On the Limitations of Abstract Argumentation

Martin Caminada [a] Yining Wu [a]

[a] *University of Luxembourg*

Abstract

In the current paper we re-examine the three-step procedure with respect to argumentation for inference. It is observed that when viewing the argumentation process in a holistic way, one encounters several problems that tend to be overlooked when restricting oneself to pure abstract argumentation. We describe three such problems, which have to do with the interaction between abstract argumentation and instantiated (structured) arguments. We argue that these problems are related to fundamental limitations in the approach of abstract argumentation.

1 Introduction

The field of formal argumentation is based on the idea that (nonmonotonic) reasoning can be performed by constructing and evaluating arguments, which are composed of a number of reasons that together support a particular claim. Arguments distinguish themselves from proofs by the fact that they are defeasible, that is, the validity of their conclusions can be disputed by other arguments. Whether a claim can be accepted therefore depends not only on the existence of an argument that supports this claim, but also on the existence of possible counter arguments, that can then themselves be attacked by counter arguments, etc.

Nowadays, much research on the topic of argumentation is based on the abstract argumentation theory of [8]. The central concept in this work is that of an *argumentation framework*, which is essentially a directed graph in which the arguments are represented as nodes and the attack relation is represented by the arrows. Given such a graph, one can then examine the question on which set(s) of arguments can be accepted: answering this question corresponds to defining an *argumentation semantics*. It is important to keep in mind that the issue of argumentation semantics is only one specific aspect (although an important one) in the overall theory of formal argumentation. For instance, if one wants to use argumentation theory for the purpose of (nonmonotonic) entailment, one can distinguish three steps (see Figure 1). First of all, one would use an underlying knowledge base to generate a set of arguments and determine in which ways these arguments attack each other (step 1). The result is an argumentation framework, to be represented as a directed graph in which the internal structure of the arguments, as well as the nature of the attack relation has been abstracted from. Based on this argumentation framework, the next step is to determine the sets of arguments that can be accepted, using a pre-defined criterion called an argumentation semantics (step 2). After the set(s) of accepted arguments have been identified, one then has to identify the set(s) of accepted conclusions (step 3), for which there exist various approaches.

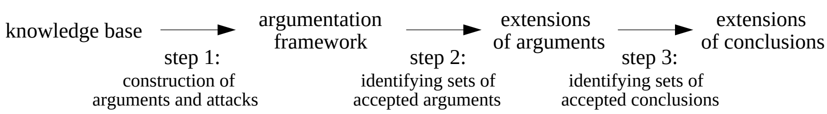

Figure 1: Argumentation for inference

Despite its advantages, the argumentation approach to non-monotonic reasoning also has important difficulties that are often overlooked by those studying purely abstract argumentation. The point is that in step 1 of the overall argumentation process, one constructs arguments that have a logical content. Yet, in step 2, one selects the sets of accepted arguments (argument-based extensions) purely based on some topological principle of the resulting graph, without looking what is actually inside of the arguments. The abstract level (step 2) is essentially about how to apply a semantics "blindly", without looking at the logical content of the arguments. But if one cannot see what is inside of the arguments, then how can one make sure that the selected set of arguments makes sense from a logical perspective? For instance, how can one be sure that the conclusions yielded by these sets of arguments (step 3) will be consistent? Or, alternatively, how does one know that these conclusions will actually be closed under logical entailment?

Issues like that of consistency and closure of argumentation-based entailment cannot be handled purely at the level of any of the individual three steps in the argumentation process. Instead, they require a carefully selected *combination* of how to carry out *each* of these individual steps, as has been pointed out in [5, 6].

The remaining part of this paper is structured as follows. In Section 2 we provide an overview of the three-step argumentation process, using a reference formalism for fully instantiated argumentation. Then, in Section 3, we use our reference formalism to state some particularly troublesome open issues in non-abstract formal argumentation. In Section 4 we round off by critically examining some commonly made implicit assumptions about the nature and scope of abstract argumentation theory.

2 An Overview of the Three-Step Argumentation Process

In the current section, we provide a brief overview of a Dung-style argumentation formalism called the ASPIC system. We have chosen to apply the version of the ASPIC system described in [6] rather than the more extended version that was subsequently published in [14]. This has been done because the simpler version of [6] is sufficient to illustrate the kind of problems we are interested in. Throughout this paper, we assume the presence of a logical language $\mathcal{L}$ that is closed under classical negation.

