FULL NONMONOTONICITY: A NEW PERSPECTIVE IN DEFEASIBLE REASONING

Pietro Baroni*, Giovanni Guida*, Silvano Mussi°*

* Università di Brescia, Dipartimento di Elettronica per l'Automazione, Via Branze 38, 25123 Brescia, Italy
tel: +39 30 3715469 fax: +39 30 380014 - e-mail: {baroni, guida}@bsing.ing.unibs.it

° Consorzio Interuniversitario Lombardo per l'Elaborazione Automatica, Via Sanzio 4, 20090 Segrate (MI), Italy
tel: +39 2 26922434 fax: +39 2 2135520 - e-mail: mussi@icil64.cilea.it

1 Introduction: the concept of defeasible reasoning

The capability of drawing defeasible conclusions in presence of partial information is a crucial factor of intelligent behavior. To achieve this capability, human beings resort to a particular kind of knowledge, called *default knowledge*. The most significant property of default knowledge is that it can be exploited in the reasoning process even if there is only partial information about the satisfaction of the preconditions which allow its application, on condition that there is no reason to believe that such preconditions are not satisfied. If new information becomes available from which the falsity of such preconditions can be deduced, the conclusions derived from the application of default knowledge have to be retracted. This particular form of reasoning, involving the use of default knowledge, may be called *defeasible reasoning*.

In order to build automated reasoning systems including defeasible reasoning capabilities, many extensions of classical logic have been proposed as models of defeasible reasoning. These proposals, even if differing in many important technical details, share a common conceptual ground, since they all rely substantially on the same conceptual model of defeasible reasoning activity. Among the most notable and classic proposals in this field we mention default logic [7] and nonmonotonic logic [5].

However, this conceptual model suffers from some important limitations, which severely restrict its applicability scope and prevent it (as well as the approaches grounded on it) to correctly capture and represent some very general and common cases of practical defeasible reasoning. In order to overcome these limitations, both a more general conceptual model of defeasible reasoning activity and a formalism capturing the new concepts introduced are needed.

Following the track of some previous investigations in this area [1] [2], this paper points out a further limitation of most known models of defeasible reasoning which (as to our knowledge) has never been highlighted before.

The paper is organized as follows. In section 2 we briefly review some approaches to defeasible reasoning and identify the conceptual model underlying them. In section 3 we define a property for defeasible reasoning formalisms, called full nonmonotonicity: we point out that this property has received a very limited attention in the past and that most approaches to defeasible reasoning fail to satisfy this property. In section 4, we present a common sense example of defeasible reasoning (let’s call it the "ill secretary" example) that requires the full nonmonotonicity property and that, therefore, can not be adequately represented in most classical defeasible reasoning models. Finally, in section 5, we describe a new approach to modeling defeasible reasoning, based on the concept of A- uncertainty [2]. We discuss the conceptual advantages of this approach and we show that it can be properly used to deal with the "ill secretary" example and with similar kinds of defeasible reasoning activity.

2 Defeasible reasoning: a quick survey

This section presents a very synthetic review of some approaches to defeasible reasoning. Since there is a huge number of papers and different approaches about this topic, we focus on some of the most well-known ones and mention only their basic aspects, which are however sufficient for the subsequent discussion.

2.1 Default logic

In default logic [7], default knowledge is represented through specific inference rules called *defaults*. A default is an expression of the form: p(x) : j(x)/ c(x), where p(x) is the prerequisite, j(x) is the justification, and c(x) is the consequent of the default. The meaning of the formula is: if p(x) is known and if j(x) is consistent with what is known, then it is possible to deduce c(x). A default is called normal if j(x) = c(x).

A typical example of normal default is: bird(x) : fly(x) / fly(x), which means: if x is a bird and it is consistent with other available information to assume that x flies, then infer that x flies.

2.2 Non monotonic logic and autoepistemic logic

In nonmonotonic logic [5], the concept of "conceivable", is represented through a modal operator M. The formula Mp means that p is conceivable, that is equivalent to state that $\neg$ p is not provable.

Default knowledge is represented by means of implication relations of the form:

$p(x) \land Mc(x) \rightarrow c(x)$.

Autoepistemic logic [6] is a derivation of nonmonotonic logic, which moves the attention from reasoning about the conceivability of propositions to reasoning about what is believed about propositions.

