The Role of Dialectics
in Defeasible Argumentation[1] [2]

Guillermo R. Simari   Carlos I. Chesñevar   Alejandro J. García

Grupo de Investigación en Inteligencia Artificial (GIIA)
Departamento de Matemática, Universidad Nacional del Sur
Av.Alem 1253 – (8000) Bahía Blanca, ARGENTINA
FAX: (54) (91) 553933
Phone: (54) (91) 20776 (ext.354)
e-mail: {grs|ccchesne|ccgarcia}@criba.edu.ar

CONTENT AREAS: knowledge representation, nonmonotonic reasoning, argumentative systems

Abstract

In *A Mathematical Treatment of Defeasible Reasoning* [8], or *MTDR*, a clear and theoretically sound structure for a reasoning system was introduced. Since its publication other proposals have been advanced [2, 9, 6, 1], some of them containing valuable observations on this foundation.

This paper presents further developments based on the *MTDR* framework. Two main results are shown. Firstly, several alternative implementations of *MTDR* were based on dialectical concepts, which needed proper formalization. Secondly, the confrontation of the resulting formalism with the above-mentioned work of other researchers has shown that some of the original definitions needed to be honed to avoid *fallacious reasoning*.

The resulting, evolved, system presented here exhibits robust behavior and we contend that it establishes the ideal basis for the implementation of several extensions of defeasible reasoning, which are already being pursued.

1. Introduction

In *A Mathematical Treatment of Defeasible Reasoning* [8], or *MTDR*, a clear and theoretically sound structure for an argument-based reasoning system was introduced. Since its publication other proposals have been advanced [2, 9, 6, 1], some of them containing valuable observations on this foundation. This paper presents further developments based on the *MTDR* framework.

An Argumentative System [8, 4, 2, 9, 1] is a formalization of the process of defeasible reasoning. An argument A for a hypothesis h is a tentative piece of reasoning an agent would be inclined to accept, all things considered, as an explanation for h. In presence of new incoming information, the argument A may lose support or become weakened, and therefore the hypothesis h may no longer be regarded as valid. In that manner nonmonotonicity arises.

Research on many aspects of Argumentative Systems has recently produced some interesting results. In this direction, the works of Loui [3], Simari [8], García [2], Prakken [6] and the research on defeasible argumentation of Vreeswijk [9] are most relevant.

An important aspect of argumentative systems concerns the question of how the hypothesis supported by an argument becomes accepted as part of the current knowledge of the agent. Some approaches are based on multiple extensions [9, 6], while others consider consistency checking. In this respect, the framework presented in *MTDR* accepts an argument A as a defeasible reason for a conclusion h if A is a *justification* for h. The justification process involves the construction of an *acceptable argument* for h from the system's knowledge base. To decide the acceptability of this argument, possible *counterarguments* are generated, which will be also tested for acceptability. Those counterarguments that become *accepted* will be then compared with the original argument using a preference relation (specificity), which defines a partial ordering among arguments. The original argument will be considered a *justification* if it is "better" (*i.e.*, *strictly more specific*) than every acceptable counterargument.

Thus, the characterization of defeasible inference turns out to be conceptually simple. Nevertheless, the confrontation of the ideas presented in *MTDR* with several alternative implementations, as well as the work of the researchers mentioned above, have shown that some important definitions need to be honed to avoid *fallacious reasoning*. To handle such undesirable cases properly, some refinements became necessary.

This paper presents a refined, dialectics-based approach to defeasible argumentation, defined in terms of the basic concepts and definitions of *MTDR*. This approach allowed us to detect the necessity of introducing some refinements and modifications on the original framework, which do not affect the structure of the system itself, but rather turn the process of finding a justification into a more robust one. The resulting, evolved, system shows correct behavior and we contend that it represents the ideal basis for the implementation of different systems for defeasible reasoning, which are already being pursued.

The paper is structured as follows. Section 2 gives an overview of the main ideas and definitions of the *MTDR* argumentative system (a complete description can be found elsewhere [8]). In section 3 we will introduce some new, evolved definitions that characterize the formalism in dialectical terms. Section 4 describes a technical problem, namely the possibility of introducing mutually defeating arguments in the reasoning process. We will show how this problem can appear in more general forms. As a result, we will be able to characterize different patterns of flawed, fallacious argumentation. In section 5, we will see how the former definitions can be strenghtened with the help of dialectical concepts, allowing us to cope with the technical problems presented before. Finally, we will detail the most important conclusions that have been obtained.