Definition 1 *Let $\mathcal{L}$ be a logical language closed under classical negation. We define $- : \mathcal{L} \rightarrow \mathcal{L}$ to be a function such that $-\psi = \phi$ iff $\psi = \neg\phi$, and $-\psi = \neg\psi$ otherwise.*

The ASPIC assumes a knowledge base (*defeasible theory*) that consists of a set of strict rules and a set of defeasible rules. Throughout the current paper we assume that the set of strict rules is consistent (one cannot infer both ϕ and $-\phi$ by applying strict rules only) and closed under transposition [5, 6].[1]

Definition 2 (defeasible theory) *Let $\mathcal{L}$ be a logical language that is closed under classical negation and let $\psi_1, \ldots, \psi_n, \phi \in \mathcal{L}$. A* defeasible theory *is a pair $(\mathcal{S}, \mathcal{D})$ where $\mathcal{S}$ is a set of strict (non-defeasible) rules of the form $\psi_1, \ldots, \psi_n \rightarrow \phi$ and $\mathcal{D}$ is a set of defeasible rules of the form $\psi_1, \ldots, \psi_n \Rightarrow \phi$.*

Unlike the extended version of the ASPIC system, Definition 2 does not explicitly distinguish the seperate concept of premises. These, however, can simply be modelled as strict rules with empty antecedents.

2.1 Step 1: constructing the argumentation framework

Given a knowledge base in the form of a defeasible theory, the question becomes how to construct the associated argumentation framework. The idea is to construct arguments by chaining together the strict and defeasible rules of the defeasible theory, starting with rules with empty antecendents. The conclusion of an argument (`Conc`) is the consequent of the top-rule of the argument. `SubArgs` returns all sub-arguments of the argument, and the functions `StrRules` and `DefRules` return all its strict and defeasible rules, respectively.

Definition 3 (arguments) *Let $(\mathcal{S}, \mathcal{D})$ be a defeasible theory. The following are arguments under this theory:*

strict construction *if $A_1, \ldots, A_n$ ($n \geq 0$) are arguments and there exist a strict rule $\texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \rightarrow \phi \in \mathcal{S}$ then $A_1, \ldots, A_n \rightarrow \phi$ (A) is an argument with:*

- $\texttt{Conc}(A) = \phi$
- $\texttt{StrRules}(A) = \texttt{StrRules}(A_1) \cup \ldots \cup \texttt{StrRules}(A_n) \cup \{\texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \rightarrow \phi\}$
- $\texttt{DefRules}(A) = \texttt{DefRules}(A_1) \cup \ldots \cup \texttt{DefRules}(A_n)$
- $\texttt{SubArgs}(A) = \texttt{SubArgs}(A_1) \cup \ldots \cup \texttt{SubArgs}(A_n) \cup \{A\}$

defeasible construction *if $A_1, \ldots, A_n$ ($n \geq 0$) are arguments and there exist a defeasible rule $\texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \Rightarrow \phi \in \mathcal{D}$ then $A_1, \ldots, A_n \Rightarrow \phi$ (A) is an argument with:*

[1] A strict rule s_2 is a transposition of a strict rule s_1 iff s_1 is of the form $\psi_1, \ldots, \psi_n \rightarrow \phi$ ($n \geq 1$) and s_2 is of the form $\psi_1, \ldots, \psi_{i-1}, -\phi, \psi_{i+1}, \ldots \psi_n \rightarrow -\psi_i$ for some $1 \leq i \leq n$. Closedness of the strict rules under transposition is one of the requirements for the argumentation formalism to entail reasonable conclusions. We refer to [5, 6] for details.

- $\texttt{Conc}(A) = \phi$
- $\texttt{StrRules}(A) = \texttt{StrRules}(A_1) \cup \ldots \cup \texttt{StrRules}(A_n)$
- $\texttt{DefRules}(A) = \texttt{DefRules}(A_1) \cup \ldots \cup \texttt{DefRules}(A_n) \cup \{\texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \Rightarrow \phi\}$
- $\texttt{SubArgs}(A) = \texttt{SubArgs}(A_1) \cup \ldots \cup \texttt{SubArgs}(A_n) \cup \{A\}$

We say that an argument A is strict *iff* $\texttt{DefRules}(A) = \emptyset$.

We use restricted rebutting [6] to define the notion of rebutting. The idea is that an argument can only be rebutted on the consequent of a defeasible rule.

