An Axiomatic Account of Formal Argumentation

Martin Caminada
Institute of Information and Computing Sciences
Universiteit Utrecht, Utrecht Netherlands
martinc@cs.uu.nl

Leila Amgoud
IRIT - CNRS
118, route de Narbonne, 31062 Toulouse, France
amgoud@irit.fr

Abstract

Argumentation theory has become an important topic in the field of AI. The basic idea is to construct arguments in favor and against a statement, to select the "acceptable" ones and, finally, to determine whether the statement can be accepted or not. Dung's elegant account of abstract argumentation (Dung 1995) may have caused some to believe that defining an argumentation formalism is simply a matter of determining how arguments and their defeat relation can be constructed from a given knowledge base. Unfortunately, things are not that simple; many straightforward instantiations of Dung's theory can lead to very unintuitive results, as is discussed in this paper.

In order to avoid such anomalies, in this paper we are interested in defining some rules, called *rationality postulates* or *axioms*, that govern the well definition of an argumentation system. In particular, we define two important rationality postulates that any system should satisfy: the *consistency* and the *closeness* of the results returned by that system. We then provide a relatively easy way in which these quality postulates can be warranted by our argumentation system.

Introduction

Argumentation has become an Artificial Intelligence keyword for the last fifteen years, especially in sub-fields such as non monotonic reasoning, inconsistency-tolerant reasoning, multiple-source information systems, natural language processing and human-machine interface also in connection with multi-agents systems (Amgoud & Cayrol 2002; Prakken & Sartor 1997; Rahwan *et al.* 2004; Gómez & Chesñevar 2003).

Argumentation is a promising model for reasoning. It follows three steps: i) to construct arguments in favor and against a statement, ii) to select the "acceptable" ones and, finally, iii) to determine whether the statement can be accepted or not. It may also be considered as a different method for handling uncertainty. The basic idea behind argumentation is that it should be possible to say more about the certainty of a particular fact than the certainty quantified with a degree in [0, 1]. In particular, it should be possible to assess the reason why a fact holds, in the form of arguments, and combine these arguments to evaluate the certainty. Indeed, the process of combination may be viewed as a kind of

reasoning about arguments themselves in order to determine the most acceptable of them.

One of the most abstract argumentation systems is Dung's one. It has been shown that several formalisms for non monotonic reasoning can be expressed in terms of that argumentation system (Dung 1995). Since its original formulation, Dung's system has become very popular and different instantiations of it have been defined. This may have caused some to believe that defining an argumentation formalism is simply a matter of defining how arguments and their defeat relation can be constructed from a knowledge base. Unfortunately, things are not that simple. Some instantiations of Dung's system such as the Prakken and Sartor's system (Prakken & Sartor 1997) can lead to very unintuitive results. The same problem occurs also in systems which are not based on the Dung's system, such as (García & Simari 2004).

In order to avoid such anomalies, the aim of this paper is twofold: on the one hand, like in the field of belief revision, where the well-known AGM-postulates serve as general properties a system for belief revision should fulfill, we are interested in defining some *principles* (called here *rationality postulates* or *axioms*) that any argumentation system should fulfill. These postulates will govern the well definition of an argumentation system and will ensure the correctness of its results. In this paper we focus particularly on two important postulates: the *closeness* and the *consistency* of the results that an argumentation system may return. These postulates are violated in systems such as (Prakken & Sartor 1997; Governatori *et al.* 2004; García & Simari 2004). On the other hand, we study various ways in which these postulates can be warranted in the argumentation system developed in (Amgoud *et al.* 2004).

This paper is organized as follows: in the second section we introduce an argumentation system which is an instantiation of the Dung system. Through this system, we will illustrate what is going wrong in most systems such as (Prakken & Sartor 1997; Governatori *et al.* 2004; García & Simari 2004). In the third section, we introduce the two rationality postulates. In the fourth section, we propose two possible solutions that warrant the satisfaction of the postulates. Finally, the fifth section is devoted to some concluding remarks and perspectives.

Note that for the lack of space, all the proofs can be found

in a technical report written by the authors of this paper.

An abstract argumentation system

In what follows, we present an instantiation of the Dung's system developed in (Amgoud *et al.* 2004). For the sake of simplicity, priorities are not handled here. We will assume arguments to consist of *trees* of strict and defeasible rules. This choice is somewhat arbitrary, it would be equally possible to define arguments as *lists* of *strict* and *defeasible* rules, and still obtaining the same basic problems and possible solutions as is the case for arguments as trees. In what follows, $\mathcal{L}$ is a set of literals and $\mathcal{K}$ is a subset of $\mathcal{L}$. We assume the availability of a function "$-$", which works with literals, such that $-p = \neg p$ and $--p = p$ (where p is an atomic proposition). $\mathcal{S}$ is a set of strict rules of the form $\phi_1, \ldots, \phi_n \rightarrow \psi$ (meaning that if $\phi_1, \ldots, \phi_n$ hold, then *without exception* it holds that ψ) and $\mathcal{D}$ is a set of defeasible rules of the form $\phi_1, \ldots, \phi_n \Rightarrow \psi$ (meaning that if $\phi_1, \ldots, \phi_n$ hold, then it *usually* holds that ψ) with ϕ_i, ψ elements of $\mathcal{L}$. From a defeasible theory $(\mathcal{K}, \mathcal{S}, \mathcal{D})$, arguments can be built as follows:

