The Rôle of Defeasible Reasoning in the Modelling of Scientific Research Programmes

Claudio Delrieux
Universidad Nacional del Sur
Alem 1253 - (8000) Bahia Blanca - ARGENTINA
TE: (54) (291) 4595101 Ext. 3381 - e-mail: claudio@acm.org

Abstract

Classical studies in the theory of science concentrated on a logical analysis of scientific theories, with the unfortunate consequence of leaving aside many interesting epistemological problems (the scientific discovery problem, or the representation of sociological and pragmatical drives, among others). In the 70's a new epistemological view emerged, which suggested a departure from logical positivism. However, either the formalisms were still close to standard logic, or the presentations were too asystematic, and the formalization of scientific procedure was still obscure. In this paper we present a defeasible reasoning system that is an adequate model for representing scientific theories. Scientific reasoning is regarded mostly as a design *process. The system incorporates incomplete or uncertain evidence about a given situation as information provided by more or less trustable sources, to extend the reasoning context. This context is then used as a basis for a defeasible reasoning process. Then we show how our system can provide an adequate basis for a scientific reasoning model. In particular, we concentrate on the epistemology of the scientific research programmes.*

Keywords: Knowledge Representation, Defeasible Reasoning, Theory of Science.

1 Introduction

Most epistemologic studies in the past Century concentrated on a logical analysis of science, perhaps due to the influx of the Vienna Circle. This had unfortunate consequences. First, the cientific discovery problem (one of the most interesting and important epistemologic inquiries) was left aside because of its apparent nonlogical nature. Second, the fixation with deduction produced logical reconstructions of scientific theories that were in general unrealistic, and the representation of the scientific method was too rigid. Third, many important issues, for example the sociological drive or the strategic and pragmatic dimension, could not find its way in the formalism. This trend was modifyed in the 70's, when work by Khun and Feyerabend among others suggested a radical departure from logical positivism. This slowly encouraged to consider the use of inference mechanisms that deviate from standard logical practice (retrodiction, hypothetical reasoning, dialectics, and many others) as adequate scientific procedure.

However, these proposals fall short to overcome some of the difficulties mentioned above, perhaps because the formalisms are still too close to standard logic, and the tentative nature of scientific knowledge is somewhat overlooked. This is unfortunate, because we have a good tradition in defeasible reasoning systems that can bring impetus to a new, computational, theory of science. Defeasible reasoning is concerned with tentative knowledge representation and best conclusion inference under circumstances. In defeasible reasoning we can find at least three different branches: representation of uncertain evidence (see [16, 17, 12]), reasoning with default rules (see [5, 11, 15], and the use of ampliative inference rules (abduction, induction, analogy, etc., see [1, 9, 13]). These formalisms were developed independently, and little if any application in scientific reasoning was considered.

For this reason, here we present a reasoning system aimed to provide a formally and pragmatically adequate solution to these and other representation and reasoning issues. The system regards incomplete or uncertain evidence about a given situation as information provided by more or less trustable sources. This knowledge, together with the deductive knowledge of the context, is used as a basis for a deductive inference process, provided that no contradiction arises. In case of contradiction, the least trustable knowledge is discarded. The extended context is then used as a basis for a defeasible reasoning process. Ampliative inference patterns are also considered. The system has a semantic characterization of the set of conclusions, and a derivation procedure is proven sound and complete with respect to this semantic. The derivation procedure leads straightforwardly to a tractable computational implementation. Then we show how our system can provide an adequate basis for a scientific reasoning model. In particular, we concentrate on the epistemology of the scientific research programmes proposed by Lakatos.

2 Some pragmatic considerations

In the description of our reasoning system we will incorporate some notions borrowed from the theory of knowledge, in particular, we will use extensively the distinction between *de dicto* and *de re* modalities. These modalities qualify the attribution of a property p to an individual x. In a *de dicto* modal sentence, the sentence itself is qualified (*i.e.*, we have "$\Diamond(x \in p)$"), and thus mainstream mathematical modal logics correspond to *de dicto* modalities[1]. In our pragmatic analysis, we can interpret this sentence as *"(I believe that) I see* $p(x)$*"*.