Definition 4 *Let A and B be arguments. A rebuts B on B' iff* $\texttt{Conc}(A) = \phi$ *and* $B' \in \texttt{SubArgs}(B)$ *such that B' is of the form $B''_1, \ldots, B''_n \Rightarrow -\phi$*

The definition of undercutting, taken from [6], applies the objectification operator ($\lceil \ldots \rceil$) introduced by Pollock [13]. The idea is to translate a meta-level expression (in our case: a rule) to an object-level expression (in our case: an element of $\mathcal{L}$).

Definition 5 *Let A and B be arguments. A undercuts B on B' iff $\exists B' \in$ $\texttt{SubArgs}(B)$ such that B' is of the form $B''_1, \ldots, B''_n \Rightarrow \psi$ and* $\texttt{Conc}(A) = \neg\lceil B''_1, \ldots,$
$B''_n \Rightarrow \psi\rceil$.

Definition 6 *Let A and B be arguments. A attacks B iff A rebuts B or A undercuts B.*

Definition 7 (argumentation framework) *Let $(\mathcal{S}, \mathcal{D})$ be a defeasible theory. The associated argumentation framework is a tuple (Ar, att) where Ar is the set of all arguments that can be constructed following Definition 3 and att is the attack relation between the arguments following Definition 6.*

2.2 Step 2: applying abstract argumentation semantics

Given the argumentation framework as provided at the end of step 1 (Definition 7) the next question then becomes how to determine the associated sets of arguments that can collectively be accepted. Follow Dung's approach, determining these sets (extensions) is done without looking at the logical content of the arguments.

Definition 8 (defence / conflict-free)
Let (Ar, att) be an argumentation framework, $A \in Ar$ and $\mathcal{A}rgs \subseteq Ar$.
We define A^+ as $\{B \in Ar \mid A\ att\ B\}$ and $\mathcal{A}rgs^+$ as $\{B \in Ar \mid A\ att\ B$ for some $A \in \mathcal{A}rgs\}$.
We define A^- as $\{B \in Ar \mid B\ att\ A\}$ and $\mathcal{A}rgs^-$ as $\{B \in Ar \mid B\ att\ A$ for some $A \in \mathcal{A}rgs\}$.
$\mathcal{A}rgs$ is conflict-free *iff $\mathcal{A}rgs \cap \mathcal{A}rgs^+ = \emptyset$.*
$\mathcal{A}rgs$ defends *an argument A iff $A^- \subseteq \mathcal{A}rgs^+$.*
We define $F : 2^{Ar} \to 2^{Ar}$ as the function such that: $F(\mathcal{A}rgs) = \{A \mid A$ is defended by $\mathcal{A}rgs\}$.

Definition 9 (acceptability semantics) *Let (Ar, att) be an argumentation framework. A conflict-free set $\mathcal{A}rgs \subseteq Ar$ is called*

- *an* admissible set *iff $\mathcal{A}rgs \subseteq F(\mathcal{A}rgs)$.*
- *a* complete extension *iff $\mathcal{A}rgs = F(\mathcal{A}rgs)$.*
- *a* grounded extension *iff $\mathcal{A}rgs$ is a minimal complete extension.*
- *a* preferred extension *iff $\mathcal{A}rgs$ is a maximal complete extension.*
- *a* stable extension *iff $\mathcal{A}rgs$ is a complete extension that attacks every argument in $Ar \backslash \mathcal{A}rgs$.*

2.3 Step 3: determining the sets of justified conclusions

Depending on the particular abstract argumentation semantics, step 2 provides zero or more extensions of arguments. However, what one is often interested in for practical purposes are not so much the arguments themselves, but the *conclusions* supported by these arguments. That is, for each set (extension) of arguments, one needs to identify the associated set (extension) of conclusions.

Definition 10 *Let $\mathcal{A}rgs$ be a set of arguments whose structure complies with Definition 3. We define* $\texttt{Concs}(\mathcal{A}rgs)$ *as* $\{\texttt{Conc}(A) \mid A \in \mathcal{A}rgs\}$.

Definition 10 makes it possible to refer to the extensions of conclusions under various argumentation semantics. For instance, the extensions of conclusions under preferred semantics are simply the associated conclusions (Definition 10) of each preferred extension of arguments.