In autoepistemic approach, the reason why we infer that Tom can fly from the fact it is a bird is that if Tom could not fly, we would know it. So, in absence of more specific information, we assume we can reason as if no further interesting information could come, because we believe that if it could, we knew it. The modal operator of nonmonotonic logic, which will be denoted here as Lp, assumes therefore the meaning of "p is believed".

Default knowledge can then be represented as:

$p(x) \land \neg c(x) \rightarrow L \neg c(x)$,

which is the formal translation of " if the premise hold and the consequence does not, I would know it".

2.3 Circumscription

Circumscription [3] is an alternative technique which aims to formally represent the intuitive tendency to minimize the assumptions concerning some "abnormal" notion, restricting abnormalities only to cases where they are manifest. Circumscription can be used to model defeasible reasoning [4] by introducing a predicate ab, denoting abnormality, which has to be circumscribed, and representing default knowledge by implication rules having the form:

$p(x) \land \neg ab(x) \rightarrow c(x)$.

2.4 The underlying conceptual model: rules and exceptions

Given this very quick survey, it is in any case possible to identify a basic conceptual model of defeasible reasoning shared by these approaches (and by their innumerable technical variants appeared in the literature).

This model relies on the idea that the rules, i.e. the pieces of knowledge used in reasoning activity such as "birds fly" (or, more formally, IF bird(x) THEN flies(x)), may admit exceptions and that it is impossible to include an explicit and detailed representation of all these exceptions within the rules themselves.

In these cases, the premise of a rule is only partially specified (if this is the case, the rule is called a "default rule") and the fact that an individual satisfies the premise does not guarantee that the individual is not an exception. Since, during reasoning activity, the fact that the premise holds is sufficient to enable the application of a default rule (as well as of other "normal" rules), it may happen that a rule is applied, inadvertently, to an "exceptional" individual.

Subsequently, when the fact that the individual was an exception is noticed, the application of the rule reveals incorrect and the consequences of the application have to be retracted.

To put it in other words, it may be stated that the premise of a default rule is sufficient to partition the set AI of all individuals in two subsets: a set RNA of individuals to which the rule certainly does not apply (i.e. those individuals which do not satisfy the premise) and a set RPA of individuals to which the rule could possibly be applied. This second set is larger than the precise set RMA of individuals to which the rule must be applied, since it includes also a set EXC of exceptions. However, determining exactly the set RMA and/or verifying a priori if and individual belongs to RMA, is considered impossible (or unpractical): the set RPA is thus used to discriminate the individuals to which the rule is, defeasibly, applied. Of course, if subsequently it emerges that an individual to which the rule was applied belongs to EXC, the derived conclusions have to be retracted.

This basic conceptual model lies on the foundations of all the approaches to defeasible reasoning introduced above.

As far as the concept of exception is concerned, two additional remarks are worth to be introduced, since they will be recalled in the following.

First of all it has to be noted that, within the set RPA, only the distinction between RMA and EXC is considered: no further conceptual refinements (i.e. no subpartitions) are envisaged for these sets.

Moreover the approaches differ in the way the fact that an individual is an exception is noticed. In default logic, nonmonotonic logic, and autoepistemic logic the concept of exception is strictly related to the negation of the consequent of the rule. In other words, in order to retract the application of a rule such as IF bird(x) THEN flies(x) to an individual, let say Tweety, it is necessary to derive $\neg$flies(x), through some other reasoning path (for instance IF ostrich(x) THEN $\neg$flies(x)). Though rather widespread, the identification of exceptions with the negation of the

consequent suffers from some significant conceptual limitations (the interested reader may refer to [2] for a thorough discussion about this aspect).
In circumscription, the representation of exceptions is related to the concept of abnormality, which is formally represented by a predicate ab(x), which is true if the individual x has to be regarded as abnormal, i.e. as an exception.
Default rules have the form IF bird(x) AND ¬ab(x) THEN flies(x), and the consequences of the application of a rule have to be retracted if, through some reasoning path, the abnormality of the individual is asserted. This view of exceptions, which does not strictly relies on the negation of the consequence, should be regarded, in our view, as preferable. However, it has to be remarked that, in the examples presented in literature we are aware of, actually abnormality always corresponds to the negation of the consequent.