In a *de re* modal sentence, the attribution of the property p to the individual x is qualified (*i.e.*, we have "$x \Diamond \in p$"), and thus the sentence is plain first order (nonmodal). We interpret this sentence as *"(I believe that) normally* $p(x)$*"*. As we can see, the differences between these modalities were overlooked in the development of mathematical logic, and then are inexpressible in a first order logic language.

In this work, however, we must emphasize the distinction from the standpoint of the theory of knowledge (see [2]). A *de dicto* sentence refers to a given state of affairs. A *de dicto* belief, then, should be held about *particular* situations, (*ground* atomic sentences or evidence). It seems absurd to hold a *de dicto* belief about nonexistent indi-

[1] In first order formalizations of modal logic, it is usual to designate as *de dicto* a sentence in which the scope of the modal operator includes the scope of the (standard first order) quantifier (*v.gr.* $\Diamond\forall X.p(X)$.), and as *de re* when the contrary is the case [18].

viduals or about general (*nonground*) knowledge. On the other hand, *de re* beliefs qualify the inherence of one property in another, and then, a *de re* belief is about sets of individuals (*i.e.*, general knowledge), existent or not. We can summarize the point in the following table.

	de dicto belief	*de re* belief
Particular	(I see that) Opus is a penguin.	(Normally,) Opus is a penguin.
General	(I see that) birds fly.	(Normally,) birds fly.

In the development of our reasoning system, both knowledge representation and inference issues are concerned. For the purposes of our work, we must first consider the pragmatic treatment of qualified knowledge. The backbone of the reasoning system is a set of deductively firm knowledge represented as a first order theory. In addition, our system will accept *de dicto* knowledge as a set of ground atomic sentences that *extend the context* of the theory with *prima facie* evidence about certain individuals (as can be seen in the table above, accepting *de dicto* general knowledge can be a quite outrageous jump [3]). Certain prototypical properties that have been recurrently observed in practice can be represented as defeasible rules. This *de re* knowledge can be regarded as abstractions of a set of states of affairs that lead to useful but unsound generalizations. These generalizations assume the form of *prima facie* material implications or inference rules that *expand the inference capabilities* of the theory (holding *de re* particular beliefs is nonsense, as can be seen also from the table above). Now it is plain to see that *default rules* of standard nonmonotonic reasoning correspond to *de re* knowledge, and that uncertain evidence or *plausible knowledge* corresponds to *de dicto* knowledge. Defeasible reasoning with default rules, then, can be regarded as a logical system that incorporates representation and reasoning with *de re* knowledge.

3 A Reasoning System

We will now describe a reasoning system for defeasible reasoning under uncertain evidence. The system first considers a set of (possibly contradictory) evidence from information sources to extend the reasoning context, and then triggers a defeasible reasoning process to expand the set of conclusions. The system can be considered an improvement upon the proposals of Rescher [16] and Roos [17], since it handles also deductive knowledge of the context, and defeasible knowledge represented as default rules. Knowledge $\mathcal{K}$ is represented in a deductively closed first order theory. A set of uncertain evidence is acquired from a finite set of information sources $\mathcal{I}$. A partial plausibility relation is established among $\mathcal{I}$.