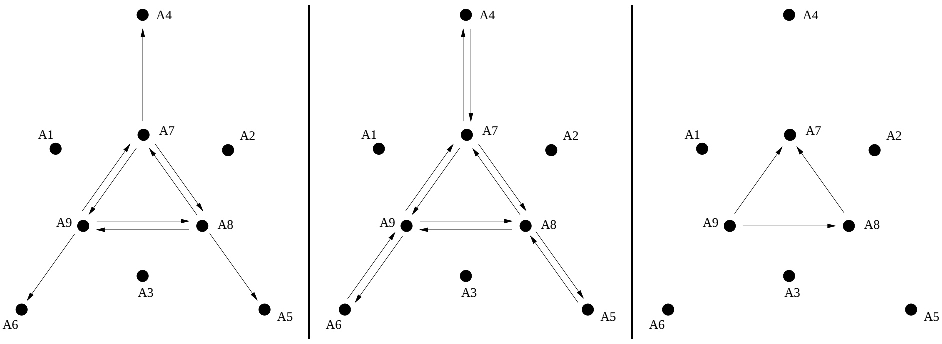

Figure 2: Three argumentation frameworks.

2.4 An example of the three step procedure

To illustrate the three step procedure of applying argumentation theory for the purpose of non-monotonic entailment, consider the example of a defeasible theory $(\mathcal{S}, \mathcal{D})$ with
$\mathcal{S} = \{\rightarrow jw; \quad \rightarrow mw; \quad \rightarrow sw; \quad mt, st \rightarrow \neg jt; \quad jt, st \rightarrow \neg mt; \quad jt, mt \rightarrow \neg st\}$ and
$\mathcal{D} = \{jw \Rightarrow jt; \quad mw \Rightarrow mt; \quad sw \Rightarrow st\}$ (notics that $\mathcal{S}$ is indeed closed under transposition in the sense of [5, 6]).
One can interpret this example as follows. John, Mary and Suzy want to go cycling in the coutryside ($\rightarrow jw$; $\rightarrow mw$; $\rightarrow sw$). They have a tandem bicycle, on which they all want to grab a seat ($jw \Rightarrow jt$; $mw \Rightarrow mt$; $sw \Rightarrow st$). However, since the tandem only has two seats, they cannot all three be on it ($mt,st \rightarrow \neg jt$; $jt,st \rightarrow \neg mt$; $jt,mt \rightarrow \neg st$). Using Definition 3, one can construct the following arguments:

$A_1 :\rightarrow jw$	$A_2 :\rightarrow mw$	$A_3 :\rightarrow sw$
$A_4 : A_1 \Rightarrow jt$	$A_5 : A_2 \Rightarrow mt$	$A_6 : A_3 \Rightarrow st$
$A_7 : A_5, A_6 \rightarrow \neg jt$	$A_8 : A_4, A_6 \rightarrow \neg mt$	$A_9 : A_4, A_5 \rightarrow \neg st$

Using the notion of attack specified in Definition 6, one obtains the argumentation framework on the left of Figure 2. Here, the grounded extension is $\{A_1, A_2, A_3\}$, yielding the associated set of conclusions $\{jw, mw, sw\}$. There are three preferred extensions (that are also stable and semi-stable): $\{A_1, A_2, A_3, A_5, A_6, A_7\}$, $\{A_1, A_2, A_3, A_4, A_6, A_8\}$ and $\{A_1, A_2, A_3, A_4, A_5, A_9\}$, yielding three associated sets of conclusions:
$\{jw, mw, sw, \neg jt, mt, st\}$, $\{jw, mw, sw, jt, \neg mt, st\}$, and $\{jw, mw, sw, jt, mt, \neg st\}$.

3 Open Issues

The first open issue to be discussed has to do with the interaction between abstract argumentation semantics and the consistency of the entailed conclusions. Caminada and Amgoud [6] specify the rationality postulates of *direct consistency*, *indirect consistency* and *closure*. To explain these postulates, first recall that each extension of arguments yields an associated extension of conclusions (step 3). Direct consistency requires that there is no extension of conclusions E such that $\phi, \neg\phi \in E$ (for some $\phi \in \mathcal{L}$). Closure requires that each extension of conclusions is closed under the strict rules, that is: $E = Cl_{\mathcal{S}}(E)$.[2] Indirect consistency requires that the closure of each extension of conclusions is directly consistent, that is, there is no extension of conclusions E such that $\phi, \neg\phi \in Cl_{\mathcal{S}}(E)$ (for some $\phi \in \mathcal{L}$).

