Defeasible Reasoning + Partial Models: a Formal Framework for the Methodology of Research Programs

Fernando Tohmé
CONICET
Departamento de Economía — Universidad Nacional del Sur — 8000 Bahía Blanca - Argentina — email: ftohme@criba.edu.ar
Claudio Delrieux
Departamento de Ingeniería Eléctrica y de Computadoras — Universidad Nacional del Sur — 8000 Bahía Blanca - Argentina — email: usdelrie@criba.edu.ar
Otávio Bueno
Department of Philosophy — University of South Carolina — Columbia, SC 29208 - USA — email: obueno@sc.edu

1 Introduction

Scientific reasoning may use any type of inference in which the derived consequences are less than conclusive, but it can be clearly distinguished from induction and other kinds of reasoning by its explicit emphasis in corrigibility. That is, while other types of reasoning also involve fallible conclusions, the key here is in the use of procedures to correct the conclusions once it becomes clear that they are incorrect. In this sense, we present in this paper a broad framework of scientific reasoning that incorporates corrigibility at the same time that it retains many desirable properties from classical logic. In fact, one particular reason for keeping the classical forms of inferences is that they are useful to compute consequences, even though, from an epistemic point of view, they may be revoked.

The framework to be presented in this paper may serve the purpose of providing a blueprint for computational implementations of reasoning, which have a wide range of potential applications, but the prime example can be found in scientific research. Scientific theories are not static entities made-up once and for all. On the contrary, they evolve in time, according to their ability to generate new problems and to yield confirmable predictions. In particular, the improved theories must accommodate either the presence of new phenomena or the failure of predictions made by the former theories [Hempel 1965]. A crucial issue for any account of science is to develop a clear and articulated picture of how theories are reformed in presence of new information. In this sense Lakatos' methodology of *scientific research programs* stands as a remarkable example of this [Lakatos 1978]. Since in many ways its underlying pattern of reasoning departs from the current practice in logic (by allowing inconsistencies to play a role in theory development) and given its emphasis on heuristics, a possible formalization may be based on the methods of knowledge representation and reasoning (KR&R) of Artificial Intelligence (AI), in particular on the methods of defeasible reasoning [Poole 1985], [Loui 1986], [Simari-Loui 1992], [Vreeswijk 1997], [Dung 1995]. In fact, most if not all the goals sought in KR&R parallel well-known objectives pursued in scientific inquiries [Stalnaker 1984]. In particular, it seems worth to combine the accuracy of the AI methods with Lakatos' methodology, which has shown to be rigorous but flexible enough to adapt itself to very different contexts. Due to its emphasis on procedures to guide the evolution of programs–the positive and negative heuristics– this methodology is just an instance of pragmatic reasoning. Therefore, it seems a particularly adequate testbed for the framework of pragmatic reasoning presented in this paper.

The formalization sought here has to represent the features of ampliative reasoning used in the process of changing (or adapting) scientific theories. One of these features is the ability to draw conclusions even if it is known that the body of knowledge is not certain. The epistemic assumption of the provisionality of the conclusions and the disposition to revoke them at the incorporation of new information demand a formalism that overcomes an important limitation of the classical approach, namely, its inability to change a conclusion even if new relevant information is incorporated to a body of knowledge. This desideratum has been accomplished by the syntactic methods of *defeasible reasoning*. They are called *syntactic* because they derive consequences just from the expressions in the formal language of the knowledge base. This characterization makes these syntactic methods amenable for computational implementations with sound formal foundations. But defeasible reasoning systems (unlike many variants of non-monotonic logics) lack model-theoretic counterparts. Therefore, the usual metalogical properties of completeness or soundness cannot be assessed. This aspect of computational defeasible reasoning systems constitutes a drawback for the formalization of a methodology for reasoning. In fact, even if one takes a non-realist stance towards science, it is clear that statements refer to something, be it the "real world" or a class of both actual and potential observations and experimental results.

However, syntactic defeasible reasoning has been shown to be useful for providing at least a linguistic characterization of how scientific theories evolve. It seems to us that it should be kept as a notational and computational tool in the analysis of the dynamics of scientific theories. But its lack of a clear semantics affects its usefulness as a foundation for an integrated framework of scientific reasoning. Therefore, one of the main goals of this paper is to show how to complement the use of defeasible reasoning with model-theoretic counterparts.

The model theoretic approach of *partial structures* and *quasi-truth* seems appropriate in this context. A relatively recent development, the conception of partial structures is based on the idea that some sentences may not have an exact truth value due to a lack of information [Mikenberg-da Costa-Chuaqui 1986], [Bueno-de Souza 1996], [da Costa, Bueno, French 1998]. Instead of remaining silent about their validity this approach assumes that these sentences can be interpreted in some structures that may deemed *partial*, in the sense that further information can be added to them. It is said that a sentence is quasi-true if there exists a possible extension of current partial structures in which the sentence in question is in fact true. This is particularly appropriate for the representation of scientific change since sometimes new evidence reduces the number or size of the extensions in which the original claims are true.

As is apparent from the previous discussion, defeasible reasoning and the model-theoretic approach of partial structures seem to address similar issues from different angles. We argue, in fact, that the conception of partial structures provides a foundation for the semantics for defeasible reasoning. Moreover, the computational procedure of defeasible reasoning seems to provide a blueprint for the implementation of reasoning procedures based on partial structures.

This combination of approaches to reasoning with less than certain information becomes clearly fruitful in the arena of scientific change, which constitutes the prime example of pragmatic reasoning. We intend to show how both defeasible reasoning and partial structures may be applied to represent Lakatos' methodology. We claim that the status of a research program after the incorporation of new evidence depends on its ability to build undefeated arguments for the new information. If the new evidence implies either an anomaly or an indetermination in the program, its models have to be modified to keep the statements in the *hard core* of the program.

In outline, this is the structure of the rest of the paper. In section 2 we will present a characterization of defeasible reasoning. In section 3 we describe the main features of the partial structures approach. In section 4 we will develop an enhanced version of defeasible/pragmatic reasoning with a partial structures-based model theory. Finally, in sections 5 and 6 we will introduce a simplified version of Lakatos' methodology, formalize it by means of the enhanced version of defeasible/pragmatic reasoning and apply the resulting framework to assess research programs in cosmology.

2 Defeasible Reasoning

Defeasible reasoning is concerned with finding *warrant* for sentences. What distinguishes this type of reasoning from classical logic is that it is fundamentally procedural: the warrant of sentences is defined in terms of the procedure followed to support them [Poole 1985], [Loui 1991], [Simari-Loui 1992], [Vreeswijk 1997]. The procedure of justification is called *argumentation.* The idea is that arguments or *defeasible proofs* are derived for and against a sentence. Arguments are very similar to classical deductions. In fact, the syntactic apparatus of classical logic is applied to the construction of arguments. Arguments can be partially ordered in terms of their "conclusive strength". If two arguments $\mathcal{H}_1$ and $\mathcal{H}_2$ support A and $\neg A$, respectively, and $\mathcal{H}_1 \lhd \mathcal{H}_2$ (where $\lhd$ represents the relation *less strong*) then we say that $\neg A$ is justified.[1]Notice that the notion of justification here is sensitive to the arguments found for and against sentences. That is, if there exist another pair of arguments $\mathcal{H}_1^{'}$, $\mathcal{H}_2^{'}$ such that $\mathcal{H}_1^{'}$ also supports A while $\mathcal{H}_2^{'}$ supports $\neg A$, but $\mathcal{H}_2^{'} \lhd \mathcal{H}_1^{'}$, then A is justified as well. This shows that more than justifications are needed in order to warrant a sentence.

To make this discussion more precise we will present a system of defeasible reasoning which follows closely, albeit with some significant differences, the presentation in [Simari-Loui 1992] and [García-Simari 2004]. Let us consider a finite set $\mathcal{L}$ of sentences in a first order language. $\mathcal{L}$ is constructed applying a finite set of rules of inference on a finite knowledge base $\mathcal{K}$ of sentences. Given a set of sentences $\Lambda \subseteq \mathcal{L}$, we assume two types of rules:

– *strict rules of inference* $\Lambda \Rightarrow A$, where $A \in \mathcal{L}$. It indicates that a sentence A is added to $\mathcal{L}$, without the possibility of being withdrawn due the incorporation of new information.
– *defeasible rules of inference* $\Lambda \succ\!\!- A$. Here $\succ\!\!-$ represents a less-than-conclusive consequence relation, according to which a sentence A is incorporated to $\mathcal{L}$, but it is known that it may be revoked by the incorporation of new information.

