PLAUSIBILITY AS TRUTH
A 'local' pattern for non-monotonic reasoning

Silvana Badaloni* **- Alberto Zanardo****

Introduction. Several non-monotonic logics existing in the literature are conceived to deal with defeasible inferences of different kinds in the same way. As a consequence, the inner structures of the particular patterns of non-monotonic reasoning are hardly captured and represented. We believe that this is a limitation of these logics; for instance, it seems to us misleading to treat in the same way the usual deduction about Tweety, the deduction that there is no flight from Venice to Pittsburgh, based on the fact that no such flight appears on the time-table, and the deduction that a given ticket will not win the lottery, based on the fact that this event has a very low probability.

We think that the understanding of the inner structures of various different non-monotonicity patterns is a crucial point in the research on non-monotonicity and that such an understanding can be made deeper if specific formal languages are determined in which those patterns can be expressed. In other words, we hold that a general logical system for non-monotonic reasoning should be composed of many subsystems, each of which is capable of dealing 'locally' with the various aspects of this phenomenon; the language should be sufficiently expressive to distinguish the particular local situations, and hence to determine the particular deductive subsystem to put into action.

The aim of the present work is to analyze the features which are peculiar to those defeasible inferences generated by assertions like

(1) "generally, or typically, the elements of the set A have the property Q"

and to show that this kind of non-monotonicity can be dealt with by slightly extending first-order logic; in particular, the logical approach we are proposing is based on an effective way to produce defeasible inferences, without involving any reference to the models of the theory or to the global notion of consistency (and hence of deduction).

---

* Dipartimento di Elettronica e Informatica, Via Gradenigo 6A, I-35131 Padova. email: badaloni@ladseb.pd.cnr.it

** Dipartimento di Matematica Pura ed Applicata, Via Belzoni 7 - I-35131 Padova. email: alberto@pdmat1.unipd.it

The logical system we consider, called HP-logic (*high plausibility logic*) [Badaloni & Zanardo, 91,94], and is obtained from the usual first-order logic by adding binary determiners first and, then, by changing the usual semantics for the constants.

Determiners. The formulas constructed by using the binary determiners have the form $\Pi x(\alpha,\beta)$ and their truth depends on whether a given relation holds between the set of individuals fullfilling $\alpha(x)$ and that of the individuals fullfilling $\beta(x)$ (see [van Benthem 86] ch.s 1,2 and [Westerståhl 89] section 2). The formula $\Pi x(\alpha,\beta)$ will be viewed as an extensional translation of the natural language assertion

(2) "generally, or typically, the individuals having the property (determined by) $\alpha(x)$ also have $\beta(x)$".

In other words, referring to the canonical Tweety example of non-monotonic reasoning based on assertions like (1), we hold that the assertion "Generally, birds can fly" expresses a relation between properties, rather than a property of a single arbitrary bird; in other words, it asserts that

(3) "the property of being able to fly is highly plausible with respect to the property of being a bird",

which, from an extensional point of view, becomes:

(4) "the set of birds that can fly constitutes a significantly large subset of the set of all birds".

That is, high plausibility is reduced to a relation between sets and, in this context, the assertion (4) will be expressed by $\Pi x(Bird(x),Fly(x))$; thus, the determiner Πx permits us to express the high plausibility (of a property with respect to another) within the formal language itself, which substantially remains an extensional first order language. Of course, the crucial problem is to define suitably the relation "to constitute a significantly large subset of", that is to define the semantical properties of the determiner Πx [Badaloni & Zanardo, 94].

Non standard evaluation of the constants. To conclude "Tweety can fly" from the premises "Generally birds fly" and "Tweety is a bird", we must consider also that the only knowledge we have about Tweety is that it is a bird; indeed, if further information is added (e.g. that Tweety is a penguin and no penguin flies) the conclusion could no longer be possible. Thus, the assertion "Tweety is a bird" can be rewritten as

(5) "Tweety is a bird" and "the most specific thing we know about it is that it is a bird", or, simply,

(6) "all we know about Tweety is that it is a bird".

Taking into account the second part of (5), we have that, as far as we know, Tweety could be any kind of bird, that is, any bird could be the individual denoted by the term 'Tweety'. This fact justifies the departure from the usual semantics for first order logic; every individual constant is evaluated on a set of individuals: the set of all possible denotations of that constant in the present situation.