In the example on the left of Figure 2, the extensions of conclusions generated by grounded semantics, as well as the extensions of conclusions generated by preferred semantics, satisfy direct consistency, indirect consistency and closure. This is not a coincidence. In fact, it follows from [6] that, when constructing the argumentation framework (step 1) as is done by Definition 3 and Definition 6, then applying any *admissibility-based* semantics (that is, a semantics where each extension of arguments is also an admissible set) will yield conclusions that satisfy the postulates of direct consistency, indirect consistency and closure.[3]

The results of Caminada and Amgoud are limited in the sense that they only apply to admissibility-based semantics. However, several semantics have been specified in the literature that are not admissibility-based. Examples of such are stage semantics [19] and CF2 semantics [2]. Do these semantics also yield conclusions that satisfy direct consistency, indirect consistency and closure?

[2] $Cl_{\mathcal{S}}(E)$ is the smallest set such that $E \subseteq Cl_{\mathcal{S}}(E)$ and if $\phi_1, \ldots, \phi_n \rightarrow \psi \in \mathcal{S}$ and $\phi_1, \ldots, \phi_n \in Cl_{\mathcal{S}}(E)$ then $\psi \in Cl_{\mathcal{S}}(E)$.

[3] Although Proposition 8 and Theorem 4 of [6] state this property only for semantics whose extensions are complete, the proofs actually only require these extensions to be admissible.

Consider the case of *naive semantics*, whose extensions are simply the maximal conflict-free sets of the argumentation framework. Needless to say, naive extensions are not necessarily admissible. In the case of the argumentation framework at the left of Figure 2, $\{A_1, A_2, A_3, A_4, A_5, A_6\}$ is one of the naive extensions, even though it is not an admissible set. This extension of arguments yields the associated set of conclusions $\{jw, mw, sw, jt, mt, st\}$. Hence, it puts three people on a two-person tandem. This does not only violate closure, but also violates indirect consistency, because $Cl_{\mathcal{S}}(\{jw, mw, sw, jt, mt, st\}) = \{jw, mw, sw, jt, mt, st, \neg jt, \neg mt, \neg st\}$. Hence, naive semantics does not satisfy the rationality postulates, at least not in combination with carrying out step 1 and step 3 as specified in both the current paper (as well as in [6] or [14]).

Although the argumentation framework at the left of Figure 2 is a counter example against closure and indirect consistency under naive semantics, it is not a counter example against closure and indirect consistency under stage semantics or CF2 semantics. This is because in this argumentation framework, the stage extensions as well as the CF2 extensions coincide with the preferred extensions, which are known to yield conclusions that satisfy the rationality postulates. For stage semantics it is, however, possible to come up with a slightly more complex example where a stage extension does yield inconsistent conclusions. Such an example[4] could be constructed by taking the argumentation framework at the left of Figure 2 and adding three self-attacking arguments A_{10}, A_{11} and A_{12} where A_{10} is also attacked by A_4 , A_{11} by A_5 and A_{12} by A_6 . Such arguments could be constructed by using the notion of self-undercut. In that case, $\{A_1, A_2, A_3, A_4, A_5, A_6\}$ would indeed be a stage extension, yielding conclusions that violate closure and indirect consistency.

Although admissibility appears to be a necessary condition for entailing conclusions that satisfy the rationality postulates, it is difficult to a priori rule out the existence of a non-admissibility based semantics whose entailment indeed *does* satisfy the rationality postulates. For instance, until now we have been unable to find any counter examples against CF2 semantics, nor a proof of CF2 satisfying the rationality postulates. This leads to the following open issue.

Open Issue 1 *Are there any non-admissibility based semantics whose entailment satisfies the rationality postulates?*

The results of Caminada and Amgoud seem to suggest that, as far as the rationality postulates are concerned, all admissibility-based semantics are equal. This, however, is only partly the case. The key point is that although all admissibility-based semantics indeed satisfy the rationality postulates when step 1 and step 3 are carried out as specified in the current paper (as well as in [6] and [14]), only some of these semantics still satisfy the postulates when these steps are changed.

To illustrate why it might be desirable to change the way step 1 (AF construction) is carried out, consider the fact that according to Definition 4, one can only rebut an argument based on the consequent of a *defeasible* rule. Now consider an argument that produces a conclusion by first applying a defeasible rule (say $a \Rightarrow b$), and then applying a strict rule (say $b \rightarrow c$). To some extent, the conclusion c could be seen as defeasible, since at least one defeasible rule has been involved in its entailment, even though this defeasible rule was not the last one (as is required by the concept of restricted rebutting, as implemented by Definition 4). One may wonder what would happen if one were to alter the definition of rebutting, to make it possible to rebut any conclusion that has been entailed using at least one defeasible rule. Such a definition would look as follows.

