On the evaluation of argumentation formalisms ⋆

Martin Caminada [a] Leila Amgoud [b]

[a] *Institute of Information and Computing Sciences, Universiteit Utrecht, Utrecht, The Netherlands*
martinc@cs.uu.nl

[b] *Institut de Recherche en Informatique de Toulouse, 118 route de Narbonne, 31062 Toulouse Cedex 9, 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 original statement can be accepted or not.

Several argumentation systems have been proposed in the literature. Some of them, the so-called *rule-based systems*, use a particular logical language with *strict* and *defeasible rules*. While these systems are useful in different domains (e.g. legal reasoning), they unfortunately 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 principles, called *rationality postulates*, that can be used to judge the quality of a rule-based argumentation system. In particular, we define two important rationality postulates that should be satisfied: the *consistency* and the *closure* of the results returned by that system.

We then provide a relatively easy way in which these rationality postulates can be warranted for a particular rule-based argumentation system developed within a European project on argumentation.

Key words: Formal argumentation, Nonmonotonic logic, Commonsense reasoning

⋆ This work has been supported by the EU-ASPIC project

1 INTRODUCTION

Agents express claims and judgments when engaged in decision making, drawing conclusions, imparting information, and when persuading and negotiating with other agents. Information may be uncertain and incomplete, or there may be relevant but partially conflicting information. Also, in multi-agents systems, conflicts of interest are inevitable. To address these problems, agents can use argumentation, a process based on the exchange and valuation of arguments for and against opinions, proposals, claims and decisions.

Argumentation, in its essence, can be seen as a particular useful and intuitive paradigm for doing nonmonotonic reasoning. The advantage of argumentation is that the reasoning process is composed of modular and quite intuitive steps, and thus avoids the monolithic approach of many traditional logics for defeasible reasoning. The process of argumentation starts with the construction of a set of arguments based on a given knowledge base. As some of these arguments may attack each other, one needs to apply a criterion for determining the sets of arguments that can be regarded as "acceptable": the argument-based *extensions*. The last step is then to examine whether a particular statement can be regarded as *justified*. This can for instance be the case if every extension contains an argument which has this statement as its conclusion. An interesting property of the argumentation approach is that it can be given dialectical proof procedures that are quite close to the process by which humans would discuss an issue. The similarity with human-style discussions gives formal argumentation an advantage that can be useful in many contexts.

Argumentation has developed into an important area of study in artificial intelligence over the last fifteen years, especially in sub-fields such as nonmonotonic reasoning (e.g. [19,25,26,28,43,45]), multiple-source information systems (e.g. [7,9,21]), decision making (e.g. [2,11,12,20,32,31,30]), and modeling interactions between agents (e.g. [3,8,10,14,18,35–38,41]). Several argumentation systems have been developed for handling inconsistency in knowledge bases (e.g. [5,15–17,29,33,34,39,42,44]), in other words for *inference*. All these systems are built around a logical language and an associated consequence relation that is used for defining an argument. Some of these systems, called *rule-based systems*, use a particular logical language defined over a set of literals, and two kinds of rules: *strict rules* and *defeasible* ones. Arguments and conflicts among them are first identified, and then an acceptability semantics (e.g. Dung's semantics) is applied in order to determine the "acceptable" arguments. Examples of such systems are Prakken and Sartor's system [42], Garcia and Simari's system [33], Governatori et al.'s system [34], and Amgoud et al.'s system [4]. Such systems are suitable in some domains like legal reasoning, where knowledge cannot be represented in a classical propositional language for instance. Unfortunately, existing rule-based systems fail to meet the objectives of an inference system, and can lead to very unintuitive results. Indeed, with these systems it may be the case that an agent believes that "if a then it is always the case that b", and the system returns as output a but not b. Worse yet, if the agent also believes that "if c then it is always the case that $\neg b$", the system may return a and c, which means that the output of the system is indirectly inconsistent.

In what follows, we will focus only on rule-based argumentation systems. In order to avoid anomalies like the ones discussed above, the aim of this paper is twofold: on the one hand, as 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 *rationality postulates*) that any rule-based argumentation system should obey. These postulates will govern the sound definition of an argumentation system and will avoid anomalous results. In this paper we focus particularly on two important postulates: the *closure* and the *consistency* of the results that an argumentation system may produce. These postulates are violated in systems such as [4,33,34,42]. On the other hand, we study various ways in which these postulates can be warranted in the argumentation system developed in [4], as well as in various other systems.

This paper is structured as follows. First, in section 2, we recall the basic concepts behind argumentation theory. We present the abstract argumentation framework of Dung [28], as well as one particular instantiation of it, for which we have chosen the ASPIC argumentation formalism [4]. In section 3, we show some examples that yield very unintuitive and undesirable results, not only for the ASPIC argumentation system, but also for various other argumentation formalisms. Then, in section 4, we state a number of postulates, based on the analysis of the examples in section 3, that we think any rule-based argumentation formalism should satisfy. Section 5 proposes a number of generic solutions which can be applied to the argumentation formalism described in section 2, as well as to other argumentation formalisms where similar problems occur (such as [34,42,33]). Two main solutions are suggested, each of which satisfies all the earlier mentioned rationality postulates. The first approach is applicable to formalisms that make use of classical logic, the other one is applicable to formalisms that do not. Section 6 then contains an overview of the main results of this paper, as well as some open research issues.

2 ARGUMENTATION PROCESS

Argumentation can be seen as a reasoning process consisting of the following four steps:

(1) Constructing *arguments* (in *favor* of / *against* a "statement") from a knowledge base.
(2) Determining the different *conflicts* among the arguments.
(3) Evaluating the *acceptability* of the different arguments.
(4) Concluding, or defining the *justified conclusions*.

Some argumentation formalisms also allow arguments to be of different strengths, but for the sake of simplicity we will not address this issue in the current paper. Many argumentation formalisms are built around an underlying logical language $\mathcal{L}$ and an associated notion of logical consequence, defining the notion of argument. Argument construction is a monotonic process: new knowledge cannot rule out an argument but only gives rise to new arguments which may interact with the first argument. Since the knowledge bases may give rise to inconsistent conclusions, the arguments may be conflicting too. Consequently, it is important to determine among all the available arguments, the ones that are ultimately acceptable. In [28], an argumentation system is defined as follows:

Definition 1 (Argumentation system) *An* argumentation system *is a pair* $\langle \mathcal{A}, Def \rangle$ *where* $\mathcal{A}$ *is a set of arguments and* $Def \subseteq \mathcal{A} \times \mathcal{A}$ *is a defeat relation. We say that an argument* A defeats *an argument* B *iff* $(A, B) \in Def$ *(or* $A\ Def\ B$*).*

Starting from the set of all (possibly conflicting) arguments, it is important to know which of them can be relied on for inferring conclusions and for making decisions. To answer this question, different attempts for defining *semantics* for the notion of acceptability have been made. Some approaches return a unique set of acceptable arguments, called an *extension*, giving a unique status to each argument, whereas others return several extensions, allowing multiple status for arguments. In [28] different semantics for the notion of acceptability have been proposed. These last have been recently refined in [13,24]. In what follows, only Dung's semantics are recalled for illustration purposes.

Definition 2 (Conflict-free, Defense) *Let* $\mathcal{A}$ *and* $\mathcal{B}$ *be sets of arguments, and let* $\mathcal{B} \subseteq \mathcal{A}$.

- $\mathcal{B}$ *is* conflict-free *iff there exist no* A, B *in* $\mathcal{B}$ *such that* $A\ Def\ B$.
- $\mathcal{B}$ defends *an argument* A *iff for each argument* $B \in \mathcal{A}$, *if* $B\ Def\ A$, *then there exists an argument* C *in* $\mathcal{B}$ *such that* $C\ Def\ B$.

Definition 3 (Acceptability semantics) *Let* $\mathcal{B}$ *be a conflict-free set of arguments, and let* $\mathcal{F}\colon 2^{\mathcal{A}} \mapsto 2^{\mathcal{A}}$ *be a function such that* $\mathcal{F}(\mathcal{B}) = \{A \mid \mathcal{B} \text{ defends } A\}$.

- $\mathcal{B}$ *is* admissible *iff it is conflict-free and defends every element in* $\mathcal{B}$.
- $\mathcal{B}$ *is a* complete extension *iff* $\mathcal{B} = \mathcal{F}(\mathcal{B})$.
- $\mathcal{B}$ *is a* grounded extension *iff it is the minimal (w.r.t. set-inclusion) complete extension.*

- $\mathcal{B}$ *is a* preferred extension *iff it is a maximal (w.r.t. set-inclusion) complete extension.*
- $\mathcal{B}$ *is a* stable extension *iff it is a preferred extension that defeats w.r.t.* Def *all arguments in* $\mathcal{A}\backslash\mathcal{B}$.

Note that a unique grounded extension always exists, although it may be the empty set. It contains all the arguments which are not defeated, as well as the arguments which are defended directly or indirectly by non-defeated arguments.
In the remainder of this paper we use the expression "Dung's standard semantics" to refer to complete, grounded and preferred semantics. We use the unqualified term "extension" to refer to a complete, grounded or preferred extension.

Dung's abstract argumentation theory leaves open the question of how arguments actually look like, how they are constructed from a knowledge base and the conditions under which one argument defeats the other. Several formalisms, such as [4,42,34] aim to fill this gap.

In this paper we have chosen to treat one particular argumentation formalism called ASPIC system [4], as an illustration of how Dung's abstract argumentation formalism can be applied for reasoning in the presence of inconsistency, or for inference. The choice of ASPIC formalism is, we must admit, somewhat arbitrary. We have chosen it mainly because of its relative simplicity, and the fact that we have been closely connected to its development. In fact, much of the current paper is a result of an analysis of the difficulties we encountered when constructing the formalism, difficulties that turned out also to play a role in other formalisms for argumentation and nonmonotonic reasoning.

In what follows, $\mathcal{L}$ is a set of literals. We assume the availability of a function "$-$", which works with $\mathcal{L}$, such that $-\psi = \phi$ iff $\psi = \neg\phi$ and $-\psi = \neg\phi$ iff $\psi = \phi$.

A *strict rule* is an expression of the form $\phi_1, \ldots, \phi_n \longrightarrow \psi$ $(n \geq 0)$, indicating that if $\phi_1, \ldots, \phi_n$ hold, then *without exception* it holds that ψ. A *defeasible rule* is an expression of the form $\phi_1, \ldots, \phi_n \Longrightarrow \psi$ $(n \geq 0)$, indicating that if $\phi_1, \ldots, \phi_n$ hold, then it *usually* holds that ψ. For both a strict and defeasible rule it holds that each ϕ_i $(1 \leq i \leq n)$ as well as ψ are elements of $\mathcal{L}$.

Definition 4 (Theory) *A* defeasible theory $\mathcal{T}$ *is a pair* $\langle \mathcal{S}, \mathcal{D} \rangle$ *where* $\mathcal{S}$ *is a set of strict rules and* $\mathcal{D}$ *is a set of defeasible rules.*

Definition 5 (Closure of a set of formulas) *Let* $\mathcal{P} \subseteq \mathcal{L}$. *The* closure *of* $\mathcal{P}$ *under the set* $\mathcal{S}$ *of strict rules, denoted* $Cl_{\mathcal{S}}(\mathcal{P})$, *is the smallest set such that:*

- $\mathcal{P} \subseteq Cl_{\mathcal{S}}(\mathcal{P})$.
- *if* $\phi_1, \ldots, \phi_n \longrightarrow \psi \in \mathcal{S}$ *and* $\phi_1, \ldots, \phi_n \in Cl_{\mathcal{S}}(\mathcal{P})$ *then* $\psi \in Cl_{\mathcal{S}}(\mathcal{P})$.

If $\mathcal{P} = Cl_{\mathcal{S}}(\mathcal{P})$, *then* $\mathcal{P}$ *is said to be* closed *under the set* $\mathcal{S}$.

Definition 6 (Consistent set) *Let* $\mathcal{P} \subseteq \mathcal{L}$. $\mathcal{P}$ *is* consistent *iff* $\nexists\ \psi, \phi \in \mathcal{P}$ *such that* $\psi = -\phi$, *otherwise it is said to be* inconsistent.

From a defeasible theory $\langle \mathcal{S}, \mathcal{D} \rangle$, arguments can be built. Before defining the arguments, we first introduce some functions. The function `Conc` returns the "top" conclusion of an argument (i.e. the last conclusion), `Sub` returns all its sub-arguments and finally the functions `StrictRules` and `DefRules` return respectively all the strict rules and the defeasible rules used in an argument.

In what follows, an argument has a *deductive* form and is constructed in a recursive way by applying one or more strict or defeasible rules. In order to distinguish them from the strict and defeasible object level rules, we use short arrows for the strict and defeasible argument construction rules.

