A formalization of defeasible argumentation using labelled deductive systems

(Preliminary report)

Carlos Iván Chesñevar[1]　　Guillermo Ricardo Simari

Instituto de Ciencias e Ingeniería de Computación (ICIC)
Grupo de Investigación en Inteligencia Artificial (GIIA)
Departamento de Ciencias de la Computación
Universidad Nacional del Sur
Av.Alem 1253 – (8000) Bahía Blanca – República Argentina
Fax: (54)(91)595136 – Tel.: (54)(91)595135
Email: {ccchesne,grs}@criba.edu.ar

Key words: artificial intelligence, defeasible reasoning, argumentative systems, labelled deductive systems

Abstract

Argumentative systems [SL92, Vre93, Che96] are formalizations of defeasible reasoning [Pol87, Nut88]. An *argument* is a tentative piece of reasoning an intelligent agent can use to reach a given conclusion. In case there is information available supporting *counterarguments* which *defeat* the argument, its conclusion will no longer be valid. In order to determine whether a conclusion h is *justified belief*, it is necessary to consider a tree-like structure (having an argument A for h as its root), in which defeaters for A, defeaters for these defeaters, and so on, must be taken into account. If the argument A prevails over all its associated defeaters, then A is called a *justification* for h.

Currently there exist several alternative formalizations of defeasible argumentation. Recent work [PS96, KT96, BDKT97] has shown that defeasible argumentation constitutes a point of confluence for the characterization of different approaches to nonmonotonic reasoning (NMR). From the early '90 there have been several attempts to find a unified logical framework for NMR. In this respect, the *labelled deductive systems* [Gab96a] (or LDS) constitute an attractive approach, allowing to characterize different logics by introducing *labels* as part of the logic's object language and keeping a single inference mechanism for all logics.

This paper presents a formal approach for characterizing defeasible argumentation in terms of LDS. Inference rules are presented in the style of natural deduction, and they capture the process of defeasible argumentation as defined in the *MTDR* framework [SL92, SCG94]. We contend that this approach makes easier to state and prove properties and characteristics of defeasible argumentation within a logical-deductive setting.

[1] Becario de Perfeccionamiento del Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), República Argentina.

A formalization of defeasible argumentation using labelled deductive systems

(Preliminary report)

1 Introduction and motivations

Argumentative systems [SL92, Vre93, Che96] are formalizations of defeasible reasoning [Pol87, Nut88]. An *argument* A is a tentative piece of reasoning an intelligent agent can use to support a given hypothesis h. In case there is information available supporting *counterarguments* which *defeat* this argument, the hypothesis h will no longer be valid. In order to determine whether a conclusion h is *justified belief*, it is necessary to consider a tree-like structure (having an argument A for h as its root), in which defeaters for A, defeaters for these defeaters, and so on, must be taken into account. If the argument A prevails over all its associated defeaters, then A is called a *justification* for h.

Currently there exist several alternative formalizations of defeasible argumentation. Recent work [PS96, KT96, BDKT97] has shown that defeasible argumentation constitutes a point of confluence for the characterization of different approaches to non-monotonic reasoning (NMR). From the early '90 there have been several attempts to find a unified logical framework for NMR. In this respect, *labelled deductive systems* [Gab96b] (or LDS) constitute an attractive approach, allowing to characterize different logics by introducing *labels* as part of the logic's object language and keeping a single inference mechanism for all logics.

This paper introduces a formal approach for characterizing defeasible argumentation in terms of LDS. Inference rules are presented in the style of natural deduction, capturing most of the process of defeasible argumentation as defined in the *MTDR* framework [SL92, SCG94]. We contend that this approach makes easier to state and prove properties and characteristics of defeasible argumentation within a logical-deductive setting.

The paper is structured as follows. Section 2 introduces the concept of labelled deductive systems. Then, in section 3, we analyze how to define the MTDR framework in terms of an LDS. In our discussion we assume that the reader is familiar with this framework (which is also summarized in the appendix at the end of this paper). Finally, in section 4 we consider the main conclusions that have been obtained, as well as research directions for future work.

2 Logical systems as LDS

Following Gabbay [Gab96b], a *logic* can be seen as a pair $(\mid\!\sim, S_{\mid\sim})$, where $\mid\!\sim$ is a consequence relation on a language L and $S_{\mid\sim}$ is an *labelled deductive system* or LDS for short. The need for introducing LDS is based on the fact that a consequence relation defines a binary relation on finite sets of formulas Δ and Γ, written as $\Gamma \mid\!\sim \Delta$. The LDS $S_{\mid\sim}$ provides the proof system for $\mid\!\sim$.

A LDS constitutes a proof system defined as a triple $(\mathcal{A}, L, \mathcal{R})$, where L is some logical language (involving connectives and wffs), $\mathcal{A}$ is an algebra (with some operations) of *labels*, and $\mathcal{R}$ is a discipline of labelling formulas of the logic. Formulas are labelled according to a family of deduction rules, and with agreed ways of propagating the labels via the application of the deduction rules. We replace the traditional notion of consequence between formulas of the form $A_1, \ldots, A_n \mid\!\sim B$ by the notion of consequence between labelled formulas $t_1 : A_1, \ldots, t_n : A_n \mid\!\sim s : B$. Depending on the logical system involved, the intuitive meaning of the labels varies.

Definition 2.1 (Algebraic LDS**)** *Let* $\mathcal{A}$ *be a first-order language,* $\mathcal{A} = (A, R_1, \ldots, R_k, f_1, \ldots, f_m)$ *where* A *is the set of terms of the algebra, and* R_i*'s are predicate symbols on* A*, and* f_i*'s are function symbols on* A *of various arities. Elements of* A *can be thought of as* atomic labels*. Functions can help generate more labels, and predicates give additional structure to the labels. A diagram of labels is a set* D *containing elements generated from* A *by the function symbols together with formulas of the form* $R(t_1, \ldots, t_m)$*, where* $t_i \in D$ *and* R *is a predicate symbol of the algebra. Let* L *be a predicate language with connectives* $\sharp_1, \ldots, \sharp_n$ *of various arities, with quantifiers and the same set of atomic terms* A *as the algebra. We define the notions of* label *and* declarative unit*, as follows:*

1. *An* atomic label *is any* $t \in A$*. A* label *is any term generated from the atomic labels by the symbols* $f_1, \ldots, f_m$*.*
2. *A* formula *is any formula of* L*.*
3. *A* declarative unit *is a pair* $t : A$*, where* t *is a label, and* A *is a formula.*

Natural deduction allows us to characterize a logical system in terms of a set of *inference rules*. No axioms are needed, since they can be expressed as inference rules with no preconditions (they can always be applicable). The lack of axioms is compensated by a clean application of the deduction theorem. New formulas can be introduced as hypotheses (assumptions) whenever necessary along a given proof. Those assumptions should be later discharged. Natural deduction allows us to represent introduction and discharge of assumptions using *boxes* as a pictorial notation. Every time a new assumption is introduced, a box is opened. In order to discharge assumptions, rules which have boxes as part of their premises can be used. Hence a box contains a fragment of a proof, in which it became necessary to introduce some assumptions (those which 'opened' the box). Some rules will allow to discharge assumptions by 'closing' boxes.