Definition 11 *Let A and B be arguments constructed according to Definition 3. We say that A* unrestrictively rebuts *B iff $\texttt{Conc}(A) = \phi$ and $\exists B' \in \texttt{SubArgs}(B)$ with B' is a non-strict argument and $\texttt{Conc}(B') = -\phi$.*

Although one could argue that the concept of unrestricted rebutting (Definition 11) is to some extent more intuitive than the concept of restricted rebutting (Definition 4), it does, however, introduce problems when it comes to satisfying the rationality postulates. Consider again the example treated earlier, of the three friends that collectively want to ride a two person tandem. If we were to construct the argumentation framework (step 1) using the concept of unrestricted rebutting instead of restricted rebutting, then not only would argument A_7 attack argument A_4, this attack would even be symmetric, because A_4 would also attack A_7. The same would hold for A_8 and A_5, and for A_9 and A_6, yielding the argumentation framework in the middle of Figure 2.

[4] This counter example was presented at COMMA 2010 and is available at: http://www.ing.unibs.it/comma2010/presentations/P15-Caminada.pdf

In the argumentation framework in the middle of Figure 2, there exist four preferred extensions. The first three are the same as for the argumentation framework on the left of Figure 2: $\{A_1, A_2, A_3, A_5, A_6, A_7\}$, $\{A_1, A_2, A_3, A_4, A_6, A_8\}$ and $\{A_1, A_2, A_3, A_4, A_5, A_9\}$. The trouble is related to the fourth extension: $\{A_1, A_2, A_3, A_4, A_5, A_6\}$. This extension yields conclusions $\{jw, mw, sw, jt, mt, st\}$ whose closure under the strict rules is $\{jw, mw, sw, jt, mt, st, \neg jt, \neg mt, \neg st\}$, which violates the rationality postulates of closure and indirect consistency. Hence, when applying unrestricted rebutting (Definition 11), it is no longer the case that every admissibility-based semantics yields conclusions that satisfy the rationality postulates. It can be observed that this is a counter-example not only against preferred semantics, but also against stable and semi-stable semantics.

The question then becomes whether there are any abstract argumentation semantics that *do* satisfy the rationality postulates even when the principle of unrestricted rebutting is applied. For instance, in the argumentation framework in the middle of Figure 2, grounded semantics seems to function fine. The grounded extension is $\{A_1, A_2, A_3\}$, which yields conclusions $\{jw, mw, sw\}$, thus satisfying (direct and indirect) consistency, as well as closure. Caminada and Amgoud prove that this is a general phenomenon. When constructing the argumentation framework using the principle of unrestricted rebutting, applying grounded semantics will always satisfy the rationality postulates [6].

Grounded semantics, however, has been criticized for taking an overly sceptical approach. In this context, additional unique-status semantics have been specified like ideal [7] and eager [4]. Although both of these semantics satisfy the rationality postulates in the context of the particular argumentation framework in the middle of Figure 2, it is not clear whether this is also a general phenomenon. This leads to the following open issue.

Open Issue 2 *Are there any abstract argumentation semantics, apart from grounded, that satisfy the rationality postulates when applying unrestricted rebutting?*

The interaction between the abstract level (step 2) and the argumentation framework construction level (step 1) is also problematic for approaches that try to do knowledge representation purely on the abstract level, without explicitly taking into account the entailment of *conclusions* by applying one or more *reasons* (as represented by the sets of strict and defeasible rules). An example of such an approach is value-based argumentation [3].

The idea of value-based argumentation is to assign to each argument a particular *value*. Different audiences can assign different preferences regarding these values. If, to a particular audience, the value of an argument A is lower than the value of an argument B, then any attack of A to B that might have existed in the argumentation framework is effectively neutralized, and is in essence erased from the argumentation framework as far as the audience is concerned.