Definition 7 (Argument) *Let* $\langle \mathcal{S}, \mathcal{D} \rangle$ *be a defeasible theory. An* argument A *is:*

- $A_1, \ldots, A_n \rightarrow \psi$ *if* $A_1, \ldots, A_n$, *with* $n \geq 0$, *are arguments such that there exists a strict rule* $\texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \longrightarrow \psi$.
  $\texttt{Conc}(A) = \psi$,
  $\texttt{Sub}(A) = \texttt{Sub}(A_1) \cup \ldots \cup \texttt{Sub}(A_n) \cup \{A\}$.
  $\texttt{StrictRules}(A) = \texttt{StrictRules}(A_1) \cup \ldots \cup \texttt{StrictRules}(A_n) \cup \{\texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \longrightarrow \psi\}$,
  $\texttt{DefRules}(A) = \texttt{DefRules}(A_1) \cup \ldots \cup \texttt{DefRules}(A_n)$.
- $A_1, \ldots, A_n \Rightarrow \psi$ *if* $A_1, \ldots, A_n$, *with* $n \geq 0$, *are arguments such that there exists a defeasible rule* $\texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \Longrightarrow \psi$.
  $\texttt{Conc}(A) = \psi$,
  $\texttt{Sub}(A) = \texttt{Sub}(A_1) \cup \ldots \cup \texttt{Sub}(A_n) \cup \{A\}$,
  $\texttt{StrictRules}(A) = \texttt{StrictRules}(A_1) \cup \ldots \cup \texttt{StrictRules}(A_n)$,
  $\texttt{DefRules}(A) = \texttt{DefRules}(A_1) \cup \ldots \cup \texttt{DefRules}(A_n) \cup \{\texttt{Conc}(A_1), \ldots \texttt{Conc}(A_n) \Longrightarrow \psi\}$.

$\mathcal{A}rg$ *denotes the set of all arguments that can be built from the theory* $\langle \mathcal{S}, \mathcal{D} \rangle$. *Let* $A, A' \in \mathcal{A}rg$.

- A' *is a* subargument *of* A *iff* $A' \in \texttt{Sub}(A)$.
- A' *is a* direct subargument *of* A *iff* $A' \in \texttt{Sub}(A)$, $\nexists A'' \in \mathcal{A}rg$, $A'' \in \texttt{Sub}(A)$, $A' \in \texttt{Sub}(A'')$, $A \neq A''$, *and* $A' \neq A''$.
- A *is an* atomic argument *iff* $\nexists A' \in \mathcal{A}rg$, $A' \neq A$, *and* $A' \in \texttt{Sub}(A)$.

Let us illustrate the above definition with the following example.

Example 1 *Let* $\mathcal{S} = \{\longrightarrow a;\ \longrightarrow d\}$ *and* $\mathcal{D} = \{a \Longrightarrow b;\ d \Longrightarrow \neg b\}$*. The following arguments can be built:*

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

A_1 *and* A_2 *are atomic arguments.* A_1 *is a direct subargument of* A_3*, and* A_2 *is a direct subargument of* A_4*.*

An argument may be either strict if no defeasible rule is involved in it, or defeasible otherwise. Formally:

Definition 8 (Strict vs. defeasible argument) *Let* A *be an argument.* A *is* strict *iff* $\texttt{DefRules}(A) = \emptyset$*, otherwise* A *is called* defeasible*.*

Generally arguments may be in conflict with each other in different manners. 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 9 (Rebutting) *Let* $A, B \in \mathcal{A}rg$*.* A rebuts B *iff* $\exists\ A' \in \texttt{Sub}(A)$ *with* $\texttt{Conc}(A') = \phi$ *and* $\exists\ B' \in \texttt{Sub}(B)$ *with* B' *a non-strict argument and* $\texttt{Conc}(B') = -\phi$*.*

Example 2 *Let* $\mathcal{S} = \{\longrightarrow a;\ \longrightarrow t;\ a \longrightarrow b\}$*,* $\mathcal{D} = \{b \Longrightarrow c;\ t \Longrightarrow \neg b;\ \neg b \Longrightarrow d\}$*. The argument* $[[[\rightarrow a] \rightarrow b] \Rightarrow c]$ rebuts $[[[\rightarrow t] \Rightarrow \neg b] \Rightarrow d]$*. The reverse is not true.*

The above definition puts strict arguments above defeasible ones in the sense that a strict argument can rebut a defeasible one, but the reverse cannot be the case. Note that this definition of rebutting is more general than the classical one defined in [29]. Indeed, in [29], an argument is supposed to have only one conclusion. The intermediate consequences obtained when building that argument are not taken into account. However, in [4] arguments may disagree not only on their conclusions, but also on their intermediate consequences.

Two arguments may also conflict if one of them uses a defeasible rule whose applicability is disputed by the other argument. In the following definition, $\lceil . \rceil$ stands for the objectivation operator [40], which converts a meta-level expression (in our case: a defeasible rule) into an object-level expression (in our case: a literal). This is needed because, syntactically, the conclusion of a rule can only be a *literal*, whereas with undercutting one wants to express the inapplicability of a *rule*.

Definition 10 (Undercutting) *Let A and B be arguments. A* undercuts B *iff $\exists\ B' \in \texttt{Sub}(B)$ of the form $B''_1, \ldots, B''_n \Rightarrow \psi$ and $\exists\ A' \in \texttt{Sub}(A)$ with* $\texttt{Conc}(A') = \neg\ulcorner\texttt{Conc}(B''_1), \ldots, \texttt{Conc}(B''_n) \Longrightarrow \psi\urcorner$.

As an example to illustrate the difference between rebutting and undercutting, consider argument A: "The object is red because John says it looks red." A rebutter of A could be (B_1) "The object is not red because Suzy says it looks blue." An undercutter of A could be [40] (B_2) "The object is merely illuminated by a red light." This, of course, is not a reason for it not being red, but merely indicates that the fact that it looks red is no longer a reason for it actually being red.

The two relations: *undercut* and *rebut* are brought together is the definition of "defeat" as follows: [1]

Definition 11 (Defeat) *Let A and B be elements of $\mathcal{A}rg$. We say that A* defeats *B iff*

(1) A rebuts B*, or*
(2) A undercuts B*.*

The ASPIC system, built from a theory $\mathcal{T} = \langle \mathcal{S}, \mathcal{D} \rangle$, is a pair $\langle\mathcal{A}\text{rg, Defeat}\rangle$, where $\mathcal{A}$rg is the set of arguments built from $\mathcal{T}$ using Definition 7, and Defeat is the relation given in the above definition 11. For determining among elements of $\mathcal{A}$rg the acceptable arguments, any of Dung's standard semantics (Definition 3) can be applied. We will write $E_1, \ldots, E_n$ to denote the different extensions under one of those semantics.

We can show that if an argument is in a given extension, then all its sub-arguments are also in that extension.

Proposition 1 *Let $\langle\mathcal{A}rg,\ Defeat\rangle$ be an argumentation system, and let E_1, $\ldots$, E_n be its different extensions under one of Dung's standard semantics. $\forall E_i \in \{E_1,\ \ldots,\ E_n\}$, $\forall A \in E_i$, $\texttt{Sub}(A) \subseteq E_i$.* [2]

The last step of an argumentation process consists of determining, among all the conclusions of the different arguments, the ones that can ultimately be accepted: the *justified conclusions*. Let `Output` denote this set of justified conclusions. One way of defining `Output` is to consider the conclusions that

[1] In the original ASPIC system, it is also possible to take into account the relative strength of the arguments when determining when argument A defeats argument B. For reasons of simplicity, argument strength is not treated in the current discussion. In [6], it has been shown that it is straightforward to extend the system to handle preferences.
[2] Proofs for propositions and theorems can be found in the appendix of the paper.

are supported by at least one argument in each extension. The idea is that one should not only define rationality postulates for each individual extension, but also for the overall justified conclusions, thus the need for `Output`.

Definition 12 (Justified conclusions) *Let $\langle \mathcal{A}rg, \mathit{Defeat} \rangle$ be an argumentation system, and $\{E_1, \ldots, E_n\}$ $(n \geq 1)$ be its set of extensions under one of Dung's standard semantics.*

- $\mathtt{Concs}(E_i) = \{\mathtt{Conc}(A) \mid A \in E_i\}$ $(1 \leq i \leq n)$.
- $\mathtt{Output} = \cap_{i=1 \ldots n} \mathtt{Concs}(E_i)$.

It should be noticed that `Output` is defined using a skeptical attitude. This is a deliberate choice, since basing `Output` on a credulous attitude can result in inconsistencies, even in the case where each individual extension has consistent conclusions. In the remainder of this paper, we are interested in both the conclusions of an individual extension ($\mathtt{Concs}(E_i)$) as well as in the overall justified conclusions (`Output`).

It should also be noticed that for simplicity we do not consider the case where there are no extensions. This, for instance, rules out a treatment of stable semantics in this paper.

Let us consider the following illustrative example as an illustration of the above definitions.

Example 3 *Let $\mathcal{S} = \{\longrightarrow a;\ \longrightarrow d\}$ and $\mathcal{D} = \{a \Longrightarrow b;\ d \Longrightarrow \neg b\}$. The following arguments can be constructed:*

$A_1 :\ [\rightarrow a]$ $\quad$ $A_3 :\ [A_1 \Rightarrow b]$

$A_2 :\ [\rightarrow d]$ $\quad$ $A_4 :\ [A_2 \Rightarrow \neg b]$

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 each extension. Consequently, a and d will be considered as justified conclusions.

3 SOME PROBLEMS IN ARGUMENTATION FRAMEWORKS

In this section, we start first by proving some interesting properties of the formalism described in the previous section, especially regarding the consistency of its conclusions. It will then be argued that these properties may not be enough to warrant a good quality of the formalism. It turns out that there exist anomalies that occur not only in the above described first version of the ASPIC formalism, but also in several other of today's argumentation formalisms. Before discussing all these issues in detail, let us first introduce a notion that is useful for the rest of the paper, that of consistency of a set $\mathcal{S}$ of strict rules.

Definition 13 (Consistent set of strict rules) *Let $\mathcal{S}$ be a set of strict rules. $\mathcal{S}$ is said to be* consistent *iff $\nexists$ A, $B \in \mathcal{A}rg$ such that A and B are strict arguments and* $\texttt{Conc}(A) = -\texttt{Conc}(B)$.

In the remainder of this paper, we will use the pair $\langle \mathcal{A}, Def \rangle$ to refer to any argumentation system that is built around a defeasible theory $\mathcal{T}$. The structure of arguments and the conflict relation are unspecified. This means that arguments in $\mathcal{A}$ may be defined for instance as a tree, a sequence, etc. Similarly, one may consider any definition of the relation Def. Moreover, this argumentation system may use any acceptability semantics, i.e. Dung's standards ones or their different refinements or alternatives proposed in the literature.

3.1 Consistency

The ASPIC system, like many other formalisms in the field of argumentation and defeasible reasoning, satisfies the requirement that each extension has consistent conclusions.

Proposition 2 *Let $\langle \mathcal{A}rg$, Defeat$\rangle$ be an argumentation system built from a theory $\langle \mathcal{S}, \mathcal{D} \rangle$ with $\mathcal{S}$ consistent, and $E_1, \ldots, E_n$ its different extensions under one of Dung's standard semantics.* $\texttt{Concs}(E_i)$ *is consistent for each $1 \leq i \leq n$.*

We can verify that if the sets of conclusions of the different extensions are consistent, then the output of the system is also consistent. Note that this result is general in the sense that it does not depend on the particular definitions of argument structure and defeat of the ASPIC system.

Proposition 3 *Let $\mathcal{T}$ be a defeasible theory, $\langle \mathcal{A}, Def \rangle$ be an argumentation system built from $\mathcal{T}$. Let $E_1, \ldots, E_n$ be its extensions under one of Dung's standard semantics, and* `Output` *be as in definition 12.*
If $\texttt{Concs}(E_i)$ *is consistent for each $1 \leq i \leq n$ then* `Output` *is consistent.*

From Proposition 2 and Proposition 3, we can then deduce that the output of the ASPIC system is consistent.

Property 1 *Let $\mathcal{T}$ be a defeasible theory with $\mathcal{S}$ consistent, $\langle \mathcal{A}rg$, Defeat$\rangle$ be an argumentation system built from $\mathcal{T}$. Then,* `Output` *is consistent.*

The sole fact that a formalism for defeasible reasoning or argumentation returns consistent results may in many cases not be enough to warrant the absence of other anomalies. To make this point more clear, it is interesting to consider the following example.

Example 4 (Married John) *Let $\mathcal{S} = \{\longrightarrow wr;\ \longrightarrow go;\ b \longrightarrow \neg hw;\ m \longrightarrow hw\}$ and $\mathcal{D} = \{wr \Longrightarrow m;\ go \Longrightarrow b\}$ 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:*

$A_1 :\ [\rightarrow wr]$ $\qquad A_4 :\ [A_2 \Rightarrow b]$

$A_2 :\ [\rightarrow go]$ $\qquad A_5 :\ [A_3 \rightarrow hw]$

$A_3 :\ [A_1 \Rightarrow m]$ $\qquad A_6 :\ [A_4 \rightarrow \neg hw]$

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. If one applies, for instance, grounded semantics, the grounded extension then becomes $\{A_1, A_2, A_3, A_4\}$. Consequently, `Output` $= \{wr, go, m, b\}$*, this means that both m ("John is married") and b ("John is a bachelor") are considered* justified.

Example 4 clearly shows that counter-intuitive conclusions may be inferred from a defeasible theory using the above argumentation framework. As a consequence, the closure of the set of inferences under the set of strict rules may be inconsistent. In the previous example, the closure of `Output` $(= \{wr, go, m, b\})$ under the set of strict rules is $\{wr, go, m, b, hw, \neg hw\}$ which is inconsistent. To some extent, the problem can be identified as m and b being incompatible without the entailment mechanism being strong enough to detect this. If, for instance, in the previous example it would be allowed to apply contraposition on $m \longrightarrow hw$ and $b \longrightarrow \neg hw$ then counterarguments against m and b could be constructed, which would prevent them to follow from the same extension.

