Arguments and Defeat in Argument-Based Nonmonotonic Reasoning

Bart Verheij

University of Limburg, Department of Metajuridica
P.O. Box 616, 6200 MD Maastricht, The Netherlands
bart.verheij@metajur.rulimburg.nl, http://www.cs.rulimburg.nl/~verheij/

Abstract. Argument-based formalisms are gaining popularity as models of nonmonotonic reasoning. Central in such formalisms is a notion of argument. Arguments are formal reconstructions of how a conclusion is supported. Generally, an argument is defeasible. This means that an argument supporting a conclusion does not always justify its conclusion: the argument can be defeated. Whether a conclusion supported by an argument is justified depends on the structure of the argument and on the other arguments available.

In this paper, we argue for four points that are refinements of how arguments and defeat have been used in argument-based nonmonotonic reasoning. First we argue that an argument can be defeated because it contains a weak sequence of steps; second that arguments accrue, which means that arguments for a conclusion reinforce each other; third that defeat can be compound, which means that groups of arguments can defeat other groups of arguments; fourth that defeated arguments must be distinguished from not yet considered arguments. In related work these points are overlooked, or even denied. We describe a formalism that incorporates them.

1 Introduction

Recently, several formalisms for nonmonotonic reasoning have been proposed that are argument-based. In this paper, we argue for the following points that have been overlooked, or even denied.

1. An argument can be defeated because it contains a weak *sequence* of steps.
2. Arguments *accrue*, i.e., arguments for a conclusion reinforce each other.
3. Defeat can be *compound*, i.e., groups of arguments can defeat other groups of arguments.
4. *Defeated* arguments must be distinguished from *not yet considered* arguments.

In the next section, we discuss what distinguishes argument-based formalisms from other formalisms for nonmonotonic reasoning. Section 3 contains the main points of the paper. In section 4, a formalism is described that incorporates them. In the last section, we summarize the conclusions of the paper.

2 Why argument-based?

Argument-based formalisms can be distinguished from nonmonotonic logics in general by a notion of 'argument'. An argument is a reconstruction of how a conclusion is supported. Arguments can consist of several steps from their premises to their conclusion. In this sense, arguments are similar to proofs. Unlike proofs, however, arguments are not strict, but defeasible. Arguments can be defeated by other arguments. If an argument is defeated, it does not justify its conclusion.

So, there is a close relation between the defeat of arguments and the justification of conclusions. This results in two main reasons to take the arguments into account to find out which conclusions are justified. First, the *structure* of the argument determines whether it is defeated or not. Second, whether an argument is defeated is determined by *other* available arguments.

The structure of an argument. An argument can be defeated if it is not sufficiently cogent to support its conclusion. The cogency of an argument is influenced by its structure. For instance, an argument is more cogent if it contains less weak steps, and if it contains more information to support its conclusion. When one only considers conclusions and single argument steps, the influence of the structure of arguments is overlooked.

Other arguments. An argument can be defeated by other arguments. For instance, there can be an exception to the conclusion of an argument or to a step in an argument. Arguments can attack other arguments, resulting in the defeat of the attacked arguments. Arguments can also reinforce each other, so that they remain undefeated. As a result, it is not sufficient to consider arguments in isolation.

Examples of argument-based formalisms are those described by Loui (1987), Pollock (1987-1994), Nute (1988), Lin (1993), Vreeswijk (1991, 1993), Dung (1993), Prakken (1993), and Verheij (1995).

3 The defeat of arguments

In this section, we discuss four points concerning the defeat of arguments that form the crux of this paper.

3.1 Defeat by sequential weakening

In most argument-based formalisms, an argument can be defeated in two ways: at the conclusion (or an intermediate conclusion) and at a step. Let's for instance consider the argument that it will be a sunny day since the weather forecast says so. If we look out the window and see that it is raining, the conclusion simply is false. In this case we say that the argument is defeated *at the conclusion*. If we learn that we mistakenly read the weather forecast in yesterday's paper, the argument step breaks down: yesterday's forecast does not say much about the weather today. In this case we say that the argument is defeated *at the step*.

We think that there is a third way in which an argument can be defeated: *at a sequence of steps*. The reason for this is that an argument gets weaker at each step. If the chain of steps gets too long, it breaks down.