2. The *MTDR* framework

We will briefly introduce the main concepts and definitions of the *MTDR* framework (see [8] for further details). 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, called the *context*, 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 $(\mathcal{K}, \Delta)$, called *Defeasible Logic Structure*, 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.

DEFINITION **2.1** Let Γ be a subset of $\mathcal{K} \cup \Delta^{\downarrow}$. A ground literal h is a *defeasible consequence* of Γ, abbreviated $\Gamma \hspace{0.1em}|\!\!\sim h$, if and only if there exists a finite sequence $B_1, \ldots, B_n$ such that $B_n = h$ and for $1 \le i < n$, 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 $\mathcal{K} \cup A \hspace{0.1em}|\!\!\sim h$ distinguishing the set A of defeasible rules used in the derivation from the context $\mathcal{K}$.

DEFINITION **2.2** Given a context $\mathcal{K}$, a set Δ of defeasible rules, and a ground literal h in the language $\mathcal{L}$, we say that a subset A of $\Delta^{\downarrow}$ is an *argument structure* for h in the context $\mathcal{K}$ (denoted by $\langle A, h \rangle_{\mathcal{K}}$, or just $\langle A, h \rangle$) if and only if: 1) $\mathcal{K} \cup A \hspace{0.1em}|\!\!\sim h$, 2) $\mathcal{K} \cup A \hspace{0.1em}|\!\!\not\sim \bot$ and 3) $\nexists A' \subset A$, $\mathcal{K} \cup A' \hspace{0.1em}|\!\!\sim h$. Given an argument structure $\langle A, h \rangle$, 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 $S \subseteq A$.

EXAMPLE **2.1** Given $(\mathcal{K}, \Delta)$, $\mathcal{K} = \{d {\rightarrow} b, d, f, l\}$, $\Delta = \{b \wedge c {\succ\!\!-} h, f {\succ\!\!-} c, l \wedge f {\succ\!\!-} \neg c\}$, we say that $\langle \{f {\succ\!\!-} c, b \wedge c {\succ\!\!-} h\}, h \rangle$ is an argument structure for h.

DEFINITION **2.3** Two argument structures $\langle A_1, h_1 \rangle$ 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 **2.4** Given two argument structures $\langle A_1, h_1 \rangle$ and $\langle A_2, h_2 \rangle$, we say that $\langle A_1, h_1 \rangle$ *counterargues* $\langle A_2, h_2 \rangle$ at literal h, denoted $\langle A_1, h_1 \rangle \overset{h}{\otimes\!\!\longrightarrow} \langle A_2, h_2 \rangle$, if and only if there exists a subargument $\langle A, h \rangle$ of $\langle A_2, h_2 \rangle$ such that $\langle A_1, h_1 \rangle \bowtie \langle A, h \rangle$. $\langle A, h \rangle$ will be also called the *disagreement subargument*. If $\langle A_1, h_1 \rangle \overset{h}{\otimes\!\!\longrightarrow} \langle A_2, h_2 \rangle$, we also say that $\langle A_1, h_1 \rangle$ is a *counterargument* for $\langle A_2, h_2 \rangle$.

DEFINITION **2.5** Let $\mathcal{D} = \{a \in \mathcal{L} : a \text{ is a ground literal and } \mathcal{K} \cup \Delta^{\downarrow} \hspace{0.1em}|\!\!\sim a\}$, and let $\langle A_1, h_1 \rangle$ and $\langle A_2, h_2 \rangle$ be two argument structures. We say that A_1 for h_1 is *strictly more specific than* A_2 for h_2, denoted $\langle A_1, h_1 \rangle \succ_{\mathbf{spec}} \langle A_2, h_2 \rangle$, if and only if
i) $\forall S \subseteq \mathcal{D}$ if $\mathcal{K}_G \cup S \cup A_1 \hspace{0.1em}|\!\!\sim h_1$ and $\mathcal{K}_G \cup S \hspace{0.1em}|\!\!\not\sim h_1$, then $\mathcal{K}_G \cup S \cup A_2 \hspace{0.1em}|\!\!\sim h_2$.
ii) $\exists S \subseteq \mathcal{D}$ such that $\mathcal{K}_G \cup S \cup A_2 \hspace{0.1em}|\!\!\sim h_2$, $\mathcal{K}_G \cup S \hspace{0.1em}|\!\!\not\sim h_2$ and $\mathcal{K}_G \cup S \cup A_1 \hspace{0.1em}|\!\!\not\sim h_1$.

