, "La logique de la probabilité." In: *Actes du Congrès Inter. de Philosophie Scientifique*, Paris, 1935, pp 565–573. Hermann et Cie.
- De Finetti, B, 1937, "La prévision, ses lois logiques et ses sources subjectives." English translation in *Studies in Subjective Probability*, H Kyburg and HE Smokler, eds., Wiley, 1964.
- Dubois, D, 1986, "Belief structures, possibility theory and decomposable confidence measures on finite sets." *Computers and Artificial Intelligence* 5(5) 403–416.
- Dubois, D, Godo, L, López de Màntaras, R and Prade, H, 1993, "Qualitative reasoning with imprecise probabilities." *J. Intelligent Information Systems*, 3, pp 319–363.
- Dubois, D, Lang, J and Prade, H, 1991, *Possibilistic logic*. Tech. Report IRIT/91-98-R. (In DM Gabbay et al., eds.), *Handbook of Logic in AI and Logic Programming*, 3. Oxford University Press, 1994, pp 439–513.
- Dubois, D and Prade, H, 1988a, *Possibility Theory*. Plenum Press.
- Dubois, D and Prade, H, 1988b, "An introduction to possibilistic and fuzzy logics." In P Smets et al. (eds.), *Non-Standard Logics for Automated Reasoning*, pp 287–326. Academic Press.
- Dubois, D and Prade, H, 1991, "Epistemic entrenchment and possibilistic logic." *Artificial Intelligence* 50 223–239.
- Dubois, D and Prade, H, 1993, "A glance at non-standard models and logics of uncertainty and vagueness." In: JP Dubucs (ed.), *Philosophy of Probability*, pp 169–222. Kluwer Academic.
- Dubois, D and Prade, H, 1994a, "Focusing versus updating in belief function theory." In: RR Yager et al. (eds.), *Advances in the Dempster–Shafer Theory of Evidence*, pp 71–95. Wiley.

- Dubois, D and Prade, H, 1994b, "Conditional objects as nonmonotonic consequence relations." In: *Proc. KR'94*.
- Dubois, D, Prade, H and Smets, P, 1993, "Representing partial ignorance." In: *Post-UAI-93 Workshop on Higher Order Uncertainty*, Washington, DC. Revised version to appear in *IEEE Trans. on Systems, Man and Cybernetics*.
- Elkan, C, 1993, "The paradoxical success of fuzzy logic." *Proc. AAAI'93*, pp 698–703.
- Fariñas del Cerro, L, Herzig, A and Lang, J, 1992, "From ordering-based nonmonotonic reasoning to conditional logics." In: *Proc. ECAI'92*, pp 314–318. Extended version. *Artificial Intelligence* 66, 375–393.
- Fine, TL, 1973, *Theories of Probability*. Academic Press.
- Fonck, P, 1993, *Réseaux d'inférence pour le raisonnement possibiliste*. Thèse de Docteur es Sciences, Université de Liège, Belgium.
- Gabbay, DM, 1985, "Theoretical foundations for non-monotonic reasoning in expert systems." In: KR Apt (ed.), *Logics and Models of Concurrent Systems*, pp 439–457. Springer-Verlag.
- Gärdenfors, P, 1988, *Knowledge in Flux*. The MIT Press.
- Gärdenfors, P and Makinson, D, 1994, "Non-monotonic inference based on expectations." *Artificial Intelligence* 65, 197–245.
- Geffner, H, 1992, *Default Reasoning: Causal and Conditional Theories*. The MIT Press.
- Goldszmidt, M, Morris, P and Pearl, J, 1990, "A maximum entropy approach to nonmonotonic reasoning." In: *Proc. AAAI'90*, 646–652.
- Goldszmidt, M and Pearl, J, 1992, "Rank-based systems." In: *Proc. KR'92*, pp 661–672.
- Goodman, IR, Nguyen, HT and Walker, EA, 1991, *Conditional Inference and Logic for Intelligent Systems*. North-Holland.
- Kraus, K, Lehmann, D and Magidor, M, 1990, "Nonmonotonic reasoning, preferential models and cumulative logics." *Artificial Intelligence* 44 167–207.
- Kyburg, Jr. HE, 1974, *The Logical Foundations of Statistical Inference*. Reidel.
- Lauritzen, SL and Spiegelhalter, DJ, 1988, "Local computations with probabilities on graphical structures and their application to expert systems." *J. Royal Statist. Society* 50(2) 157–224.
- Léa Sombé, 1990, *Reasoning Under Incomplete Information in Artificial Intelligence*. Wiley.
- Lehmann, D, 1993, *Another perspective on default reasoning*. Inst. Computer Science, Hebrew University, Jerusalem.
- Lehmann, D and Magidor, M, 1992, "What does a conditional knowledge base entail?" *Artificial Intelligence* 55 1–60.
- Lewis, D, 1973, *Counterfactuals*. Basil Blackwell.
- Makinson, D and Gärdenfors, P, 1991, "Relations between the logic of theory change and nonmonotonic logic." In: A Furmann and M Moreau (eds.), *The Logic of Theory Change*, 185–205. Springer-Verlag.
- Nilsson, N, 1986, "Probabilistic logic." *Artificial Intelligence* 28 71–87.
- Pearl, J, 1988, *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann.
- Pearl, J, 1990, "System Z: A natural ordering of defaults with tractable applications to default reasoning." In: M Vardi (ed.), *Proc. TARK'90*, pp 121–135. Morgan Kaufmann.
- Poole, DL, 1985, "On the comparison of theories: Preferring the most specific explanation." In: *Proc. IJCAI'85*, pp 144–147.
- Rescher, N, 1969, *Many-Valued Logic*. McGraw-Hill.
- Shafer, G, 1976, *A Mathematical Theory of Evidence*. Princeton University Press.
- Shenoy, PP and Shafer, G, 1990, "Axioms for probability and belief-function propagation." In: RD Shachter et al. (eds), *Uncertainty in Artificial Intelligence 4*, pp 169–198. North-Holland.
- Shoham, Y, 1988, *Reasoning About Change*. The MIT Press.
- Smets, P, 1988, "Belief functions." In: P Smets et al. (eds), *Non-Standard Logics for Approximate Reasoning*, pp 253–286. Academic Press.
- Spohn, W, 1988, "Ordinal conditional functions: A dynamic theory of epistemic states." In: WL Harper and B Skyrms (eds.), *Causation in Decision, Belief Change, and Statistics*, pp 105–134. Kluwer Academic.
- Thöne, H, Güntzer U and Kießling, 1992, "Towards precision of probabilistic bounds propagation." In: *Proc. 8th Conf. on Uncertainty in AI*, pp 315–322.
- Zadeh, LA, 1965, "Fuzzy sets." *Information and Control* 8 338–353.
- Zadeh, LA, 1978, "Fuzzy sets as a basis for a theory of possibility." *Fuzzy sets and Systems* 1 3–28.
- Zadeh, LA, 1985, "Syllogistic reasoning in fuzzy logic and its application to usuality and reasoning with dispositions." *IEEE Trans. SMC* 15(6) 745–763.