3 Characterizing *MTDR* as an LDS

In any argumentative framework several representational levels (or layers) can be distinguished [PV98]. First, we may consider a *logical layer*, which involves providing a suitable object language for representing knowledge (e.g. some extension of classical logic). That object language, together with some inference rules, should enable the construction of *arguments*. Using this level as a basis, a *dialectical layer* can be defined, in which conflict relationships between arguments can be formally stated[2].

In [PV98], the authors discuss two additional layers, the *procedural layer* (which involves the definition of protocols for dispute), and the *strategic layer* (which considers some

[2] Actually some approaches start by taking this level as a basis for further analysis (for example, Dung's approach [Dun93] is based on a 2-uple $(Args, <)$, where $Args$ denotes the set of all possible arguments, and $<$ denotes a preference relationship on members of $Args$)

kind of heuristics for dispute). In our approach, we will only take into consideration the first two layers. Next we will discuss the first layer, which involves representing knowledge and performing inference in order to build arguments.

3.1 The logical layer

We will assume that the reader is familiar with the MTDR framework (see appendix). We will define an object language $\mathcal{L}_{da}$ which will allow us to represent knowledge for defeasible argumentation. Following Gabbay [Gab96b], formulas in $\mathcal{L}_{da}$ will be labelled. We will provide an algebra $\mathcal{A}$ of labels. Atomic labels will be associated with wffs in the agent's knowledge base, denoted $(\mathcal{K}, \Delta)$. Every wff in $(\mathcal{K}, \Delta)$ will have the form $l : f$, where l denotes an *atomic label*, and f is a wff in $\mathcal{L}_{da}$.

Definition 3.1 (Language $\mathcal{L}_{da}$) *Let $\mathcal{L}$ be a propositional language. We will define an object language for defeasible argumentation $\mathcal{L}_{da}$ as a subset of $\mathcal{L}$. Wffs in $\mathcal{L}_{da}$ can be* facts, strong rules *and* defeasible rules*.*

- Facts*, which correspond to literals in classical logic.*
- Strong rules*, having the form $a_1, a_2, \ldots, a_k \rightarrow b$, where $a_1, a_2, \ldots, a_k, b$ are literals. Strong rules should be understood as material implications in classical logic.*
- Defeasible rules*, having the form $a_1, a_2, \ldots, a_k \succ\!\!- b$, where $a_1, a_2, \ldots, a_k, b$ are literals, and " $\succ\!\!-$ " is a meta-relation linking the antecedent $a_1, a_2, \ldots, a_k$ with the consequent b.*

The agent's knowledge base is a pair $(\mathcal{K}, \Delta)$ involving *non-defeasible knowledge* $\mathcal{K}$, and *defeasible knowledge* Δ. $\mathcal{K}$ is a consistent set of facts and strong rules (*i.e.*, $\mathcal{K} \not\vdash \bot$), and Δ is a finite set of defeasible rules. Every wff in $(\mathcal{K}, \Delta)$ will be labelled. The *algebra of labels* for defeasible argumentation $\mathcal{A}_{arg}$ will be defined as follows:

Definition 3.2 (Algebra $\mathcal{A}_{arg}$) *Let $\mathcal{A}_{arg} = (A, \otimes, *, (,))$, where $\otimes$, $*$ are predicate symbols on A. Members of A will be called* atomic *labels. They will be associated with each fact, strong rule and defeasible rule in $\mathcal{L}_{da}$. These labels will be denoted with subscripted Greek letters α, β and γ. For the sake of simplicity, we will use $\alpha_1, \ldots, \alpha_k$ for facts, $\beta_1, \ldots, \beta_j$ for strong rules, and $\gamma_1, \ldots, \gamma_j$ for defeasible rules. More complex labels involving predicate symbols will result from applying inference rules on wffs in $(\mathcal{K}, \Delta)$.*

Example 3.1 *Consider a knowledge base with the facts e_1 and e_2, the strong rule $e_1 \rightarrow h_2$, and the defeasible rules $e_2 \succ\!\!- h_1$, $h_1, h_2 \succ\!\!- q$, $q \succ\!\!- \neg e_1$. We can represent that KB as a pair $(\mathcal{K}, \Delta)$ where*

$$\begin{array}{llcll} \mathcal{K} = \{ & \alpha_1 : e_1, & & \Delta = \{ & \gamma_1 : e_2 \succ\!\!- h_1, \\ & \alpha_2 : e_2, & \textit{and} & & \gamma_2 : h_1, h_2 \succ\!\!- q, \\ & \beta_1 : e_1 \rightarrow h_2 \quad \}, & & & \gamma_3 : q \succ\!\!- \neg e_1 \quad \} \end{array}$$

Let $(\mathcal{K}, \Delta)$ be a knowledge base. We will define a consequence relation $\mathrel{|\!\!\sim}_{A}$, which will allow us to derive new wffs from those in $(\mathcal{K}, \Delta)$. The consequence relation $\mathrel{|\!\!\sim}_{A}$ will also propagate labels along with wffs in $\mathcal{L}_{da}$. Complex labels will be built from atomic ones, giving account for the steps needed to conclude a new wff from $(\mathcal{K}, \Delta)$. Some distinguished labels will be called *arguments*. We will also define a preference ordering on labels, according to which some arguments can be considered to be preferred over others.