DEFINITION **2.6** Given two argument structures $\langle A_1, h_1 \rangle$ 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 $\langle A_1, h_1 \rangle \gg_{\mathbf{def}} \langle A_2, h_2 \rangle$, 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 than $\langle A, h \rangle$, or
2. $\langle A_1, h_1 \rangle$ is unrelated by specificity to $\langle A, h \rangle$.
If $\langle A_1, h_1 \rangle \gg_{\mathbf{def}} \langle A_2, h_2 \rangle$, we will also say that $\langle A_1, h_1 \rangle$ is a *defeater* for $\langle A_2, h_2 \rangle$. In case (1) $\langle A_1, h_1 \rangle$ will be called a *proper* defeater, and in case (2) a *blocking* defeater. [3]

---

[3] During this research, it was observed that it is convenient to distinguish between these two kinds of defeaters. In *MTDR* [8], this distinction was not drawn.

EXAMPLE **2.2** Given $(\mathcal{K}, \Delta)$ as defined in example 2.1, we have the following relations between arguments.

$$\langle \{l \wedge f \succ\!\!- \neg c\}, \neg c \rangle \bowtie \langle \{f \succ\!\!- c\}, c \rangle$$
$$\langle \{l \wedge f \succ\!\!- \neg c\}, \neg c \rangle \overset{c}{\otimes\!\!\rightarrow} \langle \{f \succ\!\!- c, b \wedge c \succ\!\!- h\}, h \rangle$$
$$\langle \{l \wedge f \succ\!\!- \neg c\}, \neg c \rangle \succ_{\mathbf{spec}} \langle \{f \succ\!\!- c\}, c \rangle$$
$$\langle \{l \wedge f \succ\!\!- \neg c\}, \neg c \rangle \gg_{\mathbf{def}} \langle \{f \succ\!\!- c, b \wedge c \succ\!\!- h\}, h \rangle$$

3. Defeasible argumentation in terms of dialectical trees

We will accept an argument A as a defeasible reason for a conclusion h if A is a *justification* for h. The justification process involves the construction of an acceptable argument for h. To decide the acceptability of an argument A for a literal h, its associated counterarguments B_1, $B_2, \ldots B_k$ will be obtained, each of them being a (defeasible) reason for rejecting A. If some B_i is supported on "better" (or unrelated) evidence than A, then B_i will be a candidate for defeating A. As we have mentioned before, specificity is the preference criterion for deciding between two conflicting arguments. When specificity cannot decide, a blocking situation occurs.

Since counterarguments are also arguments, the former analysis should be in turn carried out on them. Now, B_i will defeat A unless there exists an argument C_j (which corresponds to one of the counterarguments $C_1, C_2, \ldots C_r$ associated to B_i) that defeats B_i. In that case, we will be forced to reject B_i, and hence our original argument A would be *reinstated*. If it turns out that there is a defeater D_k among the counterarguments of C_j, then B_i would be reinstated as defeater for A. Thus, the acceptance of an argument A will result from a recursive procedure, in which arguments, counterarguments, counter-counterarguments, and so on, should be taken into account. The above description leads in a natural way to the use of trees to organize that dialectical analysis.

In order to accept an argument A for a conclusion h, a tree structure can be generated. Every inner node in this tree will represent a defeater (proper or blocking), and the root of the tree will correspond to the original argument A. Nodes in this tree can in turn be recursively labeled as defeated or undefeated nodes. If all children nodes of the root turn out to be labeled as defeated, we say that A is an *acceptable argument*. The procedure just described closely resembles an argumentation between two parties, *i.e.*, it is a *dialectical* process. Next, we will formalize the concepts examined so far.