3 The full nonmonotonicity property

Defeasible reasoning is commonly denominated also nonmonotonic reasoning. This denominations refers to a technical property of the logical formalisms which aim to model defeasible reasoning activity.
In fact, classical logical formalisms, which are admittedly unable to represent defeasible reasoning, satisfy a property called *monotonicity*, which, in a restricted form, can be defined as follows: given three different sets A, B, C of expressions, if $A \vdash B$, then $(A \cup C) \vdash B$.

Of course, in any approach to model defeasible reasoning, monotonicity property can not hold, since the acquisition of new information (e.g. the fact that Tweety is an ostrich, represented by C) in addition to previously available one (e.g. the fact that Tweety is a bird, represented by A) may deny some previously deduced conclusions (e.g. the fact that Tweety flies, represented by set B).

In order to introduce now our definition of full nonmonotonicity and to compare it with nonmonotonicity in its usual understanding, it is necessary to set up an introductory conceptual background.
In general, in any formal reasoning context, we have some general rules defining how expressions (which are the basic instrument to represent facts about the world) should be formed (e.g. rules defining Well Formed Formulas in a logic formalism).
These rules allow to generate the (possibly infinite) set E of all expressions one might consider when reasoning about some piece of the world. Of course, (and happily) the reasoning activity normally needs to deal only with a very restricted subset of E.
At any stage of the reasoning activity, the set E can be partitioned in two subsets: the set AE of expressions for which a truth value (namely, TRUE or FALSE in classical logic) has been explicitly asserted (AE stands for *asserted expressions*) and the set UE of expressions to which a truth value has not been explicitly ascribed (UE stands for *unasserted expressions*). The set UE includes both the expressions to which it is actually impossible to ascribe a truth value, for instance due to a lack of information, and the (very numerous) expressions which are not considered in the reasoning activity, though their truth value could be derived given the available information (for instance, if a proposition p is asserted to be TRUE, also any expression of the type $p \vee q$, could be explicitly asserted to be true, but, of course, the bothersome and useless activity of deriving all these assertions is normally avoided).
Moreover, the set AE can be further partitioned in two subsets, the set TE of expressions whose asserted truth value is TRUE, and the set FE of expressions whose asserted truth value is FALSE.
Each step of the reasoning activity can be seen, in general, as a modification of the partitions defined above, which moves one or more expressions from a partition to another one.
In classical reasoning, this modification is always and only the shift of expressions from UE to AE.
In fact the monotonicity property implies both that asserted expressions can not revert to unasserted and that the truth value ascribed to an expression can not be changed. So the monotonicity property implies both that the set AE is monotonically growing and that there are no cross shifts between TE and FE.
That being stated, an approach can be defined nonmonotonic if it fails to satisfy either one of the properties entailed by monotonicity. Therefore, we can define two types of nonmonotonicity (denoted as AU- and TF-respectively) for a reasoning formalism:

- a reasoning formalism is said *AU-nonmonotonic*, if it allows an expression to move from AE to UE;
- a reasoning formalism is said *TF-nonmonotonic*, if it allows cross movements of expressions between TE and FE.

As we will discuss later, most approaches to defeasible reasoning are only TF-nonmonotonic.
We can now define the property of *full nonmonotonicity* as follows: a reasoning formalism is said fully nonmonotonic, if it is both AU-nonmonotonic and TF-nonmonotonic.

The existence of two types of nonmonotonicity has never been pointed out in previous literature. As a matter of fact, the existence of two distinct types of exceptions, leading respectively to AU- and TF-nonmonotonicity is suggested by previous considerations. However such distinction can not be explicitly captured by any of the formalisms presented in section 2, because they envisage only one representation for default rules and, consequently, for exceptions. More precisely, recalling that default logic, nonmonotonic logic, and autoepistemic logic allow revision only when the negation of the consequent has been explicitly asserted, it is easy to show that such formalisms enjoy only TF-nonmonotonicity. On the other hand, since the concept of abnormality in circumscription is more general (and more ambiguous), this formalism is, at least in theory, open to full nonmonotonicity. However, since no explicit distinction between different classes of abnormalities is envisaged in this formalism, the actual achievement of full nonmonotonicity in circumscription is left to the user of the formalism rather than being enforced by proper structures within the formalism itself.
The meaning and the importance of full nonmonotonicity property in common sense reasoning will be better illustrated by the example presented in the following section.