The above example is problematic not only in the ASPIC system. In fact, the defeasible logic of Donald Nute, as described in [34] suffers from exactly the same problem. When one translates the example to Nute's particular syntax, one obtains essentially the same result: m and b are justified, and hw and $\neg hw$ are left undecided.

It should be noted that another argumentation formalism, stated by Prakken and Sartor [42], is defined in such a way to avoid the problematic outcome of example 4. When translated to the formalism of [42], example 4 no longer yields m and b as justified conclusions. This is implemented by extending the notion of defeat. Informally, argument A rebuts argument B in [42] iff it is possible to "add" strict rules to A and B (like $m \longrightarrow hw$ to A and $b \longrightarrow \neg hw$ to B) such that the extended versions of A and B have opposite conclusions.

Although Prakken and Sartor's solution works for the case of example 4, there exist other examples where their approach still yields anomalies. A relatively straightforward case is the following.

Example 5 *Let* $\mathcal{S} = \{\longrightarrow a;\ \longrightarrow d;\ \longrightarrow c;\ b, e \longrightarrow \neg c\}$ *and* $\mathcal{D} = \{a \Longrightarrow b;\ d \Longrightarrow e\}$*. Here, the arguments*
$A = [[\rightarrow a] \Rightarrow b]$,
$B = [[\rightarrow d] \Rightarrow e]$,
$C = [\rightarrow c]$
do not have any defeaters. This means that A*,* B *and* C *are in any Dung-style extension. Therefore, the propositions* b*,* e *and* c *are considered justified. Note that although there exists a strict rule* $b, e \longrightarrow \neg c$*,* $\neg c$ *is not a justified conclusion. This shows that the justified conclusions are not closed under strict rules. Worse yet, the closure of the justified conclusions under strict rules may even be inconsistent.*

The last formalism to be discussed is that of García and Simari [33]. It is interesting to notice that this formalism can properly handle both example 4 and example 5. It essentially does so by considering two arguments to be conflicting (disagreeing) iff from their respective conclusions, an inconsistency can be derived using strict rules only. Although this indeed yields the desired results in example 4 and example 5, there still exist examples that are not handled correctly.

Example 6 *Let* $\mathcal{S} = \{\longrightarrow a;\ \longrightarrow d;\ \longrightarrow g;\ b, c, e, f \longrightarrow \neg g\}$ *and* $\mathcal{D} = \{a \Longrightarrow b;\ b \Longrightarrow c;\ d \Longrightarrow e;\ e \Longrightarrow f\}$*. Now, consider the following arguments:*
$A = [[\rightarrow a] \Rightarrow b]$
$B = [[\rightarrow d] \Rightarrow e]$
$C = [[A \Rightarrow c]$
$D = [[B \Rightarrow f]$
The arguments A*,* B*,* C *and* D *do not have any defeaters. To see why, consider for instance argument* D*.* D *has no defeaters because there is no argument that can produce a literal (conclusion) that disagrees with* f*. Similar observations also hold for* A*,* B *and* C*. Because* A*,* B*,* C *and* D *do not have defeaters, they are automatically ultimately acceptable. This means that the literals* b*,* c*,* e *and* f *are justified (as well as the facts* a *and* g*). This means that the closure of the justified conclusions under strict rules is again inconsistent!*

3.3 *Discussion*

The way the above mentioned argumentation systems deal with the critical examples is unsatisfactory from a conceptual point of view. Suppose, for instance, a user wants to use an inference engine of Defeasible Logic [34]. For this, he provides the inference engine with a set of strict and defeasible rules. Suppose one of the strict rules is of the form: "if m then it is *always* the case that hw". Then he may be very surprised to find that the outcome of the inference engine contains m but not hw. Worse yet, if the user tries to do his own reasoning based on the inference engine's output ("My inference engine says m and I know that m always implies hw, so it must hold that hw. My inference engine also says that b and I know that b always implies $\neg hw$, so it must hold that $\neg hw$.") then the outcome is directly inconsistent.

The problem with the above examples is that the language used 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 4, for instance, it should simply not be possible to conclude that John is both married and bachelor, as deriving these conclusions means that problems of *inconsistency* and *non-closure* appear.

4 RATIONALITY POSTULATES

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

The idea of closure 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 $a \longrightarrow b$ ("if a then it is also *without exception* 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.

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

Postulate 1 (Closure) *Let $\mathcal{T}$ be a defeasible theory, $\langle \mathcal{A}, Def \rangle$ be an argumentation system built from $\mathcal{T}$.* `Output` *is its set of justified conclusions, and $E_1, \ldots, E_n$ its extensions under a given semantics. $\langle \mathcal{A}, Def \rangle$ satisfies* closure *iff:*

(1) $\texttt{Concs}(E_i) = Cl_{\mathcal{S}}(\texttt{Concs}(E_i))$ *for each* $1 \leq i \leq n$.
(2) $\texttt{Output} = Cl_{\mathcal{S}}(\texttt{Output})$.

The first 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 closure is an important property, one should search for ways to alter or constrain one's argumentation formalism in such a way that its resulting extensions and conclusions satisfy closure.

It can be shown that if the different sets of conclusions of the extensions are closed, then the set `Output` is also closed.

Proposition 4 *Let $\mathcal{T}$ be a defeasible theory, $\langle \mathcal{A}, Def \rangle$ be an argumentation system built from $\mathcal{T}$. Let $E_1, \ldots, E_n$ be its extensions under a given semantics, and* `Output` *be as in definition 12.*
If $\texttt{Concs}(E_i) = Cl_{\mathcal{S}}(\texttt{Concs}(E_i)$ *for each* $1 \leq i \leq n$, *then* $\texttt{Output} = Cl_{\mathcal{S}}(\texttt{Output})$.

Another important property of an argumentation system is *direct consistency.* An argumentation system satisfies direct consistency if its set of justified conclusions and the different sets of conclusions corresponding to each extension are consistent. Formally:

Postulate 2 (Direct Consistency) *Let $\mathcal{T}$ be a defeasible theory, $\langle \mathcal{A}, Def \rangle$ be an argumentation system built from $\mathcal{T}$.* `Output` *is its set of justified conclusions, and $E_1, \ldots, E_n$ its extensions under a given semantics. $\langle \mathcal{A}, Def \rangle$ satisfies* direct consistency *iff:*

(1) $\texttt{Concs}(E_i)$ *is consistent for each* $1 \leq i \leq n$.
(2) `Output` *is consistent.*

Most argumentation systems satisfy the above postulate of direct consistency. Unfortunately, they often violate the postulate of *indirect* consistency. By indirect consistency we mean that (1) the closure under the set of strict rules of the set of justified conclusions is consistent, and (2) for each extension, the closure under the set of strict rules of its conclusions is consistent. When this postulate is violated, it means that undesirable conclusions can be inferred.

Postulate 3 (Indirect Consistency) *Let $\mathcal{T}$ be a defeasible theory, $\langle \mathcal{A}, Def \rangle$ be an argumentation system built from $\mathcal{T}$.* `Output` *is its set of justified conclusions, and $E_1, \ldots, E_n$ its extensions under a given semantics. $\langle \mathcal{A}, Def \rangle$ satisfies* indirect consistency *iff:*

(1) $Cl_{\mathcal{S}}(\texttt{Concs}(E_i))$ *is consistent for each* $1 \leq i \leq n$*.*
(2) $Cl_{\mathcal{S}}(\texttt{Output})$ *is consistent.*

Again, we can show that if all the extensions produce a consistent closed output, then the closure of the set `Output` is consistent. Formally:

Proposition 5 *Let* $\mathcal{T}$ *be a defeasible theory,* $\langle \mathcal{A}, Def \rangle$ *be an argumentation system built from* $\mathcal{T}$*. Let* $E_1, \ldots, E_n$ *be its extensions under a given semantics, and let* `Output` *be as in definition 12.*
If $Cl_{\mathcal{S}}(\texttt{Concs}(E_i))$ *is consistent for each* $1 \leq i \leq n$*, then* $Cl_{\mathcal{S}}(\texttt{Output})$ *is consistent.*

Another straightforward result is that, if indirect consistency is satisfied by an argumentation system, then direct consistency is also satisfied by that system.

Proposition 6 *If an argumentation system* $\langle \mathcal{A}, Def \rangle$ *satisfies indirect consistency, then it also satisfies direct consistency.*

In addition to the above result, one can show that a formalism that satisfies closure as well as direct consistency also satisfies indirect consistency.

Proposition 7 *Let* $\langle \mathcal{A}, Def \rangle$ *be an argumentation system. If* $\langle \mathcal{A}, Def \rangle$ *satisfies closure and direct consistency, then it also satisfies indirect consistency.*

So far, we have identified a number of rationality postulates and examined the effects of their violation. Table 1 provides a brief summary of these effects.

Postulate	Violation can result in
Direct consistency	Absurdities
Indirect consistency	Users not being allowed to apply modus ponens on strict rules
Closure	Conclusions that should come out appear to be missing

Table 1
The effects of violated postulates.

As for direct consistency, the situation is straightforward. When direct consistency is violated, two contradictory statements (say ψ and $\neg\psi$) are justified at the same time, which is clearly an absurdity. As for indirect inconsistency — which is for instance violated in the original "Married John" example (Example 4) — the situation is somewhat more complex. It can be the case that a formalism satisfies direct consistency but violates indirect consistency (an example would be the Defeasible Logic of Donald Nute [33]). In that case, the users of an implementation of such a system would be disallowed from doing their own reasoning based on its outcome. That is, one may not take the outcome of the formalism and apply modus ponens using the strict rules, as otherwise absurdities may result.

As for the property of closure, the basic idea is that the conclusions of the formalism should be "complete". It should not be the case that the user must do its own reasoning (take the outcome of the formalism and apply modus ponens using the strict rules) to derive statements that the formalism apparently "forgot" to entail. A formalism that satisfies closure has done all of this work by itself.

5 POSSIBLE SOLUTIONS

The aim of this section is to "repair" the ASPIC system defined in [4], by providing two solutions that satisfy the three rationality postulates discussed in the previous section. Thus, in all what follows, we will handle only the system of [4]. Nevertheless, the proposed solutions could also be implemented in [34], [42] and [33].

According to Proposition 2 and Property 1, it is clear that this system already satisfies direct consistency.

Property 2 *Let $\langle \mathcal{A}rg, Defeat \rangle$ be an argumentation system built from a theory $\langle \mathcal{S}, \mathcal{D} \rangle$ with $\mathcal{S}$ consistent. $\langle \mathcal{A}rg, Defeat \rangle$ satisfies direct consistency (i.e Postulate 2).*

However, as shown through Example 4, the ASPIC system violates closure and indirect consistency. A possible analysis of Example 4 is that some strict rules are missing. That is, if the rules $\neg hw \longrightarrow \neg m$ and $hw \longrightarrow \neg b$ (which are the contraposed versions of the existing rules $m \longrightarrow hw$ and $b \longrightarrow \neg hw$) are added to $\mathcal{S}$, then one can, for instance, construct a counter-argument against $[[\rightarrow go] \Rightarrow b]$: $[[[[\rightarrow wr] \Rightarrow m] \rightarrow hw] \rightarrow \neg b]$. The basic idea is then to make explicit in $\mathcal{S}$ this 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 the defeasible theory $\langle Cl(\mathcal{S}), \mathcal{D} \rangle$ satisfies closure and consistency.

5.1 Strict rules closed under classical entailment

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

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

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} satisfies the following three properties, which follow from the nature of classical logic:

Property 3 *Let $\mathcal{S}$ be a set of strict rules and let $\mathcal{S}_1, \mathcal{S}_2 \subseteq \mathcal{S}$.*

(1) $\mathcal{S} \subseteq Cl_{pp}(\mathcal{S})$
(2) If $\mathcal{S}_1 \subseteq \mathcal{S}_2$ *then* $Cl_{pp}(\mathcal{S}_1) \subseteq Cl_{pp}(\mathcal{S}_2)$
(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 closure (postulate 1), direct consistency (postulate 2) and indirect consistency (postulate 3) are warranted for the ASPIC system.

Theorem 1 *Let $\langle \mathcal{A}rg,\ Defeat \rangle$ be an argumentation system built from the defeasible theory $\langle Cl_{pp}(\mathcal{S}), \mathcal{D} \rangle$ such that $Cl_{pp}(\mathcal{S})$ is consistent.* `Output` *is its set of justified conclusions and E its grounded extension. Then, $\langle \mathcal{A}rg,\ defeat \rangle$ satisfies closure and indirect consistency.*

To illustrate how Cl_{pp} works, consider again example 4.
Example 4 – continued: *Let $\mathcal{S} = \{\longrightarrow wr;\ \longrightarrow go;\ m \longrightarrow hw;\ b \longrightarrow \neg hw\}$ and $\mathcal{D} = \{wr \Longrightarrow m;\ go \Longrightarrow b\}$.*
Under $\langle Cl_{pp}(\mathcal{S}), \mathcal{D} \rangle$ the following arguments can be constructed:

A_1: $[\rightarrow wr]$
A_2: $[\rightarrow go]$
A_3: $[A_1 \Rightarrow m]$
A_4: $[A_2 \Rightarrow b]$
A_5: $[A_3 \rightarrow hw]$
A_6: $[A_4 \rightarrow \neg hw]$
A_7: $[A_5 \rightarrow \neg b]$ *(using the rule $hw \longrightarrow \neg b$)*
A_8: $[A_6 \rightarrow \neg m]$ *(using the rule $\neg hw \longrightarrow \neg m$)*

Now the argument A_3 has a defeater which is A_8. Since A_8 also defeats A_3, the two arguments will not be in the grounded extension. Consequently, m is no longer a justified conclusion. Similarly, the two arguments A_4 and A_7 are conflicting. Therefore, b is not justified either. Thus, only the premises wr and go are considered justified in this example.