Definition 1 (Argument). *Let* $(\mathcal{K}, \mathcal{S}, \mathcal{D})$ *be a defeasible theory. An* argument *is defined as follows:*

Premises: *if* $\phi \in \mathcal{K}$ *then* ϕ *is an argument (A) with:*
- $\text{Conc}(A) = \phi$
- $\text{StrictRules}(A) = \emptyset$
- $\text{DefRules}(A) = \emptyset$
- $\text{SubArgs}(A) = \{A\}$

Strict construction: *if* $A_1, \ldots, A_n$ $(n \geq 0)$ *are arguments and* $\mathcal{S}$ *contains a strict rule* $\text{Conc}(A_1), \ldots, \text{Conc}(A_n) \rightarrow \psi$ *then* $A_1, \ldots A_n \rightarrow \psi$ *(A) is an argument with:*
- $\text{Conc}(A) = \psi$
- $\text{StrictRules}(A) = \text{StrictRules}(A_1) \cup \ldots \cup \text{StrictRules}(A_n) \cup \{\phi_1, \ldots, \phi_n \rightarrow \psi\}$
- $\text{DefRules}(A) = \text{DefRules}(A_1) \cup \ldots \cup \text{DefRules}(A_n)$
- $\text{SubArgs}(A) = \text{SubArgs}(A_1) \cup \ldots \cup \text{SubArgs}(A_n) \cup \{A\}$

Defeasible construction: *if* $A_1, \ldots, A_n$ $(n \geq 0)$ *are arguments and* $\mathcal{D}$ *contains a defeasible rule* $\text{Conc}(A_1), \ldots, \text{Conc}(A_n) \Rightarrow \psi$ *then* $A_1, \ldots A_n \Rightarrow \psi$ *(A) is an argument with:*
- $\text{Conc}(A) = \psi$
- $\text{StrictRules}(A) = \text{StrictRules}(A_1) \cup \ldots \cup \text{StrictRules}(A_n)$
- $\text{DefRules}(A) = \text{DefRules}(A_1) \cup \ldots \cup \text{DefRules}(A_n) \cup \{\phi_1, \ldots, \phi_n \Rightarrow \psi\}$
- $\text{SubArgs}(A) = \text{SubArgs}(A_1) \cup \ldots \cup \text{SubArgs}(A_n) \cup \{A\}$

An argument A is strict *iff* $\text{DefRules}(A) = \emptyset$. $\mathcal{A}$ *denotes the set of all arguments that can be built from* $(\mathcal{K}, \mathcal{S}, \mathcal{D})$.

Since the knowledge bases are generally inconsistent, the arguments may be conflicting. The first kind of conflicts concerns the conclusions of the arguments. Indeed, two arguments may conflict with each other if they support contradictory conclusions.

Definition 2 (Rebutting). *Let* $A, B \in \mathcal{A}$. A *rebuts* B *iff* $\exists A' \in \text{SubArgs}(A)$ *with* $\text{Conc}(A') = \psi$ *and* $\exists$ *a non-strict argument* $B' \in \text{SubArgs}(B)$ *with* $\text{Conc}(B') = -\psi$.

Two arguments may also conflict if one of them uses a defeasible rule of which the applicability is disputed by the other argument. In the following definition, $\lceil . \rceil$ stands for the objectivation operator (Pollock 1995), which converts a meta-level expression into an object-level expression (in our case: a literal).

Definition 3 (Undercutting). *Let* $A, B \in \mathcal{A}$. A *undercuts* B *iff* $\exists B_1, \ldots, B_n \Rightarrow \psi \in \text{SubArgs}(B)$ *and* $\exists A' \in \text{SubArgs}(A)$ *with* $\text{Conc}(A') = \neg \lceil \text{Conc}(B_1), \ldots, \text{Conc}(B_n) \Rightarrow \psi \rceil$.

The two above relations are brought together in a unique relation of "defeat". Formally:

Definition 4 (Defeat). *Let* A *and* B *be two arguments.* A *defeats* B *iff* A *rebuts* B *or* A *undercuts* B.

The second step of an argumentation process consists of computing the *acceptable* arguments. Dung has defined different acceptability semantics.