DEFINITION **1** *An* **Information Structure** *is a pair* $\langle \mathcal{I}, \prec_{\mathcal{P}} \rangle$*, where* $\mathcal{I}$ *is a set of* **Information Sources** $\mathcal{I} = \{I_1, I_2, \cdots, I_k\}$ *and* $\prec_{\mathcal{P}}$ *is a partial order in* $\mathcal{I}$ *named* **Plausibility Relation**. $\mathcal{I}$ *contains an element* $I_\top$ *such that* $\forall I_i \in \mathcal{I}. I_i \prec_{\mathcal{P}} I_\top$*, and an element* $I_\bot$ *such that* $\forall I_i \in \mathcal{I}. I_\bot \prec_{\mathcal{P}} I_i$*. Every* I_i *provides a finite set of ground literals* l*, plausible at level* I_i*. If* l_i *is such a literal, we call the pair* $\langle l_i, I_i \rangle$ *an* **Evidence Item**. *The* **Evidence Set** $\mathcal{E}$ *is the union of all evidence items.* □

The restriction of l_i to be ground literals is rooted with the epistemological considerations discussed above, a distinction not made by Rescher and Roos. To make the formalization concise, we will overload the relation $\prec_{\mathcal{P}}$, relating sets of evidence under

plausibility, since it is usual that a reasoning line is rooted in several ground literals. Roughly, the plausibility of a set S of evidence items is the subset of lower bounds of S under plausibility, *i.e.*, the "weakest" items where an attack on S can be spotted.

DEFINITION **2** *Given a set $S \subseteq \mathcal{E}$, the* **Plausibility** *of S, denoted as $\mathcal{P}_S$ is the set $\mathcal{P}_S = \{I_i | \exists \langle l_i, I_i \rangle \in S \ \& \ \nexists \langle l_j, I_j \rangle \in S . I_j \prec_{\mathcal{P}} I_i\}$. Given two evidence sets $\mathcal{E}_1$ and $\mathcal{E}_2$, we say that $\mathcal{E}_1$ is more plausible than $\mathcal{E}_2$ (denoted as $\mathcal{E}_2 \prec_{\mathcal{P}} \mathcal{E}_1$) if and only if every evidence item in $\mathcal{E}_1$ is at least as plausible as every evidence item in $\mathcal{E}_2$, and there exists at least one evidence item in $\mathcal{E}_1$ that is strictly more plausible than every evidence item in $\mathcal{E}_2$.* □

Most of the times $(\mathcal{K} \cup \mathcal{E})$ contains contradictions[2]. Then, we must find a set $\mathcal{K}_{\mathcal{E}} \subseteq \mathcal{E}$ of *accepted evidence* such that $(\mathcal{K} \cup \mathcal{K}_{\mathcal{E}})$ is free from contradictions, and that $\mathcal{K}_{\mathcal{E}}$ is a maximally plausible subset of $\mathcal{E}$ with respect to $\mathcal{K}$. This can be characterized as the intersection of all the maximally plausible consistent subsets (MPCS) of $\mathcal{E}$ (with respect to $\mathcal{K}$), which are based on the linear extensions of $\prec_{\mathcal{P}}$.

DEFINITION **3** *Given a linear extension e of $\prec_{\mathcal{P}}$, a* **Maximally Plausible Consistent Subset** *(MPCS) of $\mathcal{E}$ (with respect to $\mathcal{K}$) is a set $\mathcal{E}^e$ such that*

- $\mathcal{E}^e \subseteq \mathcal{E}$ *($\mathcal{E}^e$ is a subset of $\mathcal{E}$),*
- $(\mathcal{E}^e \cup \mathcal{K}) \nvdash \bot$ *($\mathcal{E}^e$ is consistent with $\mathcal{K}$),*
- $\forall \langle l_i, I_i \rangle \in \mathcal{E}^e . \forall \langle l_j, I_j \rangle \in (\mathcal{E}/\mathcal{E}^e) . I_i \not\prec_{\mathcal{P}} I_j$ *(no item in $\mathcal{E}^e$ is less plausible than every item in $\mathcal{E}/\mathcal{E}^e$),*
- $\nexists \mathcal{E}' . \mathcal{E}^e \subset \mathcal{E}' \subseteq \mathcal{E}, (\mathcal{E}' \cup \mathcal{K}) \nvdash \bot$ *($\mathcal{E}^e$ is maximal),*