The previous example illustrates that, although the closure of strict rules under Cl_{pp} operator solves the issue of closure and indirect consistency under grounded semantics, the problem is still open for the other acceptability semantics (like complete and preferred semantics) This can be seen by again examining the example of "Married John".
Example 4 – continued: *As said before, the argument A_4 has a unique defeater which is A_7 and A_3 has one defeater which is A_8. However, A_4 defeats A_7 and A_3 defeats A_8. Thus, the set $\{A_3, A_4\}$ is an admissible extension since it defends itself against all its defeaters (A_7, A_8). And because $\{A_3, A_4\}$ is admissible, there also exists a preferred extension (a superset of $\{A_3, A_4\}$) with conclusions b and also m. This means that this preferred extension does not satisfy closure. Moreover, the closure under the strict rules of its conclusions is inconsistent. However, note that since we are using a skeptical reasoning, neither m nor $\neg m$ (resp. neither b nor $\neg b$) can be inferred from this theory. Thus, the problem concerns only the results returned by individual extensions, and not the output of the system.*

To solve this problem, an alteration to the core formalism is necessary, in particular to the notion of rebutting. The idea is to consider a restricted notion of rebutting so that an argument can only be rebutted on the consequent of one of its *defeasible* rules. This can be stated as follows:

Definition 15 (Restricted rebutting) *Let A and B be arguments. A* restrictively rebuts *B on (A', B') iff $A' \in \texttt{Sub}(A)$ such that $\texttt{Conc}(A') = \phi$ and $B' \in \texttt{Sub}(B)$ such that B' is of the form $B''_1, \ldots, B''_n \Rightarrow -\phi$.*

Note that the restricted version of rebut is a special case of the unrestricted version of rebut.

Property 4 *Let A and B be arguments. If A restrictively rebuts B, then A rebuts B. The reverse is not always true.*

Let us consider the following counter-example:
Example 4 – continued: *In the previous example, the argument A_4 rebuts A_7, but A_4 does not restrictively rebuts A_7.*

We now consider the following argumentation system: $\langle \mathcal{A}\text{rg}, \textit{Defeat}_r \rangle$ such that $\textit{Defeat}_r$ is defined as follows:

Definition 16 (Restricted Defeating) *Let A and B be arguments. We say that A* defeats$_r$ *B iff*

(1) A restrictively rebuts B*, or*
(2) A undercuts B*.*

Before showing how restricted rebut can help to solve the issue of postulates, let us first introduce an important result. In fact, it can be verified that when "restricted rebutting" is used instead of "rebutting", [3] then the argumentation formalism immediately satisfies *closure*, without the need to compute any closure of the set $\mathcal{S}$. On the other hand, when "rebutting" is used, it is *direct consistency* that is immediately satisfied, as shown by Property 2.

Proposition 8 *Let $\langle \mathcal{A}rg,\ Defeat_r \rangle$ be an argumentation system built from a theory $\langle \mathcal{S}, \mathcal{D} \rangle$, and $E_1, \ldots, E_n$ its complete extensions. Then, $\langle \mathcal{A}rg,\ Defeat_r \rangle$ satisfies closure.*

Now let us consider again the problem of example 4.
Example 4 – continued: *Using the restricted version of defeat, the argument A_4 does not defeat A_7 and A_3 does not defeat A_8. Thus, the set $\{A_3, A_4\}$ is no longer an admissible extension since it does not defend itself against all its defeaters (A_7, A_8).*

We will now show that if we consider the Cl_{pp} operator and the "restricted rebutting" then the two remaining postulates (closure and indirect consistency) are satisfied under each of Dung's standard semantics.

Theorem 2 *Let $\langle \mathcal{A}rg,\ Defeat_r \rangle$ be an argumentation system built from the theory $\langle Cl_{pp}(\mathcal{S}), \mathcal{D} \rangle$ such that $\mathcal{S}$ is consistent,* `Output` *its set of justified conclusions and $E_1, \ldots, E_n$ its extensions under one of Dung's standard semantics.*
Then, $\langle \mathcal{A}rg,\ Defeat_r \rangle$ satisfies direct consistency and indirect consistency.

In the previous example, it can be seen that Cl_{pp} can generate a rule (in this case: $\neg hw \longrightarrow \neg m$) 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 $b \longrightarrow \neg m$, which corresponds

[3] Applying "restricted rebutting" instead of (unrestricted) "rebutting" also affects the validity of some of the results that have been obtained until now. Proposition 2, for instance, is not valid under restricted rebutting (Theorem 2 and 4 will repair this). Proposition 1, however, can also quite easily be proved under restricted rebutting. Furthermore, results that do not depend on the particular way in which defeat is defined (like Proposition 3, 4, 5, 6 and 7) remain valid under restricted rebutting.

to applying transitivity on the rules $b \longrightarrow \neg hw$ and $\neg hw \longrightarrow \neg m$. Worse yet, Cl_{pp} may also generate rules which are actually *harmful* for obtaining an intuitive outcome. An example of such a rule is $p, \neg p \longrightarrow \neg q$. To see why this is harmful, consider the case of two arguments for conflicting conclusions (like the Nixon diamond) p and $\neg p$. With strict rules as classical entailment, one can then combine these arguments to form an argument that can defeat an arbitrary statement (like q), as $p, \neg p \vdash \neg q$. This phenomenon is particularly problematic under grounded semantics [39,40] but also plays a role under preferred semantics [22]. Although an approach is given in [22] we will not go into details here.

5.2 Strict rules closed under transposition

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 closure and consistency, but at the same time does 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 should not generate too little, but it should not generate too much either.

We are now about to define a second closure operator Cl_{tp} that is significantly weaker than our first one (Cl_{pp}). Our discussion starts with the observation that a strict rule (say $\phi_1, \ldots, \phi_n \longrightarrow \psi$), 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 ($\phi_1, \ldots, \phi_{i-1}, \neg\psi, \phi_{i+1}, \ldots, \phi_n \longrightarrow \neg\phi_i$). This leads to the following definition.

Definition 17 (Transposition) *A strict rule s is a* transposition *of $\phi_1, \ldots, \phi_n \longrightarrow \psi$ iff $s = \phi_1, \ldots, \phi_{i-1}, \neg\psi, \phi_{i+1}, \ldots, \phi_n \longrightarrow \neg\phi_i$ for some $1 \leq i \leq n$.*

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

Definition 18 (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 $Cl_{tp}(\mathcal{S}) = \mathcal{S}$.*

It is easily verified that with the Cl_{tp} operator, example 4 (Married John) is handled correctly. More generally, the use of such an operator allows the three rationality postulates to be satisfied.

Theorem 3 *Let* $\langle\mathcal{A}rg, Defeat\rangle$ *be an argumentation system built from* $\langle Cl_{tp}(\mathcal{S}), \mathcal{D}\rangle$ *where* $Cl_{tp}(\mathcal{S})$ *is consistent,* `Output` *its set of justified conclusions and* E *its* grounded *extension. Then,* $\langle\mathcal{A}rg, Defeat\rangle$ *satisfies closure and indirect consistency.*

Note that $\langle\mathcal{A}\text{rg}, Defeat\rangle$ satisfies also consistency (according to Property 2) since the unrestricted version of the rebutting relation is considered here. As for the Cl_{pp} operator, the Cl_{tp} operator by itself is not enough to guarantee the closure and indirect consistency of an argumentation system for the other acceptability semantics (like complete and preferred semantics). Let us consider another example to illustrate this issue.

Example 7 *Let* $\mathcal{S} = \{\longrightarrow a;\ \longrightarrow b;\ \longrightarrow c;\ \longrightarrow g;\ d, e, f \longrightarrow \neg g\}$ *and* $\mathcal{D} = \{a \Longrightarrow d;\ b \Longrightarrow e;\ c \Longrightarrow f\}$.
Now, consider the following arguments:
$A_1 :\ [[\rightarrow a] \Rightarrow d]$
$A_2 :\ [[\rightarrow b] \Rightarrow e]$
$A_3 :\ [[\rightarrow c] \Rightarrow f]$
One can easily verify that without Cl_{tp}*, the arguments* A_1*,* A_2 *and* A_3 *do not have any counter-arguments (which makes them members of each Dung-style extension). However, if one would replace the defeasible theory* $\langle\mathcal{S}, \mathcal{D}\rangle$ *by* $\langle Cl_{tp}(\mathcal{S}), \mathcal{D}\rangle$*, then counter-arguments against* A_1*,* A_2 *and* A_3 do *exist. For instance,* $A_4 = [[[\rightarrow b] \Rightarrow e], [[\rightarrow c] \Rightarrow f], [\rightarrow g] \rightarrow \neg d]$ *defeats* A_1 *(because* $e, f, g \longrightarrow \neg d \in Cl_{tp}(\mathcal{S})$*).*
The counter-arguments against A_1*,* A_2 *and* A_3 *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* $\{A_1, A_2, A_3\}$ *is admissible. Even though* A_4 *defeats* A_1*,* A_1 *also defeats* A_4*, and similar observations can also be made with respect to* A_2 *and* A_3*. And because* $\{A_1, A_2, A_3\}$ *is admissible, there also exists a preferred extension (a superset of* $\{A_1, A_2, A_3\}$*) with conclusions* d*,* e*,* f *and also* g*. This means that this preferred extension does not satisfy closure. 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 closure and indirect consistency under grounded semantics, the problem is still open for preferred semantics. For this, we will consider again the argumentation system $\langle\mathcal{A}\text{rg}, Defeat_r\rangle$ with the restricted version of rebutting.
To see how the restricted rebut can help to solve the issue of postulates, consider again the problem of example 7.

Example 7 – continued: *Again consider the following arguments:*
$A_1 : [[\to a] \Rightarrow d]$
$A_2 : [[\to b] \Rightarrow e]$
$A_3 : [[\to c] \Rightarrow f]$
Under the restricted version of rebutting, it holds that $\{A_1, A_2, A_3\}$ *is not an admissible set under* $\langle Cl_{tp}(\mathcal{S}), \mathcal{D}\rangle$*. For instance, the argument* $[[[\to b] \Rightarrow e], [[\to c] \Rightarrow f], [\to g] \to \neg d]$ *(*A_4*) now rebuts* A_1 *but* A_1 *does not rebut* A_4*, nor does any other argument in* $\{A_1, A_2, A_3\}$ *defeat* A_4*. Thus* $\{A_1, A_2, A_3\}$ *is not admissible in* $\langle Cl_{tp}(\mathcal{S}), \mathcal{D}\rangle$ *under the restricted definition of rebutting.*

We will now show that if we consider the transposition closure Cl_{tp} and the restricted version of the rebutting relation then direct and indirect consistency are satisfied under each of Dung's standard semantics.

Theorem 4 *Let* $\langle \mathcal{A}rg,\ Defeat_r\rangle$ *be an argumentation system built from the theory* $\langle Cl_{tp}(\mathcal{S}), \mathcal{D}\rangle$ *with* $\mathcal{S}$ *is consistent.* `Output` *its set of justified conclusions and* $E_1, \ldots, E_n$ *its extensions under one of Dung's standard semantics. Then,* $\langle \mathcal{A}rg,\ Defeat_r\rangle$ *satisfies direct consistency and indirect consistency.*

Note that $\langle \mathcal{A}$rg, *Defeat*$\rangle$ satisfies also closure as shown by Proposition 8 in the Appendix.

5.3 Conclusions

So far, we have proposed two solutions for satisfying the three rationality postulates for the argumentation framework proposed in [4]. Table 2 summarizes the different results obtained concerning direct consistency, indirect consistency and closure. In what follows, we will use the wording "any extension" in order to refer to all Dung's standard semantics.

Type of rebutting	Direct consistency	Indirect consistency	Closure
Rebut	Under any semantics	Under grounded extension	Under grounded extension
Restricted Rebut	Under any semantics	Under any semantics	Under any semantics

Table 2
Consistency and closure with the two closure operators.

6 SUMMARY AND DISCUSSION

Although various systems for formal argumentation have been defined during recent years, many of them can easily produce results that, when given a closer inspection, are very problematic to serve as a basis for beliefs, or for any other purpose that allows for introspection on the results.

In order to avoid such problems, the aim of this paper is to define a number of postulates that any argumentation system should satisfy. These postulates should warrant that an argumentation formalism is well-defined and guarantee some basic suitability of its outputs. We have focused on three important postulates: the *closure*, the *direct consistency* and the *indirect consistency* of the results of a system. These are violated by several argumentation systems such as [33,34,42]. We then studied ways in which these postulates can be 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. Thus, the contribution of this paper is not to state an entirely new formalism for argumentation and defeasible reasoning. Instead, we have stated a number of general approaches (like transposition) that can be applied to a wide variety of argumentation formalisms, including [4,33,34,42].