DEFINITION **3.1** Let $\langle A, h \rangle$ be an argument structure. A *dialectical tree* for $\langle A, h \rangle$, denoted $\mathcal{T}_{\langle A, h \rangle}$, is recursively defined as follows:

1. A single node containing an argument structure $\langle A, h \rangle$ with no defeaters (proper or blocking) is by itself a dialectical tree for $\langle A, h \rangle$. This node is also the root of the tree.
2. Suppose that $\langle A, h \rangle$ is an argument structure with defeaters (proper or blocking) $\langle A_1, h_1 \rangle$, $\langle A_2, h_2 \rangle$, $\ldots$, $\langle A_n, h_n \rangle$. 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 dialectic trees for $\langle A_1, h_1 \rangle$, $\langle A_2, h_2 \rangle$, $\ldots$, $\langle A_n, h_n \rangle$, *i.e.*, $\mathcal{T}_{\langle A_1, h_1 \rangle}$, $\mathcal{T}_{\langle A_2, h_2 \rangle}$, $\ldots$, $\mathcal{T}_{\langle A_n, h_n \rangle}$.

According to the preceding definition, every argument has an associated dialectical tree. Nodes in this tree correspond to directly related arguments via the defeat relation. This characterization leads us to a labeling procedure, after which we can conclude whether the root of the dialectical tree corresponds indeed to an acceptable argument.

DEFINITION **3.2** Let $\mathcal{T}_{\langle A, h \rangle}$ be a dialectical tree for an argument structure $\langle A, h \rangle$. The nodes of $\mathcal{T}_{\langle A, h \rangle}$ can be recursively labeled as *undefeated nodes* (U-nodes) and *defeated nodes* (D-nodes) as follows:

1. Leaves of $\mathcal{T}_{\langle A, h \rangle}$ are *U-nodes*.
2. Let $\langle B, q \rangle$ be an inner node of $\mathcal{T}_{\langle A, h \rangle}$. Then $\langle B, q \rangle$ will be an *U-node* iff every child of $\langle B, q \rangle$ is a *D-node*. $\langle B, q \rangle$ will be a *D-node* iff it has at least an *U-node* as a child.

This definition suggests a bottom-up labeling procedure, through which we are able to determine if the root of a dialectical tree turns out to be labeled as defeated or undefeated.

DEFINITION **3.3** Let $\langle A, h \rangle$ be an argument structure, and let $\mathcal{T}_{\langle A, h \rangle}$ be its associated dialectical tree. We will say that A is a *justification* for h (or simply $\langle A, h \rangle$ is a *justification*) iff the root node of $\mathcal{T}_{\langle A, h \rangle}$ is an U-node.

According to this definition, an argumentative knowledge-based system has four possible answers for a given query h.

- YES, if there is a justification $\langle A, h \rangle$.
- NO, if for every possible argument structure $\langle A, h \rangle$, there exists a justification for at least one proper defeater of $\langle A, h \rangle$.
- UNKNOWN, if there exists no argument structure $\langle A, h \rangle$.
- UNDEFINED, if for every possible argument structure $\langle A, h \rangle$, there are no proper defeaters for $\langle A, h \rangle$, but there exists at least one blocking defeater for $\langle A, h \rangle$.

Now we will introduce two additional concepts, already suggested in [8], which will prove to be useful in what follows.

DEFINITION **3.4** Let $\langle A_0, h_0 \rangle$ be an argument structure, and let $\mathcal{T}_{\langle A_0, h_0 \rangle}$ be its associated dialectical tree. Then every path λ in $\mathcal{T}_{\langle A_0, h_0 \rangle}$ from the root $\langle A_0, h_0 \rangle$ to a leaf $\langle A_n, h_n \rangle$, denoted $\lambda = [\, \langle A_0, h_0 \rangle, \langle A_1, h_1 \rangle, \langle A_2, h_2 \rangle, \ldots, \langle A_n, h_n \rangle]$, constitutes an *argumentation line* for $\langle A_0, h_0 \rangle$.

DEFINITION **3.5** Let $\mathcal{T}_{\langle A_0, h_0 \rangle}$ be a dialectical tree, and let $\lambda = [\, \langle A_0, h_0 \rangle, \langle A_1, h_1 \rangle, \langle A_2, h_2 \rangle, \ldots, \langle A_n, h_n \rangle]$ be an argumentation line for $\langle A_0, h_0 \rangle$. Then every $\langle A_i, h_i \rangle$ in λ can be labeled as a *supporting* or *interfering* argument as follows

1. $\langle A_0, h_0 \rangle$ is a supporting argument in λ, and
2. If $\langle A_i, h_i \rangle$ is a supporting (interfering) argument in λ, then $\langle A_{i+1}, h_{i+1} \rangle$ is an interfering (supporting) argument in λ.