Definition 5 (Defence/conflict-free). *Let* $S \subseteq \mathcal{A}$.
- S defends *an argument A iff each argument that defeats A is defeated by some argument in S.*
- S *is* conflict-free *iff there exist no A_i, A_j in S such that A_i defeats A_j.*

Definition 6 (Acceptability semantics). *Let S be a conflict-free set of arguments and let $F: 2^{\mathcal{A}} \to 2^{\mathcal{A}}$ be a function such that $F(S) = \{A \mid A$ is defended by $S\}$.*
- S *is* admissible *iff $S \subseteq F(S)$.*
- S *is a* complete *extension iff $S = F(S)$.*
- S *is a* preferred *extension iff S is a maximal (w.r.t set $\subseteq$) complete extension.*
- S *is a* grounded *extension iff it is the smallest (w.r.t set $\subseteq$) complete extension.*

Note that there is only one grounded extension. It contains all the arguments which are not defeated and also the arguments which are defended directly or indirectly by non-defeated arguments.

The last step of an argumentation process consists of determining, among all the conclusions of the different arguments, the "good" ones called *justified conclusions*. Let Output denote this set of justified conclusions. One way of defining Output is to consider the conclusions which are supported by at least one argument in each extension.

Definition 7 (Justified conclusions). *Let $(\mathcal{A}, \text{defeat})$ be an argumentation system and $\{E_1, \ldots, E_n\}$ be its set of extensions (under a given semantics).* $\text{Output} = \{\psi | \forall E_i, \exists A \in E_i$ *such that* $\text{Conc}(A) = \psi\}$.

Example 1. *Let* $\mathcal{K} = \{a; d\}$, $\mathcal{S} = \emptyset$, $\mathcal{D} = \{a \Rightarrow b; d \Rightarrow \neg b\}$. *The following arguments can be constructed:*

$A_1 : [a]$ $A_3 : [A_1 \Rightarrow b]$
$A_2 : [d]$ $A_4 : [A_2 \Rightarrow \neg b]$

Argument A_3 defeats A_4 and vice versa. However, the arguments A_1 and A_2 do not have any defeaters. Thus, they belong to any extension. Consequently, a and b will be considered as justified conclusions.

Let us now consider the following *interesting* example.

Example 2 (Married John). *Let* $\mathcal{K} = \{wr; go\}$, $\mathcal{S} = \{b \to \neg hw; m \to hw\}$ *and* $\mathcal{D} = \{wr \Rightarrow m; go \Rightarrow b\}$ *with:* $wr =$ *"John wears something that looks like a wedding ring", m = "John is married", hw = "John has a wife", go = "John often goes out until late with his friends", b = "John is a bachelor". The following arguments can be constructed:*

$A_1 : [wr]$ $\qquad$ $A_4 : [A_2 \Rightarrow b]$
$A_2 : [go]$ $\qquad$ $A_5 : [A_3 \to hw]$
$A_3 : [A_1 \Rightarrow m]$ $\qquad$ $A_6 : [A_4 \to \neg hw]$

The argument A_5 defeats the argument A_6 and vice versa. However, the arguments A_1, A_2, A_3 and A_4 do not have any defeaters. Thus, they belong to the grounded extension $\{A_1, A_2, A_3, A_4\}$. Consequently, the set of justified conclusions is Output $= \{b; m; wr; go\}$. *This means that John is both married (m) and a bachelor (b), even though from the content of the knowledge base (the strict rules $m \to hw$ and $b \to \neg hw$) it should be clear that b and m cannot hold together.*

Example 2 shows clearly that counter-intuitive conclusions may be inferred from a base using the above argumentation system. As a consequence, the *closure under the set of strict rules* of the set of conclusions may be inconsistent. In the above example, the closure under the strict rules of $\{b; m; wr; go\}$ is $\{b; m; wr; go; hm; \neg hm\}$ which is directly inconsistent. One may argue that the grounded extension is "indirectly" inconsistent. Moreover, the set of conclusions may also be not *closed under the set of strict rules*. For instance, hw and $\neg hw$ which are in the closure of the set of justified conclusions are not in the set itself.

The above example returns counter-intuitive results in our reference formalism, as well as in (Governatori *et al.* 2004). It should be noticed, however, that the problem is not limited to these particular two systems. The following example, for instance, is going wrong in (Prakken & Sartor 1997; García & Simari 2004; Governatori *et al.* 2004).

Example 3. *Let* $K = \{a; d; g\}$, $\mathcal{S} = \{b, c, e, f \to \neg g\}$ *and* $\mathcal{D} = \{a \Rightarrow b; b \Rightarrow c; d \Rightarrow e; e \Rightarrow f\}$.
The following arguments can be constructed:

$A_1 : [a]$ $\qquad$ $A_5 : [A_4 \Rightarrow c]$
$A_2 : [d]$ $\qquad$ $A_6 : [A_2 \Rightarrow e]$
$A_3 : [g]$ $\qquad$ $A_7 : [A_6 \Rightarrow f]$
$A_4 : [A_1 \Rightarrow b]$ $\qquad$ $A_8 : [A_4, A_5, A_6, A_7 \to \neg g]$

Here, argument A_8 is defeated by A_3. The arguments A_1, A_2, A_3, A_4, A_5, A_6 and A_7 do not have any defeaters, thus they belong to the grounded extension. Therefore, the propositions a, b, c, d, e, f and g are considered justified. Notice that although there exists a strict rule $b, c, e, f \to \neg g$, $\neg g$ is not a justified conclusion. This shows that the justified conclusions are not closed strict rules.