Definition 3.3 (Consequence relation $\mathrel{|\!\!\sim}_{A}$) *Let $\mathcal{L}_{da}$ be our object language for defeasible argumentation. We will define the consequence relation $\mathrel{|\!\!\sim}_{A}$ based on the language $\mathcal{L}_{da}$ as a set of inference rules { $\wedge$-introduction, $\wedge$-elimination1, $\wedge$-elimination2, $\otimes$-introduction, $*$-introduction, *-elimination in argument (1), *-elimination in argument (2), projection, $\bot$-introduction in argument, $\bot$-propagation }.*

1.
$$\frac{\epsilon_1 : h_1 \qquad \epsilon_2 : h_2}{(\epsilon_1, \epsilon_2) : h_1, h_2}\ \wedge\textit{-introduction}$$

2.
$$\frac{(\epsilon_1, \epsilon_2) : h_1, h_2}{\epsilon_1 : h_1}\ \wedge\textit{-elimination1}$$

3.
$$\frac{(\epsilon_1, \epsilon_2) : h_1, h_2}{\epsilon_2 : h_2}\ \wedge\textit{-elimination2}$$

4.
$$\frac{\epsilon : h_1 \ldots h_k \qquad \gamma : h_1 \ldots h_k \succ\!\!\!-\, q}{\gamma \otimes \epsilon : q}\ \otimes\textit{-introduction}$$

5.
$$\frac{\epsilon : h_1 \ldots h_k \qquad \beta : h_1 \ldots h_k \rightarrow q}{\beta * \epsilon : q}\ \textit{*-introduction}$$

6.
$$\frac{\beta * \epsilon : h}{\epsilon : h}\ \textit{*-elimination in arguments (1)}$$

7.
$$\frac{\gamma \otimes \epsilon * \epsilon' : q \quad \boxed{\begin{array}{c} \epsilon * \epsilon' : h \\ \vdots \\ \epsilon' : h \end{array}}}{\gamma \otimes \epsilon' : q}\ \textit{*-elimination in argument (2)}$$

8.
$$\frac{\epsilon : h}{h}\ \textit{projection (assuming } \epsilon \textit{ is free of occurrences of } \otimes\textit{)}$$

9.
$$\frac{\epsilon : q \quad \boxed{\begin{array}{c} q \\ \vdots \\ \bot \end{array}}}{\epsilon : \bot}\ \bot\textit{-introduction in argument}$$

10.
$$\frac{\epsilon : \bot}{\beta \otimes \epsilon : \bot}\ \bot\textit{-propagation}$$

11.
$$\frac{\gamma \otimes \epsilon : b \qquad \gamma : a_1, a_2, \ldots, a_k \succ\!\!\!-\, b}{\epsilon : a_i \ \textit{(for } i = 1 \ldots k\textit{)}}\ \otimes\textit{-elimination}$$

Rules 1, 2 and 3 allow to introduce and eliminate conjunction when performing inference.[3] Rule 4 allows the elimination of $\succ\!\!\!-$ by infering a new wff which introduces a more complex label (using $\otimes$).[4] Rule 11, on its turn, accounts for eliminating a single occurence

[3] The conjunction $a_1 \wedge a_2 \wedge a_k$ will be denoted $a_1, a_2, \ldots, a_k$.

[4] Predicate symbols in the algebra of labels are introduced using infix notation.

of $\otimes$.[5] Rules 6 and 7 allow eliminating occurences of $*$ in labels. Every occurrence of $*$ accounts for the application of material implication (rule 5). Rule 8 states that any wff $\epsilon : h$ whose label is $\otimes$-free can be considered to hold in classical logic (*i.e.*, no defeasible rules were used in its derivation). Rule 9 accounts for introducing inconsistency. Inconsistency propagates from one wff to another, when application of defeasible rules is involved (rule 10).

Following [SL92], we want to capture the notion of an *argument* for a given literal q as a subset $A \subset \Delta$ such that a) $\mathcal{K} \cup A$ allow us to infer q; b) $\mathcal{K} \cup A$ is consistent (in the sense that $\mathcal{K} \cup A$ should not derive p and $\neg p$). We introduce inference rules which allow us to characterize the notion of argument as a distinguished label. Thus, rules 6 and 7 allow us to discard material implications from labels. Two different inference rules are needed: in the first case, we are discarding the last application of the inference rule 5; in the second case, we discard some previous introduction of material implication.

Rule 8 establishes that any wff $\epsilon : h \in \mathcal{K}$ can be considered as holding in classical logic. Rule 9 defines a special notion of *inconsistency* in our setting. Whenever having a formula $\epsilon : q$ allows us to derive inconsistency from q (within classical logic),[6] we will say that the wff $\epsilon : \bot$ also holds.

We will distinguish those labelled wffs which are "consistent" (in the sense of rule 9) and free of ocurrences of $*$. This kind of labelled wffs will be called *arguments*.

Definition 3.4 (Completed argument. Argument) *Let $\epsilon : h$ be a wff in $\mathcal{L}_{da}$ such that $(\mathcal{K}, \Delta) |\!\sim_{A} \epsilon : h$, and $(\mathcal{K}, \Delta) \not|\!\sim_{A} \epsilon : \bot$. Then $\epsilon : h$ is called an* completed argument. *If $\epsilon : h$ is closed under application of rules 6 and 7, then $\epsilon : h$ is called an* argument.

Example 3.2 *Let $(\mathcal{K}, \Delta)$ be defined as in example 3.1. Then $(\mathcal{K}, \Delta) |\!\sim_{A} \beta_1 * \alpha_1 : h_2$, $(\mathcal{K}, \Delta) |\!\sim_{A} \gamma_1 \otimes \alpha_2 : h_1$, $(\mathcal{K}, \Delta) |\!\sim_{A} (\gamma_1 \otimes \alpha_2, \beta_1 * \alpha_1) : h_1, h_2$, and $(\mathcal{K}, \Delta) |\!\sim_{A} (\gamma_2 \otimes (\gamma_1 \otimes \alpha_2, \beta_1 * \alpha_1)) : q$.*

Note that, according to definition 3.4, the wff $\mathcal{A}_1 : q$ with $\mathcal{A}_1 = (\gamma_2 \otimes (\gamma_1 \otimes \alpha_2, \beta_1 * \alpha_1))$, is a completed argument. It also holds that $(\mathcal{K}, \Delta) |\!\sim_{A} \mathcal{A}_2 : q$, with $\mathcal{A}_2 = (\gamma_2 \otimes (\gamma_1 \otimes \alpha_2, \alpha_1))$. Note that $\mathcal{A}_2 : q$ is an argument.