It should be mentioned that the problem of rationality postulates is not necessarily connected to argumentation formalisms that use Dung-style semantics. For instance, as was explained in section 3, the formalism of [33] violates the rationality postulates of closure and indirect consistency, even though it does not use any of Dung's standard semantics. The point is that many of the problematic examples, as discussed in section 3 arise because some arguments do not have counterarguments (like is for instance the case in the Married John example), although intuitively they should have. In almost all semantics that we know of, such arguments without defeaters are in each extension; this is not only the case for the standard semantics (i.e. grounded, preferred, complete and stable), but also for semi-stable [24], ideal [1] and CF2 [13].

As for the proposed solutions, in section 5 it was stated that the approach of transposition (Cl_{tp}) in combination with restricted rebutting satisfies the rationality postulates for any of Dung's standard semantics. Actually, one could even generalize this result. As the proof of Theorem 4 works for each semantics of which the extensions are a non-empty subset of the complete extensions, this not only includes preferred, complete or grounded semantics, but also for instance ideal semantics [1] or even relatively new approaches like semi-stable semantics [24]. For all these semantics, the approach of transposition in combination with restricted rebutting satisfies the rationality postulates discussed in this paper.

Although in most of this paper, the rationality postulates are discussed using Dung's analysis of formal argumentation [28], it should be mentioned that the issue of rationality postulates is not necessarily bound to it. In fact, the rationality postulates are applicable to any formalism for defeasible reasoning (be it argumentation or a more traditional nonmonotonic logic) that uses a knowledge base containing strict and defeasible rules to entail extensions with associated conclusions. This includes approaches like [39] and [33]. The reason why we have selected Dung's approach to illustrate some of the problems is mainly because it is relatively well-known among researchers in formal argumentation and nonmonotonic logic.

A different issue is that of computational complexity of the proposed solutions. While it is true that the approach of propositional closure (Cl_{pp}) involves the usual issues of computational complexity that are associated with propositional logic, the approach of transposition (Cl_{tp}) can be seen as a more lightweight approach. It should be observed that a strict rule has at most k transpositions, where k is the size of its body. If we assume that a set $\mathcal{S}$ of n strict rules has an average body size of k then this generates at most n times k transpositions of $\mathcal{S}$. Thus, generating the necessary transpositions is a task that is *linear* to the size of $\mathcal{S}$. Experiences from the first experimental implementations [4] using Cl_{tp} indicate that the added computational complexity from transposition does not cause any serious problems.

At a first sight, the rationality postulate of closure seems to require the property of logical omniscience, but this is not necessarily the case. The point is that an implementation of an argumentation formalism may very well be query-based. For semantics like grounded, complete or preferred, it is very well possible to answer the question whether a formula p follows from at least one extension, or even from every extension [42,46] without actually computing all extensions under the semantics in question. With the query-based approach, logical omniscience is not an issue since one only generates the conclusions and arguments that one actually needs in order to answer a query.

The approach of closing the strict rules under transposition (Cl_{tp}) is to some extent comparable with the approach of Clark completion for logic programs [27]. Both make explicit information that was left implicit in the original formalization. The difference, however, is that Clark completion is related to the Closed World Assumption. That is, if something does not follow from the knowledge base, then it is assumed not to hold. Transposition, on the other hand, is based on the idea that something may not be explicitly in the original knowledge base, but it still should be assumed to hold since it logically follows from this knowledge base.

[4] See http://aspic.acl.icnet.uk/

A topic related to the one discussed in the current paper is whether one can state rationality postulates not so much with respect to the conclusions of the argumentation formalism, but with respect to the argument-based semantics applied by it. It is interesting to see that this topic has caught some recent attention. Caminada, for instance, is able to capture the traditional Dung-style semantics (grounded, preferred, complete and stable) as well as the newly invented semi-stable semantics essentially in one postulate [23]. Baroni and Giacomin apply a total of nine postulates with which they are able not only to evaluate the traditional Dung-style semantics, but also non admissibility based semantics, such as CF2 [13]. Thus, the approach of applying postulates in formal argumentation has many useful applications.

Acknowledgments

This work was supported by the Commission of the European Communities under contract IST-2004-002307, ASPIC project "Argumentation Service Platform with Integrated Components".

We would like to thank Gerard Vreeswijk for his comments and discussions, and Matt South for implementing transposition and restricted rebutting in his implementation of Argumentation System (http://aspic.acl.icnet.uk), as well as the anonymous reviewers for their useful comments.

References

[1] J. Alferes, P. Dung, and L. Pereira. Scenario semantics of extended logic programs. In A. Nerode and L. Pereira, editors, *Proc. 2nd International Workshop on Logic Programming and Non-monotonic Reasoning*, pages 334–348. MIT Press, 1993.

[2] L. Amgoud. A general argumentation framework for inference and decision making. In *Proceedings of the 21st Conference on Uncertainty in Artificial Intelligence, UAI'05*, pages 26–33, 2005.

[3] L. Amgoud, S. Belabes, and H. Prade. Towards a formal framework for the search of a consensus between autonomous agents. In *Proceedings of the 4th International joint Conference on Autonomous Agents and Multi-Agent Systems, AAMAS'2005*, pages 537–543, 2005.

[4] L. Amgoud, M. Caminada, C. Cayrol, , M. Lagasquie, and H. Prakken. Towards a consensual formal model: inference part. Technical report, In Deliverable D2.2: Draft Formal Semantics for Inference and Decision-Making. ASPIC project, 2004. www.argumentation.org.

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

[6] L. Amgoud and C. Cayrol. A reasoning model based on the production of acceptable arguments. *Annals of Mathematics and Artificial Intelligence*, 34:197–216, 2002.

[7] L. Amgoud and S. Kaci. An argumentation framework for merging conflicting knowledge bases: The prioritized case. In *8th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty, ECSQARU'2005*, pages 527–538, 2005.

[8] L. Amgoud, N. Maudet, and S. Parsons. Arguments, dialogue, and negotiation. In *Proceedings of the 14th European Conference on Artificial Intelligence*, pages 338–342, 2000.

[9] L. Amgoud and S. Parsons. An argumentation framework for merging conflicting knowledge bases. In *Proceedings of International Conference on Logics in Artificial Intelligence*, pages 27–37, 2002.

[10] L. Amgoud and H. Prade. Reaching agreement through argumentation: A possibilistic approach. In *9 th International Conference on the Principles of Knowledge Representation and Reasoning, KR'2004*, pages 175–182, 2004.

[11] L. Amgoud and H. Prade. Using arguments for making decisions: A possibilistic logic approach. In *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence, UAI'04*, pages 10–17, 2004.

[12] L. Amgoud and H. Prade. Explaining qualitative decision under uncertainty by argumentation. In *Proceedings of the 21st National Conference on Artificial Intelligence, AAAI'06*, pages 219–224, 2006.

[13] P. Baroni and M. Giacomin. Scc-recursiveness: a general schema for argumentation semantics. *Artificial Intelligence*, 168(1-2):165–210, 2005.

[14] T. J. M. Bench-Capon. Persuasion in practical argument using value-based argumentation frameworks. *Journal of Logic and Computation*, 13(3):429–448, 2003.

[15] S. Benferhat, D. Dubois, and H. Prade. Argumentative inference in uncertain and inconsistent knowledge bases. In D. Heckerman and A. Mamdani, editors, *Proc. of the 9^{th} UAI*, pages 411–419, Washington, DC, 1993. Morgan-Kaufmann.

[16] P. Besnard and A. Hunter. A logic-based theory of deductive arguments. *Artificial Intelligence*, 128 (1-2):203–235, 2001.

[17] P. Besnard and A. Hunter. Practical first-order argumentation. In *Proceedings of the 20th American National Conference on Artificial Intelligence (AAAI'05)*, pages 590–595, 2005.

[18] E. Black and A. Hunter. A generative inquiry dialogue system. In *Proceedings of the 6th International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS'07)*, 2007.

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

[20] B. Bonet and H. Geffner. Arguing for decisions: A qualitative model of decisionmaking. In F. J. e. E. Horwitz, editor, *Proc. 12th Con. on Uncertainty in Artificial Intelligence(UAI'96)*, pages 98–105, Portland, Oregon, 1996.

[21] R. Brena, C. Chesevar, and J. Aguirre. Argumentation-supported information distribution in a multiagent system for knowledge management. In *2^{nd} International Workshop on Argumentation in Multiagent Systems (ArgMAS 2005),*.

[22] M. Caminada. Contamination in formal argumentation systems. In *Proceedings of the 17th Belgium-Netherlands Conference on Artificial Intelligence (BNAIC)*, pages 59–65, 2005.

[23] M. Caminada. On the issue of reinstatement in argumentation. In M. Fischer, W. van der Hoek, B. Konev, and A. Lisitsa, editors, *Logics in Artificial Intelligence; 10th European Conference, JELIA 2006*, pages 111–123. Springer, 2006. LNAI 4160.

[24] M. Caminada. Semi-stable semantics. In P. Dunne and T. Bench-Capon, editors, *Computational Models of Argument; Proceedings of COMMA 2006*, pages 121–130. IOS Press, 2006.

[25] C. Cayrol and M.-C. Lagasquie-Schiex. Graduality in argumentation. *Journal of Artificial Intelligence Research*, 23:245–297, 2005.

[26] C. I. Chesñevar, A. Maguitman, and R. P. Loui. Logical models of arguments. *ACM Computing Surveys*, 32(4):337–383, 2000.

[27] K. Clark. Negation as failure. In H. Gallaire and J. Minker, editors, *Logic and data bases*, pages 293–322. Plenum Press, New York, NY, 1978.

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

[29] M. Elvang-Gøransson, J. Fox, and P. Krause. Dialectic reasoning with inconsistent information. In D. Heckerman and A. Mamdani, editors, *Proc. of the 9^{th} UAI*, pages 114–121, Washington, DC, 1993. Morgan-Kaufmann.

[30] G. T. F. and N. Karacapilidis. The zeno argumentation framework. In *Proceedings of the sixth international conference on Artificial intelligence and law*, pages 10 – 18. ACM Press, 1997.

[31] J. Fox and P. McBurney. Decision making by intelligent agents: logical argument, probabilistic inference and the maintenance of beliefs and acts. In *Proc. 9th International Workshop on Non-Monotonic Reasoning (NMR'2002)*, Toulouse, France, April 2002.

[32] J. Fox and S. Parsons. On using arguments for reasoning about actions and values. In *Proceedings of the AAAI Spring Symposium on Qualitative Preferences in Deliberation and Practical Reasoning, Stanford*, 1997.

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

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

[35] D. Hitchcock, P. McBurney, and S. Parsons. A framework for deliberation dialogues. In H. V. Hansen, C. W. Tindale, J. A. Blair, and R. H. Johnson, editors, *Proceedings of the Fourth Biennial Conference of the Ontario Society for the Study of Argumentation (OSSA 2001)*, Windsor, Ontario, Canada, 2001.

[36] A. Kakas and P. Moraitis. Adaptive agent negotiation via argumentation. In *Proceedings of the 5th International joint Conference on Autonomous Agents and Multi-Agent Systems, AAMAS'2006*, pages 384–391, 2006.

[37] R. P. Loui. Process and policy: resource-bounded non-demonstrative reasoning. *Computational Intelligence*, 14:1–38, 1998.

[38] P. McBurney, S. Parsons, and M. Wooldridge. Desiderata for agent argumentation protocols. In C. Castelfranchi and W. L. Johnson, editors, *Proceedings of the First International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 2002), Bologna, Italy*, pages 402–409, New York City, NY, USA, 2002. ACM Press.

[39] J. L. Pollock. How to reason defeasibly. *Artificial Intelligence*, 57:1–42, 1992.

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

[41] H. Prakken. Relating protocols for dynamic dispute with logics for defeasible argumentation. *Synthese*, 127:187–219, 2001.

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

[43] H. Prakken and G. A. W. Vreeswijk. Logics for defeasible argumentation. In D. Gabbay and F. Günthner, editors, *Handbook of Philosophical Logic*, volume 4, pages 219–318. Kluwer Academic Publishers, Dordrecht/Boston/London, second edition, 2002.

[44] G. Simari and R. Loui. A mathematical treatment of defeasible reasoning and its implementation. *Artificial Intelligence*, 53:125–157, 1992.

[45] G. A. W. Vreeswijk. Abstract argumentation systems. *Artificial Intelligence*, 90:225–279, 1997.

[46] G. A. W. Vreeswijk and H. Prakken. Credulous and sceptical argument games for preferred semantics. In *Proceedings of the 7th European Workshop on Logic for Artificial Intelligence (JELIA-00)*, number 1919 in Springer Lecture Notes in AI, pages 239–253, Berlin, 2000. Springer Verlag.

Appendix

The following two lemmas follow directly from [28].