The problem with the above examples is that the considered language is not expressive enough to capture all the different kinds of conflicts that may exist between arguments. As a consequence of missing some conflicts, the conclusions may be *counter-intuitive*. In example 2, for instance, it is not possible to conclude at the same time that John is both married and a bachelor. Since conclusions may be counter-intuitive, problems of *inconsistency* and *non-closure* appear.

Rationality postulates

Like any reasoning model, an argumentation-based system should satisfy some principles which guarantee the good quality of the system. The aim of this section is to present and to discuss two important postulates: *consistency* and *closeness*, that any argumentation-based system should satisfy in order to avoid the problems discussed in the previous section.

The idea of closeness is that the answer of an argumentation-engine should be closed under strict rules. That is, if we provide the engine with a strict rule $a \to b$ ("if a then it is also *unexceptionally* the case that b"), together with various other rules, and our inference engine outputs a as justified conclusion, then it should also output b as justified conclusion. Consequently, b should also be supported by an acceptable argument. Before stating the postulate, let's first define the closure of a set of formulas.

Definition 8 (Closure of a set of formulas). *Let* $F \subseteq \mathcal{L}$. F *is* closed *iff for every rule* $\phi_1, \ldots, \phi_n \to \psi$ *in* $\mathcal{S}$ *with* $\phi_1, \ldots, \phi_n \in F$, *it holds that* $\psi \in F$.

We say that an argumentation system satisfies closeness if its set of justified conclusions, as well as the set of conclusions supported by each extension are closed.

Postulate 1 (Closeness). *Let* $(\mathcal{A}, defeat)$ *be an argumentation system built from a defeasible theory* $(\mathcal{K}, \mathcal{S}, \mathcal{D})$. Output *is its set of justified conclusions, and $E_1, \ldots, E_n$ its extensions. $(\mathcal{A}, defeat)$ satisfies* closeness *iff:*

1. Output *is closed.*
2. $\forall E_i$, $\{\texttt{Conc(A)} | A \in E_i\}$ *is closed.*

The second condition says that every extension should be closed in the sense that an extension should contain all the arguments acceptable w.r.t it.

As closeness is an important property, one should search for ways to alter or constrain his argumentation formalism in such a way that its resulting extensions and conclusions satisfy closeness.

Another important property of an argumentation system is *consistency*. It should not be the case that an extension or the set of justified conclusions supports opposite statements. This is of great importance since it guarantees that the argumentation system delivers *safe* conclusions. Let's first define the consistency of a set of formulas.

Definition 9 (Consistency of a set of formulas). *Let* $F \subseteq \mathcal{L}$. F *is* consistent *iff* $\neg \exists \psi, \chi \in F$ *such that* $\psi = -\chi$.

An argumentation system satisfies consistency if its set of justified conclusions, and the different sets of conclusions corresponding to each extension are consistent.

Postulate 2 (Consistency). *Let* $(\mathcal{A}, defeat)$ *be an argumentation system built from a defeasible theory* $(\mathcal{K}, \mathcal{S}, \mathcal{D})$. Output *is its set of justified conclusions, and $E_1, \ldots, E_n$ its extensions. $(\mathcal{A}, defeat)$ satisfies* consistency *iff:*

1. Output *is consistent.*
2. $\forall E_i$, $\{\texttt{Conc(A)} | A \in E_i\}$ *is consistent.*

The consistency of the extensions is verified in our formalism since complete extensions should be conflict-free, and thus cannot contain two arguments that rebut each other in the sense of definition 2.

Possible Solutions

In this section we propose two possible solutions for ensuring the closeness and the consistency of the argumentation system proposed in the previous section.

A possible analysis of example 2 and example 3 is that some strict rules are missing. That is, if the rules $\neg hw \rightarrow \neg m$ and $hw \rightarrow \neg b$ (which are the contraposed versions of the existing rules $m \rightarrow hw$ and $b \rightarrow \neg hw$) are added to $\mathcal{S}$, then one can, for instance, construct a counterargument against $[(\rightarrow go) \Rightarrow b]$: $[(((\rightarrow wr) \Rightarrow m) \rightarrow hw) \rightarrow \neg b]$. The basic idea is then to make explicit in $\mathcal{S}$ these implicit information by computing a closure of the set $\mathcal{S}$. The question then becomes whether it is possible to define a *closure operator Cl* on $\mathcal{S}$ such that the outcome makes sure that the argumentation system built on $(\mathcal{K}, Cl(\mathcal{S}), \mathcal{D})$ satisfies closeness and consistency.

One way to define a closure operator given a set of strict rules would be to convert the strict rules to material implications, calculate their closure under propositional logic, and convert the result back to strict rules again. In what follows, $\vdash$ denotes classical inference.

