The influence of defeated arguments in defeasible argumentation

Bart Verheij
University of Limburg, Department of Metajuridica
P.O. Box 616, 6200 MD Maastricht, The Netherlands
fax: +31 43 256538
email: bart.verheij@metajur.rulimburg.nl

Abstract

Formal defeasible argumentation is currently the subject of active research. Formalisms of defeasible argumentation are characterized by a notion of defeasible argument. The influence of arguments on which conclusions can be drawn distinguishes formalisms of defeasible argumentation from nonmonotonic logics. This influence occurs for two reasons: by the structure of an argument, and by interaction with other arguments.

In the process of argumentation not all arguments are available at once. At each stage of argumentation new arguments are taken into account. In defeasible argumentation, where arguments can be defeated by other arguments, this results in the possible change of the status of arguments, depending on which arguments have been considered. Existing formalisms of defeasible argumentation do not provide a process view on argumentation, or overlook the influence on this process of the defeated arguments that have been taken into account.

In this paper we argue that a model of the process of argumentation requires that the arguments that are defeated at some stage of argumentation cannot simply be ignored. Otherwise, different stages of argumentation cannot be distinguished, and orders of argumentation can disappear.

Keywords: defeasible argumentation, nonmonotonic logic

1 Introduction

Currently the formal study of defeasible argumentation gets much attention [BoToKo93, Du93, HaVe94, Li93, Po94, Pr93a, Pr93b, Vr93, Ve95]. Formalisms of defeasible argumentation can be distinguished from nonmonotonic logics in general. The principal distinction is that a notion of *defeasible argument* is central. An argument is like a proof: It represents the argumentation from premises to conclusion, *including the intermediate steps*. Unlike proofs, however, defeasible arguments are not strict, and can become *defeated* by other arguments.

There are two main reasons to take arguments into account in defeasible argumentation. First, the *structure* of an argument influences whether it is defeated or not. Second, whether an argument is defeated is influenced by *other arguments*.

- The structure of an argument

  The quality of an argument is influenced by its structure. For instance, an argument with a number of weak steps is worse than an argument with fewer weak steps, and an argument that uses more information is better than (or as good as) an argument that uses less. When one only looks at conclusions and reasons (the direct

predecessors of a conclusion in an argument) the influence of the structure of arguments is overlooked. Therefore, it is natural to determine which arguments can be based on a given set of information first, and then which conclusions are justified by those arguments.

- Other arguments

  By the interaction of arguments some of the arguments taken into account are undefeated, others defeated. For instance, arguments can impair other arguments, resulting in the defeat of the latter. Arguments can also reinforce each other, so that they remain undefeated. Because only undefeated arguments justify their conclusions, and the interaction of arguments influences which arguments are defeated and which undefeated, arguments have to be considered when one determines the conclusions that follow from given information.

Not all formalisms mentioned use arguments for both reasons. Together these reasons can be used to distinguish formalisms of defeasible argumentation from others (see also section 6).

In the next section we explain why defeated arguments cannot be ignored at a stage in the process of argumentation. In the sections thereafter we describe a formalism appropriate for modeling the influence of defeated arguments. After a brief comparison with some other formalisms of defeasible argumentation, we end with the conclusion of this paper.

2 How defeated arguments influence argumentation

An overlooked aspect of defeasible argumentation is the influence of the defeated arguments that have been taken into account. A good example can be given in case of *accrual* of arguments.[1] Arguments for a conclusion accrue, if they reinforce each other, and therefore together give better support to their conclusion than on their own.

We give an example. Suppose we have the situation that there are three arguments, denoted α_1, α_2 and β, available to a reasoner. It can be the case that the arguments α_1 and α_2 are both on their own defeated by β, but together remain undefeated, and even defeat β. So, we have the following situation:

- The argument β defeats the argument α_1, if α_1 and β are the arguments considered.
- The argument β defeats the argument α_2, if α_2 and β are the arguments considered.
- The arguments α_1 and α_2 defeat the argument β, if α_1, α_2 and β are the arguments considered.[2]

---

[1] The term is used by Pollock [Po91]. Even though Pollock finds it a natural supposition that arguments accrue, he surprisingly rejects it. Verheij [Ve95] argues that arguments do accrue, and provides a formal model for it. This paper extends that model.

[2] The following natural language example is taken from Verheij [Ve94]. 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).

There are six orders in which the arguments can be taken into account by a reasoner, such as α_1, α_2, β or α_2, β, α_1. In figure 1, the six orders are shown in one diagram. Each corner of the block represents a stage in the process of argumentation, and has a label representing which arguments have been considered then. The 0 represents that no argument has been considered yet. An argument in brackets is defeated. Each arrow denotes that an argument is being 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.