Lemma 1 *Let $\langle \mathcal{A}, Def \rangle$ be an argumentation framework and $\mathcal{B} \subseteq \mathcal{A}$. If $\mathcal{B}$ is admissible, then $\mathcal{B} \subseteq \mathcal{F}(\mathcal{B})$.*

Proof *Suppose that $\mathcal{B}$ is admissible. Now take an arbitrary argument $A \in \mathcal{B}$. As A is defended by $\mathcal{B}$ (because $\mathcal{B}$ is admissible), it is also in $\mathcal{F}(\mathcal{B})$. Therefore, $\mathcal{B} \subseteq \mathcal{F}(\mathcal{B})$.* ■

Lemma 2 *Let $\langle \mathcal{A}, Def \rangle$ be an argumentation framework and $\mathcal{B} \subseteq \mathcal{A}$. If $\mathcal{B}$ is admissible, then $\mathcal{F}(\mathcal{B})$ is also admissible.*

Proof *Suppose $\mathcal{B}$ is admissible. In order to prove that $\mathcal{F}(\mathcal{B})$ is also admissible, we have to prove two things:*

(1) $\mathcal{F}(\mathcal{B})$ is conflict-free.
Suppose $\mathcal{F}(\mathcal{B})$ is not conflict-free. That is, there exists some $A, B \in \mathcal{F}(\mathcal{B})$ such that A defeats B. The fact that $B \in \mathcal{F}(\mathcal{B})$ means that $\mathcal{B}$ must defend B against A. That is, $\mathcal{B}$ contains some C that defeats A. But the fact that $A \in \mathcal{F}(\mathcal{B})$ means that $\mathcal{B}$ must defend A against C. Therefore, $\mathcal{B}$ must contain some D that defeats C. But then $\mathcal{B}$ would not be conflict-free. Contradiction.

(2) for each $A \in \mathcal{F}(\mathcal{B})$: A is defended by $\mathcal{F}(\mathcal{B})$.
This follows almost immediately from Lemma 1.

■

Proposition 1. Let $\langle \mathcal{A}\text{rg}, \text{Defeat} \rangle$ be an argumentation system and $E_1, \ldots, E_n$ its different extensions under one of Dung's standard semantics.
$\forall E_i \in \{E_1, \ldots, E_n\}$, $\forall\ A \in E_i$, $\texttt{Sub}(A) \subseteq E_i$

Proof *As every complete, grounded or preferred extension is also a complete extension, we only have to prove this under complete semantics. Let E be a complete extension. Suppose that $A \in E$ and $A' \in \texttt{Sub}(A)$. Suppose also that $A' \notin E$. Since E is a complete extension, then this means that either $E \cup \{A'\}$ is not conflict-free, or E does not defend A'.*

Case 1: *Suppose that $E \cup \{A'\}$ is not conflict-free. This means that $\exists B \in E$ such that B defeats A', or A' defeats B.*
Suppose that B defeats A', thus, B rebuts A' or B undercuts A' on $A'' \in \texttt{Sub}(A')$. However, $A'' \in \texttt{Sub}(A)$. This means that B defeats A to. This means that E is not conflict-free. Contradiction with the fact that E is a

complete extension.
Suppose that A' defeats B, thus, A' rebuts B or A' undercuts B on $B' \in$ $\texttt{Sub}(B)$. This means also that A defeats B on B' since $A' \in \texttt{Sub}(A)$ and $B' \in \texttt{Sub}(B)$ (according to the definitions of Rebutting and Undercutting). This means that E is not conflict-free. Contradiction with the fact that E is a complete extension.

Case 2: *Suppose that E does not defend A'. This means that $\exists B \in \mathcal{A}rg$ such that B defeats A' on some $A'' \in \texttt{Sub}(A')$ and $\nexists C \in E$ such that C defeats B. Since $A'' \in \texttt{Sub}(A)$ it holds that B defeats A. Since E is a complete extension and $A \in E$, then E defends A against B. Contradiction.*

■

Proposition 2. Let $\langle \mathcal{A}\text{rg}, \text{Defeat}\rangle$ be an argumentation system built from a theory $\langle \mathcal{S}, \mathcal{D}\rangle$ with $\mathcal{S}$ consistent, and $E_1, \ldots, E_n$ its different extensions under one of Dung's standard semantics. $\texttt{Concs}(E_i)$ is consistent for each $1 \leq i \leq n$.

Proof *Let E be a complete extension. Suppose that $\{\texttt{Conc}(A) \mid A \in E\}$ is inconsistent. This means that $\exists A, B \in E, \texttt{Conc}(A) = -\texttt{Conc}(B)$. Since E is a complete extension, E is conflict-free. This means that A does not defeat B and B does not defeat A. According to the definition of defeat, this means that A does not rebut B and B does not rebut A. Consequently, A and B are strict arguments (according to the definition of rebutting). Thus,* $\texttt{StrictRules}(A) \cup \texttt{StrictRules}(B)$ *is inconsistent. However,* $\texttt{StrictRules}(A) \cup \texttt{StrictRules}(B) \subseteq \mathcal{S}$, *and $\mathcal{S}$ is consistent. Contradiction.* ■

Proposition 3. Let $\mathcal{T}$ be a defeasible theory, $\langle \mathcal{A}, Def\rangle$ be an argumentation system built from $\mathcal{T}$. Let $E_1, \ldots, E_n$ be its extensions under one of Dung's standard semantics, and $\texttt{Output}$ be as in definition 12.
If $\texttt{Concs}(E_i)$ is consistent for each $1 \leq i \leq n$ then $\texttt{Output}$ is consistent.

Proof *Suppose that $\forall E_i$, $\{\texttt{Conc}(A) | A \in E_i\}$ is consistent. Suppose also that* $\texttt{Output}$ *is inconsistent. According to Definition 6 this means that $\exists \psi, -\psi \in$* $\texttt{Output}$. *According to Definition 12, it holds that $\forall E_i$, $\exists A_i, B_i \in E_i$ such that* $\texttt{Conc}(A_i) = \psi$ *and* $\texttt{Conc}(B_i) = -\psi$. *This means that $\forall E_i$, $\{\texttt{Conc}(A) | A \in E_i\}$ is inconsistent. Contradiction.* ■

Proposition 4. Let $\mathcal{T}$ be a defeasible theory, $\langle \mathcal{A}, Def\rangle$ be an argumentation system built from $\mathcal{T}$. Let $E_1, \ldots, E_n$ be its extensions under one of Dung's standard semantics, and $\texttt{Output}$ be as in definition 12.
If $\texttt{Concs}(E_i) = Cl_{\mathcal{S}}(\texttt{Concs}(E_i)$ for each $1 \leq i \leq n$ then $\texttt{Output} = Cl_{\mathcal{S}}(\texttt{Output})$.

Proof *Suppose that $\forall E_i$, $\{\texttt{Conc}(A) | A \in E_i\} = Cl_{\mathcal{S}}(\{\texttt{Conc}(A) | A \in E_i\})$. Suppose also that* $\texttt{Output} \neq Cl_{\mathcal{S}}(\texttt{Output})$ *then $\exists\ \psi \in Cl_{\mathcal{S}}(\texttt{Output})$ such that $\psi \notin \texttt{Output}$.*

Case 1: $\exists \phi_1, \ldots, \phi_n \longrightarrow \psi \in \mathcal{S}$ *such that* $\phi_1, \ldots, \phi_n \in \texttt{Output}$.
Since $\phi_1, \ldots, \phi_n \in \texttt{Output}$ *then for each* ϕ_k *(*$1 \leq k \leq n$*) it holds that* $\forall E_i$, $\exists A_j \in E_i$ *with* $\texttt{Conc}(A_j) = \phi_k$.
Then $\forall E_i$, $\phi_1, \ldots, \phi_n \in \{\texttt{Conc}(A) | A \in E_i\}$.
However, $\forall E_i$, $\{\texttt{Conc}(A) | A \in E_i\} = Cl_{\mathcal{S}}(\{\texttt{Conc}(A) \mid A \in E_i\})$. *This means that* $\forall E_i$, $\psi \in \{\texttt{Conc}(A) \mid A \in E_i\}$. *Consequently,* $\psi \in \texttt{Output}$. *Contradiction*

Case 2: *By induction, the above reasoning is generalized to the case where* $\phi_1, \ldots, \phi_n \in Cl_{\mathcal{S}}(\texttt{Output})$.

■

Proposition 5. Let $\mathcal{T}$ be a defeasible theory, $\langle \mathcal{A}, Def \rangle$ be an argumentation system built from $\mathcal{T}$. Let $E_1, \ldots, E_n$ be its extensions under one of Dung's standard semantics and let `Output` be as in definition 12.
If $Cl_{\mathcal{S}}(\{\texttt{Concs}(E_i))$ is consistent for each $1 \leq i \leq n$, then $Cl_{\mathcal{S}}(\texttt{Output})$ is consistent.

Proof *Suppose that* $\forall E_i$, $Cl_{\mathcal{S}}(\{\texttt{Conc}(A) | A \in E_i\})$ *is consistent. Suppose also that* $Cl_{\mathcal{S}}(\texttt{Output})$ *is inconsistent. This means that* $\exists \psi, -\psi \in Cl_{\mathcal{S}}(\texttt{Output})$.

Case 1: $\exists \psi, -\psi \in Cl_{\mathcal{S}}(\texttt{Output})$ *means that* $\exists \phi_1, \ldots, \phi_n \longrightarrow \psi \in \mathcal{S}$ *such that* $\phi_1, \ldots, \phi_n \in \texttt{Output}$ *and* $\exists \phi'_1, \ldots, \phi'_m \longrightarrow -\psi \in \mathcal{S}$ *such that* $\phi'_1, \ldots, \phi'_m \in \texttt{Output}$. *This means that* $\phi_1, \ldots, \phi_n, \phi'_1, \ldots, \phi'_m \in \{\texttt{Conc}(A) \mid A \in E_i\}$, $\forall E_i$. *As a consequence,* $\psi, -\psi \in Cl_{\mathcal{S}}(\{\texttt{Conc}(A) | A \in E_i\})$. *This means that* $Cl_{\mathcal{S}}(\{\texttt{Conc}(A) | A \in E_i\})$ *is inconsistent. Contradiction.*

Case 2: $\exists \psi, -\psi \in Cl_{\mathcal{S}}(\texttt{Output})$ *means that* $\exists \phi_1, \ldots, \phi_n \longrightarrow \psi \in \mathcal{S}$ *such that* $\phi_1, \ldots, \phi_n \in Cl_{\mathcal{S}}(\texttt{Output})$ *and* $\exists \phi'_1, \ldots, \phi'_m \longrightarrow -\psi \in \mathcal{S}$ *such that* $\phi'_1, \ldots, \phi'_m \in Cl_{\mathcal{S}}(\texttt{Output})$.
By induction, we apply the reasoning of case 1.

■

Proposition 6. If an argumentation system $\langle \mathcal{A}, Def \rangle$ satisfies indirect consistency, then it also satisfies direct consistency.

Proof *According to Definition 5,* $\texttt{Output} \subseteq Cl_{\mathcal{S}}(\texttt{Output})$. *Therefore, if* $Cl_{\mathcal{S}}(\texttt{Output})$ *is consistent, then* `Output` *is also consistent. Similarly, since* $\{\texttt{Conc}(A) \mid A \in E_i\} \subseteq Cl_{\mathcal{S}}(\{\texttt{Conc}(A) \mid A \in E_i\})$, *then if* $Cl_{\mathcal{S}}(\{\texttt{Conc}(A) \mid A \in E_i\})$ *is consistent, then* $\{\texttt{Conc}(A) \mid A \in E_i\}$ *is also consistent. Consequently, if an argumentation system satisfies indirect consistency, then it also satisfies direct consistency.* ■

Proposition 7. Let $\langle \mathcal{A}, Def \rangle$ be an argumentation system. If $\langle \mathcal{A}, Def \rangle$ satisfies closure and direct consistency, then it also satisfies indirect consistency.

Proof *Suppose that the argumentation system satisfies closure, then* `Output` $= Cl_{\mathcal{S}}($`Output`$)$. *Suppose also that the system satisfies direct consistency, then* `Output` *is consistent. Consequently,* $Cl_{\mathcal{S}}($`Output`$)$ *is also consistent.*
Similarly, we have $\{\texttt{Conc}(A) | A \in E_i\} = Cl_{\mathcal{S}}(\{\texttt{Conc}(A) | A \in E_i\})$ *(due to the closure of the system). Moreover,* $\{\texttt{Conc}(A) | A \in E_i\}$ *is consistent (because of direct consistency). Thus,* $Cl_{\mathcal{S}}(\{\texttt{Conc}(A) | A \in E_i\})$ *is also consistent. Consequently, the system satisfies indirect consistency.* ■

Lemma 3 *Let P be a set of propositions that is closed under propositional entailment (that is: $Cn(P) = P$). It holds that $Cn(Prop(Rules(P))) = P$.*

Proof *We have to prove two things:*

(1) $Cn(Prop(Rules(P) \subseteq P$
First of all, it should be mentioned that from the definitions of Prop and Rules it follows that $Prop(Rules(P)) \subseteq P$. As in propositional logic Cn is a monotonic function, it also holds that $Cn(Prop(Rules(P))) \subseteq Cn(P)$. As we have that $Cn(P) = P$, it also holds that $Cn(Prop(Rules(P))) \subseteq P$.

(2) $P \subseteq Cn(Prop(Rules(P)))$
Let $\phi \in P$. Let ϕ^{CNF} be a proposition in Conjunctive Normal Form that is logically equivalent to ϕ. Assume, without loss of generality, that ϕ^{CNF} is of the form $(p_1 \vee \ldots \vee p_n) \wedge \ldots \wedge (q_1 \vee \ldots \vee q_m)$. As P is closed under propositional entailment, it holds that $\phi^{CNF} \in P$. This means that P also contains the formulas $\neg p_1 \wedge \ldots \wedge \neg p_{n-1} \supset p_n$, ..., $\neg q_1 \wedge \ldots \wedge \neg q_{m-1} \supset q_m$. These formulas, by definition of Rules and Prop, will also be in $Prop(Rules(P))$. Together, these formulas entail ϕ^{CNF} and therefore also ϕ. Therefore, $\phi \in Cn(Prop(Rules(P)))$.