Definition 10 (Propositional operator). *Let $\mathcal{S}$ be a set of strict rules and $\mathcal{P} \subseteq \mathcal{L}$. We define the following functions:*

- $Prop(\mathcal{S}) = \{\phi_1 \wedge \ldots \wedge \phi_n \supset \psi \mid \phi_1, \ldots, \phi_n \rightarrow \psi \in \mathcal{S}\}$
- $Cn_{prop}(\mathcal{P}) = \{\psi \mid \mathcal{P} \vdash \psi\}$
- $Rules(\mathcal{P}) = \{\phi_1, \ldots, \phi_n \rightarrow \psi \mid \phi_1 \wedge \ldots \wedge \phi_n \supset \psi \in \mathcal{P}\}$

The propositional closure of $\mathcal{S}$ is $Cl_{pp}(\mathcal{S}) = Rules(Cn_{prop}(Prop(\mathcal{S})))$.

First of all, it can easily be seen that Cl_{pp} is indeed a closure operator. That is, it satisfies the following three properties:

Property 1. *Let $\mathcal{S}$ be a set of strict rules.*

1. $\mathcal{S} \subseteq Cl_{pp}(\mathcal{S})$
2. *if $\mathcal{S}_1 \subseteq \mathcal{S}_2$ then $Cl_{pp}(\mathcal{S}_1) \subseteq Cl_{pp}(\mathcal{S}_2)$, $(\mathcal{S}_1, \mathcal{S}_2 \subseteq \mathcal{S})$*
3. $Cl_{pp}(Cl_{pp}(\mathcal{S})) = Cl_{pp}(\mathcal{S})$

Furthermore, by using $Cl_{pp}(\mathcal{S})$ instead of just $\mathcal{S}$, one guarantees that under grounded semantics the postulates closeness (postulate 1) and consistency (postulate 2) are warranted for the argumentation system presented in the previous section.

Theorem 1. *Let $(\mathcal{A}, defeat)$ be an argumentation system built from $(\mathcal{K}, Cl_{pp}(\mathcal{S}), \mathcal{D})$. $(\mathcal{A}, defeat)$ satisfies closeness and consistency under the grounded extension.*

To illustrate how Cl_{pp} works, consider again example 2.

Example 4 (Married John, continued). *Let $\mathcal{K} = \{wr; go\}$, $\mathcal{S} = \{m \rightarrow hw; b \rightarrow \neg hw\}$ and $\mathcal{D} = \{wr \Rightarrow m; go \Rightarrow b\}$. Under $(\mathcal{K}, Cl_{pp}(\mathcal{S}), \mathcal{D})$ the following argument can be constructed: $[[wr] \Rightarrow m]$. However, since*

$Cl_{pp}(\mathcal{S})$ *also contains the rule $\neg hw \rightarrow \neg m$ it is now possible to construct the following counterargument: $[[[[go] \Rightarrow b] \rightarrow \neg hw] \rightarrow \neg m]$. Thus, the two arguments will not be in the grounded extension. Consequently, m is no longer a justified conclusion. Similarly, the two following conflicting arguments can be constructed: $[[go] \Rightarrow b]$ and $[[[[wr] \Rightarrow m] \rightarrow hw] \rightarrow \neg b]$ (since $hw \rightarrow \neg b$ is now in $Cl_{pp}(\mathcal{S})$). Thus, b is not justified.*

In the above example, it can be seen that Cl_{pp} can generate a rule (in this case: $\neg hw \rightarrow \neg m$) that is needed to obtain an intuitive outcome. As a side effect, Cl_{pp} also generates many rules that are not actually needed to obtain the intuitive outcome. An example of such a rule is $b \rightarrow \neg m$, which corresponds to applying transitivity on the rules $b \rightarrow \neg hw$ and $\neg hw \rightarrow \neg m$. Worse yet, Cl_{pp} may also generate rules which are actually *harmful* for obtaining an intuitive outcome. An example of such a rule is $p, \neg p \rightarrow q$. Rules like these, which are generated regardless of the content of $\mathcal{S}$, may give birth to self-defeating arguments, which may prevent some "good" arguments from becoming acceptable. This problem can be illustrated with the following example.

Example 5. *Let $\mathcal{K} = \{a; b; c\}$, $\mathcal{S} = \emptyset$ and $\mathcal{D} = \{a \Rightarrow d; b \Rightarrow \neg d; c \Rightarrow e\}$. This allows us to construct, among others, the following arguments: $A = [[a] \Rightarrow d]$, $B = [[b] \Rightarrow \neg d]$, and $C = [[c] \Rightarrow e]$. Intuitively, one may wish to have e justified, d and $\neg d$ not justified. Unfortunately, this is not the case because there now exists $D = [A, B \rightarrow \neg e]$ which defeats C. Note that D is a self-defeating argument. To see why this is a legally constructed counterargument, first consider the fact that, under propositional logic, it holds that $d, \neg d \vdash \neg e$. Therefore, there exists a rule of the form $d, \neg d \rightarrow \neg e$. It is this rule that is applied in argument D to combine A and B to obtain $\neg e$. Thus, D defeats C and consequently, the argument C is not acceptable and e is not justified.*