To understand the problems related to the current approach of value-based argumentation, consider again the argumentation framework at the left of Figure 2. Arguments A_4, A_5 and A_6 seem to promote the value of "happiness", since these are about people using the tandem to enjoy a day off in the country side. Assume that the arguments A_1, A_2, A_3, A_7, A_8 and A_9 all have distinct values. Now consider an audience that, perhaps on philosophical grounds, puts the value of happiness above all other values. Furthermore, for the sake of argument, assume that this audience puts the value of A_9 above the value of A_8, and the value of A_8 above the value of A_7. Then, to this audience, the attack from A_7 to A_4 is effectively removed, as well as the attack from A_8 to A_5, the attack from A_9 to A_6, the attack from A_8 to A_9, the attack from A_7 to A_8 and the attack from A_7 to A_9. This yields the argumentation framework on the right of Figure 2. Here, there exist just one preferred extensions of arguments: $\{A_1, A_2, A_3, A_4, A_5, A_6, A_9\}$, which yields the associated extension of conclusions: $\{jw, mw, sw, jt, mt, st, \neg st\}$, hence violating direct consistency, indirect consistency and closure.

Here we see the limitations of pure abstract argumentation theories. Value based argumentation, like any other theory that is based on the paradigm of pure abstract argumentation, applies its principles "blindly", without taking into account the actual logical contents of the arguments. But if one selects sets of arguments without looking at their logical contents, then how does one know that their collective conclusions will be consistent, or satisfy any other reasonable properties?

It appears that there are two possible ways towards resolving the consistency problem of value-based argumentation. The first would be to have the values not specified on the abstract level, but instead on the level of the knowledge base, as some kind of audience-dependent ordering on the defeasible rules. In that way, one could apply existing work like [12] or [14] to obtain entailment that satisfies the rationality postulates. Another way would be to apply the values not by altering the argumentation framework (removing attacks) but by selecting particular extensions of the *unaltered* argumentation framework. The idea is that one extension is considered more desirable than the other if the values the former extension promotes are considered to be higher than the values the latter extension. This approach allows the application of a standard admissibility-based semantics on the unaltered argumentation framework. The results of [6, 14] then state that each resulting extension will satisfy the rationality postulates. The idea is then to select among extensions that already satisfy these postulates those which we like most desirable according to our values.[5]

The solutions proposed in the previous paragraph are in essence no more than a rough sketch. The relevant open issue can be described as follows.

Open Issue 3 *In which way can the approach of value-based argumentation be repaired in order to yield consistent conclusions?*

4 Discussion

The examples in the previous section illustrate a key problem in much of today's argumentation research, of which we will now take a closer look. The work on abstract argumentation theory [8] started partly as an abstraction of existing formalisms for nonmonotonic reasoning. Dung himself carried out this type of research with respect to logic programming under stable [9] and well-founded semantics [18], as well as with respect to Default Logic [17] and Pollock's inductive version of the OSCAR system [13]. Other research that follows this line includes the argument-based interpretation of Nute's Defeasible Logic [10] and the argument-based interpretation of logic programming under the 3-valued stable model semantics [21]. The basic procedure in this type of research is that one starts with an existing formalism for nonmonotonic reasoning, and reinterprets it through the three-step argumentation procedure.[6] The abstract level (step 2) thus serves as an abstraction of an *actual* formalism for non-monotonic reasoning.

Over the years, however, the research attention appears to have shifted. Instead of regarding argumentation theory in a holistic way, taking into account all three steps of the process, researchers have become more and more focussed on just one step in the overall entailment process: the abstract level (step 2). When examining much of today's work in formal argumentation, one can sense a widespread feeling among researchers that the abstract level is the thing that matters most, and that any improvements and enhancements (like values [3], bipolarity [1] or alternative semantics [19, 2]) should be implemented by taking the abstract level as a basis. This has led to a new kind of research that has produced new abstract theories without specifying any real *instances* of these abstract theories.

Value-based argumentation is essentially a modification of Dung's abstract argumentation theory, but unlike Dung's original work, it does not specify how value-based argumentation frameworks can be generated from an underlying knowledge base, or what is the overall entailment yielded by them. Similarly, stage semantics is purely specified at the abstract level, which distinguishes it from other semantics like grounded, stable and preferred, which are abstractions of actual underlying systems (like logic programming under well-founded semantics [18, 8], Default Logic [17, 8] or Pollock's 1995 version of the OSCAR system [13, 11]).

The basic assumption made by much of today's abstract argumentation research is that by specifying things on the abstract level, one obtains a theory that is very general, since it is not bound to a particular argument form, making it possible to interoperate with any arbitrary form of arguments. The experiences with stage semantics and value-based argumentation, however, indicate that this is not the case. One cannot expect any arbitrary formalism for abstract argumentation to be applied in the context where arguments consist of reasons (rules) that support claims (conclusions), at least not when one expects the outcome to satisfy some reasonable properties (like consistency and closure). Hence, the fact that a theory is abstract does not necessarily mean that it is general. In many cases it simply means that the authors did not give the issue of instantiation any consideration.