Strict rules represent strongly accepted connections among sentences while defeasible rules indicate that the corresponding connections are provisionally accepted but admittedly with exceptions, which may force the retraction of their conclusions. Each sentence in $\mathcal{L}$ is supported by an *argument* based on the sentences in $\mathcal{K}$ and obtains by means of a recursive application of rules of inference:[2]

Definition 1 *The pair $\langle \mathcal{H}, A \rangle$ upon the knowledge base $\mathcal{K}$ is called an argument for A, where:*

– *$A \in \mathcal{L}$.*
– *$\mathcal{H}$ is a finite tree, with nodes in $\mathcal{L}$. The root node is A and each node B has a set of children Λ iff it exists a rule $\Lambda \Rightarrow B$ or a rule $\Lambda \succ\!\!- B$. The leaves of $\mathcal{H}$ are sentences in $\mathcal{K}$.*
– *For at least one node B in $\mathcal{H}$, and its children Λ, there exists a defeasible rule of inference $\Lambda \succ\!\!- B$.*

The notion of strength of an argument can have several characterizations. We will consider here the following, which is intended to capture the notion of *specificity*:

Definition 2 *Given two arguments $\langle \mathcal{H}, A \rangle$ and $\langle \mathcal{H}^{'}, A^{'} \rangle$, we say that the first one has less probative strength ($\langle \mathcal{H}, A \rangle \lhd \langle \mathcal{H}^{'}, A^{'} \rangle$) iff*

[1] This relation can be easily extended to cover the case in which the supported sentences are not contradictory but their conjunction may yield an inconsistency.

[2] Notice that $\mathcal{L}$ is far smaller than the deductive closure of $\mathcal{K}$ under the classical consequence relation $\vdash$. Even so, the complexity of the whole process of argumentation over a first-order language is beyond tractability. However, as it will be seen in the examples, actual applications will be carried out in the domain of Horn clauses. The general properties characterized in this paper are general enough to cover both tractable and non-tractable cases.

- *there exists a node* A^1 *in* $\mathcal{H}$ *such that there exists a subset of nodes* Λ^1 *of the tree* $\mathcal{H}^{'}$ *and a strict rule* $\Lambda^1 \Rightarrow A^1$*,*
- *there exists a node* A^2 *in* $\mathcal{H}^{'}$ *such that there is no subset of nodes* Λ^2 *of the tree* $\mathcal{H}$ *and a strict rule* $\Lambda^2 \Rightarrow A^2$*.*

This relation allows to define the stronger relationship of *defeat*:[3]

Definition 3 *Given* $\langle \mathcal{H}, A\rangle$ *and* $\langle \mathcal{H}^{'}, A^{'}\rangle$*, we say that* $\langle \mathcal{H}^{'}, A^{'}\rangle$ *defeats* $\langle \mathcal{H}, A\rangle$ *(symbolized* $\langle \mathcal{H}, A\rangle \ll \langle \mathcal{H}^{'}, A^{'}\rangle$*) if:*

- *there exists a subtree* $\mathcal{H}^1$ *of* $\mathcal{H}$*, with root* $\neg A^{'}$ *and* $\langle \mathcal{H}^1, \neg A^{'}\rangle \lhd \langle \mathcal{H}^{'}, A^{'}\rangle$
- *it does not exist a subtree* $\mathcal{H}^2$ *of* $\mathcal{H}^{'}$*, with root* $\neg A$ *and* $\langle \mathcal{H}^2, \neg A\rangle \lhd \langle \mathcal{H}, A\rangle$

The following property is straightforward:

Proposition 1 $\ll$ *is an* asymmetric *relationship.*

Proof: *Immediate, from the definition of* $\ll$*: if* $\langle \mathcal{H}, A\rangle \ll \langle \mathcal{H}^{'}, A^{'}\rangle$ *it does not exist a subtree* $\mathcal{H}^1$ *of* $\mathcal{H}^{'}$*, with root* $\neg A$ *such that* $\langle \mathcal{H}^1, \neg A\rangle \lhd \langle \mathcal{H}, A\rangle$*. Therefore* $\langle \mathcal{H}^{'}, A^{'}\rangle \not\ll \langle \mathcal{H}, A\rangle$*.* □

Argumentation is the process by which arguments are generated for and against sentences. The goal is to decide if an initial sentence is warranted or not. A search procedure defines the notion of *warranted sentence*. Given the set of arguments on $\mathcal{L}$, denoted $\mathcal{ARG}$, this procedure generates alternatively arguments supporting a sentence and arguments against it (or against sentences in its support):

Procedure 1 *To decide if a sentence A is warranted or not, an inductive procedure must be applied. Let us define its base case:*

- $\mathbf{S}upp^0(A) = \{\langle \mathcal{H}, A\rangle : \langle \mathcal{H}, A\rangle \in \mathcal{ARG}\}$.
- $\mathbf{A}nt^0(A) = \{\langle \mathcal{H}^{'}, A^{'}\rangle : \langle \mathcal{H}^{'}, A^{'}\rangle \in \mathcal{ARG} \setminus \mathbf{S}upp^0(A)$ *and there exists a* $\langle \mathcal{H}, A\rangle \in \mathbf{S}upp^0(A)$ *such that* $\langle \mathcal{H}, A\rangle \ll \langle \mathcal{H}^{'}, A^{'}\rangle\}$.

Then, for $k > 0$

- $\mathbf{S}upp^k(A) = \{\langle \mathcal{H}^k, A^k\rangle : \langle \mathcal{H}^k, A^k\rangle \in \mathcal{ARG} \setminus \cup_{i=0}^{i=k-1}(\mathbf{S}upp^i(A) \cup \mathbf{A}nt^i(A))$ *and there exists a* $\langle \mathcal{H}^{k-1}, A^{k-1}\rangle \in \mathbf{A}nt^{k-1}(A)$ *such that* $\langle \mathcal{H}^{k-1}, A^{k-1}\rangle \ll \langle \mathcal{H}^k, A^k\rangle\}$.
- $\mathbf{A}nt^k(A) = \{\langle \mathcal{H}^{k^{'}}, A^{k^{'}}\rangle : \langle \mathcal{H}^{k^{'}}, A^{k^{'}}\rangle \in \mathcal{ARG} \setminus (\cup_{i=0}^{i=k}\mathbf{S}upp^i(A) \cup \cup_{i=0}^{i=k-1}\mathbf{A}nt^i(A))$ *and there exists a* $\langle \mathcal{H}^k, A^k\rangle \in \mathbf{S}upp^k(A)$ *such that* $\langle \mathcal{H}^k, A^k\rangle \ll \langle \mathcal{H}^{k^{'}}, A^{k^{'}}\rangle\}$.

Let $N = \min\{k : \mathbf{S}upp^k(A) = \emptyset$ *or* $\mathbf{A}nt^k(A) = \emptyset\}$*. Two cases are possible:*

- *if* $\mathbf{S}upp^N(A) = \emptyset$ *then A is said* non-warranted*. Else,*
- *if* $\mathbf{A}nt^N(A) = \emptyset$*, A is said* warranted.

One of the properties of arguments generated in the search is that there are no cycles. That means that if an argument $\langle \mathcal{H}^m, A^m\rangle$ is in either $\mathbf{S}upp^m(A)$ or $\mathbf{A}nt^m(A)$ it cannot reappear in $\mathbf{S}upp^k(A)$ or $\mathbf{A}nt^k(A)$, for $k > m$. This requirement, together with the fact that $\mathcal{ARG}$ is finite,[4] ensures the following result:

Lemma 1 *The search procedure applied to a sentence* $A \in \mathcal{L}$ *terminates in finite time, indicating that A is warranted or not warranted.*

[3] The characterization of $\ll$ given here in terms of specificity is just one possible instance of the notion of defeat: one argument may defeat another for formal reasons (like specificity) or because of some other reason like having a better evidential support, more plausibility, etc.

[4] Because $\mathcal{L}$ and the set of rules of inference are both finite.