It also holds that $(\mathcal{K}, \Delta) |\!\sim_{A} \mathcal{A}_3 : \neg e_1$, with $\mathcal{A}_3 = (\gamma_3 \otimes (\gamma_2 \otimes (\gamma_1 \otimes \alpha_2, \beta_1 * \alpha_1)))$. However, $\mathcal{A}_3 : \neg e_1$ is not an argument, since $(\mathcal{K}, \Delta) |\!\sim_{A} \mathcal{A}_3 : \bot$.

3.2 The dialectical layer

One interesting feature of labelled deductive systems is the possibility of defining a new LDS in terms of another one, *i.e.*, adding a new labelling discipline to labelled formulas from an LDS L_1, getting a new LDS L_2. We will use this feature in order to define a dialectical layer for defeasible argumentation based on the logical layer introduced before. In the previous section we presented an LDS $((\mathcal{K}, \Delta), |\!\sim_{A})$ which allowed us to characterize the notion of argument as a distinguished labelled formula, in which defeasible rules were used consistently with the knowledge available in $\mathcal{K}$. Next we will show how to capture the notion of *defeat* among arguments (see def. A.5), as well as the notion of *dialectical tree* (see def. A.6).

[5] This rule is needed when considering *subarguments* of a given argument.

[6] We assume that all inference rules for classical logic are available.

3.2.1 Capturing defeat among arguments

A setting for reasoning with defeasible argumentation involves *inconsistent* information in the agent's knowledge base. An argument [7] $\langle A, h\rangle$ can be *defeated* by another argument $\langle B, q\rangle$. Defeaters can on its turn be defeated by other arguments, so that a so-called *dialectical tree* (see def. A.6) results.

In order to capture when an argument defeats another, we will use the following lemma which provides a syntactic criterion for finding the defeaters associated with a given argument.

Lemma 3.1 *(*Pruning Lemma*).*[8] *Let $\langle A, h\rangle$, $\langle B, j\rangle$ be arguments, such that $\langle B, j\rangle \gg_{\mathsf{def}} \langle A, h\rangle$. Then B is also an argument for a ground literal q, such that q is the complement of some consequent of a defeasible rule in A, and $\langle B, q\rangle$ is a defeater.*

This lemma states that in order to find a defeater for a given argument $\langle A, h\rangle$, it is only necessary to compute those defeaters whose conclusions correspond to the complement of the defeasible rules in A. Thus, given a defeater $\langle B, j\rangle$ for $\langle A, h\rangle$, either $\langle B, j\rangle$ attacks $\langle A, h\rangle$'s conclusion, or $\langle B, j\rangle$ attacks some inner literal corresponding to the conclusion of a defeasible rule in A This allows us to formalize defeat by using only two inference rules.

1. $$\frac{\mathcal{A}_1 : b \qquad \mathcal{A}_2 : \neg b \qquad \phi(\mathcal{A}_2, \mathcal{A}_1)}{\mathcal{A}_2 \gg_{\mathsf{def}} \mathcal{A}_1 : (\mathcal{A}_2 : \neg b)} \; \gg_{\mathsf{def}}\text{-introduction 1}$$

2. $$\frac{\mathcal{A}_3 : \neg q \quad \mathcal{A}_3 \gg_{\mathsf{def}} \mathcal{A}_2 : (\mathcal{A}_3 : \neg q) \quad \boxed{\begin{array}{c} \mathcal{A}_1 : b \\ \vdots \\ \mathcal{A}_2 : q \end{array}}}{\mathcal{A}_3 \gg_{\mathsf{def}} \mathcal{A}_1 : (\mathcal{A}_3 : \neg q)} \; \gg_{\mathsf{def}}\text{-introduction2}$$

where $\phi(\mathcal{A}_1, \mathcal{A}_2)$ denotes that $\mathcal{A}_1$ is a label to be *preferred over* $\mathcal{A}_2$. The preference ordering on labels could be defined using some extra-logical criterion (e.g. specificity).

3.2.2 Dialectical proof theory

The process of determining whether a given argument is a justification can be understood as a *dialogue* between two parties, *proponent* (PRO) and *opponent* (OPP). Informally, an argument is considered to be justified if PRO can make OPP run out of moves against every possible attack. A dialogue is thus an alternate sequence of moves performed by PRO and OPP. PRO introduces the main argument at issue. A party 'wins' a dialogue iff the other party cannot move.

We will formalize a dialogue starting with argument $\mathcal{A}_1 : h_1$ as a label $\mathcal{T}$ attached to the wff $\mathcal{A}_1 : h_1$. Hence we will define an LDS in which wffs are arguments, whereas labels correspond to the steps of a dialogue which has that argument as its root. The label will stand for the 'dialectical tree' computed so far, identifying whether the main argument at issue has been defeated or not. A wff of the form $PRO(\mathcal{A}_1 \star (\mathcal{T}_1 \ldots \mathcal{T}_k)) : (\mathcal{A}_1 : h_1)$ will identify a dialogue in which the argument $\mathcal{A}_1 : h_1$ is supported by PRO, where $\mathcal{A}_1 \star (\mathcal{T}_1 \ldots \mathcal{T}_k)$ represents the associated dialectical tree, and $\mathcal{T}_1 \ldots \mathcal{T}_k$ stands for their immediate subtrees.

[7] Ocassionally we will use MTDR notation for denoting arguments (see def. A.2).

[8] Proof of this lemma in this paper can be found in [Che96].