The issue of instantiation is more than just a technical concern. Argumentation, as it happens in the world around us, is almost never completely abstract. How often does one open a newspaper or magazine and read completely abstract arguments? Instead, the arguments one encounters in daily life consist of *reasons* that

[5] A similar approach, in the context of preferences, was recently proposed in [20].

[6] For instance, one can start by interpreting a default theory as a knowledge base. Based on this one then starts to construct arguments (step 1), applies stable semantics (step 2) and determines the conclusions associated with the stable extensions of arguments (step 3). One has successfully modelled default logic iff the stable extensions of conclusions (the result of step 3) are precisely the same as the default extensions in Reiter's original theory.

support particular *claims*. These reasons can formally be modelled in the form of *rules*, that are instances of underlying *argument schemes* [15]. If our community's abstract argumentation theories are fundamentally unable to interact in a meaningful way with combinations of reasons that support claims, then how can one defend that these theories have any real value for modelling the notion of argument?

In our opinion, what is needed is a fundamental rethinking of the way formal argumentation research is carried out. Instead of taking an *abstraction* as the basis for enhancement (as was for instance done in [19, 3]), it seems to make more sense to start with a *fully instantiated system* that explicitly takes into account reasons (rules) and claims (conclusions). It is such a fully instantiated system that should serve as a basis to define further enhancements and additional functionality. An example of such research is [12], which takes the fully instantiated theory of [16] as a basis and broadens it by making it applicable to various semantics that were not originally supported by [16]. Although it is always possible to subsequently define abstractions of fully instantiated systems, it is clear that the fully specified system should come *before* the abstraction, and not the other way around, in order to avoid problems as illustrated in the current paper.

References

[1] L. Amgoud, C. Cayrol, and M.-C. Lagasquie-Schiex. On the bipolarity in argumentation frameworks. In *Proc. of the 10th NMR workshop, Uncertainty Framework subworkshop*, pages 1–9, 2004.

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

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

[4] M.W.A. Caminada. Comparing two unique extension semantics for formal argumentation: ideal and eager. In *Proc. of the 19th Belgian-Dutch Conference on Artificial Intelligence* , pages 81–87, 2007.

[5] M.W.A. Caminada and L. Amgoud. An axiomatic account of formal argumentation. In *Proceedings of the AAAI-2005*, pages 608–613, 2005.

[6] M.W.A. Caminada and L. Amgoud. On the evaluation of argumentation formalisms. *Artificial Intelligence*, 171(5-6):286–310, 2007.

[7] P. M. Dung, P. Mancarella, and F. Toni. Computing ideal sceptical argumentation. *Artificial Intelligence*, 171(10-15):642–674, 2007.

[8] 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.

[9] M. Gelfond and V. Lifschitz. The stable model semantics for logic programming. In *Proc. of the 5th International Conference/Symposium on Logic Programming*, pages 1070–1080. MIT Press, 1988.

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

[11] H. Jakobovits and D. Vermeir. Robust semantics for argumentation frameworks. *Journal of logic and computation*, 9(2):215–261, 1999.

[12] S. Modgil. Reasoning about preferences in argumentation frameworks. *Artif. Intell.*, 173:901–1040, 2009.

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

[14] H. Prakken. An abstract framework for argumentation with structured arguments. *Argument and Computation*, 1(2):93–124, 2010.

[15] H. Prakken. On the nature of argument schemes. In *Dialectics, Dialogue and Argumentation. An Examination of Douglas Walton's Theories of Reasoning and Argument*, pages 167–185. College Pub., 2010

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

[17] R. Reiter. A logic for default reasoning. *Artificial Intelligence*, 13:81–132, 1980.

[18] A. van Gelder, K.A. Ross, and J.S. Schlipf. The well-founded semantics for general logic programs. *J. ACM*, 38(3):620–650, 1991.

[19] B. Verheij. Two approaches to dialectical argumentation: admissible sets and argumentation stages. In *Proc. of the Eighth Dutch Conference on Artificial Intelligence (NAIC'96)*, pages 357–368, 1996.

[20] T. Wakaki. Preference-based argumentation capturing prioritized logic programming. In *Proceedings ArgMAS 2010*, 2010.

[21] Y. Wu, M.W.A. Caminada, and D.M. Gabbay. Complete extensions in argumentation coincide with 3-valued stable models in logic programming. *Studia Logica*, 93(1-2):383–403, 2009.