An extreme example of this is the Sorites paradox. The basis of the paradox is the argument step that taking a grain of sand from a heap leaves you with a heap. In principle, this argument step can be repeated many times. This leads to a long argument that supports the conclusion that you are left with a heap of sand. But some reflection shows that in the end the argument becomes unacceptable: after taking away the last grain of sand we are certainly not anymore left with a heap. An explanation of this paradox is that the more steps the argument contains the weaker it becomes, until it does not anymore justify the conclusion, and is defeated.

We call this the *sequential weakening* of arguments. Only Vreeswijk's (1991) formalism allows for defeat by sequential weakening. This is however hidden in his conclusive force relation, and left implicit.

3.2 Accrual of arguments

The following example is taken from Verheij (1994, 1995). Assume that John has robbed someone, so that he should be punished (α_1). Nevertheless, a judge decides that he should not be punished, because he is a first offender (β). Or, assume that John has injured someone, and should therefore be punished (α_2). Again, the judge decides he should not be punished, being a first offender (β). Now assume John has robbed and injured someone at the same time, so that there are two arguments for punishing him (α_1, α_2). In this case, the judge might decide that John should be punished, even though he is a first offender (β).

This is an example of what Pollock (1991) has called the *accrual* of arguments. The arguments α_1 and α_2 together give better support to the conclusion that John must be punished than on their own. As a result, they can on their own be defeated by the argument β, but together remain undefeated. We have the following situation:

- The argument β defeats the argument α_1, if α_1 and β are the arguments available.
- The argument β defeats the argument α_2, if α_2 and β are the arguments available.
- The arguments α_1 and α_2 defeat the argument β, if α_1, α_2 and β are the arguments available.

Even though Pollock (1991) finds it a natural supposition that arguments reinforce each other in such a way, he surprisingly rejects it. We do not agree, and think that arguments can accrue. Pollock's main point against the accrual of arguments is the following thought experiment. He asks to imagine a linguistic community in which speakers tend to confirm each other's statements, only when they are fabrications. So, in this community it is not true that arguments, based on speakers' testimonies, accrue. Indeed, two equal testimonies reduce their value to zero.

In our opinion, this is not an argument against the accrual of arguments in general, but only an example that shows that defeat information can be overruled by more specific defeat information. *Normally*, different arguments for a conclusion make the conclusion more plausible. *In exceptional situations*, however, such as in Pollock's thought experiment, this is not the case.

The idea to incorporate accrual of arguments in a formalism for defeasible reasoning is inspired by the research on Reason-Based Logic (Hage, 1993; Hage and Verheij, 1994; Verheij, 1994).

3.3 Compound defeat

In other formalisms, only single arguments defeat single arguments. We think however that defeat can be *compound*, which means that groups of arguments defeat other groups of arguments. We give two situations that involve compound defeat. First, defeat by accruing arguments is compound. Second, skeptical defeat is compound.

Defeat by accruing arguments. We have already seen an example of defeat by accruing arguments in the previous subsection. In the example two arguments were on their own defeated by another, but could together defeat the latter. In this case, a group of arguments defeats another argument, and the defeat is compound.

Skeptical defeat. The second reason why we think that defeat can be compound involves our view on the distinction between skeptical and credulous reasoning. In nonmonotonic reasoning it can be the case that incompatible conclusions are supported. Now, a skeptical reasoner withholds from drawing a conclusion, while a credulous reasoner considers both conclusions as separate possibilities.

Both options are reasonable, and all reasoning formalisms we know make a choice: they are either skeptical or credulous. We propose a formalism in which no choice is made and both skeptical and credulous reasoning can be modeled. We can do this because we think that skeptical reasoning involves compound defeat: the arguments with incompatible conclusions are together defeated. In this case, a group of arguments is defeated at once, and the defeat is compound.

Vreeswijk's (1991) formalism suggests a restricted form of compound defeat. Among a group of arguments that leads to a contradiction one argument is defeated, if it is not better (with respect to a given conclusive force relation) than the other arguments in the conflict. So, the group of undefeated arguments can be considered to defeat the defeated argument. This is however left implicit.

Our notion of compound defeat should not be confused with Pollock's (1994) notion of *collective* defeat, that is only a variant of the choice for skeptical reasoning.

3.4 Defeated vs. not yet considered arguments

Argumentation is a process. Not all information, in the form of arguments, is available at once. At each stage of argumentation new arguments are taken into account. If arguments are defeasible, this results in the possible change of the status of arguments, depending on which arguments have been considered.