Proof: *Suppose that the procedure does not terminate. Then, for each argument in* $\mathbf{S}upp^0(A)$, $\langle \mathcal{H}^0, A\rangle$, *a sequence* $\{\langle \mathcal{H}^k, A^k\rangle\}_{k=0}^{\infty}$ *can be generated, where* $\langle \mathcal{H}^k, A^k\rangle \ll \langle \mathcal{H}^{k+1}, A^{k+1}\rangle$. *As cyclicity is not allowed, there does not exist a repeated argument in the sequence. Absurd, because* $\mathcal{ARG}$ *is finite.*

Therefore, there exists an argument $\langle \mathcal{H}^l, A^l\rangle$, *such that there is no* $\langle \mathcal{H}^{l+1}, A^{l+1}\rangle$ *verifying that* $\langle \mathcal{H}^l, A^l\rangle \ll \langle \mathcal{H}^{l+1}, A^{l+1}\rangle$. *If l is an even number,* $\langle \mathcal{H}^l, A^l\rangle$ *is in* $\mathbf{S}upp^N(A)$ *while* $\mathbf{A}nt^N(A) = \emptyset$ *and it follows that A is warranted. If l is odd, it follows that A is not warranted.*□

A well-known example shows how the search procedure can be applied:

Example 1 Consider the following sentences:

- *Birds fly.*
- *Penguins do not fly.*
- *Penguins are Birds.*

We represent these conditional expressions by means of the following *rules of inference* (where 'a' represents any of the finite number of *constants* in the language):

- $B(a) \succ\!\!- F(a)$
- $P(a) \succ\!\!- \neg F(a)$
- $P(a) \Rightarrow B(a)$

Suppose furthermore that we are informed that Opus is a penguin, i.e., we have that $P(Opus)$.

We want to see if $F(Opus)$ is a warranted sentence. We can generate an argument for it:

- $arg_0 = \langle\{P(Opus) \Rightarrow B(Opus), B(Opus) \succ\!\!- F(Opus)\}, F(Opus)\rangle$

Now we can generate an argument against $F(Opus)$:

- $arg_1 = \langle\{P(Opus) \succ\!\!- \neg F(Opus)\}, \neg F(Opus)\rangle$

Notice that we just used classical logic to derive both conclusions. But in the classical framework this would mean that the information about *Opus* is inconsistent. Here, instead, we can see that $arg_0 \lhd arg_1$, because from $P(Opus)$ in arg_0 it can be derived $B(Opus)$ (via $P(Opus) \Rightarrow B(Opus)$), while from $P(Opus), B(Opus)$ no node in the tree in arg_1 can be inferred. As the conclusions of both arguments are contradictory it follows that $arg_0 \ll arg_1$.

Moreover, no argument arg_2 can be generated against arg_1, so $F(Opus)$ is not warranted.

Applying the procedure to $\neg F(Opus)$ we have that:

- $arg_0^{'} = \langle\{P(Opus) \succ\!\!- \neg F(Opus)\}, \neg F(Opus)\rangle$

and the argument against it is:

- $arg_1^{'} = \langle\{P(Opus) \Rightarrow B(Opus), B(Opus) \succ\!\!- F(Opus)\}, F(Opus)\rangle$

But it is clear that (as $arg_0^{'} = arg_1$ and $arg_1^{'} = arg_0$) $arg_0^{'} \not\ll arg_1^{'}$, and therefore $\neg F(Opus)$ is warranted.

An interesting feature of the search procedure is that it still works in the case of *bounded resources* (time, computational power, etc.). The procedure can stop at any stage and provide a *tentative* answer. This aspect of defeasible reasoning makes it appropriate for representing ampliative inference where both computation and corrigibility can force revision. Computation is used to derive classically valid arguments for and against a given conclusion, and therefore indicating the necessity of a revision. Since conclusions are assumed to be revokable according to the defeat relation among arguments, the consequent correction yields a revised conclusion.

As is clear from the presentation, defeasible reasoning does not have any semantic features. But the example showed that an interpretation helps to assess the meaning of warranted sentences. To remedy this gap we will look for a model-theoretic counterpart of defeasible reasoning. A good candidate is the semantic approach based on partial structures.

3 Partial Structures and Quasi-Truth

Scientific and commonsense reasoning frequently have a semantic nature. That is, conclusions are accepted as long as they hold in their intended domain of reference. In the philosophy of science this realization lead many authors to a model-theoretic view of science (e.g. [Suppe 1977], [da Costa - French 1990]). Pragmatic reasoning in this framework involves the epistemic decision to accept conclusions even if it is known that their interpretations in the intended domain are not completely determined. That is, conclusions may be provisionally accepted so long they hold in the fragment of the domain whose knowledge is well established. This approach is known as the *partial structures* approach [Mikenberg-da Costa-Chuaqui 1986], [da Costa- French 1989], [Bueno-de Souza 1996], [da Costa, Bueno, French 1998], [French 2000].

Given a domain of knowledge, consisting of a class of objects, $\mathcal{D}$, and the relations among them, it is frequent the case in which it is unknown whether each relation is verified by each relevant tuple of objects. It is said that such relations are *partial*:

Definition 4 *A partial relation R of arity n can be represented as $R = \langle R^1, R^2, R^3 \rangle$, such that $R^i \cap R^j = \emptyset$ for $i, j : 1, 2, 3$, $i \neq j$, and $R^1 \cup R^2 \cup R^3 = \mathcal{D}^n$, where*

- *R^1 is the set of n-tuples that we know belong to R.*
- *R^2 is the set of n-tuples that we know do not belong to R.*
- *R^3 is the set of n-tuples that are not known either to belong to R or not.*

That is, a partial relation partitions $\mathcal{D}^n$ into three sets. One is the set of elements that belong to the relation. The second set is the set of elements that are not in the relation. The third set consists of those elements that are not known to be in the relation or in its complement. It is clear that when no uncertainty exist, $R^3 = \emptyset$ and R verifies the usual definition of a relation.[5]

Structures which include partial relations are said *pragmatic*:

Definition 5 *A pragmatic structure is of the form $\langle \mathcal{D}, \{R_i\}_{i\in I}, \mathcal{K}\rangle$, where $\mathcal{D} \neq \emptyset$ is the universe of the structure, $\{R_i\}_{i\in I}$ is the family of partial relations over $\mathcal{D}$ and $\mathcal{K}$ is the set of sentences accepted in the structure, called* primary *sentences.*[6]

The set of primary sentences $\mathcal{K}$ includes all the statements that belong to a given knowledge base. They may be universal statements or just atomic sentences (i.e. they may represent either *laws* or *facts*). $\mathcal{K}$ constraints the possible extensions of a partial structure, because the derived sentences in any feasible extension should be consistent with $\mathcal{K}$. Each possible extension is called a *normal* structure:

Definition 6 *Given a pragmatic structure $\mathcal{P} = \langle \mathcal{D}, \{R_i\}_{i\in I}, \mathcal{K}\rangle$, a structure $\mathcal{P}' = \langle \mathcal{D}', \{R_i'\}_{i\in I}\rangle$ is said a $\mathcal{P}$-normal structure if*

1. *$\mathcal{D} = \mathcal{D}'$.*
2. *Every constant is interpreted as the same object in $\mathcal{P}$ and $\mathcal{P}'$.*
3. *Each R_i' extends R_i, i.e. $R_i^1 \subseteq R_i'^1$, $R_i^2 \subseteq R_i'^2$ and $R_i'^3 = \emptyset$.*

But only some of the normal structures are admissible, namely those that support sentences that are consistent with the primary sentences $\mathcal{K}$. These sentences are said *quasi-true* (or *pragmatically* true):

Definition 7 *Given a $\mathcal{P}$-normal structure $\mathcal{P}'$, we say that it is* admissible *if $\mathcal{P}' \models \alpha$ for each $\alpha \in \mathcal{K}$.*[7] *A sentence $\phi \in \mathcal{L}$ is quasi-true (quasi-false) relative to $\mathcal{P}$ according to an admissible extension $\mathcal{P}'$ if $\mathcal{P}' \models \phi$ ($\mathcal{P}' \not\models \phi$).*

---

[5] It is customary to describe such a relation only by means of R^1 since R^2 is its complement.

[6] These sentences belong to a first-order language $\mathcal{L}$ such that its interpretation in $\langle \mathcal{D}, \{R_i\}_{i\in I}, \mathcal{K}\rangle$ consist in associating each constant of $\mathcal{L}$ with an element in $\mathcal{D}$ while each predicate symbol of arity n is associated to a relation R_i of the same arity.

[7] $\models$ is the classical (or Tarskian) model-theoretic consequence relation.

4 A Partial Structures Semantics for Defeasible Reasoning

Although the defeasible reasoning system described in Section 2 is a purely syntactic procedure and does not constitute a logical system, it can be seen as the basis for one. Most authors in the field of defeasible reasoning consider that systems based in such procedures are more than enough to capture the intricacies of rational reasoning[8]. Despite that, since we intend to use defeasible reasoning as the computational representation of the evolution of scientific theories, its connections with more traditional approaches to the evolution of theories must be clarified. In this sense, a semantic foundation for defeasible reasoning would allow us to make it compatible with the model-theoretic approach to science. A semantics provides a formal account of the notational procedures used in inquiry processes, since it allows to see more clearly the structure behind the dynamics of reasoning. We are interested in providing a link between defeasible reasoning and the view of science based on partial structures, since both illuminate specific components of scientific inquiry and therefore provide complementary insights on its dynamics.