where $\vdash$ is the classical consequence relation. The set of **Accepted Evidence** *$\mathcal{K}_{\mathcal{E}}$ is the intersection of the MPCS's induced under every linear extension of $\prec_{\mathcal{P}}$. Finally, the set of* **conclusions** *is the deductive closure of the accepted evidence (regarding each evidence item as a plain literal) together with knowledge $\mathcal{K}$. Abusing on language we denote the set of conclusions as $Th(\mathcal{K} \cup \mathcal{K}_{\mathcal{E}})$.* □

A proof procedure for this reasoning system that is sound and complete with respect to the previous definitions can been given (space considerations do not allow to do this, the readers can consult [4]). This proof procedure leads naturally to a computational implementation.

Example [Cascaded Ambiguities [8]]

Our knowledge about political attitudes can be sumarized as follows:

Republicans are not pacifists	$r(X) \succ\!\!\!- \neg p(X)$
Quakers are pacifists	$q(X) \succ\!\!\!- p(X)$
Republicans are football fans	$r(X) \succ\!\!\!- ff(X)$
Football fans are belicists	$ff(X) \succ\!\!\!- b(X)$
Pacifists are not belicists	$p(X) \succ\!\!\!- \neg b(X)$
Nixon is a Republican	$\langle I_p, r(nixon) \rangle$
Nixon is a Quaker	$\langle I_q, q(nixon) \rangle$

Now what can we conclude about Nixon's belicism? We may consider two extensions $E^1 = \{ff(nixon), b(nixon)\}$ and $E^2 = \{p(nixon), \neg b(nixon)\}$, which are mutually incompatible. If we are in disposition to assign a better plausibility to any of her infor-

[2]Here and in what follows we will abuse on language. A proper notation should be "$\mathcal{K} \cup \{l_i | \langle l_i, I_i \rangle \in \mathcal{E}\}$ contains contradictions" but, whenever it is clear from context, we will refer implicitly to the literals in an evidence set with the same symbol as the evidence set itself.

mants I_p and I_q, then there will be only one preferred extension.

In this example, some reasoning systems arrive only to the first conclusion (*i.e.*, that Nixon is belicist) because the ambiguity in $p(nixon)$ "blocks" the second extension. In our system this is never the case. If we trust more in any of our informants that in the other, then our decision about $p(nixon)$ reduces to Nixon's Diamond, (as was solved above), and our decision about $b(nixon)$ can be inferred in consequence.

Now that the system has expanded the context with contradiction-free evidence, we will briefly state the general towards embedding defeasible reasoning in the system. Consider that our system believes that normally birds fly. But it also receives the report from a not quite trustable source that *opus* is a bird but it does not fly. Should the system accept that report? If so, what can we conclude? This is a simple situation of the more complex cases that can arise in default reasoning with uncertain evidence. In our system we consider that plausible information should be given precedence with respect to defeasibile conclusions. This criterion is similar to the adopted by Loui [11] in his *defeat among arguments*, and in Prakken [15] in the context of modeling issues of legal reasoning. In Loui, when combining kinds of defeaters, the use of more evidence is the most important defeater. Following Prakken, legal criteria assign precedence to *Lex Superior* with respect to *Lex Specialis*.

As we can see, plausibility ordering can be seen both as an evidence importance criterion (since using more plausible evidence is in a sense to use "better" evidence), and as a superiority criterion (since using more plausible evidence is using knowledge from a superior stance). Then, the general form of default reasoning under plausible grounds is that a plausibility analysis comes first and the default reasoning (*i.e.*, reasoning with default rules) comes next. Put simply, our first considers the reports from the information sources to find the set $\mathcal{K}_\mathcal{E}$ of accepted information. Then it reasons defeasibly with the set of default rules, on the enlarged (but contradiction-free) context ($\mathcal{K} \cup \mathcal{K}_\mathcal{E}$). If multiple extensions arise, then a preference can be established among extensions, based on the plausibility of the subset of $\mathcal{K}_\mathcal{E}$ used to generate each extension.