Definition 3.5 (Consequence relation $\mathrel{|\!\!\sim}_{D}$) *Let $Args_{(\mathcal{K},\Delta)}$ (or Args for short) be the set of all possible* arguments *that can be built from a given knowledge base $(\mathcal{K},\Delta)$, i.e., $Args = \{\ \mathcal{A} : h \in (\mathcal{K},\Delta)^{\mathrel{|\!\!\sim}_{A}}\ :\ \mathcal{A} : h$ is an argument $\}$. Let $Labels(Args)$ be the set of all labels in $Args$. Consider the algebra ($Labels(Args)$, $\gg_{\mathsf{def}}$, $\star$, PRO, OPP), where $\gg_{\mathsf{def}}$, $\star$, PRO and OPP are predicates on $Labels(Args)$. We will define the consequence relation $\mathrel{|\!\!\sim}_{D}$ based on Args as a set of inference rules $\{$ $\gg_{\mathsf{def}}$-introduction (1), $\gg_{\mathsf{def}}$-introduction (2), PR1, PR2, OP2, PR3, OP3$\}$.*

Rules $\gg_{\mathsf{def}}$-introduction (1) and $\gg_{\mathsf{def}}$-introduction (2) capture the definition of defeat, as discussed above. Rule PR1 corresponds to the situation in which PRO introduces an argument $\mathcal{A}_1 : h_1$, starting a new dialogue. The dialogue consists of only the original argument $\mathcal{A}_1 : h_1$; hence we infer $PRO(\mathcal{A}_1)$ as associated dialectical tree.

A dialogue can be attacked in two ways: either by attacking the root (the main argument at issue), or by attacking some subdialogue. Rules OP2 (analogously PR2) corresponds to the first situation. If a dialogue starting with $\mathcal{A}_1 : h_1$ has been performed to a certain extent and it is being won by PRO, (i.e. $PRO(\mathcal{A}_1{\star}(\mathcal{T}_1,\ldots,\mathcal{T}_n)) : (\mathcal{A}_1 : h_1)$), and OPP can defeat $\mathcal{A}_1 : h_1$ by introducing $\mathcal{A}_2 : h_2$, then we get a new, expanded dialectical tree, in which PRO 'loses' (i.e. $\neg PRO(\mathcal{A}_1{\star}(\mathcal{T}_1,\ldots,\mathcal{T}_n,\mathcal{A}_2)) : (\mathcal{A}_1 : h_1)$).

When attacking a subdialogue, the situation is as follows. It may be the case that PRO is losing the dialogue (i.e., $\neg PRO(\mathcal{A}_1{\star}(\mathcal{T}_1,\ldots,\mathcal{T}_i\ldots\mathcal{T}_n)) : (\mathcal{A}_1 : h_1)$) because OPP has defeated PRO in the previous move. In that case, PRO can attack (if possible) the winning subdialogue that supports the OPP's winning position (i.e., $OPP(\mathcal{T}_i) : (\mathcal{A}_2 : q)$). By doing so, a new dialogue results, in which PRO is the winner again (i.e., $PRO(\mathcal{A}_1{\star}(\mathcal{T}_1,\ldots,\mathcal{T}_i',\ldots,\mathcal{T}_n)) : (\mathcal{A}_1 : h_1)$). A similar situation arises when PRO is in a winning situation, and OPP has to find how to 'reinstate' some subdialogue in order not to lose (see rule OP3).

1. Proponent introduces first argument

$$\frac{\mathcal{A}_1 : h_1}{PRO(\mathcal{A}_1) : (\mathcal{A}_1 : h_1)}\ \mathrm{PR1}$$

2. Attacking proponent's argument

$$\frac{PRO(\mathcal{A}_1{\star}(\mathcal{T}_1,\ldots,\mathcal{T}_n)) : (\mathcal{A}_1 : h_1) \qquad \mathcal{A}_2{\gg_{\mathsf{def}}}\mathcal{A}_1 : (\mathcal{A}_2 : q)}{\neg PRO(\mathcal{A}_1{\star}(\mathcal{T}_1,\ldots,\mathcal{T}_n,\mathcal{A}_2)) : (\mathcal{A}_1 : h_1)}\ \mathrm{(OP2)}$$

3. Attacking opponent's argument

$$\frac{OPP(\mathcal{A}_1{\star}(\mathcal{T}_1,\ldots,\mathcal{T}_n)) : (\mathcal{A}_1 : h_1) \qquad \mathcal{A}_2{\gg_{\mathsf{def}}}\mathcal{A}_1 : (\mathcal{A}_2 : q)}{\neg OPP(\mathcal{A}_1{\star}(\mathcal{T}_1,\ldots,\mathcal{T}_n,\mathcal{A}_2)) : (\mathcal{A}_1 : h_1)}\ \mathrm{(PR2)}$$

4. Proponent attacks subdialogue

$$\frac{\neg PRO(\mathcal{A}_1{\star}(\mathcal{T}_1,\ldots,\mathcal{T}_i\ldots\mathcal{T}_n)) : (\mathcal{A}_1 : h_1) \quad \boxed{\begin{array}{c} OPP(\mathcal{T}_i) : (\mathcal{A}_2 : q) \\ \vdots \\ \neg OPP(\mathcal{T}_i') : (\mathcal{A}_2 : q) \end{array}}}{PRO(\mathcal{A}_1{\star}(\mathcal{T}_1,\ldots,\mathcal{T}_i',\ldots,\mathcal{T}_n)) : (\mathcal{A}_1 : h_1)}\ \mathrm{PR3}$$

5. Opponent attacks subdialogue

$$\frac{PRO(\mathcal{A}_1\star(\mathcal{T}_1,\ldots,\mathcal{T}_i\ldots\mathcal{T}_n)) : (\mathcal{A}_1 : h_1) \qquad \boxed{\begin{array}{c} \neg OPP(\mathcal{T}_i) : (\mathcal{A}_2 : q) \\ \vdots \\ OPP(\mathcal{T}_i') : (\mathcal{A}_2 : q) \end{array}}}{\neg PRO(\mathcal{A}_1\star(\mathcal{T}_1,\ldots,\mathcal{T}_i',\ldots,\mathcal{T}_n)) : (\mathcal{A}_1 : h_1)} \ \text{OP3}$$

These rules suffice for defining a dialectical exchange between PRO and OPP. The ability to prove that $PRO(\mathcal{T}) : \mathcal{A}_1 : h_1$ holds, whereas $\neg PRO(\mathcal{T}\,') : \mathcal{A}_1 : h_1$ does not (being $\mathcal{T}\,'$ such that it contains $\mathcal{T}$ as a sub-label), corresponds to the usual notion of *justified argument* in MTDR.