An overlooked aspect of argumentation is the influence of the defeated arguments that have been considered. A good example can be given in case of accrual of arguments (section 3.2). Suppose we have again the situation that there are three arguments, denoted α_1, α_2 and β, available to a reasoner, and that the arguments α_1 and α_2 are both on their own defeated by β, but together remain undefeated, and even defeat β.

There are several orders in which the arguments can be taken into account by a reasoner, such as first α_1, then α_2, and finally β or first α_2, then β, and finally α_1. In figure 1, these orders are shown in a diagram. Each node in the diagram represents an argumentation stage and has a label representing which arguments are undefeated at that stage. The 0 represents that no argument has been considered yet. Each arrow denotes that a new argument is taken into account. For instance, if at stage α_2 the argument β is taken into account, α_2 is defeated, and only β is not. If at stage α_2 the argument α_1 is taken into account, both α_2 and α_1 are undefeated.

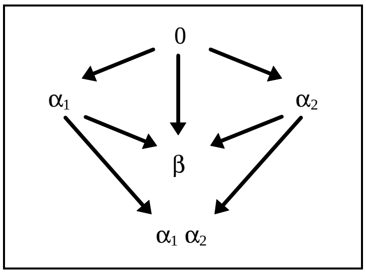

Fig. 1. A first attempt

Figure 1 is wrong, for two reasons:

1. *It does not properly represent all different stages of the argumentation process.*

For instance, the stage that only β has been taken into account is represented by the same node as the stage that both α_2 and β have been taken into account.

2. *Orders of argumentation have disappeared.*
   For instance, there is no arrow from β to α_1 α_2, because it has become unclear what it means to go from stage β to stage α_1 α_2 by taking one extra argument into account.

The picture is wrong because defeated arguments are not distinguished from not yet considered arguments. Figure 2 is the right picture. Each corner of the 'block' in the picture again represents a stage in the argumentation process, and has a label representing which arguments have been considered at that stage. The arguments in brackets are defeated. Again, each arrow denotes that a new argument is taken into account. For instance, the arrow from α_2 to β (α_2) means that the argument α_2 becomes defeated after β has been taken into account.

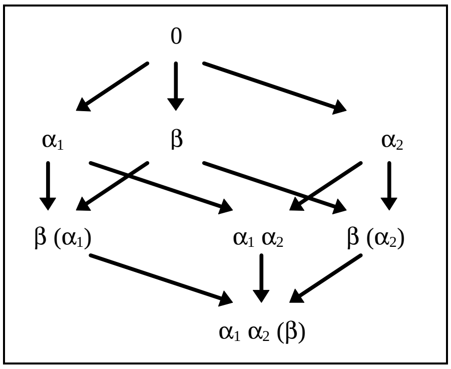

Fig. 2. The right picture

In this picture all different stages of the argumentation process can be distinguished and no orders of argumentation have disappeared. The intermediate stages β (α_1), β (α_2), and α_1 α_2 (β) dissolve the problems.

In our formalism, argumentation stages are represented as in figure 2; they contain not only the undefeated, but also the defeated arguments at the stage. Other formalisms either do not treat argumentation as a sequence of stages, or neglect the influence of the defeated arguments.

4 An appropriate formalism

In this section, we describe a formalism that incorporates the main points of this paper. We start with the formal definition of *arguments* that represent how conclusions are supported. Then we formally define *defeaters* that represent when arguments defeat other arguments. In the last subsection, we formally define *argumentation stages*. They represent which arguments have been considered at a stage of the argumentation process, and which of them are defeated at that stage.

4.1 Arguments

Our notion of an argument is related to that of Lin (1993) and Vreeswijk (1991, 1993), and is basically a tree of sentences in some language. Our approach to argumentation is independent of the choice of a language. Therefore, we treat a language as a set without any structure. A language does not even contain an element to denote negation or contradiction.

Definition 1. A *language* is a set, whose elements are the *sentences* of the language.

Lin (1993), Vreeswijk (1991, 1993) and Dung (1993) do more or less the same. Lin and Shoham use a language with negation, and Vreeswijk one with contradiction. Dung even goes a step further, and uses completely unstructured arguments.

The structure of an argument is like a proof. An argument supports its conclusion (relative to its premises), but unlike a proof, an argument is defeasible. Any argument can be defeated by other arguments. Each argument has a *conclusion* and *premises*. An argument can contain arguments for its conclusion. Arguments contain *sentences*, and have *initial* and *final* parts. A special kind of argument is a *rule*.