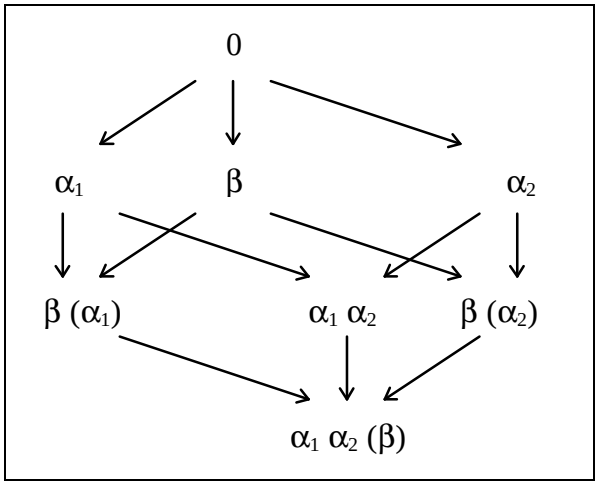

Figure 1: Considering arguments in different orders

Now we come to the influence of the defeated arguments. If we would forget about the defeated arguments, as in other work on defeasible argumentation, we are left with the following (wrong) picture. Again, each corner represents a stage of argumentation, but this time only the undefeated arguments at that stage are denoted. Just as in the previous figure, each arrow means that one argument is being taken into account.

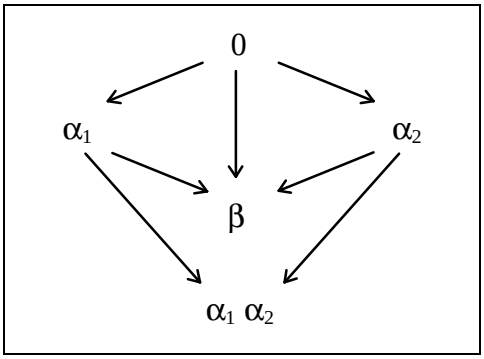

Figure 2: The wrong picture

This picture is wrong for two reasons:

- Different stages of argumentation have collapsed into one.
  For instance β, β (α_1) and β (α_2) can no longer be distinguished.
- An order of argumentation has disappeared.
  There is no direct arrow from β to α_1 α_2, because it has become unclear what it means to go from β to α_1 α_2 by taking a single extra argument into account. In the correct picture the intermediate stages β (α_1) and β (α_2) dissolve this problem.

We summarize the influence of defeated arguments:

If the defeated arguments taken into account are overlooked, different stages of argumentation cannot be distinguished, and orders of argumentation can disappear.

3 Arguments

We start with the formal definition of an *argument*. Our notion of an argument is related to that of Lin and Shoham [LiSh89] and Vreeswijk [Vr91, Vr93], and is basically a tree of sentences in some language. Our approach to defeasible 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. This is not required, because in our formalism contradiction is not the trigger for defeat. We briefly come back to this in the next section.[3]

Definition 3.1 A *language* is a set, whose elements are the *sentences* of the language.

An argument is like a proof, possibly with conditions. An argument supports its conclusion (relative to its conditions), but unlike a proof, an argument is *defeasible*. Any argument can be defeated by other arguments. Each argument has a *conclusion* and *conditions* (possibly zero). 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 3.2 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
   $\mathrm{Conc}(s) = s$
   $\mathrm{Conds}(s) = \mathrm{Sents}(s) = \mathrm{Initials}(s) = \mathrm{Finals}(s) = \{s\}$
2. If A is a set of arguments in L, s an element of L, and $s \notin \mathrm{Sents}[A]$,[4] then $A \rightarrow s$ is an argument in L. In this case we define
   $\mathrm{Conc}(A \rightarrow s) = s$
   $\mathrm{Conds}(A \rightarrow s) = \mathrm{Conds}[A]$
   $\mathrm{Sents}(A \rightarrow s) = \{s\} \cup \mathrm{Sents}[A]$
   $\mathrm{Initials}(A \rightarrow s) = \{A \rightarrow s\} \cup \mathrm{Initials}[A]$
   $\mathrm{Finals}(A \rightarrow s) = \{s\} \cup \{B \rightarrow s \mid \exists f$: f is a surjective function from A onto B, such that $\forall \alpha$: $f(\alpha) \in \mathrm{Finals}(\alpha)\}$[5]

$\mathrm{Conc}(\alpha)$ is the *conclusion* of α. An element of $\mathrm{Conds}(\alpha)$, $\mathrm{Sents}(\alpha)$, $\mathrm{Initials}(\alpha)$, and $\mathrm{Finals}(\alpha)$ is a *condition*, 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 $S \rightarrow s$, where $S \subseteq L$ and $s \in L$. For each argument α we define the set of arguments $\mathrm{Subs}(\alpha)$, whose elements are the *subarguments* of α:

$\mathrm{Subs}(\alpha) = \mathrm{Initials}[\mathrm{Finals}(\alpha)]$

---

[3] Lin and Shoham [LiSh89], Vreeswijk [Vr91, Vr93] and Dung [Du93] 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.