Definition 3.6 (Justification) *Let Args be the set of arguments based on* $(\mathcal{K}, \Delta)$*, and let* $\mathcal{A}_1 : h_1 \in Args$*. We will say that* $\mathcal{A}_1 : h_1$ *is a* justified argument *(or* justification*) iff*

1. $Args \mathrel{|\!\!\sim_{D}} PRO(\mathcal{T}) : (\mathcal{A}_1 : h_1)$
2. $Args \not\mathrel{|\!\!\sim_{D}} \neg PRO(\mathcal{T}\,') : (\mathcal{A}_1 : h_1)$

where $\mathcal{T}$ *is a sub-tree (sub-label) of* $\mathcal{T}\,'$*.*

Example 3.3 *Let* $(\mathcal{K}, \Delta)$ *be a knowledge base from which the arguments* $\mathcal{A}_1 : h_1$ *,* $\mathcal{A}_2 : h_2$ *,* $\mathcal{A}_3 : h_3$ *,* $\mathcal{A}_4 : h_4$ *,* $\mathcal{A}_5 : h_5$ *,* $\mathcal{A}_6 : h_6$ *and* $\mathcal{A}_7 : h_7$ *can be inferred, such that* $\mathcal{A}_2 : h_2$ *defeats* $\mathcal{A}_1 : h_1$ *,* $\mathcal{A}_3 : h_3$ *defeats* $\mathcal{A}_2 : h_2$ *,* $\mathcal{A}_4 : h_4$ *defeats* $\mathcal{A}_1 : h_1$ *,* $\mathcal{A}_5 : h_5$ *defeats* $\mathcal{A}_4 : h_4$ *,* $\mathcal{A}_6 : h_6$ *defeats* $\mathcal{A}_3 : h_3$ *, and* $\mathcal{A}_7 : h_7$ *defeats* $\mathcal{A}_2 : h_2$ *. The proponent starts the debate introducing the argument* $\mathcal{A}_1 : h_1$ *. Proponent and opponent exchange arguments until there are no more arguments to consider. It can be shown that* $\mathcal{A}_1 : h_1$ *is a justification, by performing the following proof steps:*

1. *(PRO introduces* $\mathcal{A}_1 : h_1$ *)*
   $PRO(\mathcal{A}_1) : (\mathcal{A}_1 : h_1)$ *(by PR1).*

2. *(OPP attacks* $\mathcal{A}_1 : h_1$ *with* $\mathcal{A}_2 : h_2$ *)*
   $\neg PRO(\mathcal{A}_1\star(\mathcal{A}_2)) : (\mathcal{A}_1 : h_1)$ *(From 1, by OP2).*

3. *(PRO attacks* $\mathcal{A}_2 : h_2$ *with* $\mathcal{A}_3 : h_3$ *)*
   From 2, we know that $\neg PRO(\mathcal{A}_1\star(\mathcal{A}_2)) : (\mathcal{A}_1 : h_1)$
   Assume $OPP(\mathcal{A}_2) : (\mathcal{A}_2 : h_2)$
   By PR2, it follows that $\neg OPP(\mathcal{A}_2\star(\mathcal{A}_3)) : (\mathcal{A}_2 : h_2)$
   Then $PRO(\mathcal{A}_1\star(\mathcal{A}_2\star(\mathcal{A}_3))) : (\mathcal{A}_1 : h_1)$ *(by PR3).*

4. *(OPP attacks* $\mathcal{A}_1 : h_1$ *with* $\mathcal{A}_4 : h_4$ *)*
   $\neg PRO(\mathcal{A}_1\star(\mathcal{A}_2\star(\mathcal{A}_3), \mathcal{A}_4) : (\mathcal{A}_1 : h_1)$ *(by OP2).*

5. *(PRO attacks* $\mathcal{A}_4 : h_4$ *with* $\mathcal{A}_5 : h_5$ *)*
   By step 4, $\neg PRO(\mathcal{A}_1\star(\mathcal{A}_2\star(\mathcal{A}_3), \mathcal{A}_4)) : (\mathcal{A}_1 : h_1)$

   Assume $OPP(\mathcal{A}_4) : (\mathcal{A}_4 : h_4)$
   Then $\neg OPP(\mathcal{A}_4\star(\mathcal{A}_5)) : (\mathcal{A}_4 : h_4)$ *(by OP2)*
   Then $PRO(\mathcal{A}_1\star(\mathcal{A}_2\star(\mathcal{A}_3), \mathcal{A}_4\star(\mathcal{A}_5))) : (\mathcal{A}_1 : h_1)$

6. *(OPP attacks* $\mathcal{A}_3 : h_3$ *with* $\mathcal{A}_6 : h_6$ *)*
From step 5, $PRO(\mathcal{A}_1\star(\mathcal{A}_2\star(\mathcal{A}_3), \mathcal{A}_4\star(\mathcal{A}_5))) : (\mathcal{A}_1 : h_1\ )$
Assume $\neg OPP(\mathcal{A}_2\star(\mathcal{A}_3)) : (\mathcal{A}_2 : h_2\ )$
Assume $PRO(\mathcal{A}_3) : (\mathcal{A}_3 : h_3\ )$
Then, by OP2, $\neg PRO(\mathcal{A}_3\star(\mathcal{A}_6) : (\mathcal{A}_3 : h_3\ )$
Then $OPP(\mathcal{A}_2\star(\mathcal{A}_3\star(\mathcal{A}_6))) : (\mathcal{A}_2 : h_2\ )$
Then $\neg PRO(\mathcal{A}_1\star(\mathcal{A}_2\star(\mathcal{A}_3\star(\mathcal{A}_6)), \mathcal{A}_4\star(\mathcal{A}_5))) : (\mathcal{A}_1 : h_1\ )$

7. *(PRO attacks* $\mathcal{A}_6 : h_6$ *with* $\mathcal{A}_7 : h_7$ *)*
By step 6, $PRO(\mathcal{A}_1\star(\mathcal{A}_2\star(\mathcal{A}_3\star(\mathcal{A}_6)), \mathcal{A}_4\star(\mathcal{A}_5))) : (\mathcal{A}_1 : h_1\ )$
Assume $OPP(\mathcal{A}_2\star(\mathcal{A}_3\star(\mathcal{A}_6))) : (\mathcal{A}_2 : h_2\ )$
Then by PR2, $\neg PRO(\mathcal{A}_3\star(\mathcal{A}_6), \mathcal{A}_7) : (\mathcal{A}_3 : h_3\ )$
Then, by PR3, $PRO(\mathcal{A}_1\star(\mathcal{A}_2\star(\mathcal{A}_3\star(\mathcal{A}_6\star(\mathcal{A}_7))), \mathcal{A}_4\star(\mathcal{A}_5))) : (\mathcal{A}_1 : h_1\ )$