Definition 2. Let L be a language. An *argument* in the language L is recursively defined as follows:

1. Any element s of L is an argument in L. In this case we define

   Conc(s) = s
   Prems(s) = Sents(s) = Initials(s) = Finals(s) = {s}

2. If A is a set of arguments in L, s an element of L, and s $\notin$ Sents[A],[1] then A → s is an argument in L. In this case we define

   Conc(A → s) = s
   Prems(A → s) = Prems[A]
   Sents(A → s) = {s} $\cup$ Sents[A]
   Initials(A → s) = {A → s} $\cup$ Initials[A]
   Finals(A → s) = {s} $\cup$ {B → s | $\exists$f: f is a surjective function from A
   onto B, such that $\forall\alpha$: f(α) $\in$ Finals(α)}

Conc(α) is the *conclusion* of α. An element of Prems(α), Sents(α), Initials(α), and Finals(α) is a *premise*, a *sentence*, an *initial argument*, and a *final argument* of α, respectively. The conclusion of an initial argument of α, other than the argument α itself, is an *intermediate conclusion* of α. An argument in L is a *rule*, if it has the form S → s, where S $\subseteq$ L and s $\in$ L. For each argument α we define the set of arguments Subs(α), whose elements are the *subarguments* of α:

Subs(α) = Initials[Finals(α)]

A *proper subargument* of an argument α is a subargument other than α. If α is a subargument of β, then β is a *superargument* of α. A subargument of an argument α that is a rule is a *subrule* of α.

Notation. If A is finite, i.e. A = {α_1, α_2, ..., α_n}, we write α_1, α_2, ..., α_n → s for an argument A → s = {α_1, α_2, ..., α_n} → s, if no confusion can arise.

Intuitively, if A → s is an argument (in some language L), the elements of A are the arguments supporting the conclusion s. It may seem strange that also sentences are considered to be arguments. An argument of the form s, where s is a sentence in the language L, represents the degenerate (but in practice most common) kind of argument that a sentence is put forward without any arguments supporting it.

Some examples of arguments in the language L = {a, b, c, d} are {{a} → b} → c and {{a} → c, {b} → c} → d. They are graphically represented in figure 3.

[1] If f: V → W is a function and U $\subseteq$ V, then f[U] denotes the image of U under f.

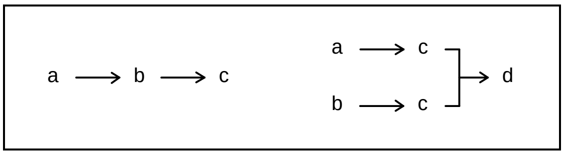

Fig. 3. Examples of arguments

The *premise*s of the argument $\{\{a\} \rightarrow c, \{b\} \rightarrow c\} \rightarrow d$ are a and b. It has d as its conclusion. Some of its initial arguments are b, $\{a\} \rightarrow c$ and the argument itself. Some of its final arguments are d, $\{c\} \rightarrow d$, and $\{c, \{b\} \rightarrow c\} \rightarrow d$. Among its subarguments are c and $\{b\} \rightarrow c$.

The structure of our arguments differs from those of Lin (1993) and Vreeswijk (1991, 1993). In these formalisms, the conclusion of an argument (or an intermediate conclusion) can only be supported by a single argument. Because we think that arguments accrue, in our formalism the same conclusion can be supported by several arguments. As a result, we can make *parallel strengthening* (and *weakening*) of an argument explicit. Intuitively, an argument becomes stronger if more arguments support its conclusion and intermediate conclusions. For instance, the argument $\{\{a\} \rightarrow c, \{b\} \rightarrow c\} \rightarrow d$ is a *strengthening* of the argument $\{\{b\} \rightarrow c\} \rightarrow d$. The former contains $\{a\} \rightarrow c$ and $\{b\} \rightarrow c$ to support the intermediate conclusion c, while the latter only contains $\{b\} \rightarrow c$.

Definition 3. Let L be a language. For any argument α in the language L we recursively define a set of arguments $\text{Weaks}(\alpha)$:

1. For $\alpha = s$, $s \in L$,

   $\text{Weaks}(s) = \{s\}$.

2. For $\alpha = A \rightarrow s$, $A \subseteq \text{Args}(L)$, $s \in L$,

   $\text{Weaks}(A \rightarrow s) = \{B \rightarrow s \mid B \subseteq \text{Weaks}[A] \text{ and } \text{Conc}[B] = \text{Conc}[A]\}$

An element of $\text{Weaks}(\alpha)$ is a *weakening* of α. A weakening of α, other than α, is a *proper* weakening of α. If α is a weakening of β, then β is a *strengthening* of α.