A way of characterizing the behavior of Procedure 1 over sentences is by means of the following notion of *model*:

Definition 8 *Given $\mathcal{L}$, a finite set of sentences in a first order language, a model for defeasible reasoning over $\mathcal{L}$ is:*

$$\langle W, \preceq, [|\ |], Ch \rangle$$

where W is a set of worlds, partially ordered by $\preceq$ and the interpretation correspondence $[|\cdot|] : \mathcal{L} \to W$ is such that:

- $[|A^{\mathcal{K}}|] = W$ *for each* $A^{\mathcal{K}} \in \mathcal{K}$,
- *given any two arguments* $\langle \mathcal{H}, A \rangle$ *and* $\langle \mathcal{H}', A' \rangle$, $\langle \mathcal{H}, A \rangle \ll \langle \mathcal{H}', A' \rangle$ *iff there exist* $w_1 \in [|A|], w_2 \in [|A'|]$ *verifying that* $w_1 \preceq w_2$.

Finally, $Ch(\cdot) : \mathcal{L} \to W$ is a choice function *that selects worlds. To characterize Ch, consider for any $w^* \in W$ an* upper-chain, $\mathbf{U}(w^*) = \{w \in W : w^* \preceq w\}$ *such that* $\mathbf{U}(w^*)$ *is linearly ordered by $\preceq$. Then,*

$$Ch(A) \;=\; \{w \in [|A|] : \text{ there exists } \mathbf{U}(w) \text{ and } |\mathbf{U}(w)| \text{ is odd}\}.$$

This means that given a set of worlds W, with a partial order $\preceq$, each sentence in the knowledge base $\mathcal{K}$ is supported by each world, while, if an argument for a sentence A is defeated by another for A' there exists a world w_1 which supports A and a world w_2 that supports A' such that w_1 has a lower ranking than w_2, according to $\preceq$. Moreover, the worlds chosen by a sentence A, $Ch(A)$ are those that support A and have an upper-chain of odd length.[9]

We claim that this semantics is sound and complete for Procedure 1:

Proposition 2 *A sentence A is warranted in $\mathcal{L}$ iff $Ch(A) \neq \emptyset$.*

Proof: $\Rightarrow$) **(Soundness)** *Suppose $Ch(A) = \emptyset$. Then, for all $w \in [|A|]$, every $\mathbf{U}(w)$ has an even number of worlds. Since A is warranted, there exists a sequence $\{\langle \mathcal{H}^k, A^k \rangle\}_{k=0}^N$ (where $A = A^0$) constructed according to Procedure 1 such that $\langle \mathcal{H}^N, A^N \rangle \in \mathbf{S}upp^N(A)$ while $\mathbf{A}nt^N(A) = \emptyset$. But this means that $|\{\langle \mathcal{H}^k, A^k \rangle\}_{k=0}^N|$ is odd. On the other hand, by construction we have that $\langle \mathcal{H}^k, A^k \rangle \ll \langle \mathcal{H}^{k+1}, A^{k+1} \rangle$ for $0 \leq k < N$. But then, by the definition of $[|\cdot|]$, there exists a $w_k \in [|A^k|]$ for each k, and $w_k \preceq w_{k+1}$ for $k < N$. Then, since $w_0 \in [|A|]$, we have $\mathbf{U}(w_0) = \{w_0, w_1, \ldots, w_N\}$, a linearly ordered set*

[8] See [Pollock 1991], [Loui 1998] or [Prakken-Vreeswijk 2001]. For some alternatives to our semantic approach see [van der Hoek et al. 1992].

[9] Notice that for each w, $w \in \mathbf{U}(w)$.

with an odd number of elements. Contradiction.
$\Leftarrow$) **(Completeness)** *Suppose that A is not warranted. Then, there exists a sequence $\{\langle\mathcal{H}^k, A^k\rangle\}_{k=0}^N$ such that $A = A^0$ and $\langle\mathcal{H}^k, A^k\rangle \ll \langle\mathcal{H}^{k+1}, A^{k+1}\rangle$ for $0 \leq k < N$, with $\langle\mathcal{H}^N, A^N\rangle \in \mathbf{A}nt^N(A)$ and $\mathbf{S}upp^{N+1}(A) = \emptyset$. Therefore, $|\{\langle\mathcal{H}^k, A^k\rangle\}_{k=0}^N|$ is even. But since $Ch(A) \neq \emptyset$ there exists $w \in [|A|]$ such that there exists a $\mathbf{U}(w)$ with an even number of elements. Then, if we define $w_0 = w$, given $\mathbf{U}(w_0) = \{w_0, w_1, \ldots, w_M\}$, we can just choose $\langle\mathcal{H}^k, A^k\rangle$ such that $w_k \in [|A^k|]$ for $k = 0, \ldots, M$ and $\langle\mathcal{H}^k, A^k\rangle \ll \langle\mathcal{H}^{k+1}, A^{k+1}\rangle$ for $0 \leq k < M$. Therefore $\{\langle\mathcal{H}^k, A^k\rangle\}_{k=0}^M$ has an odd number of elements and therefore $\langle\mathcal{H}^M, A^M\rangle \in \mathbf{S}upp^M(A)$ while $\mathbf{A}nt^M(A) = \emptyset$. That is, A is warranted. Contradiction.*□

This result shows that the semantics reflects the procedural aspects of defeasible reasoning. A sentence is warranted if it is possible to build an odd numbered sequence of arguments, each defeating the previous one and being defeated by the next. The last argument is on the support side and there is no one that defeats it. If a world is assigned to each argument in the sequence (or more precisely, to each A^k) the choice function selects the worlds that correspond to sentences that are warranted.

This semantics illuminates the order structure of defeasible reasoning, and shows why it is so amenable to computational treatment. On the other hand, since defeasible reasoning is seen here as part of a methodology of pragmatic reasoning–which is frequently carried out in interpreted domains–it seems convenient to relate the worlds in W and their corresponding ordering to some field of knowledge beyond $\mathcal{L}$. This field has to be the intended domain of the statements in $\mathcal{L}$.

In fact, partial structures may provide an alternative foundation for a semantics of defeasible reasoning. The basic idea is that an argument must be interpreted as an admissible extension of a pragmatic structure. In this sense, we have to provide an interpretation for the individuals, the predicates and the rules of inference involved in reasoning defeasibly over a set of sentences $\mathcal{L}$:

Definition 9 *Given a set of sentences $\mathcal{L}$, and the families of strict and defeasible rules of inference that define $\mathcal{L}$, $\{\Lambda \Rightarrow_i A\}_{j\in J_1}$ and $\{\Lambda \succ\!\!-_j A\}_{j\in J_2}$, and the knowledge base $\mathcal{K}$, a pragmatic structure for this system is $\mathcal{P} = \langle\mathcal{D}, \{R_i\}_{i\in I}, \mathcal{K}\rangle$ such that:*

- *Each of the constants in the sentences of $\mathcal{L}$ is assigned to an element in $\mathcal{D}$.*
- *Each predicate in an $A \in \mathcal{K}$ and in any A such that there exists a strict rule of inference $\Lambda \Rightarrow_j A$, with $\Lambda \subseteq \mathcal{K}$, corresponds to a relation R_t such that $R_t^3 = \emptyset$. In other words, the relation is total.*
- *Each predicate in an $A \in \mathcal{L} - \mathcal{K}$ for which there exists an argument $\langle\mathcal{H}, A\rangle \in \mathcal{ARG}$ but not a strict rule of inference $\Lambda \Rightarrow_j A$, with $\Lambda \subseteq \mathcal{K}$, corresponds to a partial predicate R_p, i.e. such that $R_p^3 \neq \emptyset$.*
- *The knowledge base, $\mathcal{K}$, becomes the set of primary sentences of $\mathcal{P}$.*

Then, we can proceed by defining interpretations for each sentence in $\mathcal{L}$:

Definition 10 *Given $\mathcal{L}$ and the pragmatic structure $\mathcal{P}$ a model for defeasible reasoning over $\mathcal{L}$ is:*

$$\langle W, \preceq, [|\ |], Ch\rangle$$

where each $w \in W$ is an admissible extension of $\mathcal{P}$, i.e. $w \equiv \mathcal{P}'$ for a normal extension $\mathcal{P}'$ of $\mathcal{P}$ such that $\mathcal{P}' \models \mathcal{K}$. Two different extensions can be compared by means of the partial order $\preceq$. The interpretation correspondence $[|\ |]$ is such that for any sentence $A \in \mathcal{L}$, $[|A|] = \{w : w \models A\}$, i.e. it yields the extensions that make A quasi-true relative to $\mathcal{P}$. Ch, the choice function*, yields, for each sentence A, all the admissible extensions of $\mathcal{P}$ such that there exists an odd-numbered upper-chain of extensions (ordered according to $\preceq$).*

Notice that according to the proof of Proposition 3, $Ch(A)$ includes all the extensions that contribute to warrant A. That is, $Ch(A) = \emptyset$ if A is not warranted, while $Ch(A) = \{w : w \models A\}$ if A is warranted by Procedure 1.