We will denote as S_λ and I_λ the set of all supporting and interfering arguments in λ, respectively.

As we can see from this definition, an argumentation line λ can now be thought of as an alternate sequence of supporting and interfering arguments as in any ordered debate.

4. Fallacious Argumentation

4.1. Reciprocal Defeaters. Our characterization of the notion of justification in terms of trees can help us to understand why the associated procedure will eventually terminate. The answer seems to be simple: since we start with a finite knowledge base, and arguments must be non-redundant because of the minimality condition, we can only generate a finite number of them. An argument can only have a finite number of tentative defeaters, *i.e.*, the number of nodes at any level of the dialectical tree is finite. Now, any argumentation line could at most involve every possible argument structure. Even in that case, its length would be finite, and the justification procedure should terminate.

Nevertheless, the correctness of this last statement rests on the assumption that there can be no loops in the tree. And this assumption is flawed, as we can see in the following example mentioned by Prakken in [6]:

EXAMPLE **4.1** Let $\mathcal{K} = \{a, c\}$ and $\Delta = \{a \succ \neg b, c \succ \neg d, \neg b \wedge c \succ d, a \wedge \neg d \succ b\}$ be a defeasible logic structure. Given the argument structures $\langle A, d \rangle = \langle \{a \succ \neg b, \neg b \wedge c \succ d\}, d \rangle$ and $\langle B, b \rangle = \langle \{c \succ \neg d, a \wedge \neg d \succ b\}, b \rangle$, then $\langle A, d \rangle$ defeats $\langle B, b \rangle$, and $\langle B, b \rangle$ defeats $\langle A, d \rangle$.

The situation in our example can be graphically represented as follows. [4]
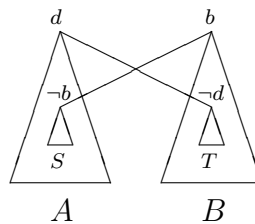
In this example, we are dealing with some kind of *reciprocal* defeaters. However, such a situation makes intuitively no sense: $\langle A, d \rangle$ claims to defeat $\langle B, b \rangle$, because the latter contains a subargument structure $\langle T, \neg d \rangle$. However, $\langle B, b \rangle$ is a defeater for a subargument structure $\langle S, \neg b \rangle$ of $\langle A, d \rangle$. As a consequence, a cyclic pattern may now arise as we generate the dialectical tree for $\langle A, d \rangle$, since we are facing an infinite argumentation line $[\langle A, d \rangle, \langle B, b \rangle, \langle A, d \rangle, \langle B, b \rangle, \ldots]$ (*i.e.*, $\langle A, d \rangle$ is defeated by $\langle B, b \rangle$, $\langle B, b \rangle$ is defeated by $\langle A, d \rangle$, and this pattern repeats indefinitely). The inference mechanism would loop forever. Therefore, there are both intuitive and technical reasons for rejecting the presence of such "reciprocal" defeaters in an argumentation line.

---

[4]Argument structures can be seen as trees and therefore can be graphically represented as triangles abstracting a tree shape. The upper vertex of the triangle will be labeled with the argument's conclusion, and the name of the argument will be associated to the triangle itself. Subarguments will be represented as smaller triangles inside a big one, which corresponds to the main argument at issue. Those arguments in disagreement will be connected with a straight line.

4.2. Circularity. In order to avoid reciprocal defeaters, a first approach would be to analyze the internal structure of the arguments involved, and find out if they defeat each other. Nevertheless, this does not suffice for avoiding undesirable situations, since there still exists the possibility of constructing cycles of any length in an argumentation line. In this respect, Loui observes [3] that, according to his formalism, the defeat relation is not acyclic. Recently, Vreeswijk [9] makes a similar remark about Loui's formalism: "Justification can be circular, ..., it is possible to construct odd cycles of arguments, in which each argument interferes with the next argument, so that, indirectly, each argument is self-defeating." The same situation arises in the $MTDR$ framework. Consider the following example:

EXAMPLE **4.2** Let $\mathcal{K} = \{e_1, e_2, e_3\}$ and $\Delta = \{e_1 \succ\!\!- p,\ e_2 \succ\!\!- q,\ e_3 \succ\!\!- r,\ e_3 \wedge p \succ\!\!- \neg r,$ $e_1 \wedge q \succ\!\!- \neg p,\ e_2 \wedge r \succ\!\!- \neg q\}$ be a defeasible logic structure. Given the argument structures $\langle A, \neg r \rangle = \langle \{e_1 \succ\!\!- p, e_3 \wedge p \succ\!\!- \neg r\}, \neg r \rangle$, $\langle B, \neg p \rangle = \langle \{e_2 \succ\!\!- q, e_1 \wedge q \succ\!\!- \neg p\}, \neg p \rangle$, $\langle C, \neg q \rangle = \langle \{e_3 \succ\!\!- r, e_2 \wedge r \succ\!\!- \neg q\}, \neg q \rangle$, we have that $\langle B, \neg p \rangle$ defeats $\langle A, \neg r \rangle$, $\langle C, \neg q \rangle$ defeats $\langle B, \neg p \rangle$ and $\langle A, \neg r \rangle$ defeats $\langle C, \neg q \rangle$. Pictorially:
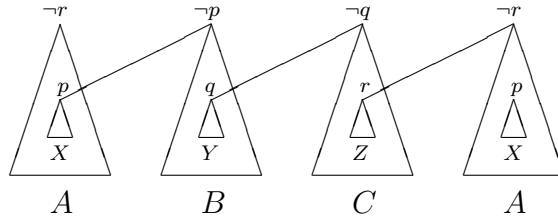

However, the ability of conceptualizing what a justification is in terms of supporting and interfering arguments will help us to detect how the fallacies come to be. We claim that the fact of accepting the argumentation line in the example as valid should be blamed for causing cycles rather than the justification procedure itself.

In the example above, the argument structure $\langle B, \neg p \rangle$ defeats $\langle A, \neg r \rangle$, but $\langle C, \neg q \rangle$ defeats $\langle B, \neg p \rangle$ allowing $\langle A, \neg r \rangle$ to be reinstated. $\langle C, \neg q \rangle$ is then defeated by $\langle A, \neg r \rangle$. We must be aware of the fact that $\langle C, \neg q \rangle$ was intended as a supporting argument for $\langle A, \neg r \rangle$, not as an interfering one, and in this way $\langle A, \neg r \rangle$ becomes an interference argument for itself. Hence, it is clear that there should exist a certain *agreement* among supporting arguments in any argumentation line. Note that the same situation arises for interfering arguments. This internal coherence is essential to the dialectical process of argumentation: supporting (interfering) arguments should not contradict each other. However, this requirement cannot be inferred from the definitions of counterargument and defeat in $MTDR$, so that they must be fixed.

Finally, we will consider a generalization of example 4.1, that is, an even cycle of defeaters in an argumentation line.

EXAMPLE **4.3** *Circular* argumentation. Consider a defeasible logic structure, with $\mathcal{K} = \{e_1, e_2, e_3, e_4\}$, and $\Delta = \{e_1 \succ\!\!- p, e_2 \succ\!\!- q, e_3 \succ\!\!- s, e_4 \succ\!\!- r, e_4 \wedge p \succ\!\!- \neg r, e_1 \wedge q \succ\!\!- \neg p, e_2 \wedge s \succ\!\!- \neg q, e_3 \wedge r \succ\!\!- \neg s\}$. Given the argument structures $\langle A, \neg r \rangle = \langle \{e_1 \succ\!\!- p, e_4 \wedge p \succ\!\!- \neg r\}, \neg r \rangle$, $\langle B, \neg p \rangle = \langle \{e_2 \succ\!\!- q, e_1 \wedge q \succ\!\!- \neg p\}, \neg p \rangle$, $\langle C, \neg q \rangle = \langle \{e_3 \succ\!\!- s, e_2 \wedge s \succ\!\!- \neg q\}, \neg q \rangle$, $\langle D, \neg s \rangle = \langle \{e_4 \succ\!\!- r, e_3 \wedge r \succ\!\!- \neg s\}, \neg s \rangle$, we have that $\langle B, \neg p \rangle$ defeats $\langle A, \neg r \rangle$, $\langle C, \neg q \rangle$ defeats $\langle B, \neg p \rangle$, $\langle D, \neg s \rangle$ defeats $\langle C, \neg q \rangle$, and $\langle A, \neg r \rangle$ defeats $\langle D, \neg s \rangle$.
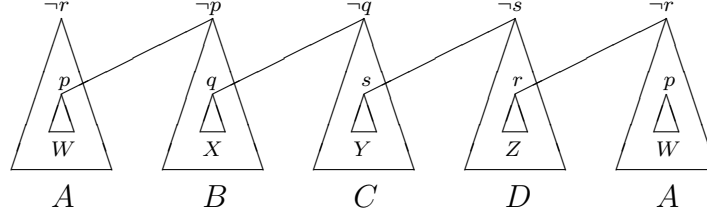
In this case, supporting (interfering) arguments are not in conflict. However, the inference mechanism takes us again to an endless argumentation line. Here, the fallacy lies on how the argumentation line was constructed. We must somehow be able to assure that every step along a given argumentation line will be *progressive* (see [7]).