4 Scientific Reasoning Models

Scientific theories are intended to establish systematic connections between phenomena of a given domain, in a way such that inference of new facts from observations may be possible. Our view of theory formation is a *design process*, where the purpose is to systematize a given domain. This systematization can be stated as a *covering* of a relevant subset of the observations, *i.e.*, a good theory is one that covers most important cases with little (if any) *ad hoc* procedure. We can distinguish at least three different levels or strata of statements within a scientific theory. The first level, $\mathcal{N}_1$ is the set of particular sentences that represent the different states of affairs that can arise in a given domain. Normally, statements in this level are ground literals. The second level $\mathcal{N}_2$ considers the empirical or accidental generalizations [14, 6]. Knowledge in this level tends to represent in a regular manner the classifications and correlation that have been observed over sets of statements of the previous level. A statement in this level assumes the form of a lawlike (universal, existential, probabilistic) statement referred to objects and properties of $\mathcal{N}_1$ (*i.e.*, observables). The third level $\mathcal{N}_3$ covers the theoretical statements, *i.e.*, non observable entities. These statements are also called (internal or first)

principles, and assume the form of universally quantified sentences,

One of the earliest models of scientific reasoning was Hempel's hypothetic-deductive (H-D) paradigm [7]. Hempel proceeded with the schema $L \vdash e$, where L –the *explanans*– is a set $L \in \mathcal{N}_2$ of general laws, and $e \in \mathcal{N}_1$ –the *explanandum*– is the fact to be explained. The nature of L is always tentative. This means that scientific theories cannot be verified but can only be rebated. There is no possible evidence set that renders true a given theory, but *"a single ugly fact can render false an otherwise beautiful theory"* [14]. This shows a pragmatic inadequacy of the H-D paradigm. In practice, scientists certainly do not abandon a fruitful theory when it is confronted with a single refutation. Moreover, as history shows, any given theory that produces positive results will not be completely abandoned.

This fact, observed by Lakatos [10], was the inspiration for his rational reconstruction of the dynamic of scientific theories. We can regard a research programme as a structured set of knowledge that includes a knowledge set that is considered to be the *kernel* of the programme. It includes a set of lawlike statements, generalizations and postulates, that theoretically shapes the programme itself. This kernel is therefore definitive, being the remainder of the knowledge structure a mechanism to protect it from refutation. This protection operates by means of a "protective belt" of ancillary hypotheses that protects the kernel from refutation. There are at least two heuristic procedures to confront a theory $\mathcal{T}$ with a given experimental result e. If e is adequately predicted (or explained), then the programme has a positive result, and then we can apply the positive heuristic, *i.e.*, try to discharge earlier auxiliary hypotheses rendering them as consequences of the kernel. If e is not adequately predicted or explained by the theory, then we can apply the negative heuristic and find an auxiliary hypothesis c that is particular to the case e such that $\mathcal{T}$ together with c fails to entail e. If that hypothesis c is incompatible with $\mathcal{T}$, something in $\mathcal{T}$ must be given up to accommodate c. It is remarkable how this view of scientific reasoning is in fact a design process, in which the drive is towards maximization of knowledge.

Criteria for theory comparison, within the behavior of a programme, are based on the epistemic importance relation. Within a given discipline, theories share a common ground of which their epistemic structures are subsets. The following example shows some different alternatives that can arise when confronted with a negative result.