In this new semantics a world is conceived now as an admissible extension of $\mathcal{P}$. Each sentence is interpreted as the set of admissible extensions in which it is true and therefore make it quasi-true in $\mathcal{P}$. The choice function, just selects for each sentence the admissible extensions (if any exists) that support its warrant.

Example 1 (revisited): The knowledge base is $\mathcal{K} = \{P(a)\}$, the set of sentences is $\mathcal{L} = \{P(a), B(a), F(a), \neg F(a)\}$, while the strict and defeasible inferences are:

– $P(a) \Rightarrow B(a)$
– $B(a) \succ\!\!- F(a)$
– $P(a) \succ\!\!- \neg F(a)$

Then, we define a pragmatic structure $\mathcal{P}$ as follows:

– $\mathcal{D} = \{Opus\}$. *Opus* is then assigned to the single constant a.
– From $P(a) \in \mathcal{K}$ we obtain the monadic predicate $P(\cdot)$, from the strict inference (in which the antecedent is in $\mathcal{K}$) $P(a) \Rightarrow B(a)$ we are led to $B(a)$ and extract the monadic predicate $B(\cdot)$. We assign $P(\cdot)$ and $B(\cdot)$ to the properties of being a *Penguin* and a *Bird*.
– From $F(a)$ and $\neg F(a)$ for which we have arguments but not strict inferences from $\mathcal{K}$, we obtain the predicate $F(\cdot)$ to which we assign the partial property *Flies*.
– The only primary sentence held in $\mathcal{P}$ is $P(a)$.

We now define the set of worlds:

– w_1:

Individual	Bird	Flies	Penguin
Opus	1	1	1

– w_2:

Individual	Bird	Flies	Penguin
Opus	1	0	1

It is clear that w_1 and w_2 are admissible extensions of $\mathcal{P}$ since $w_1 \models P(a)$ and $w_2 \models P(a)$.

The interpretation of the sentences in $\mathcal{L}$ is as follows: $[|P(a)|] = \{w_1, w_2\}$, $[|B(a)|] = \{w_1, w_2\}$, $[|F(a)|] = \{w_1\}$ and $[|\neg F(a)|] = \{w_2\}$.

Furthermore, it follows that $w_1 \preceq w_2$, because otherwise the result of Procedure 1 would not be supported. In other words, $P(a)$ and $B(a)$ are warranted because they are derived classically from $\mathcal{K}$ while, according to Procedure 1, $\neg F(a)$ is warranted and $F(a)$ is not so. Therefore, $Ch[P(a)] = \{w_1, w_2\}$, $Ch[B(a)] = \{w_1, w_2\}$, $Ch[F(a)] = \emptyset$ and $Ch[\neg F(a)] = \{w_2\}$.

This example suggests a notion of *pragmatic validity* that arises from the interplay between the semantics of partial structures and the syntactic formalism of defeasible reasoning:

Definition 11 *A sentence* $\phi \in \mathcal{L}$ *is* pragmatically valid*,* $\mathcal{P} \models_T \phi$*, if* $Ch[\phi] \neq \emptyset$ *while* $Ch[\neg\phi] = \emptyset$*.* ϕ *is* pragmatically countervalid*,* $\mathcal{P} \models_F \phi$*, if* $Ch[\neg\phi] \neq \emptyset$ *while* $Ch[\phi] = \emptyset$*. Otherwise, if* $Ch[\phi] = \emptyset$ *and* $Ch[\neg\phi] = \emptyset$*,* ϕ *is said* pragmatically non-determined*,* $\mathcal{P} \models_u \phi$*.*

That is, pragmatic validity (countervalidity) may change as a consequence of the incorporation of a new sentence to the set of primary sentences. That means that a sentence deemed pragmatically valid in a given stage may become pragmatically non-determined or even countervalid when new information is added to the set of primary sentences. This represents well what may happen in a defeasible reasoning setting: a warranted sentence may loose its warrant or even its negation may become warranted when new information is incorporated to the knowledge base.

The integration between defeasible reasoning with partial structures constitutes our main result, the construction of a framework of pragmatic reasoning. We will see in the next section how this framework can be put at work to formalize Lakatos' methodology.

5 Formalizing the Methodology of Scientific Research Programs

Imre Lakatos presented, in the early 1970s a challenge to the methodology of Karl Popper and to Thomas Kuhn's characterization of how science actually evolves. In fact, he took from both the most significative ideas but left aside the rigidity of the former and the sociological turn of the latter [Lakatos-Musgrave 1970]. On the descriptive side, he showed that in a given field of knowledge several theories may coexist, competing among them. Each theory and the methods of inquiry associated with it constitute a "program". New programs arise in time while others dissappear, due to new discoveries and insights.

Since a scientific theory is never proven to be true nor is it completely unable to yield verifiable consequences, the corresponding program remains prone to change and evolution. This, associated to the selective pressure of the competition among theories, makes scientific inquiry a clear example of pragmatic reasoning.

Thus, a scientific research program consists of a series of theories plus a range of operational procedures and mechanisms of inference. Its *hard core* is the set of assumptions considered central for the program and can be identified with the fundamental theoretical principles guiding the program's research. The final goal of the program is, in fact, to either articulate the core (amassing new evidence confirming its claims) or to protect the core from negative evidence. In this last case the *negative heuristics* builds a *protective belt* of auxiliary hypotheses that, added to the core, yields the negative evidence as a consequence. That is, if an evidence e is not a consequence of the theory, the negative heuristic should find an hypothesis h such that from the theory plus h it follows e. The *positive heuristic* seeks, instead, to systematize the protecting belt and to make it derivable from the core by means of new laws. In fact, if this goal is achieved, what formerly constituted the protecting belt becomes part of the area of knowledge dominated and systematized by the hard core of the program.

Therefore, the size of the protective belt indicates the relative degree of success of a program. This is particularly important for the competition among programs. A theory that is continuously subject to the application of the negative heuristic is prone to disappear in favor of a less *degenerative* program. Thus, *progressive* programs tend to seize the room occupied by less successful ones.

To represent all these aspects in the framework of pragmatic reasoning developed in this paper, we will give a characterization of a program, beginning with the characterization of its language.[10]. We will distinguish four types of sentences [Delrieux 2001]. The first one involves the particular statements that describe states of affairs. Any statement of this type adopts the form of an atomic formula that is interpreted as stating that its terms (representing objects) verify its propositional function (representing properties, relations, etc.). The class of statements of this type is denoted by $\mathcal{N}_1$.

A second type of statements involves those that represent empirical generalizations. That is, it includes all the law-like[11] statements relating observational terms and relations. For the connection among classes in a statement of the form *"Some objects that have properties $p_1, \ldots, p_n$ normally have property q."* we will use the *face value* defeasible rule of inference $p_1(x), \ldots, p_n(x) \succ\!\!- q(x)$. Statements of this type form the class we denote by $\mathcal{N}_2$.

Statements in $\mathcal{N}_3$ represent the *theoretical* statements that establish the connection between theoretical and observational statements. Finally, $\mathcal{N}_4$ is the set of theoretical claims, not subject to direct observation, which state (assumed) "facts"

[10] For an alternative attempt to provide a computational reconstruction of scientific reasoning see [Thagard 1993]

[11] In many fields of inquiry it is customary to use probabilistic laws. Since this introduces a higher degree of precision than what is actually needed to describe Lakatos' methodology, we will not use them in our presentation.

about the domain of interpretation. This level includes the hard core of the research program, $\mathcal{C}_0 \subseteq \mathcal{N}_4$.

In the framework developed above, $\mathcal{N}_1$ is an instance of $\mathcal{L}$, while $\mathcal{N}_2$ is the corresponding set of *defeasible rules of inference*. The statements in $\mathcal{N}_3$ can be conceived of as strict rules of inference (i.e. of the form $p_1(x), \ldots, p_n(x) \rightarrow q(x)$). Finally, the statements in $\mathcal{N}_3$ correspond to our set of primary sentences, $\mathcal{K}$.