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

[5] This means that the set B arises by replacing each argument in the set A by one of its final arguments.

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. $A = \{\alpha_1, \alpha_2, ..., \alpha_n\}$, we write $\alpha_1, \alpha_2, ..., \alpha_n \rightarrow s$ for an argument $A \rightarrow s = \{\alpha_1, \alpha_2, ..., \alpha_n\} \rightarrow s$, if no confusion can arise.

Intuitively, if $A \rightarrow s$ 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 $\{\{a\} \rightarrow b\} \rightarrow c$ and $\{\{a\} \rightarrow c, \{b\} \rightarrow c\} \rightarrow d$. They are graphically represented in figure 3.

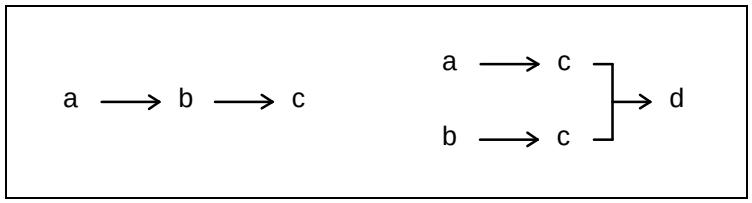

Figure 3: Examples of arguments

The conditions 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 formal structure of our arguments differs from those of Lin and Shoham [LiSh89] and Vreeswijk [Vr91, Vr93]. In these formalisms, each condition of an argument can only be supported by a single argument. Because of our belief that arguments can accrue, in our formalism conditions can be supported by several arguments.[6] As a result, we can make *weakenings* (and *strengthenings*) of an argument explicit. Intuitively, an argument becomes weaker if less arguments support its conclusion and intermediate conclusions. For instance, the argument $\{\{b\} \rightarrow c\} \rightarrow d$ is a *weakening* of the argument $\{\{a\} \rightarrow c, \{b\} \rightarrow c\} \rightarrow d$. The latter contains $\{a\} \rightarrow c$ and $\{b\} \rightarrow c$ to support the intermediate conclusion c, while the former only contains $\{b\} \rightarrow c$.

Definition 3.3 Let L be a language. For any argument α in the language L we recursively define a set of arguments Weaks(α):

1. For $\alpha = s$, $s \in L$,
   $Weaks(s) = \{s\}$.
2. For $\alpha = A \rightarrow s$, $A \subseteq Args(L)$, $s \in L$,
   $Weaks(A \rightarrow s) = \{B \rightarrow s \mid B \subseteq Weaks[A] \text{ and } Conc[B] = Conc[A]\}$

An element of Weaks(α) 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 $\{\{a\} \rightarrow c, \{b\} \rightarrow c\} \rightarrow d$.

---

[6] This is formally accomplished by making the arguments supporting a conclusion a set of arguments, instead of a sequence.

Different arguments can have the same subrules. The arguments in figure 4 all have the subrules {a} → c, {b} → c, {c} → d and {d} → e. There is some redundancy in the last argument, because the arguments {{a} → c} → d and {{b} → c} → d are weakenings of the argument {{{a} → c, {b} → c} → d} → e. In case one of the three arguments supporting d is defeated such redundancy can become meaningful.

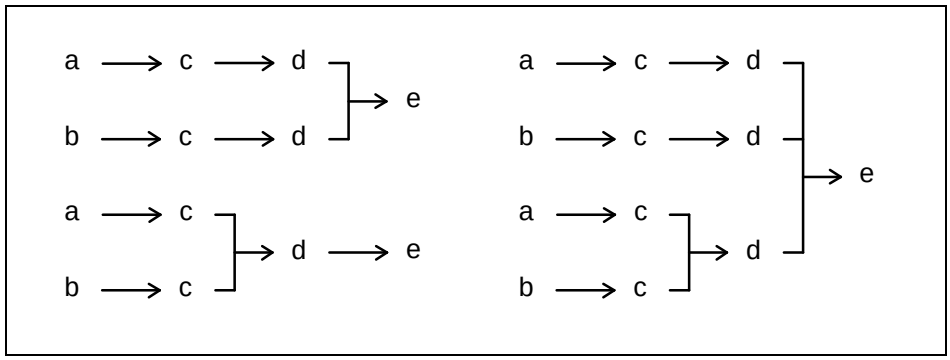

Figure 4: Different arguments with the same subrules

In other formalisms different arguments with the same subrules are not distinguished. For instance, the formal arguments of Pollock [Po87-Po94], Simari and Loui [SiLo92], Prakken [Pr93a, Pr93b], and Bondarenko *et al.* [BoToKo93] are (more or less) sets of rules, so that the distinction is concealed. In the formalisms of Lin and Shoham [LiSh89] and Vreeswijk [Vr91, Vr93], the arguments in figure 4 are not based on the same subrules, because the conditions of a rule are a sequence, and not a set. For example, they treat c → d and c, c → d as different rules.