The above example clearly yields undesirable results, even if under the grounded extensions, the system satisfies both closeness and consistency. If Nixon is both a quaker and a republican, then the issue of whether he is a pacifist or not should not influence a completely unrelated proposition (say, whether it will rain today). Indeed, in general it should not be the case that two arguments that rebut each other can keep an arbitrary argument from becoming acceptable. To solve this problem, one may think of ruling out self-defeating arguments and not considering them when computing the set of acceptable arguments. Unfortunately, this solution leads to the violate closeness. Let's take the following example:

Example 6. *Let $\mathcal{K} = \{a\}$, $\mathcal{S} = \{c, d \rightarrow \neg\lceil a \Rightarrow b\rceil\}$ and $\mathcal{D} = \{a \Rightarrow b; b \Rightarrow c; b \Rightarrow d\}$.*
Now consider the following arguments:

$A_1 : [a]$ $A_4 : [A_2 \Rightarrow d]$
$A_2 : [A_1 \Rightarrow b]$ $A_5 : [A_3, A_4 \Rightarrow \neg\lceil a \Rightarrow b\rceil]$
$A_3 : [A_2 \Rightarrow c]$

The argument A_5 is self-defeating (self-undercutting) and thus ruled out. This means that A_2, A_3 and A_4 don't have any defeaters anymore, and are thus justified. This means

that the literals c and d are also justified. Yet $\neg\lceil a \Rightarrow b\rceil$ is not justified, which violates closeness.

In the light of the above, one can observe that the approach of computing the closure of a set of strict rules requires a closure operator that generates at least those rules that are needed to satisfy closeness and consistency, but at the same time does not generate rules which can be used to build new arguments that may keep "good" arguments from becoming acceptable, and consequently keep their conclusions from becoming justified. In other words, the closure operator shouldn't generate too little, but it shouldn't generate too much either.

We are now about to define a second closure operator Cl_{tp} that is a lot weaker than our first one (Cl_{pp}). Our discussion starts with the observation that a strict rule (say $\phi_1, \ldots, \phi_n \rightarrow \psi$), when translated to propositional logic ($\phi_1 \wedge \ldots \wedge \phi_n \supset \psi$) is equivalent to a disjunction ($\neg\phi_1 \vee \ldots \vee \neg\phi_n \vee \psi$). In this disjunction, different literals can be put in front (like $\neg\phi_i$ in $\neg\phi_1 \vee \ldots \vee \neg\phi_{i-1} \vee \psi \vee \neg\phi_{i+1} \vee \ldots \vee \neg\phi_n \vee \neg\phi_i$), which can again be translated to a strict rule ($\phi_1, \ldots, \phi_{i-1}, \neg\psi, \phi_{i+1}, \ldots, \phi_n \rightarrow \neg\phi_i$). This leads to the following definition.

Definition 11 (Transposition). *A strict rule s is a transposition of $\phi_1, \ldots, \phi_n \rightarrow \psi$ iff $s = \phi_1, \ldots, \phi_{i-1}, \neg\psi, \phi_{i+1}, \ldots, \phi_n \rightarrow \neg\phi_i$ for some $1 \leq i \leq n$.*

Based on the thus defined notion of transposition, we now define our second closure operator.

Definition 12 (Transposition operator). *Let $\mathcal{S}$ be a set of strict rules. $Cl_{tp}(\mathcal{S})$ is a minimal set such that:*

- $\mathcal{S} \subseteq Cl_{tp}(\mathcal{S})$, and
- *if $s \in Cl_{tp}(\mathcal{S})$ and t is a transposition of s then $t \in Cl_{tp}(\mathcal{S})$.*

We say that $\mathcal{S}$ is closed under transposition *iff $Cl_{tp}(\mathcal{S}) = \mathcal{S}$.*

One can easily check that transposition is a special instance of contraposition. It is then easily verified that with the Cl_{tp} operator, example 2 (Married John) is handled correctly. Note also that such an operator minimizes the number of self-defeating arguments.

Lemma 1. *Let $(\mathcal{A}, \mathit{defeat})$ be an argumentation system built from $(\mathcal{K}, Cl_{tp}(\mathcal{S}), \mathcal{D})$. $(\mathcal{A}, \mathit{defeat})$ satisfies closeness and consistency under grounded semantics.*

Unfortunately, the Cl_{tp} operator by itself is not enough to guarantee the closeness and consistency of an argumentation system for the other acceptability semantics (preferred semantics, stable semantics, complete semantics). This can be seen by examining the following example.