Weakenings are in general not subarguments. For instance, $\{\{a\} \rightarrow c\} \rightarrow d$ is not a subargument of $\{\{a\} \rightarrow c, \{b\} \rightarrow c\} \rightarrow d$.

4.2 Defeaters

Arguments are defeasible. In our formalism, *all* arguments can be defeated. Except for Dung (1993), other authors have separate classes of strict and defeasible arguments. In our formalism, arguments remain undefeated, if there is no information that makes them defeated. So, if one wants a class of strict arguments, for instance, to model deductive argumentation, it can be defined, by not allowing information that leads to the defeat of the arguments in that class. In our formalism this is straightforward, because the defeat of arguments is the result of defeat information that is *explicit* and *direct*.

Explicit defeat information. Pollock's (1987-1994) defeaters, Prakken's (1993) kinds of defeat, Vreeswijk's (1991, 1993) conclusive force, and Dung's (1993) attacks are examples of explicit defeat information. Instead of hiding the information in a general procedure, for instance based on specificity, explicit information determines which arguments become defeated and which remain undefeated. Explicit defeat information is required because no general procedure can be flexible enough to be universally valid.

Direct defeat information. By direct defeat information, we mean explicit defeat information directly specifying when arguments are defeated. Pollock's defeaters and Dung's attacks are examples of direct defeat information. Explicit defeat information is not always direct. Examples of indirect defeat information are Prakken's kinds of defeat and Vreeswijk's conclusive force. In their formalisms defeat of arguments is triggered by a conflict of arguments. If there is a conflict, one of the arguments involved is selected using the defeat information. The selected argument becomes defeated, and the conflict is resolved. We think that indirect defeat information is not sufficient. An important kind of defeat requiring direct defeat information is defeat by an undercutting argument (Pollock, 1987). An undercutting argument only defeats another argument, without contradicting the conclusion.

In our formalism the defeat information is specified by explicit and direct *defeaters*. A defeater consists of two sets of arguments: The arguments in one set become defeated if the arguments in the other set are undefeated.

Definition 4. Let L be a language. A *defeater* of L has the form A (B), where A and B are sets of arguments of L, such that no argument in A has a subargument or weakening that is an element of B. The arguments in A are the *activating* arguments of the defeater. The arguments in B are its *defeated* arguments. A $\cup$ B is the *range* of the defeater.

Notation. A defeater A (B) with finite range, i.e. A = $\{\alpha_1, \alpha_2, ..., \alpha_n\}$ and B = $\{\beta_1, \beta_2, ..., \beta_m\}$, is written $\alpha_1\ \alpha_2\ ...\ \alpha_n\ (\beta_1\ \beta_2\ ...\ \beta_m)$, if no confusion can arise.

The meaning of a defeater A (B) is that if the arguments in A are undefeated, the arguments in B must be defeated. For instance, the defeater a (b $\rightarrow$ c) defeats the rule b $\rightarrow$ c, if the argument a is undefeated. By the requirement in the definition a defeater cannot defeat a subargument or strengthening of one of its own activating arguments. For instance, if the argument a $\rightarrow$ b $\rightarrow$ c is activating a defeater, it cannot defeat the argument b $\rightarrow$ c. If the argument a $\rightarrow$ c $\rightarrow$ d is activating a defeater, it cannot defeat the argument {a $\rightarrow$ c, b $\rightarrow$ c} $\rightarrow$ d.

In contrast with Pollock's (1987-1994) defeaters, and Dung's (1993) attacks, our defeaters can represent compound defeat which occurs in case of defeat by accruing arguments and in case of skeptical defeat (section 3.3). The example of accruing arguments in section 3.2 requires not only the regular defeaters β (α_1) and β (α_2), but also a defeater that represents compound defeat, namely $\alpha_1\ \alpha_2$ (β). If α and β are incompatible arguments, a credulous reasoner can use the regular defeaters α (β) and β (α), while a skeptical reasoner can use the defeater ($\alpha\ \beta$) that represents compound defeat. Defeaters of the form (α_1, α_2), where α_1 and α_2 represent different testimonies, can be used to model Pollock's (in our view mistaken) counterexample for the accrual of arguments (section 3.2).