In the preceding example, the fourth argument $\langle D, \neg s \rangle$ contains a subargument $\langle Z, r \rangle$ which is again counterarguing the first argument $\langle A, \neg r \rangle$. This situation is often described as "circular argumentation" in the literature. The original argument $\langle A, \neg r \rangle$ is being indirectly interfered by another argument $\langle D, \neg s \rangle$, but $\langle D, \neg s \rangle$ was constructed on the assumption of the falsehood of $\langle A, \neg r \rangle$'s conclusion. This kind of argumentation is not dialectically sound, and as such should be rejected. The reasons for doing so are similar to those we argued against reciprocal defeaters. Once again, this situation cannot be properly captured within *MTDR*'s framework, since it lies beyond the scope of the existing relations among arguments.

5. Avoiding Fallacies

In this section, we will show how to fix the problems discussed before. We claim that in order to do so, it suffices to introduce some additional restrictions on the original definitions, without modifying any conceptual feature in the framework. First, we will define a relation between arguments, already introduced in *MTDR*, which will allow us to accept two arguments simultaneously, without taking the risk that they contradict each other.

DEFINITION **5.1** Given two argument structures $\langle A_1, h_1 \rangle$ and $\langle A_2, h_2 \rangle$ we will say that they are *concordant* iff $\mathcal{K} \cup A_1 \cup A_2 \not\vdash \bot$. In general, a family of argument structures $\{\langle A_i, h_i \rangle\}_{i=1}^{n}$ is concordant iff $\mathcal{K} \cup \bigcup_{i=1}^{n} A_i \not\vdash \bot$.

It has been shown[8] that the family of all subarguments of a given argument structure is pairwise concordant.

We must now consider how to avoid reciprocal defeaters (*i.e.*, those pairs of arguments that defeat each other). Actually, the defeat relation is a subset of the counterargument relation [8]. This suggests that we need to take a closer look to the characterization of what a counterargument should be. We will introduce now a new, stronger definition of counterargument.

DEFINITION **5.2** 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 $\langle A_1, h_1 \rangle \overset{h}{\otimes\!\!\longrightarrow} \langle A_2, h_2 \rangle$ 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 $\langle A_2, h_2 \rangle \overset{}{\otimes\!\!\longrightarrow} \langle S, j \rangle$.

Proposition 5.1 Given two argument structures $\langle A_1, h_1 \rangle$ and $\langle A_2, h_2 \rangle$, there exists an effective procedure for determining if $\langle A_1, h_1 \rangle \overset{h}{\otimes\!\!\longrightarrow} \langle A_2, h_2 \rangle$.

Nevertheless, strengthening the notion of counterargumentation does not suffice for eliminating cyclic patterns occurring in an argumentation line. As we have already seen, it may still be the case that an argumentation line contains contradictory supporting (interfering) arguments (odd cycles), or that it does not convey any progress in the process of finding a justification (even cycles). To cope with these problems, we will strengthen the notion of argumentation line as follows.

DEFINITION **5.3** Let $\lambda = [\langle A_0, h_0 \rangle, \langle A_1, h_1 \rangle, \langle A_2, h_2 \rangle, \dots, \langle A_n, h_n \rangle]$ be an argumentation line. Then, λ will be called an *acceptable argumentation line* iff
  1. Supporting (interfering) arguments in λ are concordant pairwise, *i.e.*, $\mathcal{K} \cup A_i \cup A_j \nvdash \bot$, for every $\langle A_i, h_i \rangle, \langle A_j, h_j \rangle \in S_\lambda$ (I_λ).
  2. Let $\langle A_i, h_i \rangle$ be an argument structure in S_λ (I_λ). There is no argument $\langle A_j, h_j \rangle$ in I_λ (S_λ), such that $i < j$ and $\langle A_i, h_i \rangle$ defeats $\langle A_j, h_j \rangle$.