Example 7. *Let $\mathcal{K} = \{a; b; c; g\}$, $\mathcal{S} = \{d, e, f \rightarrow \neg g\}$ and $\mathcal{D} = \{a \Rightarrow d; b \Rightarrow e; c \Rightarrow f\}$.*
Now, consider the following arguments:
$A : [[a] \Rightarrow d]$
$B : [[b] \Rightarrow e]$
$C : [[c] \Rightarrow f]$
One can easily check that without Cl_{tp}, the arguments A, B and C do not have any counter-arguments (which makes them members of any Dung-style extension). However, if one would replace the defeasible theory $(\mathcal{K}, \mathcal{S}, \mathcal{D})$ by

$(\mathcal{K}, Cl_{tp}(\mathcal{S}), \mathcal{D})$, *then counter-arguments against A, B and C do exist. For instance, $D = [[[b] \Rightarrow e], [[c] \Rightarrow f], [g] \rightarrow \neg d]$ defeats A (because $e, f, g \rightarrow \neg d \in Cl_{tr}(\mathcal{S})$). The counter-arguments against A, B and C make sure that, under grounded semantics, neither d, e nor f is justified. At the same time, however, it must be observed that the set $\{A, B, C\}$ is admissible. Even though D defeats A, A also defeats D, and similar observations can also be made with respect to B and C. And because $\{A, B, C\}$ is admissible, there also exists a preferred extension (a superset of $\{A, B, C\}$) with conclusions d, e, f and also g. This means that this preferred extension does not satisfy closeness. Moreover, the closure under the strict rules of its conclusions is inconsistent.*

So, while the closure of strict rules under transposition solves the issue of closeness and consistency under grounded semantics, the problem is still open for preferred semantics. For this, an alteration to the core formalism is necessary, in particular to the notion of rebutting. The basic idea is that strict arguments take precedence over defeasible ones. Moreover, if two arguments (let's say A and B) are both defeasible and the top rule of A is strict and has a consequent ψ and the top rule of B is defeasible and has a consequent $-\psi$ then A takes precedence over B. This gives birth to a restricted notion of rebutting.

Definition 13 (Restricted rebut). *An argument A restrictively rebuts an argument B iff $\mathtt{Conc}(A) = \psi$ and B has a subargument of the form $B_1', \ldots, B_n' \Rightarrow -\psi$.*

It can easily be seen that the notion of restricted rebut is indeed a restricted version of "ordinary" rebut. That is, if A rebuts B under the restricted definition (definition 13), then A also rebuts B under the definition 2. The converse, however, is not true; it does in general not hold that if A rebuts B, then A rebuts B under the restricted definition. For instance, $[[\rightarrow a] \Rightarrow b]$ rebuts $[[[\rightarrow c] \Rightarrow d] \rightarrow \neg b]$ under the unrestricted definition, but not under the restricted definition. Let $R_{restricted}$ denote the restricted rebut relation and $R_{unrestricted}$ denote the unrestricted rebut relation.

Property 2. *Let A, $B \in \mathcal{A}$. If A $R_{restricted}$ B then A $R_{unrestricted}$ B. The reverse does not always hold.*

To see how the restricted rebut can help to solve the issue of postulates, again consider the problem of example 7.

Example 8 (7, continued). *Let $\mathcal{K} = \{a; b; c; g\}$, $\mathcal{S} = \{d, e, f \rightarrow \neg g\}$ and $\mathcal{D} = \{a \Rightarrow d; b \Rightarrow e; c \Rightarrow f\}$.*
Now, again consider the following arguments:
$A : [[a] \Rightarrow d]$
$B : [[b] \Rightarrow e]$
$C : [[c] \Rightarrow f]$
Under the restricted version of rebutting, it holds that $\{A, B, C\}$ is not an admissible set under $(Cl_{tp}(\mathcal{S}), \mathcal{D})$. For instance, the argument $[[[b] \Rightarrow e], [[c] \Rightarrow f], [g] \rightarrow \neg d]$ (D) now rebuts A but A does not rebut D, nor does any other argument in $\{A, B, C\}$ defeat D. Thus $\{A, B, C\}$ is not admissible in $(Cl_{tp}(\mathcal{S}), \mathcal{D})$ under the restricted definition of rebutting.

We will now show that if we consider the transposition closure Cl_{tp} and the "restricted rebutting" then the two pos-

tulates (closeness and consistency) are satisfied under any reasonable semantics. Before we do so, however, we should first make clear what we mean with "under any reasonable semantics". Surely, we want at least grounded and preferred semantics to be included. One way to achieve this is to consider the complete extensions. Dung has proved that every stable extension, preferred extension or grounded extension is also a complete extension (Dung 1995). Therefore, what we are going to prove is that with the combination of Cl_{tr} and restricted rebutting, the two postulates hold for *any* complete extension.

Theorem 2. *Let* $(\mathcal{A}, defeat)$ *be an argumentation system built on* $(\mathcal{K}, Cl_{tp}(\mathcal{S}), \mathcal{D})$. *Under restricted rebutting, every complete extension of* $(\mathcal{A}, defeat)$ *is closed.*

Now it's time for the main theorem of consistency.