$E \subseteq \mathcal{N}_1$ is said the be the set of *potential pieces of evidence*. A *theory* is $\mathcal{T} \subseteq \mathcal{N}_2 \cup \mathcal{N}_3 \cup \mathcal{C}_0$. We say that for each $e \in E \subseteq \mathcal{N}_1$, $\mathcal{T}$ *quasi-supports* the conclusion e if there exists an argument $\langle \mathcal{H}, e \rangle$ in which $\mathcal{H}$ represents a tree with terminal nodes in $\mathcal{N}_4$ while the non-terminal nodes obtain applying rules of inference in $\mathcal{N}_2$ and $\mathcal{N}_3$. We have then:

Definition 12 *$\mathcal{T}$ is said to quasi-support e if $[|e|]_{\mathcal{T}\cup\mathcal{N}_4} \neq \emptyset$, where $[|e|]_{\mathcal{T}\cup\mathcal{N}_4}$ is the interpretation of e in a model of $\mathcal{T}$ with primary sentences $\mathcal{N}_4$.*

Notice that in this characterization we distinguished the theory, $\mathcal{T}$, which includes the sentences in the hard core, from the entire set of primary sentences, $\mathcal{N}_4$. This has the purpose to make the quasi-support of a piece of evidence e not exclusively dependent on $\mathcal{T}$ and allows to use new primary sentences to change the status of quasi-support of e.

Our framework of pragmatic reasoning can be applied to characterize the result of confronting the theory with the evidence. It is customary to say that if an $e \in E$ has been already observed, $\mathcal{T}$ provides an *explanation* for it, while otherwise it yields a *prediction* of e [Hempel 1965]. But Lakatos insisted that the key point is not this temporal characterization but whether the evidence was *used* or not. We will adopt this conception, so from now on we will just say of $e \in E$ that it is a piece of evidence.

If $e \in E$ is not correctly explained or predicted by $\mathcal{T}$, i.e. if $[|e|]_{\mathcal{T}\cup\mathcal{N}_4} = \emptyset$ we apply the following strategy to ensure the quasi-support of e:

Negative Heuristics: *Find $c \in \mathcal{N}_1$ such that now $\mathcal{N}_4^{'} = \mathcal{N}_4 \cup \{c\}$ and $[|e|]_{\mathcal{T}\cup\mathcal{N}_4^{'}} \neq \emptyset$. It must also be verified that $[|e^{'}|]_{\mathcal{T}\cup\mathcal{N}_4^{'}} \neq \emptyset$ for all $e^{'} \neq e$, such that $[|e^{'}|]_{\mathcal{T}\cup\mathcal{N}_4} \neq \emptyset$. The set of these* auxiliary hypotheses, *$\bar{C} \subseteq \mathcal{N}_4 - \mathcal{C}_0$, is the* protective belt *of $\mathcal{T}$.*

In words: C "protects" $\mathcal{T}$ from rejection. If instead the evidence follows from $\mathcal{T}$, the positive heuristics pushes forward the program:

Positive Heuristics: *Two procedures may be applied in parallel:*

- *Derive a new $e \in E$ from $\mathcal{T}$.*
- *Use C to find a new set of law-like statements $\bar{C} \subseteq \mathcal{N}_2$ such that the theory is extended to $\mathcal{T}^{'} = \mathcal{T} \cup \bar{C}$. This means that the pragmatic structure corresponding to $\mathcal{T}^{'}$ (with primary sentences $\mathcal{N}_4 - \bar{C}$) has to include new relations that where not present in the pragmatic structure for $\mathcal{T}$ with primary sentences $\mathcal{N}_4$. This requires that all the previously held evidence keeps quasi-supported by the new theory, i.e. if $[|e|]_{\mathcal{T}\cup\mathcal{N}_4} \neq \emptyset$ then $[|e|]_{\mathcal{T}'\cup\mathcal{N}_4 - \bar{C}} \neq \emptyset$.*

Notice that, according to the negative heuristics, the program engrosses $\bar{C}$ while the positive heuristics "discharges" $\bar{C}$ and augments the domain in which the hard core is valid. If the trend resulting from a number of confrontations with a certain number of pieces of evidence is to continuously engross $\bar{C}$, we say that the program is in a *degenerative* phase, while if those confrontations increased the domain of validity of the hard core the program is said to be in a *progressive* phase.

Summarizing all this we have that:

Definition 13 *A program is $P = \langle \{\mathcal{T}, C\}_{i\geq 0}, E \rangle$. Each $i = 0, 1, \ldots$ indicates a stage of the program resulting from its confrontation with a given piece of evidence e_i. That is, given a new piece of evidence e_i, we have that either $\mathcal{T}_{i-1}$ quasi-supports e_i or not. In the latter case $\mathcal{T}_i = \mathcal{T}_{i-1}$ while $C_i = C_{i-1} \cup \{c_i\}$, where c_i is the auxiliary hypothesis added to $\mathcal{N}_4$ in order to quasi-support e_i. If, instead, the positive heuristic is applied, we have that $\mathcal{T}_i = \mathcal{T}_{i-1} \cup \bar{C}$ and $C_i = C_{i-1} - \bar{C}$. The set of quasi-supported evidence at stage i is $E_i \subseteq E$.*

It is immediate to see that with these procedures a piece of evidence cannot change its status from quasi-supported to not supported (by the program):

Proposition 3 *For each $i \geq 0$, we have that $E_i \subseteq E_{i+1}$.*

Proof: *Trivial. Suppose that $e \in E_i$, i.e. $\mathcal{T}_i$ quasi-supports e. Given e_{i+1}, either the positive or the negative heuristic must be applied. In both cases e will be quasi-supported by $\mathcal{T}_{i+1}$. That is, $e \in E_{i+1}$.* □

This result indicates that quasi-support is preserved from a stage to another in the program. But this only means that there will always be arguments supporting the known pieces of evidence. Our framework for pragmatic reasoning allows us to go further than that, by means of the comparison of arguments. That is, given a piece of evidence e^*, we may have several cases:

1. There is an argument arg_{e^*} supporting e^* and if there is another for $\neg e^*$, $arg_{\neg e^*}$, $arg_{\neg e^*} \ll arg_{e^*}$.
2. There is an argument $arg_{\neg e^*}$ supporting $\neg e^*$ and if there is another, for e^*, arg_{e^*}, $arg_{e^*} \ll arg_{\neg e^*}$.
3. Either there aren't arguments supporting e^* or $\neg e^*$ or, if there are, say arg_{e^*} and $arg_{\neg e^*}$, $arg_{e^*} \not\ll arg_{\neg e^*}$ and $arg_{\neg e^*} \not\ll arg_{e^*}$.

Each of this cases has consequences for the status of the program:

- **Confirmation**. If e^* is observed and we are in the conditions described in Case 1, program P gets confirmed, either strongly or partially. The ability of the program to predict or explain e^* indicates that the program is in a *progressive* phase.
- **Anomaly**. If P predicts e^* and the conditions are those of Case 2, the program faces an anomaly, either serious or a partial one. In this situation, P is not able to explain or predict the actual observation, $\neg e^*$, making necessary an addition to the protective belt in order to save the program from the anomaly. This indicates that the program may be in a *degenerative* phase.
- **Indetermination**. If e^* is observed and the conditions are those of Case 3, we are either facing a *surprising fact* (if there are no arguments for or against e^*) or a "lacuna" (if there are both arguments for and against e^* and no one defeats the other).

These cases have a counterpart in the model of the reasoning process:

- **Confirmation**. There are at least two worlds, w_{e^*} and $w_{\neg e^*}$, such that the former supports an argument for e^* while the latter supports one for $\neg e^*$ and $w_{\neg e^*} \preceq_{e^*} w_{e^*}$. e^* may be pragmatically valid or just pragmatically non-determined.
- **Anomaly**. There are at least two worlds, $w_{\neg e^*}$ and w_{e^*}, such that $w_{e^*} \preceq_{\neg e^*} w_{\neg e^*}$. e^* may be pragmatically countervalid or pragmatically non-determined.
- **Indetermination**. There aren't two worlds w_{e^*} and $w_{\neg e^*}$ such that either $w_{\neg e^*} \preceq_{e^*} w_{e^*}$ or $w_{e^*} \preceq_{\neg e^*} w_{\neg e^*}$. e^* and $\neg e^*$ are both pragmatically non-determined.

The status of the program as a result of its confrontation with the evidence indicates which heuristics must be deployed. So, in the case of anomalies, the negative heuristic must be activated. If the evidence confirms P instead, the positive heuristic should be applied.