At this point there are no more arguments to consider. Thus Args $\mathrel{|\!\sim}_D PRO(\mathcal{T}) : (\mathcal{A}_1 : h_1\ )$*, (where* $\mathcal{T} = (\mathcal{A}_1\star(\mathcal{A}_2\star(\mathcal{A}_3\star(\mathcal{A}_6\star(\mathcal{A}_7))), \mathcal{A}_4\star(\mathcal{A}_5)))$ *) and Args* $\not\mathrel{|\!\sim}_D PRO(\mathcal{T}') : (\mathcal{A}_1 : h_1\ )$*, (where* $\mathcal{T}$ *is a sub-label of* $\mathcal{T}$ *'). Hence* $\mathcal{A}_1 : h_1$ *is a justification.*

4 Conclusions and future work

We introduced a formal approach to defeasible argumentation using LDS. Following [PV98], we showed how both a logical layer and a dialectical layer could be separately defined in terms of LDS, and then interrelated.

In this first approach we did not consider some technical issues involved in MTDR, such as *circularity* and *contradictory argumentation* [SCG94]. As shown in [Che96], the notion of *commitment store* (*i.e.*, the set of wffs supported by proponent/opponent) helps avoiding contradictory argumentation, and allows us to get a pruning strategy for the dialectical tree. In our LDS-based approach to MTDR, arguments are labelled formulas. By decomposing the label structure, all defeasible rules used in a given argument can be identified. We are currently working on the definition of suitable inference rules through which the commitment store could be inferred. from the labelled wff corresponding to a particular dialogue.

Resources play also a major role in defeasible argumentation (e.g. time for building arguments, maximal number of defeaters allowed to proponent/opponent, etc.) So-called "resource logics" (in which the availability of 'resources' determines which wffs can be derived) can be naturally formalized in LDS. We think that this kind of logics can be helpful in modelling resource-bounded reasoning.

Many theoretical aspects linking defeasible argumentation and non-monotonic reasoning have been successfully formalized so far [BDKT97], and further work in this area seems to be still ahead. We think that LDS constitute a powerful, unifying framework, in which theoretical results concerning defeasible argumentation can be studied.

A The *MTDR* framework

We will briefly introduce the main concepts and definitions of the *MTDR* framework (see [SL92, SCG94, Che96] for further details).

A.1 Knowledge representation

The knowledge of an intelligent agent $\mathcal{A}$ will be represented using a first-order language $\mathcal{L}$, plus a binary meta-linguistic relation " $\succ\!\!-$ " between sets of non-ground literals of $\mathcal{L}$ which share variables. The members of this meta-linguistic relation will be called *defeasible rules*, and they have the form "$\alpha \succ\!\!- \beta$ ". The relation " $\succ\!\!-$ " is understood as expressing that "reasons to believe in the antecedent α provide reasons to believe in the consequent β". Let $\mathcal{K}$ be a consistent subset of sentences of the language $\mathcal{L}$. This set can be partitioned in two subsets $\mathcal{K}_G$, of *general* (necessary) knowledge, and $\mathcal{K}_P$, of *particular* (contingent) knowledge. The beliefs of $\mathcal{A}$ are represented by a pair $(\mathcal{K}, \Delta)$, where Δ is a finite set of defeasible rules. $\mathcal{K}$ represents the non-defeasible part of $\mathcal{A}$'s knowledge and Δ represents information that $\mathcal{A}$ is prepared to take at less than face value. $\Delta^{\downarrow}$ denotes the set of all ground instances of members of Δ.

A.2 Inference

Definition A.1 *Let Γ be a subset of $\mathcal{K} \cup \Delta^{\downarrow}$. A ground literal h is a* defeasible consequence *of Γ, abbreviated $\Gamma \mid\!\sim h$, if and only if there exists a finite sequence $B_1, \ldots, B_n$ such that $B_n = h$ and for $1 \leq i < n$, either $B_i \in \Gamma$, or B_i is a direct consequence of the preceding elements in the sequence by virtue of the application of any inference rule of the first-order theory associated with the language $\mathcal{L}$. The ground instances of the defeasible rules are regarded as material implications for the application of inference rules. We will also write $\mathcal{K} \cup A \mid\!\sim h$ distinguishing the set A of defeasible rules used in the derivation from the context $\mathcal{K}$.*

Definition A.2 *Given $(\mathcal{K}, \Delta)$, and a ground literal h in the language $\mathcal{L}$, we say that a subset A of $\Delta^{\downarrow}$ is an* argument *for h (denoted by $\langle A, h\rangle$) if and only if: 1) $\mathcal{K} \ \cup \ A \mid\!\sim h$, 2) $\mathcal{K} \ \cup \ A \not\mid\!\sim \bot$ and 3) $\nexists A' \subset A$, $\mathcal{K} \cup A' \mid\!\sim h$. Given an argument $\langle A, h\rangle$, we also say that A is an argument for h. A* subargument *of $\langle A, h\rangle$ is an argument $\langle S, j\rangle$ such that $S \subseteq A$.*

Definition A.3 *Two argument $\langle A_1, h_1\rangle$ and $\langle A_2, h_2\rangle$* disagree*, denoted $\langle A_1, h_1\rangle \bowtie \langle A_2, h_2\rangle$, if and only if $\mathcal{K} \cup \{h_1, h_2\} \vdash \bot$.*

Definition A.4 *Given two arguments $\langle A_1, h_1\rangle$ and $\langle A_2, h_2\rangle$, we say that $\langle A_1, h_1\rangle$* counterargues *$\langle A_2, h_2\rangle$, denoted $\langle A_1, h_1\rangle \overset{h}{\otimes\!\!\rightarrow} \langle A_2, h_2\rangle$ iff 1) There exists a subargument $\langle A, h\rangle$ of $\langle A_2, h_2\rangle$ such that $\langle A_1, h_1\rangle \bowtie \langle A, h\rangle$; 2) For every proper subargument $\langle S, j\rangle$ of $\langle A_1, h_1\rangle$, it is not the case that $\langle A_2, h_2\rangle \otimes\!\!\rightarrow \langle S, j\rangle$.*