Theorem 3. *Let* $(\mathcal{A}, defeat)$ *be an argumentation system built on* $(\mathcal{K}, Cl_{tp}(\mathcal{S}), \mathcal{D})$. *Under restricted rebutting, every complete extension of* $(\mathcal{A}, defeat)$ *is consistent.*

So far, what has been proved is that every complete extension is closed and consistent (assuming restricted rebutting and strict rules that are closed under transposition). The next step is to prove that the set `Output` containing the justified conclusions is also closed and consistent.

Theorem 4. *Let* $(\mathcal{A}, defeat)$ *be an argumentation system built on* $(\mathcal{K}, Cl_{tp}(\mathcal{S}), \mathcal{D})$. *Under restricted rebutting, the set* `Output` *is closed and consistent.*

From Theorem 2 and Theorem 4, it is clear that the argumentation system satisfies the closeness.

Corollary 1. *Let* $(\mathcal{A}, defeat)$ *be an argumentation system built on* $(\mathcal{K}, Cl_{tp}(\mathcal{S}), \mathcal{D})$. *Under restricted rebutting, the system* $(\mathcal{A}, defeat)$ *satisfies closeness.*

From Theorem 3 and Theorem 4, it is also clear that the argumentation system satisfies consistency.

Corollary 2. *Let* $(\mathcal{A}, defeat)$ *be an argumentation system built on* $(\mathcal{K}, Cl_{tp}(\mathcal{S}), \mathcal{D})$. *Under restricted rebutting, the system* $(\mathcal{A}, defeat)$ *satisfies consistency.*

The same results also hold if the propositional operator is combined with the restricted rebutting, even if that operator has some weak points as outlined above.

Theorem 5. *Let* $(\mathcal{A}, defeat)$ *be an argumentation system built on* $(\mathcal{K}, Cl_{pp}(\mathcal{S}), \mathcal{D})$. *Under restricted rebutting, the system* $(\mathcal{A}, defeat)$ *satisfies consistency and closeness.*

Conclusion

Argumentation theory is seen as a foundation for reasoning systems. Consequently, an increasing number of argumentation systems has been proposed. While, these systems use generally the same acceptability semantics, they differ in the way they define their logical language, the notion of argument and the defeasibility relation. These last are defined in ad hoc way and this leads the systems to encounter some problems such as returning counter-intuitive results.

In order to avoid such problems, the aim of this paper is to define some postulates or axioms that any argumentation system should satisfy. These postulates govern the well definition of an argumentation system and guarantee the safety of its outputs. We have focused on two important postulates: the *closeness* and the *consistency* of the results of a system. These last are violated by several argumentation systems such as (Prakken & Sartor 1997; Governatori *et al.* 2004; García & Simari 2004). We then studied two ways in which these postulates are warranted for an instantiation of the Dung system. In particular, we have proposed two closure operators that allow to make more explicit some implicit information. These closure operators solve the problems encountered by the argumentation systems defined in in (Prakken & Sartor 1997; Governatori *et al.* 2004; García & Simari 2004).

An extension of this work would be to explore other rationality postulates, especially for reinstating arguments, i.e. for defining the acceptable ones. Indeed, for some applications, the acceptability semantics defined in (Dung 1995) are unfortunately not suitable and new semantics are needed. One rationality postulate should guarantee that the new semantics consider non defeated arguments as acceptable.

References

Amgoud, L., and Cayrol, C. 2002. Inferring from inconsistency in preference-based argumentation frameworks. *Int. Journal of Automated Reasoning* Volume 29 (2):125–169.

Amgoud, L.; Caminada, M.; Cayrol, C.; Doutre, S.; Lagasquie, M.; Prakken, H.; and Vreeswijk, G. 2004. Towards a consensual formal model: inference part. Technical report, Deliverable D2.2: Draft Formal Semantics for Inference and Decision-Making. ASPIC project.

Dung, P. M. 1995. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence* 77:321–357.

García, A., and Simari, G. 2004. Defeasible logic programming: an argumentative approach. *Theory and Practice of Logic Programming* 4(1):95–138.

Gómez, S. A., and Chesñevar, C. I. 2003. Integrating defeasible argumentation with fuzzy art neural networks for pattern classification. In *Proc. ECML'03*.

Governatori, G.; Maher, M.; Antoniou, G.; and Billington, D. 2004. Argumentation semantics for defeasible logic. *Journal of Logic and Computation* 14(5):675–702.

Pollock, J. L. 1995. *Cognitive Carpentry. A Blueprint for How to Build a Person*. Cambridge, MA: MIT Press.

Prakken, H., and Sartor, G. 1997. Argument-based extended logic programming with defeasible priorities. *Journal of Applied Non-Classical Logics* 7:25–75.

Rahwan, I.; Ramchurn, S. D.; Jennings, N. R.; McBurney, P.; Parsons, S.; and Sonenberg, L. 2004. Argumentation-based negotiation. *Knowledge engineering review*.