It is not clear whether the negative or the positive heuristics should be used in the cases in which the evidence shows that there exist lacunae in the program. On the other hand, if a surprising fact is found, e_i, it seems that the theory must be expanded to include new statements. The new version of the theory, $\mathcal{T}_i^{'}$, might differ from $\mathcal{T}_{i-1}$ due to the incorporation of new elements to $\mathcal{N}_2$ or $\mathcal{N}_3$, i.e. because of the incorporation of new rules of inference, defeasible or strict. It must follow that $[|e_i|]_{\mathcal{T}_i^{'} \cup \mathcal{N}_4} \neq \emptyset$.

The above account of confirmation, anomalies and indetermination is useful for the comparison among programs. Let us consider $V \subseteq \mathcal{N}_1$ the set of confirming pieces of evidence for P, $F \subseteq \mathcal{N}_1$ its set of anomalies and $I \subseteq \mathcal{N}_1$ its set of indeterminations. Then given two programs, P_1 and P_2, over the same set of evidence E, we say that P_1 is more *successful* than P_2, denoted $P_2 \leq_{success} P_1$, if and only if both $V_2 \subseteq V_1$ and $F_1 \subseteq F_2$.

This last point shows that the framework of pragmatic reasoning closes a major gap in Lakatos' methodology. The framework is highly relevant to compare programs in order to determine which one is doing better and therefore which will probably gain more adherents. But this is not the only use the framework can be put into. We will now examine an application of the framework to an important domain of knowledge: cosmology.

6 Research Programs in Cosmology

Cosmology is concerned with the study of the universe as a whole. That is, it studies the origin, structure and dynamics of the universe. This inquiry has a long history, but the observation of the *redshift* of the light from distant stars led during the 1950s and the beginnings of the 1960s to the existence of only two competing programs: the *Big-Bang* and the *Steady-State* programs.

Let us denote by P_{BB} the Big-Bang program and by P_{SS} the Steady-State one. The hard core of the former consists in the idea that the universe was created in a single event, some billions of years in the past and that it has been expanding ever since. The hard core of the Steady-State program includes the idea that the universe was never created out of nothing, but that new matter and energy are continuously created everywhere and therefore expanding. That is, if $BB(u)$ represents *"The universe was created in a Big Bang"*, $SS(u)$ represents the statement *"The universe is permanently created"* and $\mathcal{PHY}$ the entire corpus of contemporary physics[12], we have that $\mathcal{PHY} \cup \{BB(u)\} \subseteq \mathcal{T}_{BB}$ and $\mathcal{PHY} \cup \{SS(u)\} \subseteq \mathcal{T}_{SS}$.[13]

If follows that both $\mathcal{PHY} \cup BB$ and $\mathcal{PHY} \cup SS(u)$ quasi-support $e(u)$, where $e(u)$ means *"The universe expands"*. The law-like expression $e(x) \succ\!\!- rs(x)$, which means that if x expands it exhibits a redshift ($rs(x)$) can be added to both theories. That is, $\mathcal{T}_{BB} = \mathcal{PHY} \cup \{BB(u)\} \cup \{e(x) \succ\!\!- rs(x)\}$ and $\mathcal{T}_{SS} = \mathcal{PHY} \cup \{SS(u)\} \cup \{e(x) > -rs(x)\}$.

Both programs, therefore, found a confirmation in the evidence of redshift. So far, both were equally successful. But in 1965, cosmic background radiation was detected. That is, that the universe has a low but uniform temperature. We represent this observation by CBR(u). The fact was that $\mathcal{PHY} \cup BB(u)$ quasi-support CBR(u) while $\mathcal{PHY} \cup SS(u)$ do not. That is, while $V_{SS} \subset V_{BB}$, $\emptyset = F_{BB} \subset F_{SS}$. In other words, $P_{SS} \leq_{success} P_{BB}$.

Although the proponents of P_{SS}, applying the negative heuristics, found eventually some auxiliary hypotheses to protect the hard core, the program never recovered from the (serious) anomaly of not being able to predict such an important consequence as the existence of background radiation [Coles-Lucchin 1995]. Therefore, P_{BB} became the dominant research program in cosmology for almost twenty years. But meanwhile, $\mathcal{PHY}$ was expanded to $\mathcal{PHY}^{'}$ by the inclusion of the theories of grand unification that conceive of the electromagnetic, weak and strong forces as results of symmetry breaks of a single grand force at different temperatures.[14]

One inference that was drawn around 1980 was that $\mathcal{PHY}^{'} \cup BB(u)$ quasi-supports $m(u)$, where $m(u)$, to be interpreted as *"Magnetic monopoles are abundant in the*

[12] This is of course an idealization. It is not even established that physics constitutes a consistent body of knowledge.

[13] $\mathcal{T}_X$ represents, of course, the theory corresponding to program P_X. We omit stage indexes, although they will be implicit in the definitions of the $\mathcal{T}$s.

[14] It is interesting to note that while $\mathcal{PHY}$ is a precondition for $\mathcal{T}_{BB}$, the success of the latter was influential in the creation of the theories of grand unification [Lederman 1995].

universe" is a sentence that can be checked out by astrophysical observations. The basic idea was that magnetic monopoles, which were never observed (magnetic poles always come in pairs), should be the result of symmetry breaks in the first instants of life of the universe in which the (relatively) paced expansion would reduce its temperature. The fact is that $m(u)$ has not been observed, becoming then a partial anomaly for P_{BB} [Guth 1997].

Some physicists postulated an alternative, called the *inflationary model*, while P_{BB} became known as the *standard model* [Coles-Lucchin 1995]. The former is defined by $\mathcal{T}_{\text{Inf}} = \{\mathcal{PHY}', BB(u), \text{Inf}(u)\}$, where $\text{Inf}(u)$ is the claim that the universe underwent a very short period of "inflation". That is, a rapid expansion that enlarged the original quantum fluctuations in the structure of the universe and did not let them smooth out. Moreover, one possible consequence of an inflationary episode is that the rapid process of cooling would have wiped out possible magnetic monopoles.

It is important to note that $\text{Inf}(u)$ constitutes in fact a set of theoretical claims that cannot be seen as part of a protective belt in P_{BB}. Therefore P_{BB} and $P_{\text{Inf}(u)}$ began a competition that has not yet a clear winner, although a certain consensus is growing that in fact $P_{\text{Inf}(u)} \leq_{success} P_{BB}$ [Earman-Mosterin 1999]. Let us see why.

One of the main reasons is that the research program of the grand unified theories has been shown to be degenerative [Penrose 1989]. It seems safe, therefore, to return back to $\mathcal{PHY}$, which together with $BB(u)$ does *not* quasi-support $m(u)$. In fact, since the premise was wrong in the argument according to which the Big Bang should have witnessed the generation of magnetic monopoles, the claim that the fact of not detecting monopoles is an "anomaly" for P_{BB} is not warranted.

Moreover, $\mathcal{T}_{\text{Inf}(u)}$ yields, as said, the prediction that the initial imperfections in the texture of the universe have been blown out, generating the seeds for galaxies and other cosmic structures. Let us represent this by means of the statement $g(u)$. But $\mathcal{T}_{BB}$ does not quasi-support $g(u)$. Since $g(u) \succ\!\!- \neg ht(u)$, where $ht(u)$ is *"The temperature of the universe is homogeneous"*, we have there the possibility of testing the claim of the inflationary model. An inflationary expansion of the early universe would have consolidated the initial anisotropies while the standard model predicts a much more homogeneous structure, since the predicted expansion, being slower, would have smoothed out initial imperfections. In fact, measurements made by the space probe COBE have shown that $\neg ht(u)$ is the case, but in a magnitude much lower than implied by $\mathcal{T}_{\text{Inf}(u)}$ [Skalski-Sukenik 1992]. Therefore, it is safe to claim that there is an indetermination about this issue.

Finally, $\mathcal{T}_{\text{Inf}(u)}$ predicts that the universe is "flat". That is, that its mass is enough to keep the expansion from accelerating. If we denote this claim by $f(u)$ we have that, again, $\mathcal{T}_{BB}$ does not quasi-support $f(u)$. In turn we have that $f(u) \succ\!\!- \bar{e}(u)$, where $\bar{e}(u)$ indicates that the expansion decreases or remains constant [Weinberg 1992]. That is, if the universe is flat it should be observed that far away events recede with less velocity than those in a closer range. But recent observations about the behavior of certain types of supernovae seem to indicate that the expansion *increases* [Hogan-Kirshner-Suntzeff 1999]. If so, this indicates that we have an anomaly for $P_{\text{Inf}(u)}$.