4 The ill secretary example

Suppose you arrive at work at 8 a.m. and you are wondering whether your secretary, say Helen, who normally arrives at 9 a.m., will come today. By default, you can assume she will, because today is a normal working day and normally Helen gives notice in advance of a programmed absence (for example, vacation).
Now imagine that a colleague of Helen tells you that yesterday Helen had an accident and she has a broken leg. Given this information, you easily deduce that she will not come, and retract your precedent assumption: this is a case of TF-nonmonotonicity, because a proposition which was before asserted as TRUE, is subsequently considered as FALSE.
Consider now a different case: again, while you are waiting for Helen, a colleague informs you that yesterday Helen went home with a bad cold. Given this new information, your confidence in Helen's coming decreases but, on the other hand, you can not exclude she will come (since she might feel better or even come with a cold). In other words, you are now completely in doubt about the fact that Helen will come and would not prefer neither the hypothesis that she will come nor that she will not. This is a case of AU-nonmonotonicity, because a proposition which was before asserted as TRUE, become subsequently not asserted any more.
It is worth to remark, referring to this simple example, the significant conceptual difference existing between TF- and AU-nonmonotonicity.
TF-nonmonotonicity refers to a case where the new incoming piece of information improves our information state about a specific subject, namely a proposition, and enables us to ascribe a more correct truth value to the proposition itself. So, hearing that Helen has a broken leg improves our information state about her coming and allows us to state that she will not come, contrary to our previous assumption, based on a less reliable information state (namely, our default knowledge).
AU-nonmonotonicity refers to a more peculiar, though very common, case, where the new incoming piece of information helps us to realize that we are more ignorant about a specific subject than we believed to be. So hearing that yesterday Helen had a cold, we realize that we are completely in doubt about the question if she will come today, and retract our previous conclusion not because it has to be replaced by a more reliable one, but because we are unable to draw any conclusion.
It may seem a little bit paradoxical that the acquisition of new information reduces the scope of our deduction capability. However, the ill secretary example shows that this case can be found in very simple (and common) examples of common sense reasoning. Moreover, the paradox reveals to be only apparent if one considers that, also in this case, the acquisition of new information increases our knowledge. Simply, in this case the new piece of knowledge we learn is our inability to make deductions. In fact, learning that we are ignorant about something, definitely increases our global state of knowledge, that includes both what we know and what we know to ignore.
Due to space limitations, we limit ourselves to present here the ill secretary example. However we claim (and the reader may easily imagine) that there are a plenty of similar examples both in daily common sense reasoning and in more specific application fields of defeasible reasoning (just think over reasoning in medicine).
In the light of this consideration, the previously showed inability of existing formalisms to explicitly deal with both TF- and AU-nonmonotonicity seems to represent a very important drawback. In the next section we present a proposal of a formalism aiming to overcome this limitation.

5 A fully nonmonotonic reasoning formalism

The approach we present here is based on an explicit representation of A-uncertainty, i.e. of the uncertainty about the applicability of a rule to an individual (for a thorough discussion about the concept of A-uncertainty see [1]). Since A-uncertainty concerns the applicability of a certain chunk of knowledge to an individual it is a property of the pair (knowledge, individual) and depends both on the features of knowledge and of individuals, so that it is possible to imagine, in principle, a different A-uncertainty assessment for each individual to which a given chunk of knowledge has to be applied. Moreover, as long as new information about the individual are acquired, the assessment of A-uncertainty relevant to the individual might need to be adjusted.
Starting from this concept, we propose here a simplified version of the formalism introduced in [2].
In our approach, uncertainty about a proposition, say the proposition A, in presence of a chunk of available evidence E, is represented by means of a pair ($\underline{bel}_E$(A, $\underline{true}$), $\underline{bel}_E$(A, $\underline{false}$)), say (bt_A, bf_A) for short. Such pair is called *belief state*, bels for short. The belief state represents how much one is authorized to believe in the association between a given proposition and its possible truth values, on the basis of the available evidence. Each one of the two elements of a belief state is a *belief degree*, represented by a real number in the interval [0, 1]. The concept of belief degree is related to the intuitive concept of amount of evidence supporting the credibility that a certain proposition should have a certain truth value. So, $\underline{bel}_E$(A, $\underline{true}$)=0 means that there is null (or negligible) evidence supporting the credibility that proposition A has the truth value $\underline{true}$ (note that this is totally different from excluding that $\underline{true}$ is a possible truth value for A).
Turning now to uncertainty about rules, given a production rule R, an individual x, and a body of evidence E, the *A-belief* of R with respect to x under E, also denoted by $\underline{A\text{-}bel}_E$(R, x), or ba_R for short, is a belief degree which provides a measure of how much one is authorized to believe that the rule is applicable to a given individual x.
As stated in [1], the cases in which a rule does not apply may be interpreted by assuming either a *conservative* attitude or an *evolutive* attitude. In fact, given the question "what do we know about the consequent in the cases where a rule does not apply to an exceptional individual ?", two answers are possible:

- according to a conservative attitude: we know nothing about the consequent,
- according to an evolutive attitude: if the individual is an exception, the consequent is false.