Our defeaters can also represent defeat by sequential weakening (section 3.1). If for instance the sequence of steps a $\rightarrow$ b $\rightarrow$ c makes an argument so weak that it must be defeated, this can be represented by the defeater (a $\rightarrow$ b $\rightarrow$ c).

4.3 Argumentation stages

We are about to define an *argumentation theory*. It formally represents which arguments are available to a reasoner, and when arguments can become defeated. Our notion of an argumentation theory is related to that of an argument system (Vreeswijk, 1991, 1993) and of an argumentation framework (Dung, 1993). A theory consists of a language, arguments, and defeaters. The language of a theory specifies the sentences that can be used in arguments. The arguments of a theory are the arguments that are available. The defeaters of a theory represent the situations in which arguments defeat other arguments.

Definition 5. An *argumentation theory* is a triple (L, Args, Defs), where

1. L is a language,
2. Args is a set of arguments in L, closed under initial arguments, and
3. Defs is a set of defeaters of L, with their ranges in Args.

For instance, a theory that represents the example of accruing arguments in section 3.2 is defined as follows:

L = $\{a_1, a_2, a, b\}$,
Args = $\{a_1, a_1 \to a, a_2, a_2 \to a, b\}$,
Defs = $\{\beta\ (\alpha_1), \beta\ (\alpha_2), \alpha_1\ \alpha_2\ (\beta)\}$, where $\alpha_1 = a_1 \to a$, $\alpha_2 = a_2 \to a$, $\beta = b$.

So we have two separate arguments α_1 and α_2 that support the conclusion a, and an argument β that supports b. The defeaters say that α_1 and α_2 are on their own defeated by β, but together defeat β. We use this theory as an illustration of the coming definitions. It is chosen, because it is a key example of accrual of arguments, and therefore suitable to show some important aspects of our formalism. It is however too simple to illustrate all aspects of the definitions.

The next definition is that of an *argumentation stage*. It can represent the arguments that at a certain stage in the process of argumentation have been taken into account, and which of them are then defeated. (Later we define argumentation stages that are *acceptable* with respect to a theory. These are the actual stages of argumentation that are made possible by an argumentation theory.) Each of the requirements in our definition corresponds to a simple intuition on stages in argumentation. For instance, one requirement is that an argument can only be taken into account if all its initial arguments already have been. The *range* of an argumentation stage consists of the arguments taken into account at that stage.

Definition 6. Let (L, Args, Defs) be an argumentation theory. An *argumentation stage* of (L, Args, Defs) has the form Σ (T), where Σ and T are subsets of Args, such that:

1. Σ is closed under initial arguments.
2. No argument can be an element of both Σ and T.
3. No proper subargument of an element of Σ can be an element of T.
4. Not all proper weakenings of an element of T that has proper weakenings can be elements of Σ.

The arguments in Σ are *undefeated*, and those in T *defeated*. The set $\Sigma \cup T$ is the *range* of Σ (T).

Remark: defeaters and argumentation stages are the same in form.

Some argumentation stages of the example theory are $a_1\ \alpha_1\ (\beta)$, $a_2\ \beta\ (a_1\ \alpha_2)$, and $a_1\ \alpha_1\ a_2\ \alpha_2\ (\beta)$.

Our definition of an argumentation stage is related to the argumentation structures of Lin and Vreeswijk. They require however that it is a set without contradicting arguments and do not include defeated arguments. Definition 6 is crucial for the point made in section 3.4: the arguments that are defeated at a stage are distinguished from the arguments not yet considered.

Which arguments of a theory become defeated and which don't is determined by its defeaters. Arguments are normally undefeated, but can at some stage of argumentation be defeated because of *relevant* defeaters. A defeater is relevant at some argumentation stage if all its arguments have been taken into account at that stage, or are parts of such arguments. Formally, this means that its range is a subset of the final parts of the arguments taken into account.

Definition 7. Let (L, Args, Defs) be an argumentation theory, A (B) a defeater in Defs, and Σ (T) an argumentation stage of (Args, Defs). A (B) is *relevant* for Σ (T), if $A \cup B \subseteq$ Finals[$\Sigma \cup T$].

So, in the example theory, β (α_2) is relevant for a_2 β (a_1 α_2), and all three defeaters of the theory are relevant for a_1 α_1 a_2 α_2 (β).

This notion of relevance of defeaters has no analogue in other formalisms. Normally, *all* defeaters are considered relevant. We can do better, because our argumentation stages explicitly represent which arguments are taken into account, including the defeated arguments.