■

Property 3. Let $\mathcal{S}$ be a set of strict rules, and let $\mathcal{S}_1, \mathcal{S}_2 \subseteq \mathcal{S}$.

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

Proof
(1) $\mathcal{S} \subseteq Cl_{pp}(\mathcal{S})$. *This follows directly from Definition 14.*
(2) If $\mathcal{S}_1 \subseteq \mathcal{S}_2$ *then* $Cl_{pp}(\mathcal{S}_1) \subseteq Cl_{pp}(\mathcal{S}_2)$.
Since $\mathcal{S}_1 \subseteq \mathcal{S}_2$ then $Prop(\mathcal{S}_1) \subseteq Prop(\mathcal{S}_2)$ (According to Definition 14 of the function Prop). Due to the monotonicity of the classical inference relation $\vdash$, we have $Cn_{prop}(Prop(\mathcal{S}_1)) \subseteq Cn_{prop}(Prop(\mathcal{S}_2))$. According-

ing to the definition of the function Rules in definition 14, we have $Rules(Cn_{prop}(Prop(\mathcal{S}_1))) \subseteq Rules(Cn_{prop}(Prop(\mathcal{S}_2)))$*. Thus,* $Cl_{pp}(\mathcal{S}_1) \subseteq Cl_{pp}(\mathcal{S}_2)$*.*

(3) $Cl_{pp}(Cl_{pp}(\mathcal{S})) = Cl_{pp}(\mathcal{S})$
From the definition of Cl_{pp} *it follows that:*
$Cl_{pp}(Cl_{pp}(\mathcal{S})) = Rules(Cn(Prop(Rules(Cn(Prop(S))))))$
As $Cn(Prop(S))$ *is closed under propositional consequence, we can apply Lemma 3. From this, it follows that:*
$Rules(Cn(Prop(Rules(Cn(Prop(S)))))) = Rules(Cn(Prop(S)))$
Applying the definition of Cl_{pp} *yields:*
$Rules(Cn(Prop(S))) = Cl_{pp}(S)$
By applying transitivity on the thus derived equations, we obtain:
$Cl_{pp}(Cl_{pp}(\mathcal{S})) = Cl_{pp}(\mathcal{S})$*.*

■

Lemma 4 *Let* $\mathcal{S}$ *be a set of strict rules.* $Cl_{pp}(\mathcal{S})$ *is closed under transposition. That is:* $Cl_{tp}(Cl_{pp}(\mathcal{S})) = Cl_{pp}(\mathcal{S})$*.*

Proof *We have to prove two things:*

(1) $Cl_{pp}(\mathcal{S}) \subseteq Cl_{tp}(Cl_{pp}(\mathcal{S}))$
This follows directly from Definition 18.

(2) $Cl_{tp}(Cl_{pp}(\mathcal{S})) \subseteq Cl_{pp}(\mathcal{S})$
Let $s \in Cl_{tp}(Cl_{pp}(\mathcal{S}))$*. Then, according to Definition 18, there are two possibilities:*
   (a) $s \in Cl_{pp}(\mathcal{S})$*. In that case, we're done.*
   (b) s *is a transposition of some rule* $s' \in Cl_{pp}(\mathcal{S})$*. Let* $s' = \phi_1, \ldots, \phi_n \longrightarrow \psi$ *and* $s = \phi_1, \ldots, \phi_{i-1}, -\psi, \phi_{i+1}, \ldots, \phi_n \longrightarrow -\phi_i$*. From the fact that* $s' \in Cl_{pp}(\mathcal{S})$ *it follows that* $s \in Cl_{pp}(Cl_{pp}(\mathcal{S}))$ *(this is because* $\phi_1 \wedge \ldots \wedge \phi_n \supset \psi \vdash \phi_1 \wedge \ldots \wedge \phi_{i-1} \wedge -\psi \wedge \phi_{i+1} \wedge \ldots \wedge \phi_n \supset -\phi_i$*). From Property 3 (*$Cl_{pp}(Cl_{pp}(\mathcal{S})) = Cl_{pp}(\mathcal{S})$*) it then follows that* $s \in Cl_{pp}(S)$*.*

■

Theorem 1. Let $\langle \mathcal{A}\text{rg}, Defeat\rangle$ be an argumentation system built from the defeasible theory $<Cl_{pp}(\mathcal{S}), \mathcal{D}>$ such that $Cl_{pp}(\mathcal{S})$ is consistent. `Output` is its set of justified conclusions and E its grounded extension.
$\langle \mathcal{A}\text{rg}, Defeat\rangle$ satisfies closure and indirect consistency.

Proof *From Lemma 4 it follows that* $<Cl_{pp}(\mathcal{S}), \mathcal{D}> = <Cl_{tp}(Cl_{pp}(\mathcal{S})), \mathcal{D}>$*. From Theorem 1 it follows that the argumentation system* $\langle \mathcal{A}rg, Defeat\rangle$ *built from* $<Cl_{tp}(Cl_{pp}(\mathcal{S})), \mathcal{D}>$ *satisfies closure and indirect consistency.* ■

Property 4. Let A and B be arguments. If A restrictively rebuts B, then A rebuts B. The reverse is not always true.

Proof *Let A and B be arguments. Suppose that A restrictively rebuts B. This means that $\exists\ A' \in \mathtt{Sub}(A)$ with $\mathtt{Conc}(A') = \phi$ and $\exists\ B' \in \mathtt{Sub}(B)$ of the form $B''_1, \ldots, B''_n \Rightarrow -\phi$. B' is a non-strict argument, moreover, $\mathtt{Conc}(B') = -\phi$. Thus, A rebuts B.* ■

Proposition 8. Let $\langle \mathcal{A}\text{rg},\ Defeat_r \rangle$ be an argumentation system built from the theory $<\mathcal{S}, \mathcal{D}>$, and $E_1, \ldots, E_n$ its complete extensions. $\langle \mathcal{A}\text{rg},\ Defeat_r \rangle$ satisfies closure.

Proof *In order to prove closure, it is sufficient to show that $\forall E_i$, $\{\mathtt{Conc}(A) | A \in E_i\} = Cl_{\mathcal{S}}(\{\mathtt{Conc}(A) | A \in E_i\})$. Because, according to Proposition 4, this means that $\mathtt{Output} = Cl_{\mathcal{S}}(\mathtt{Output})$. Consequently, the argumentation system satisfies both aspects of closure (Proposition 4).*
Let E be a complete extension. Suppose that $\{\mathtt{Conc}(A) | A \in E\} \neq Cl_{\mathcal{S}}(\{\mathtt{Conc}(A) | A \in E\})$. This means that there exist arguments $A_1, \ldots, A_n \in E$ with $\mathtt{Conc}(A_1) = \phi_1$, …, $\mathtt{Conc}(A_n) = \phi_n$ and $\exists\ \phi_1, \ldots, \phi_n \longrightarrow \psi \in \mathcal{S}$, but $A = A_1, \ldots, A_n \rightarrow \psi \notin E$. Two possible cases exist:

Case 1: *$E \cup \{A\}$ is not conflict-free. Then either $\exists B \in E$ such that B defeats A, or $\exists B \in E$ such that A defeats B.*
Suppose that $\exists\ B \in E$ such that B defeats A on a sub-argument A'. Thus, $A' \in \mathtt{Sub}(A)$. However, $\mathtt{Sub}(A) = \mathtt{Sub}(A_1) \cup \ldots \cup \mathtt{Sub}(A_n) \cup \{A\}$. According to the definition of restricted rebutting and that of undercut, the top rule of A' is defeasible. Thus, $A' \in \mathtt{Sub}(A_1) \cup \ldots \cup \mathtt{Sub}(A_n)$. Then, $A' \in \mathtt{Sub}(A_1)$, or …, or $A' \in \mathtt{Sub}(A_n)$. According to Proposition 1, since $A_i \in E$ $(1 \leq i \leq n)$, then $\mathtt{Sub}(A_i) \subseteq E$ $(1 \leq i \leq n)$. Consequently, $A' \in E$. Thus, B defeats A' (according to the definition of rebutting and undercut). Thus, E is not conflict-free. Contradiction.
Now suppose that $\exists B \in E$ such that A defeats B. As E is an admissible set, it must defend itself against A. This can only be the case if E contains some argument C such that C defeats A_1 or … or A_n (this is because C cannot defeat A on A's top-rule). But then E would not be conflict-free. Contradiction.

Case 2: *E does not defend A. This means that $\exists B \in \mathcal{A}rg$ such that B defeats A and $\nexists C \in E$ such that C defeats B. Since B defeats A, it must hold that B rebuts or undercut A on a sub-argument A' whose top rule is defeasible. Thus, B rebuts or undercut A'. However, since $A' \in \mathtt{Sub}(A)$, it must hold that $A' \in \mathtt{Sub}(A_1) \cup \ldots \cup \mathtt{Sub}(A_n)$. Then, $\exists i = 1, \ldots, n$ such that $A' \in \mathtt{Sub}(A_i)$. According to Proposition 1, since $A_i \in E$, then $\mathtt{Sub}(A_i) \subseteq E$, thus, $A' \in E$. Consequently, A' is defended by E against B. Contradiction.* ■

Theorem 2. Let $\langle \mathcal{A}\text{rg}, Defeat_r \rangle$ be an argumentation system built from the theory $<Cl_{pp}(\mathcal{S}), \mathcal{D}>$ with $\mathcal{S}$ is consistent, `Output` its set of justified conclusions and $E_1, \ldots, E_n$ its complete extensions.
$\langle \mathcal{A}\text{rg}, Defeat_r \rangle$ satisfies direct consistency and indirect consistency.

Proof *From Lemma 4 it follows that* $<Cl_{pp}(\mathcal{S}), \mathcal{D}>=<Cl_{tp}(Cl_{pp}(\mathcal{S})), \mathcal{D}>$. *From Theorem 4, it follows that the argumentation system built from* $<Cl_{tp}(Cl_{pp}(\mathcal{S}), \mathcal{D}>$ *satisfies direct consistency and indirect consistency.* ■

In order to prove Theorem 3, in particular Closure, we need first to prove the following result:

Lemma 5 *Let* $\langle \mathcal{A}rg, Defeat \rangle$ *be an argumentation system built from the defeasible theory* $<Cl_{tp}(\mathcal{S}), \mathcal{D}>$. *Let* $\mathcal{B}$ *be an admissible set of arguments under this theory. If the set* $\{\texttt{Conc}(A)$ *such that* $A \in \mathcal{B}\}$ *is closed, then the set* $\{\texttt{Conc}(A)$ *such that* $A \in \mathcal{F}(\mathcal{B})\}$ *is closed as well.*

Proof *Let* $\mathcal{B}$ *be an admissible set of arguments. Suppose that* $\{\texttt{Conc}(A)$ *such that* $A \in \mathcal{B}\}$ *is closed, and that* $\{\texttt{Conc}(A)$ *such that* $A \in \mathcal{F}(\mathcal{B})\}$ *is not closed. The fact that* $\{\texttt{Conc}(A)$ *such that* $A \in \mathcal{F}(\mathcal{B})\}$ *is not closed means that there exists some rule* $\phi_1, \ldots, \phi_n \longrightarrow \psi$ *such that* $\mathcal{F}(\mathcal{B})$ *contains arguments* $A_1, \ldots, A_n$ *with* $\texttt{Conc}(A_1) = \phi_1, \ \ldots, \ \texttt{Conc}(A_n) = \phi_n$ *but no argument with conclusion* ψ. *Now consider the argument* $A = A_1, \ldots, A_n \rightarrow \psi$. *It holds that* $A \notin \mathcal{F}(\mathcal{B})$. *This means that* A *is defeated by some argument (say* B*) that is not defeated by* $\mathcal{B}$. *The fact that* $A_1, \ldots, A_n \in \mathcal{F}(\mathcal{B})$ *means that* B *does not defeat* $A_1, \ldots, A_n$. *Therefore,* B *must have conclusion* $-\psi$.
Now, let A_i *be an arbitrary element of* $\{A_1, \ldots, A_n\}$ $(1 \leq i \leq n)$ *containing at least one defeasible rule (such an argument always exists, since otherwise* B *could not defeat* A*). Let* $B'_i = A_1, \ldots, A_{i-1}, B, A_{i+1}, \ldots, A_n \rightarrow -\texttt{Conc}(A_i)$ *(such an argument can be constructed as* $Cl_{tp}(\mathcal{S})$ *is closed under transposition). The fact that* $A_i \in F(\mathcal{B})$ *means that* $\mathcal{B}$ *contains some argument (say* A'_1*) against* B'_i. *This* A'_i *cannot defeat any of* $A_1, \ldots, A_n$ *(otherwise* $\mathcal{F}(\mathcal{B})$ *wouldn't be conflict-free, since it contains* $A_1, \ldots, A_n$ *as well as* A' *(since* $\mathcal{B} \subseteq \mathcal{F}(\mathcal{B})$ *for an admissible set* $\mathcal{B}$*)), nor does* A'_i *defeat* B *(since our assumption is that* $\mathcal{B}$ *contains no defeaters of* B*). Therefore, the only way for* A'_i *to defeat* B'_i *is to rebut* $-\texttt{Conc}(A_i)$. *That is,* A'_i *has the same conclusion as* A_i.
$\mathcal{B}$ *contains arguments with conclusions* $\phi_1, \ldots, \phi_n$. *This is because of the fact that for each* A_i *with at least one defeasible rule,* $\mathcal{B}$ *contains an argument* A'_i *with the same conclusion, and for each* A_i *without any defeasible rule,* $\mathcal{B}$ *contains* A_i *(since* A_i *is strict and* $\mathcal{B}$ *is assumed to be closed under strict rules). But the fact that* $\mathcal{B}$ *is closed under strict rules means that* $\mathcal{B}$ *also contains an argument (say* C*) with* $\texttt{Conc}(C) = \psi$. *Therefore (as* $\mathcal{B} \subseteq \mathcal{F}(\mathcal{B})$*),* $\mathcal{F}(\mathcal{B})$ *also contains* C. *Contradiction.* ■

Theorem 3. Let $\langle \mathcal{A}\text{rg}, \textit{Defeat} \rangle$ be an argumentation system built from the defeasible theory $<Cl_{tp}(\mathcal{S}), \mathcal{D}>$ such that $Cl_{tp}(\mathcal{S})$ is consistent. `Output` is its set of justified conclusions and E its grounded extension.
$\langle \mathcal{A}\text{rg}, \textit{Defeat} \rangle$ satisfies closure and indirect consistency.