Definition A.5 *Given two argument $\langle A_1, h_1\rangle$ and $\langle A_2, h_2\rangle$, we say that $\langle A_1, h_1\rangle$* defeats *$\langle A_2, h_2\rangle$ at literal h, denoted $\langle A_1, h_1\rangle \gg_{\mathsf{def}} \langle A_2, h_2\rangle$, if and only if there exists a subargument $\langle A, h\rangle$ of $\langle A_2, h_2\rangle$ such that: $\langle A_1, h_1\rangle$ counterargues $\langle A_2, h_2\rangle$ at the literal h and 1) $\langle A_1, h_1\rangle$ is* strictly more specific[9] *than $\langle A, h\rangle$, or 2) $\langle A_1, h_1\rangle$ is unrelated by specificity to $\langle A, h\rangle$. If $\langle A_1, h_1\rangle \gg_{\mathsf{def}} \langle A_2, h_2\rangle$, we will also say that $\langle A_1, h_1\rangle$ is a* defeater *for $\langle A_2, h_2\rangle$.*

We will accept an argument A as a defeasible reason for a conclusion h if A is a *justification* for h. The acceptance of the original argument A as a justification for h will result from a recursive procedure, in which arguments, counterarguments, counter-counterarguments, and so on, should be taken into account. This leads to a tree structure, called *dialectical tree.* Paths along that tree will be called *argumentation lines*, which can be thought of alternate sequences of *supporting* and *interfering* arguments in a debate.

Definition A.6 *Let $\langle A, h\rangle$ be an argument. A* dialectical tree *for $\langle A, h\rangle$, denoted $\mathcal{T}_{\langle A,\, h\rangle}$, is recursively defined as follows:*

[9] Specificity imposes a partial order on arguments, being used as a preference criterion among them [SCG94]. However, other preference criteria could also be valid.

1. *A single node containing an argument $\langle A, h\rangle$ with no defeaters is by itself a dialectical tree for $\langle A, h\rangle$.*
2. *Suppose that $\langle A, h\rangle$ is an argument with defeaters $\langle A_1, h_1\rangle$,$\langle A_2, h_2\rangle$,...,$\langle A_n, h_n\rangle$. We construct the dialectical tree for $\langle A, h\rangle$, $\mathcal{T}_{\langle A,\, h\rangle}$, by putting $\langle A, h\rangle$ in the root node of it and by making this node the parent node of the roots of the acceptable dialectical trees of $\langle A_1, h_1\rangle$,$\langle A_2, h_2\rangle$, ..., $\langle A_n, h_n\rangle$.*

Definition A.7 *Let $\mathcal{T}_{\langle A,\, h\rangle}$ be a dialectical tree for $\langle A, h\rangle$. Nodes in $\mathcal{T}_{\langle A,\, h\rangle}$ can be recursively labeled as* undefeated nodes *(U-nodes) and* defeated nodes *(D-nodes) as follows: a) Leaves in $\mathcal{T}_{\langle A,\, h\rangle}$ are* U-nodes*; b) Let $\langle B, q\rangle$ be an inner node in $\mathcal{T}_{\langle A,\, h\rangle}$. Then $\langle B, q\rangle$ will be a* U-node *iff every child node of $\langle B, q\rangle$ is a* D-node*. $\langle B, q\rangle$ will be a* D-node *iff it has at least one* U-node *as a child node.*

Definition A.8 *Let $\langle A, h\rangle$ be an argument and let $\mathcal{T}_{\langle A,\, h\rangle}$ be its associated dialectical tree.* [10] *We will say that A is a* justification *for h (or $\langle A, h\rangle$ is a* justification*) iff the root node of $\mathcal{T}_{\langle A,\, h\rangle}$ is a U-node.*

References

[BDKT97] A. Bondarenko, P.M. Dung, R.A. Kowalski, and F. Toni. An abstract, argumentation-theoretic approach to default reasoning. *Artificial Intelligence*, 93:63–101, 1997.

[Che96] Carlos I. Chesñevar. El problema de la inferencia en sistemas argumentativos: Alternativas para su solución *(msc thesis)*. December 1996.

[Dun93] Phan M. Dung. On the Acceptability of Arguments and its Fundamental Role in Nomonotonic Reasoning and Logic Programming. In *Proc. of the 13th. International Joint Conference in Artificial Intelligence (IJCAI), Chambéry, Francia*, 1993.

[Gab96a] Dov Gabbay. *Labelling Deductive Systems.* PhD thesis, Imperial College, London, England, 1996.

[Gab96b] Dov Gabbay. *Labelling Deductive Systems.* PhD thesis, Imperial College, London, England, 1996.

[KT96] Robert A. Kowalski and Francesca Toni. Abstract argumentation. *Artificial Intelligence and Law*, 4(3-4):275–296, 1996.

[Nut88] Donald Nute. Defeasible reasoning: a philosophical analysis in PROLOG. In James H. Fetzer, editor, *Aspects of Artificial Intelligence*, pages 251–288. Kluwer Academic Publishers, 1988.

[Pol87] John L. Pollock. Defeasible Reasoning. *Cognitive Science*, 11:481–518, 1987.

[PS96] Henry Prakken and Giovanni Sartor. A system for defeasible argumentation, with defeasible priorities. In *Proc. of the International Conference on Formal Aspects of Practical Reasoning, Bonn, Germany.* Springer Verlag, 1996.

[10] Actually, dialectical trees should satisfy a number of constraints for being considered *acceptable* dialectical trees (see [SCG94]). That issue, however, exceeds the scope of this paper.

[PV98] Henry Prakken and Gerhard Vreeswijk. Logical systems for defeasible argumentation (to appear). In *Handbook of Philosophical Logic, 2nd. edition.* Gabbay, 1998.

[SCG94] Guillermo R. Simari, Carlos I. Chesñevar, and Alejandro J. García. The role of dialectics in defeasible argumentation. In *Anales de la XIV Conferencia Internacional de la Sociedad Chilena para Ciencias de la Computación.* Universidad de Concepción, Concepción (Chile), November 1994.

[SL92] Guillermo R. Simari and Ronald P. Loui. A Mathematical Treatment of Defeasible Reasoning and its Implementation. *Artificial Intelligence*, 53:125–157, 1992.

[Vre93] Gerard A.W. Vreeswijk. *Studies in Defeasible Argumentation.* PhD thesis, Vrije University, Holland, 1993.