Then, the non-standard evaluation of the constants is connected with a partial information condition, in order to ascribe to them a "range of uncertainty"; that is, we want to take into account that the knowledge that we have about (the denotation of) a term does not always identify a single individual. In general, the fact that the valuation of the constant c is the set X will be viewed as the semantical counterpart of the assertion "all we know about c is that it is in X". It follows that the concept of equality between two individual terms has to be modified; a new identity symbol, $\hat{=}$, espressing "possible identity" is introduced. Intuitively, the formula $\vartheta_1 \hat{=} \vartheta_2$ means that "it is possible that ϑ_1 and ϑ_2 denote the same individual", and is true whenever the valuations of ϑ_1 and ϑ_2 are not disjoint. An important consequence of the non-standard evaluation of the constants is that HP-logic is a three-valued logic: the atomic formula P(c) has the undefined truth value whenever the valuation of c is not included in the extension of P as well as in that of not-P.

The assertions (4) and (6) yield

(7) "It is highly plausible that Tweety can fly",

which expresses the (absolute) high plausibility of the assertion "Tweety can fly".

The formalism of HP-logic allows the definition of unary operators, $\Pi_{c/x}$, such that the formula $\Pi_{c/x}\alpha(x)$ can be reasonably read "$\alpha(c)$ is highly plausible" and, in the deductive system developped in [Badaloni & Zanardo, 94], the deduction of (7) from (4) and (6) can be formalized.

High plausibility as truth. The conclusion

(8) "Tweety can fly"

can be viewed as a first-order approximation of (7): when truth is not available, it is natural to use high plausibility in its place and thus make it become truth. Thus, we have to define the conditions under which high plausibility can be replaced by truth. In [Badaloni and Zanardo 91,94], a set of formulas (the set of *meaningful formulas*) is

determined such that, for every consistent subset of it, the substitution of high plausibility by truth gives rise to a consistent set of first order formulas. Indeed, not all the formulas involving the notion of high plausibility can be meaningfully translated into first-order formulas; for instance, from the information that it "is not highly plausible that constant c has property Q" we cannot reasonably infer anything about whether c actually has property Q or not.

Non-monotonicity. In our approach, non-monotonicity is dealt with as follows. As shown above, the conclusion (8) is drawn from the set L consisting of the premises (4) and (6), via the first-order translation of (7) into (8). If we add the information that "Tweety is a penguin and no penguin flies" (which actually increases our knowledge about Tweety) to the information (4) and to "Tweety is a bird", we obtain a set L' of premises which does not contain L. The premise (6) is not in L': the added information about Tweety makes this premise false. From set L' it is possible to conclude ¬Fly(Tweety) because the valuation of Tweety is included in the set of penguins and hence in the extension of ¬Fly. Thus, in our approach, non-monotonic patterns of reasoning are possible, while the logic itself remains monotonic.

The Nixon diamond. Similarly, as far as the "Nixon diamond" is concerned, the conclusion that "it is highly plausible that Nixon is a pacifist" is possible if we assume that "generally, quakers are pacifists" and "all we know about Nixon is that he is quaker". Of course, we need the hypothesis that "all we know about Nixon is that he is a republican" in order to have the (high plausibility of the) opposite conclusion. But, if we know that "Nixon is a quaker and a republican", then both the hypotheses "all we know ..." fail and hence none of the above conclusions can be drawn. This shows how, if different typical concepts relative to the same individual may lead to conflicting conclusions: this ambiguity introduces a new kind of uncertinty, and it would be wrong to draw any conclusion.

The temporal aspect. Work in progress concerns the possibility of embedding HP-logic in a temporal framework and, as a first step in this direction, we will take into account that knowledge can change with time. This has an obvious formal counterpart as far as the knowledge of individual terms is concerned: the evaluation of the constants must be time-sensitive. As an example, if we assume that knowledge increases with time, the evaluation of a constant c at a given instant t will be a subset of the evaluation of c at any t' earlier than t.

References

[Badaloni & Zanardo 91]. 'Typicality for Plausible Reasoning', in E. Ardizzone, S. Gaglio, F. Sorbello (Eds.) *Trends in Artificial Intelligence*, Lecture Notes in Artificial Intelligence, subseries of Lecture Notes in Computer Science, Springer Verlag, 1991, pp.470-474.

[Badaloni & Zanardo 94]. 'Plausible Reasoning: a First-Order Approach', to appear, 1994.

[van Benthem 86]. J. van Benthem: *Essays in Logical Semantics*, D. Reidel Pub. Company, Dordrecht, 1986.

[Westerståhl 89]. D. Westerståhl: Quantifiers in Formal and Natural Languages, in D. Gabbay and F. Guenthner (eds.), *Handbook of Philosophical Logic Vol.IV*, D. Reidel Publishing Company, Dordrecht, 1989, pp.1-131.