The additional knowledge concerning attitude, can simply be represented as a dynamic property $\underline{att}_E$(R, x) which can assume two values, E (evolutive) or C (conservative). For each rule, a general value of $\underline{att}_E$(R,x) can be stated, which has to be modified if new specific information about the individual is acquired.
Two simple propagation schemes (see [2] for a detailed discussion) for computing the belief state of the consequence B(x), from the belief state of the premise A(x) and the A-belief of the rule R= IF A(x) THEN B(x) can be proposed:
In the evolutive attitude:

$$bt_B = bt_A \bullet ba_R \bullet (1 - bf_A) \qquad bf_B = bt_A \bullet (1 - ba_R) \bullet (1 - bf_A)$$

In the conservative attitude:

$$bt_B = bt_A \bullet ba_R \bullet (1 - bf_A) \qquad bf_B = 0.$$

The proposed formalism can be applied to the ill secretary example as follows.
The default rule considered is: R = IF workingday(today) THEN comes(Helen).
You know for certain that today is a working day, namely $\underline{bels}$(workingday(today)) = (1, 0). You know also that the rule almost certainly applies, namely ba_R = 0.99, and adopt an evolutive attitude. Therefore you obtain $\underline{bels}$(comes(Helen)) = (0.99, 0.01), expressing the fact that you are almost sure Helen will come and only a very little part of your belief is left to contingencies.
Now suppose you learn Helen has a broken leg. It is possible to model the impact of this information as a decrease of the belief in the applicability of the rule, falling to zero (ba_R = 0), while the attitude remains evolutive. In this case you obtain $\underline{bels}$(comes(Helen)) = (0, 1), expressing the fact that you are now sure that Helen will not come. So a decrease of the value of ba_R realizes TF-nonmonotonicity in our formalism.
On the other hand suppose you learn Helen yesterday had a cold. The impact of this information can be modeled again through a decrease of the belief in the applicability of the rule, (ba_R = 0), but in this case there is also a shift from the evolutive to the conservative attitude.
The different propagation formulas applied in these case give therefore: $\underline{bels}$(comes(Helen)) = (0, 0), expressing a total ignorance about the question if Helen will come or not. So the combination of a decrease of the value of ba_R along with a change from evolutive to conservative attitude realizes AU-nonmonotonicity.

Due to space limitations, we can not enter a thorough discussion about merits and flaws of the proposed formalism nor carry out a detailed comparison with other formalisms. However we remark that, differently from many approaches to defeasible reasoning we are aware of, the formalism we propose offers an explicit and cognitively plausible way of realizing the full nonmonotonicity property, which, though very important for practical applications, has escaped the attention of most researchers in the field of defeasible reasoning in the past.

References

[1] P. Baroni, G. Guida, S. Mussi, Modeling uncertain relational knowledge: the AV-production rules approach, *Proc. ECSQARU 95, 3rd European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty*, Fribourg, CH, 1995, 18-27

[2] P. Baroni, G. Guida, S. Mussi, Modeling default reasoning through A-uncertainy, *Proc. IPMU 96, International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, Granada, E, 1996, 1197-1204

[3] J. McCarthy, Circumscription - A form of non-nonotonic reasoning, *Artificial Intelligence* 13, 1980, 27-39

[4] J. McCarthy, Applications of circumscription to formalizing commonsense knowledge, *Artificial Intelligence* 28, 1986, 89-116

[5] D. McDermott and J. Doyle, Non-monotonic logic I, *Artificial Intelligence* 13, 1980, 41-72

[6] R. Moore, Autoepistemic logic, in P. Smets, A. Mamdani, D. Dubois, and H. Prade (eds.) *Non-standard logics for automated reasoning*, Academic Press, London, 1988, 105-136

[7] R. Reiter, A logic for default reasoning, *Artificial Intelligence* 13, 1980, 81-132.