Therefore, although it is too early to claim that any of these two programs has won the debate, it seems that, so far, $P_{\text{Inf}(u)} \leq_{success} P_{BB}$.[15]

7 Conclusion

We have shown that pragmatic reasoning, with its assumption of the fallibility and corrigibility of its conclusions, can be formalized combining two different approaches. One is the syntactic approach to defeasible reasoning, which assumes that the process

[15] See, however the strong statement against both hypothesis signed by a large number of prestigious scientists in [Arp et al. 2004].

of reasoning consists in the construction of arguments for and against a possible claim. Weighting the arguments according to their conclusive strength may yield warrant for their conclusions.

On the semantic side, we draw from the model-theoretic approach of partial structures. In those structures the relations may remain undefined for a fragment of their domain. Even so, sentences that may be true in an admissible extension are considered quasi-true in the original structure.

We showed that defeasible reasoning can be endowed with a partial-structure semantics in a very natural way. This framework allowed us to advance a formalization of the methodology of research programs. This methodology constitutes a clear prescription of how pragmatic reasoning may proceed in scientific matters. Possible pieces of evidence may be considered quasi-supported by a program if arguments for them can be constructed. If not, new auxiliary hypothesis can be added to the sentences accepted by the program in order to be able to construct now an argument for the controversial piece of evidence. According to the evolution of the program it is possible to compare it to a competing program.

These ideas were used to represent the evolution of three different programs in contemporary cosmology. One of them disappeared several decades ago, but the other two, the standard model and the inflationary model, keep competing with an apparent advantage for the former.

Beyond its application as a tool to make rational reconstructions in scientific matters, the methodology advanced here has many potential applications that range from Artificial Intelligence (defeasible reasoning emerged as an alternative to the *non-monotonic* logics developed in this area of Computer Science in the 1980s) to Philosophy of Science. A matter of further work is to develop an alternative methodology of scientific reasoning, superseding Lakatos' by the specification of strategies for finding alternative theories in case an anomaly arises or even for the efficient search of auxiliary hypotheses.

References

[Arp et al. 2004] Arp, H. and more than other 30 signatures, 2004, "An Open Letter to the Scientific Community", *New Scientist* **2448**, May 22 (also in `http://cosmologystatement.org/`).

[Bueno-de Souza 1996] Bueno, O. - de Souza, E., 1996, "The Concept of Quasi-Truth", *Logique et Analyse* **153-154**: 183–189.

[Coles-Lucchin 1995] Coles, P. - Lucchin, F., 1995, *Cosmology: the Origin and Evolution of Cosmic Structure*, N.Y.: John Wiley.

[da Costa- French 1989] da Costa, N.C.A. - French, S., 1989 "Pragmatic Truth and the Logic of Induction", *British Journal of Philosophy of Science* **40**:333–356.

[da Costa - French 1990] da Costa, N.C.A. - French, S., 1990 "The Model-Theoretic Approach in the Philosophy of Science", *Philosophy of Science* **57**:248–265.

[da Costa, Bueno, French 1998] da Costa, N.C.A.- Bueno, O. - French, S., 1998, "The Logic of Pragmatic Truth", *Journal of Philosophical Logic* **27**:603–620.

[Delrieux 2001] Delrieux, C., 2001, "The Role of Defeasible Reasoning in the Modelling of Scientific Research Programmes", *Proceedings of the First International Workshop on Computer Modeling of Scientific Reasoning and Applications* 861–869, Las Vegas (Nevada).

[Dung 1995] Dung, P., 1995, "On the Acceptability of Arguments and its Fundamental Role in Non-Monotonic Reasoning, Logic Programming and n-Person Games", *Artificial Intelligence* **77**:331–357.

[Earman-Mosterin 1999] Earman, J. - Mosterin, J., 1999, "A Critical Look at Inflationary Cosmology", *Philosophy of Science* **66**:1–49.

[French 2000] French, S., 2000, "The Reasonable Effectiveness of Mathematics: Partial Structures and the Application of Group Theory to Physics", *Synthese***125**: 103-120.

[García-Simari 2004] García, A. - Simari, G., 2004, "Defeasible Logic Programming: an Argumentative Approach", *Theory and Practice of Logic Programming***4**: 95–138.

[Ginsberg 1987] Ginsberg, M., 1987, (ed.), *Readings in Nonmonotonic Reasoning*, Los Altos (CA): Morgan Kaufmann.

[Guth 1997] Guth, A., 1997, *The Inflationary Universe*, Reading (MA): Addison-Wesley.

[Hempel 1965] Hempel, C., 1965, *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*, N.Y.: Free Press.

[Hogan-Kirshner-Suntzeff 1999] Hogan, C. - Kirshner, R. - Suntzeff, N., 1999, "Surveying Space-Time with Supernovae", *Scientific American* 280(1):28–33.

[Kuhn 1962] Kuhn, T., 1962, *The Structure of Scientific Revolutions*, Chicago (IL): University of Chicago Press.

[Lakatos 1978] Lakatos, I., 1978, *The Methodology of Scientific Research Programmes*, Cambridge (U.K): Cambridge University Press.

[Lakatos-Musgrave 1970] Lakatos, I. - Musgrave, A., 1970, *Criticism and the Growth of Knowledge*, Cambridge (U.K): Cambridge University Press.

[Lederman 1995] Lederman, L. - Schramm, D., 1995, *From Quarks to the Cosmos: Tools of Discovery*, N.Y.: Scientific American Library.

[Loui 1986] Loui, R., 1986, "Defeat Among Arguments: a System of Defeasible Inference", *Computational Intelligence* **3**(3):100-106.

[Loui 1991] Loui, R., 1991, "Dialectic, Computation and Ampliative Inference", in *Philosophy and AI*, Cummins, R. – Pollock, J. (eds.), Cambridge (MA): MIT Press.

[Loui 1998] Loui, R., 1998, "Process and Policy: Resource-Bounded Non-Demonstrative Reasoning", *Computational Intelligence* **14**(1): 1–38.

[Mikenberg-da Costa-Chuaqui 1986] Mikenberg, I. - da Costa, N.C.A. - Chuaqui, R., 1986, "Pragmatic Truth and Approximation to Truth", *Journal of Symbolic Logic* **51**:201–221.

[Penrose 1989] Penrose, R., 1989, "Difficulties with Inflationary Cosmology", *Annals of the New York Academy of Sciences* **271**:249–264.

[Pollock 1991] Pollock, J., 1991, "A Theory of Defeasible Reasoning", *International Journal of Intelligent Systems* **6**: 33–54.

[Polya 1954] Polya, G., 1954, *Mathematics and Plausible Reasoning*, Princeton (N.J.): Princeton University Press.

[Poole 1985] Poole, D., 1985, "On the Comparison of Theories: Preferring the Most Specific Explanation", in *Proceedings of the Ninth International Joint Conference on Artificial Intelligence*, Los Altos (CA): Morgan Kaufmann, Los Altos (CA).

[Popper 1959] Popper, K., 1959, *The Logic of Scientific Discovery*, London: Hutchinson.

[Prakken-Vreeswijk 2001] Prakken, H. - Vreeswijk, G., 2001, "Logics for Defeasible Argumentation", in *Handbook of Philosophical Logic* **Vol. 4**(second edition), Gabbay, D. – Guenthner, F. (eds.), Dordrecht (The Netherlands): Kluwer.

[Simari-Loui 1992] Simari, G. - Loui, R., 1992, "A Mathematical Treatment of Defeasible Reasoning and its Implementation", *Artificial Intelligence* **53**:127–157.

[Skalski-Sukenik 1992] Skalski, V. - Sukenik, M., 1992, "Radiation of the Cosmic Background", *Applied Space Science* **187**:155-160.

[Stalnaker 1984] Stalnaker, R., 1984, *Inquiry*, Cambridge (MA): MIT Press.

[Suppe 1977] Suppe, F., 1977, *The Structure of Scientific Theories*, Urbana (IL): University of Illinois Press.

[Thagard 1993] Thagard, P., 1993, *Computational Philosophy of Science*, Cambridge (MA): MIT Press.

[van der Hoek et al. 1992] van der Hoek, W. - Meyer, J. - Tan, Y. - Witteveen, C. (eds.), 1992, *Non-Monotonic Reasoning and Partial Semantics*, Chichester (U.K.): Ellis Horwood, Chichester, 1992.

[Vreeswijk 1997] Vreeswijk, G., 1997, "Abstract Argumentation Systems", *Artificial Intelligence* **90**:225–279.

[Weinberg 1992] Weinberg, S., 1992, *Dreams of a Final Theory*, N.Y.: Pantheon Books.