The first condition causes supporting (interfering) arguments in an argumentation line to be consistent among themselves. The second condition prevents circularity, forcing interfering arguments not to be defeated by previous arguments in a given argumentation line. Interfering arguments must be constructed considering which arguments have been already offered. It can be shown that these two conditions suffice for avoiding fallacious argumentation. From these definitions we can now give the central definition in the framework.

DEFINITION **5.4** Let $\langle A, h \rangle$ be an argument structure. An *acceptable dialectical tree* for $\langle A, h \rangle$, denoted $\mathcal{T}_{\langle A, h \rangle}$, is recursively defined as follows:
  1. A single node containing an argument structure $\langle A, h \rangle$ with no defeaters (proper or blocking) is by itself an acceptable dialectical tree for $\langle A, h \rangle$. This node is also the root of the tree.
  2. Suppose that $\langle A, h \rangle$ is an argument structure with defeaters (proper or blocking) $\langle A_1, h_1 \rangle$, $\langle A_2, h_2 \rangle$, $\dots$, $\langle A_n, h_n \rangle$. 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 $\langle A_1, h_1 \rangle$, $\langle A_2, h_2 \rangle$, $\dots$, $\langle A_n, h_n \rangle$, *i.e.*, $\mathcal{T}_{\langle A_1, h_1 \rangle}$, $\mathcal{T}_{\langle A_2, h_2 \rangle}$, $\dots$, $\mathcal{T}_{\langle A_n, h_n \rangle}$. If an unacceptable argument line gets formed, during the construction of this tree, it suffices to clip the subtree rooted in the offending argument that violates a condition in the definition of acceptable argumentation line.

In particular, introducing the definition of acceptable dialectical tree prompts the following lemma.

LEMMA **5.2** *Let $\langle A, h \rangle$ be an arbitrary argument structure. Then its acceptable dialectical tree $\mathcal{T}_{\langle A, h \rangle}$ is finite.*

6. Conclusions

MTDR's formalism proved to be robust enough for introducing significant refinements and modifications on the inference mechanism without affecting the structure of the framework itself. Thus, some flawed patterns of reasoning concerning the justification procedure could be eliminated, by just constraining the inference mechanism to a certain extent.

After imposing some restrictions on the construction of argumentation lines, we arrived at the definition of acceptable argumentation line, through which we were able to characterize an

alternate sequence of arguments and defeaters as *dialectically* sound. In this respect, dialectics provided the adequate frame for modeling and understanding anomalous situations, allowing us to recognize those aspects of the justification procedure that had to be modified.

The implementation issues deserve separate study. Computational aspects concerning the new concepts and definitions introduced in this paper are being worked on at the time.

7. References

[1] P. M. Dung. On the Acceptability of Arguments and its Fundamental Role in Nomonotonic Reasoning and Logic Programming. In *Thirteenth International Joint Conference in Artificial Intelligence, IJCAI'93*, Chambéry, France, 1993.

[2] A. J. García, C. I. Chesñevar, and G. R. Simari. Making Argument Systems Computationally Attractive. In *Proceedings of the XIII International Conference of the Chilean Society for Computer Science*, October 1993.

[3] R. P. Loui. Defeat Among Arguments: a system of defeasible inference. *Computational Intelligence*, **3** (3), 100-106, 1987.

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

[5] D. L. Poole. On the Comparison of Theories: Preferring the Most Specific Explanation. In *Proceedings of the Ninth International Joint Conference on Artificial Intelligence*, 144–147, IJCAI, 1985.

[6] H. Prakken. Logical Tools for Modelling Legal Arguments (Ph.D. Thesis). *Vrije University, Amsterdam, Holland*, 1993.

[7] N. Rescher. Dialectics: A Controversy-Oriented Approach to the Theory of Knowledge. *Ed. State University of New York Press, Albany*, 1977.

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

[9] G. A. W. Vreeswijk. Studies in Defeasible Argumentation (Ph.D. Thesis). *Vrije University, Amsterdam, Holland*, 1993.