A defeater only justifies the defeat of its defeated arguments, if its activating arguments are parts of the undefeated arguments, i.e., if they are subarguments of the undefeated arguments. The defeater is then *activated*.

Definition 8. Let (L, Args, Defs) be an argumentation theory, A (B) a defeater in Defs, and Σ (T) an argumentation stage of (Args, Defs). A (B) is *activated* in Σ (T), if it is relevant and $A \subseteq$ Finals[Σ].

In the argumentation stage a_2 β (a_1 α_2) the defeater β (α_2) is activated. In the stage a_1 α_1 a_2 α_2 (β) all three defeaters are activated.

Argumentation stages only represent actual stages of the process of argumentation, if they are *acceptable* with respect to an argumentation theory. An argumentation stage is acceptable, if

1. *The defeat of each of its defeated arguments is forced by an activated defeater.*
2. *If the defeat of an argument is forced by an activated defeater, it must actually be defeated in the stage.*
3. *No relevant defeater is unjustly ignored.*

The latter requirement requires yet another definition. Relevant defeaters must be *deactivated*, for instance, because one of its activating arguments is defeated. A defeater is deactivated, if two conditions hold. First, there must be another defeater that forces the defeat of one of its activating arguments. (It is even sufficient that the defeat of a subargument or a strengthening of one of the activating arguments is forced.) However, α_1 α_2 (β) and β (α_1) do not deactivate each other. Only the former can deactivate the latter. The reason for this is the accrual of the arguments α_1 and α_2. The defeater α_1 α_2 (β) overrules β (α_1), and therefore cannot be deactivated by it. This leads to the second condition: a defeater can only be deactivated by a defeater it does not overrule. Formally, this is captured in the following definition.

Definition 9. Let (L, Args, Defs) be an argumentation theory, A (B) and Γ (Δ) defeaters in Defs, and Σ (T) an argumentation stage of (Args, Defs). A (B) *deactivates* Γ (Δ), if both are relevant for Σ (T), and the following hold:

1. There is an element of B that is a subargument or a strengthening of an element of Γ.
2. If B is a proper subset of Γ, then A is not a subset of Δ.

We can finally define when an argumentation stage is *acceptable* with respect to an argumentation theory. The requirements in our definition have already been briefly explained just after definition 8.

Definition 10. Let (L, Args, Defs) be an argumentation theory, and Σ (T) an argumentation stage of (L, Args, Defs). Σ (T) is *acceptable* with respect to (Args, Defs), if the following hold:

1. If $\tau \in$ T, there is an activated A (B) $\in$ Defs, such that $\tau \in$ B.
2. If A (B) $\in$ Defs is activated, then B $\subseteq$ T.
3. If A (B) $\in$ Defs is relevant, but not activated, then there is an activated Γ (Δ) $\in$ Defs that deactivates A (B).

An acceptable argumentation stage of our example theory is β (a_1 α_1). The stage a_1 a_2 α_2 (β) is not acceptable, because the defeat of β is not justified, and because β (α_2) is not deactivated. It can be checked that the acceptable argumentation stages of our example theory correspond exactly to the stages represented in figure 2.

The way we define acceptable defeasible argumentation stages is related to the way Dung (1993) defines his admissible sets of arguments, and Pollock (1994) his partial status assignments. However, these do not represent *stages* in the process of argumentation, but are merely convenient formal structures on the way to the definition of extensions.

An *extension* of an argumentation theory is an acceptable stage of argumentation that has no succeeding argumentation stage, i.e. there is no argumentation stage with larger range. It must therefore be maximal with respect to set inclusion. Dung's (1993) preferred extensions and Pollock's (1994, p. 393) status assignments are defined similarly.

Definition 11. Let (L, Args, Defs) be an argumentation theory, and Σ (T) an argumentation stage of (L, Args, Defs). Σ (T) is an *extension* of (L, Args, Defs), if the following hold:

1. Σ (T) is acceptable with respect to (L, Args, Defs), and
2. There is no argumentation stage Σ' (T'), acceptable with respect to (L, Args, Defs), such that $\Sigma \cup$ T is a proper subset of Σ' $\cup$ T'.

As usual, a theory can have any number of extensions: zero, one, or several. The unique extension of our example theory is a_1 α_1 a_2 α_2 (β). The example is a real case of the accrual of the arguments α_1 and α_2, as can be seen by looking at other acceptable argumentation stages: β a_1 (α_1) and β a_2 (α_2). Here α_1 and α_2 are on their own defeated by β. The arguments α_1 and α_2 only remain undefeated if they reinforce each other.