Example
Let a theory $\mathcal{T}$ be $\mathcal{T} = \langle \{a, a \succ\!\!\!- b\}, \{\} \rangle$. This theory predicts b. If b is not experimentally observed, *i.e.*, if there is certain evidence that $\neg b$, then at least three new theories can be constructed from $\mathcal{T}$:

1) The new theory is $\mathcal{T}_1 = \langle \{a, \neg b, a \succ\!\!\!- b\}, \{a \succ \neg b, a \succ a \succ\!\!\!- b\} \rangle$.
Following $\mathcal{T}_1$, the prediction fails because a is not adequately justified, but $a \succ\!\!\!- b$ can be safely maintained. Moreover, this state of affairs suggests that presupposition that $\neg a$, which must be corroborated.

2) Here we have $\mathcal{T}_2 = \langle \{a, \neg b, a \succ\!\!\!- b\}, \{a \succ \neg b, a \succ\!\!\!- b \succ a\} \rangle$.
In $\mathcal{T}_2$, the culprit is the lawlike statement $a \succ\!\!\!- b$, which is rendered false from evidence a and $\neg b$ that can be safely maintained.

3) Other cases may exist where an auxiliary hypothesis c is proposed to protect the lawlike statement from falsation. In this cases the new theory is $\mathcal{T}_3 =< \{a, c, \neg b, a \succ\!\!\!- b, a \wedge c \succ\!\!\!- \neg b\}, \{\} >$.

Following $\mathcal{T}_3$, the statement $a \succ\!\!\!- b$ systematizes only a subset of the domain, but a more specific law $a \wedge c \succ\!\!\!- \neg b$ must exist in a way such that completes the systematization in particular situations in which c is observed.

References

[1] Craig Boutilier and Verónica Becher. Abduction as Belief Revision. *Artificial Intelligence*, 77(1):43–94, 1995.

[2] Roderick M. Chisholm. *Theory of Knowledge*. Prentice Hall, Englewood Cliffs, New Jersey, 1977.

[3] Francis Watanabe Dauer. *Critical Thinking*. Addison-Wesley, London, 1995.

[4] Claudio Delrieux. Nonmonotonic Reasoning Under Uncertain Evidence. In Fausto Giunchiglia, editor, *Artificial Intelligence: Methodology, Systems and Applications*, pages 195–204. Springer, Lecture Notes in Artificial Intelligence 1480, 1998.

[5] Matthew L. Ginsberg (ed.). *Readings in Nonmonotonic Reasoning*. Morgan Kaufmann Publishers, Los Altos, California, 1987.

[6] Carl G. Hempel. *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. The Free Press, New York, 1965.

[7] Carl G. Hempel and Paul Oppenheim. The Logic of Explanation. *Philosophy of Science*, 15:135–175, 1948.

[8] John F. Horty, Richmond H. Thomason, and David S. Touretzky. A Skeptical Theory of Inheritance in Nonmonotonic Semantic Networks. *Artificial Intelligence*, 43(1-3):311–348, 1990.

[9] Kurt Konolige. Abduction versus Closure in Causal Theories. *Artificial Intelligence*, 53(2-3):255–272, 1992.

[10] Imre Lakatos. *Proofs and Refutations. The Logic of Mathematical Discovery*. Cambridge University Press, 1976.

[11] Ronald P. Loui. Defeat Among Arguments: A System of Defeasible Inference. *Computational Intelligence*, 3(3), 1987.

[12] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann, San Mateo, California, 1988.

[13] David Poole. A Methodology for Using a Default and Abductive Reasoning System. Technical Report DCS-UW, University of Waterloo, 1988.

[14] Karl Popper. *The Logic of Scientific Discovery*. Hutchinson, London, 1959.

[15] Hendrik Prakken. *Logical Tools for Modeling Legal Argument*. PhD thesis, Vrije Universiteit, 1993.

[16] Nicholas Rescher. *Plausible Reasoning*. Van Gorcum, Dodrecht, 1976.

[17] Nico Roos. A Logic for Reasoning with Inconsistent Knowledge. *Artificial Intelligence*, 57(1):69–104, 1992.

[18] Johan van Benthem. *Intensional Logics*. CSLI/SRI International, Stanford, second edition, 1988.