Proof

Closure: *In order to prove closure, it is sufficient to show that* $\{\texttt{Conc}(A)|A \in E\} = Cl_{\mathcal{S}}(\{\texttt{Conc}(A)|A \in E\})$*. This is because, under grounded semantics, there exists exactly one grounded extension.* $\texttt{Output} = Cl_{\mathcal{S}}(\texttt{Output})$*. Consequently, the argumentation system satisfies closure.*

Let E be the grounded extension, thus $E = \bigcup^{i \geq 1} \mathcal{F}(\emptyset)$*. We prove this by induction using the inductive definition of grounded semantics. Let* $\mathcal{A}_0 = \emptyset$ *and* $\mathcal{A}_{i+1} = \mathcal{F}(\mathcal{A}_i)$ $(i \geq 0)$*.*

basis $(i = 1)$ *Let $\mathcal{A}_1$ be the set of all arguments that do not have defeaters. We now prove that $\mathcal{A}_1$ is an admissible set that satisfies closure.*

Admissible *The set of all arguments that do not have any defeaters is automatically admissible.*

Closure *Suppose the conclusions of $\mathcal{A}_1$ are not closed under strict rules. Then there exists a strict rule* $\phi_1, \ldots, \phi_n \longrightarrow \psi$ *such that $\mathcal{A}_1$ contains arguments $A_1, \ldots, A_n$ with* $\texttt{Conc}(A_1) = \phi_1, \ \ldots, \ \texttt{Conc}(A_n) = \phi_n$ *but no argument with conclusion ψ. Now consider the argument* $A = A_1, \ldots A_n \rightarrow \psi$*. It holds that* $A \notin \mathcal{A}$*. This means that A has a defeater (say B). But B does not defeat $A_1, \ldots A_n$ (as the fact that* $A_1, \ldots, A_n \in \mathcal{A}$ *means they have no defeaters). Therefore, the only way B can defeat A is by having a conclusion $\neg\psi$. It must hold that at least one of $A_1, \ldots, A_n$ contains a defeasible rule (otherwise A would be strict and have no defeaters). Let* $A_i \in \{A_1, \ldots, A_n\}$ *be an argument containing at least one defeasible rule. The fact that $Cl_{tp}(\mathcal{S})$ is closed under transposition means that $Cl_{tp}(\mathcal{S})$ also contains a rule* $\phi_1, \ldots, \phi_{n-1}, \neg\psi, \phi_{i+1}, \ldots, \phi_n \longrightarrow \neg\phi_i$*. The argument* $A_1, \ldots, A_{i-1}, B, A_{i+1}, \ldots, A_n \rightarrow \neg\phi_i$ *is now a rebutter of A_i. Contradiction.*

step $(i \geq 1)$ *Let us assume that $\mathcal{A}_i$ $(i \geq 1)$ is admissible and closed. We will now prove that* $\mathcal{A}_{i+1}$ $(= F(\mathcal{A}_i))$ *is admissible and closed.*

Admissible: *This follows directly from Lemma 2.*

Closure: *This follows directly from Lemma 5.*

Indirect Consistency: *Since the argumentation system satisfies closure (above) and direct consistency (Proposition 2), then according to Proposition 7, then it also satisfies indirect consistency.*

■

Before treating Theorem 4, we first have to give some additional terminology that is used in the proof of Theorem 4 as well as in the proof of Lemma 6.

First, we define the depth of an argument.

Definition 19 *Let A be an argument. The* depth *of A (depth(A)) is:*

- *1, if A is an atomic argument, or else*
- *1 + depth(A'), where A' is a direct subargument of A such that depth(A') is maximal.*

Next, we define the depth of a rule in an argument. The problem, however, is that a rule can occur several times in an argument. In that case, the definition below simply takes the rule with the smallest depth.

Definition 20 *Let A be an argument and r be a rule applied in the construction of A. We say that the* depth *of r in A (depth(r, A)) is:*

- *0, if r is the top-rule of A, or else*
- *1 + depth(r, A'), where A' is a direct subargument of A such that depth(r, A') is minimal.*

The next thing to define is when two subarguments are at equal level in some superargument. Again, an issue is what to do when a subargument is contained in a superargument more than once. The approach of the following definition is to see if we can find *some* occurrence of a subargument A_1 that is at the same level as *some* occurrence of a subargument A_2.

Definition 21 *Let A_1 and A_2 be arguments and $A'_1 \in \mathtt{Sub}(A_1)$ and $A'_2 \in \mathtt{Sub}(A_2)$. We say that A'_1 is* at the same level *in A_1 as A'_2 in A_2 iff:*

- *A'_1 is a direct subargument of A_1, and A'_2 is a direct subargument of A_2, or else*
- *there exists a direct subargument A''_1 of A_1 and a direct subargument A''_2 of A_2 such that $A'_1 \in \mathtt{Sub}(A''_1)$ and $A'_2 \in \mathtt{Sub}(A''_2)$, and A'_1 is at the same level in A''_1 as A'_2 in A''_2.*

We say that A_1 is at the same level as A_2 in A iff A_1 is at the same level in A as A_2 in A.

To illustrate the above definitions, consider the argument $A = [[[\rightarrow c] \rightarrow d], [\rightarrow a], [[\rightarrow a] \rightarrow b] \rightarrow e]$. Here, $depth(A) = 3$, $depth(\longrightarrow a, A) = 1$, $depth(\longrightarrow c, A) = 2$, and the arguments $[\rightarrow a]$ and $[\rightarrow c]$ are at the same level in A.

Before proving Theorem 4, namely its part concerning direct consistency, we first need to prove the following result:

Lemma 6 *Let $(\mathcal{S}, \mathcal{D})$ be a defeasible theory where $\mathcal{S}$ is closed under transposition, Ass be a nonempty set of assumptions (that is, a set of strict rules with empty antecedents $\{\longrightarrow a_1, \ldots, \longrightarrow a_n\}$) and A be a strict argument* *under* $(\mathcal{S} \cup Ass, \mathcal{D})$ *such that* A *has conclusion* c *and contains the atomic subarguments* $[\rightarrow a_1], \ldots, [\rightarrow a_n]$. *There exists a strict argument* B *under* $(\mathcal{S} \cup Ass \cup \{\longrightarrow -c\}, \mathcal{D})$ *such that* B *has conclusion* $-a_i$ $(1 \leq i \leq n)$.

Proof *We prove this by induction on the depth of* A.

basis *Let us assume that the depth of* A *is 1. In that case,* A *consists of a single rule, with empty antecedent. As the set of assumptions that is used in* A *is non-empty, it follows that this rule must be an assumption of the form* $\longrightarrow a$. *Therefore, the conclusion of* A *is* a *(that is:* $c = a$*). Then, trivially, there also exists a strict argument* B *(with* $B = [\rightarrow -a]$*) under* $(\mathcal{S} \cup Ass \cup \{\longrightarrow -a\}, \mathcal{D})$ *such that* B *has conclusion* $-a$.

step *Suppose the above lemma holds for all strict arguments of depth* $\leq j$. *We now prove that it also holds for all strict arguments of depth* $j+1$. *Let* A *be a strict argument under* $(\mathcal{S} \cup Ass, \mathcal{D})$ *of depth* $j+1$ *with conclusion* c. *Let* $\texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_m) \longrightarrow c$ *be the top-rule of* A. *Let* A_i *be a direct subargument of* A *that contains the assumption* a_i. *Because* $\mathcal{S}$ *is closed under transposition, there exists a rule* $\texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_{i-1}), -c, \texttt{Conc}(A_{i+1}), \ldots, \texttt{Conc}(A_m) \longrightarrow -\texttt{Conc}(A_i)$. *The fact that* A_i *has a depth* $\leq j$ *means that we can apply the induction hypothesis. That is, there exists a strict argument (say* B'*) under* $(\mathcal{S} \cup Ass \cup \{\longrightarrow -\texttt{Conc}(A_i)\}, \mathcal{D})$ *with conclusion* $-a_i$. *Now, in* B', *substitute* $-\texttt{Conc}(A_i)$ *by the subargument* $[A_1, \ldots, A_{i-1}, -c, A_{i+1}, \ldots A_m \rightarrow -\texttt{Conc}(A_i)]$. *The resulting argument (call it* B*) is a strict argument under* $(\mathcal{S} \cup Ass \cup \{\longrightarrow -c\}, \mathcal{D})$ *with conclusion* $-a_i$.

■

Theorem 4. Let $\langle \mathcal{A}\text{rg}, \textit{Defeat}_r \rangle$ be an argumentation system built from the theory $<Cl_{tp}(\mathcal{S}), \mathcal{D}>$ with $Cl_{tp}(\mathcal{S})$ is consistent. $\texttt{Output}$ its set of justified conclusions and $E_1, \ldots, E_n$ its extensions under one of Dung's standard semantics.
$\langle \mathcal{A}\text{rg}, \textit{Defeat}_r \rangle$ satisfies direct consistency and indirect consistency.

Proof

Direct Consistency: *In order to prove consistency, it is sufficient to show that* $\forall E_i$, $\{\texttt{Conc}(A) | A \in E_i\}$ *is consistent. This is because Proposition 3 would then imply that* $\texttt{Output}$ *is also consistent. Consequently, the argumentation system satisfies consistency.*
Let E *be a complete extension. Suppose the conclusions of* E *are not consistent. Then* E *contains an argument (say* A*) with conclusion* c *and an argument (say* B*) with conclusion* $-c$. *As* $Cl_{tp}(\mathcal{S})$ *is assumed to be consistent, at least one of these two arguments must contain a defeasible rule. Let us, without loss of generality, assume that* A *contains at least one defeasible rule. Let* d *be a defeasible rule in* A *that has minimal depth. Notice that the depth of* d *must be at least 1, for if* d *were the top-rule of* A, *then* B

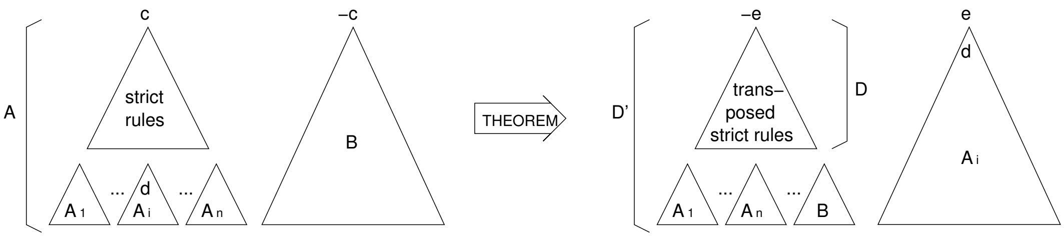

Fig. 1. Graphical representation of the proof of Theorem 4

would defeat A and E would not be conflict-free. It now holds that every rule in A with a smaller depth than d is a strict rule (see also figure 1). Let A_i *be a subargument of A that has d as its top-rule. We will now prove that there exists an argument (*D'*) in E that defeats* A_i*. Let* $A_1, \ldots, A_n$ *be the subarguments of A that are at the same level as* A_i *in A. Lemma 6 tells us that with the conclusions of* $A_1, \ldots, A_n, B$ *it is possible to construct an argument with a conclusion that is the opposite of the conclusion of* A_i*. Call this argument D. Now, let* D' *be equal to D, but with the assumptions* $\texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n), \texttt{Conc}(B)$ *substituted by the underlying arguments* $A_1, \ldots, A_n, B$*. It holds that* $D' \in E$ *(this is because each defeater of* D' *is also a defeater of* $A_1, \ldots, A_n, B \in E$*, and the fact that E is a complete extension means it defends itself against this defeater, which means that* $D' \in E$*).* D'*, however, defeats* A_i *on d, so the fact that* $D', A_i \in E$ *means that E is not conflict-free, and hence also no complete extension. Contradiction.*

Indirect Consistency: *Since the argumentation system satisfies Closure and Direct consistency, then according to Proposition 7, then it also satisfies indirect consistency.*

■