5 Conclusions

This paper is an example of the argument-based approach to nonmonotonic reasoning. We have indicated why this is a valuable approach: First, because defeat is determined by the structure of arguments, and second, because defeat is determined by other available arguments.

The main points of this paper were the following. First, arguments can be defeated by sequential weakening. Second, arguments accrue. Third, defeat can be compound. Fourth, defeated arguments must be distinguished from not yet considered arguments.

We have provided a formalism that captures these ideas. To this end, we used a definition of arguments that makes parallel strengthening explicit, a definition of defeaters that can represent defeat by sequential weakening and compound defeat, and a definition of argumentation stages that explicitly represent the defeated arguments that have been considered.

Acknowledgments

This research was partly financed by the Foundation for Knowledge-based Systems (SKBS) as part of the B3.A project. SKBS is a foundation with the goal to improve the level of expertise in the Netherlands in the field of knowledge-based systems and to promote the transfer of knowledge in this field between universities and business companies. I thank Jaap Hage, Arno Lodder and Gerard Vreeswijk for our stimulating and lively arguments about argumentation.

References

1. Bondarenko, A., Toni, F. and Kowalski, R. A. (1993). An assumption-based framework for non-monotonic reasoning. *Logic programming and non-monotonic reasoning. Proceedings of the second international workshop* (eds. L. M. Pereira and A. Nerode), pp. 171-189. The MIT Press, Cambridge (Massachusetts).
2. Dung, P. M. (1993). On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming, and human's social and economical affairs.
3. Hage, J. and Verheij, B. (1994). Reason-Based Logic: a logic for reasoning with rules and reasons. To appear in *Law, Computers and Artificial Intelligence*.
4. Lin, F. (1993). An argument-based approach to nonmonotonic reasoning. *Computational Intelligence*, Vol. 9, No. 3, pp. 254-267.
5. Loui, R. P. (1987). Defeat among arguments: a system of defeasible inference. *Computational Intelligence*, Vol. 3, No. 2, pp. 100-106.
6. Nute, D. (1988). Defeasible reasoning: a philosophical analysis in Prolog. *Aspects of Artificial Intelligence* (ed. James H. Fetzer), pp. 251-288. Kluwer Academic Publishers, Dordrecht.
7. Pollock, J. L. (1987). Defeasible reasoning. *Cognitive Science* 11, pp. 481-518.
8. Pollock, J. L. (1991). Self-defeating arguments. *Minds and Machines* 1, pp. 367-392.
9. Pollock, J. L. (1994). Justification and defeat. *Artificial Intelligence* 67, pp. 377-407.
10. Poole, D. (1988). A logical framework for default reasoning. *Artificial Intelligence* 36, pp. 27-47.
11. Prakken, H. (1993). A logical framework for modelling legal argument. *The Fourth International Conference on Artificial Intelligence and Law. Proceedings of the Conference*, pp. 1-9. ACM, New York.
12. Simari, G. R. and Loui, R. P. (1992). A mathematical treatment of defeasible reasoning and its applications. *Artificial Intelligence* 53, pp. 125-157.
13. Touretzky, D. S., Horty, J. F., and Thomason, R. H. (1987). A clash of intuitions: the current state of nonmonotonic multiple inheritance systems. *IJCAI 87; Proceedings of the Tenth International Joint Conference on Artificial Intelligence* (ed. J. McDermott), pp. 476-482. Morgan Kaufmann Publishers, Los Altos (California).
14. Verheij, H. B. (1994). Reason Based Logic and legal knowledge representation. *Proceedings of the Fourth National Conference on Law, Computers and Artificial Intelligence* (eds. I. Carr and A. Narayanan), pp. 154-165. University of Exeter.
15. Verheij, B. (1995). The influence of defeated arguments in defeasible argumentation. Accepted for the *Second World Conference on the Fundamentals of Artificial Intelligence (WOCFAI 95)*.
16. Vreeswijk, G. (1991). Abstract argumentation systems: preliminary report. *Proceedings of the First World Conference on the Fundamentals of Artificial Intelligence* (eds. D. M. Gabbay and M. De Glas), pp. 501-510. Angkor, Paris.
17. Vreeswijk, G. (1993). *Studies in defeasible argumentation*. G. A. W. Vreeswijk, Amsterdam.