4 Defeaters

As said before, in defeasible argumentation, arguments are *defeasible*. In our formalism, *all* arguments are defeasible. Except for Dung [Du93], other authors have separate classes of strict and defeasible arguments. 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 easy, because the defeat of arguments is the result of defeat information that is *explicit* and *direct*.

- Explicit defeat information

  Pollock's defeaters [Po87-Po94], Prakken's kinds of defeat [Pr93a, Pr93b], Vreeswijk's conclusive force [Vr91, Vr93], and Dung's attacks [Du93] 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 information specifying conditions that directly imply the defeat of one or more arguments. Pollock's defeaters and Dung's attacks are examples of direct defeat information. 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 [Po97]. An undercutting argument only defeats another argument, without contradicting the conclusion.

In our formalism the defeat information is specified by explicit and direct *defeaters*. In contrast with Pollock's defeaters [Po87-Po94], and Dung's attacks [Du93], our defeaters can explicitly represent *compound defeat*: a set of arguments for a conclusion can defeat another set of arguments, instead of only a single argument another single argument. This is needed for the adequate modeling of accrual of arguments [Ve95]. 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 4.4 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

1. All arguments in A have the same conclusion.
2. All arguments in B have the same conclusion.
3. 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.[7]

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 third requirement in the definition a defeater cannot defeat a subargument or strengthening of one of its activating arguments. For instance, if the argument $a \rightarrow b \rightarrow c$ is activating in a defeater, it cannot defeat the argument $b \rightarrow c$. If the argument $a \rightarrow c \rightarrow d$ is activating in a defeater, it cannot defeat the argument $\{a \rightarrow c, b \rightarrow c\} \rightarrow d$.

As said, our defeaters can represent compound defeat which is needed in case of accrual of arguments. For instance, the example in section 2 requires not only the regular defeaters $\beta\ (\alpha_1)$ and $\beta\ (\alpha_2)$, but also a defeater that represents compound defeat, namely $\alpha_1\ \alpha_2\ (\beta)$.

5 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 [Vr91, Vr93] and of an argumentation framework [Du93]. A theory consists of a language, arguments, and defeaters. The language of a theory specifies the sentences that can be used in

[7] Because arguments are their own subarguments and strengthenings, the sets A and B can have no elements in common.

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 5.1 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,[8] and
3. Defs is a set of defeaters of L, with their ranges in Args.[9]

For instance, a theory that represents the example in section 2 is defined as follows:

$L = \{a_1, a_2, a, b\}$,

$Args = \{a_1, a_1 \rightarrow a, a_2, a_2 \rightarrow a, b\}$,

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

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 the influence of defeated arguments. 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.) Our definition of an argumentation stage is related to the argumentation structures of Lin and Shoham, and Vreeswijk. They require however that it is a set without contradicting arguments (see the discussion of direct defeat information in section 4). 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 5.2 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.

---

[8] The set of arguments is not closed under 'rule application', as in other formalisms. Although the arguments of an ideal reasoner would be, this is in general an unreasonable assumption.

[9] The defeaters of a theory do not necessarily agree with each other. For instance, both α (β) and β (α) can be defeaters of a theory. A classic example of such a situation is the Nixon diamond.

Some argumentation stages of the example theory are $a_1\ \alpha_1\ (\beta)$, $a_2\ \beta\ (a_1\ \alpha_2)$, and $a_1\ \alpha_1\ a_2\ \alpha_2\ (\beta)$. Definition **5.2** is crucial: it reflects that there can be defeated arguments taken into account at a stage of argumentation.

Which arguments of a theory become defeated and which not 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 5.3** 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 $A \cup B \subseteq \text{Finals}[\Sigma \cup T]$.

So, in the example theory, $\beta\ (\alpha_2)$ is relevant for $a_2\ \beta\ (a_1\ \alpha_2)$, and all three defeaters of the theory are relevant for $a_1\ \alpha_1\ a_2\ \alpha_2\ (\beta)$.

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 represent which arguments are taken into account.

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

> **Definition 5.4** 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 $A \subseteq \text{Finals}[\Sigma]$.

In the argumentation stage $a_2\ \beta\ (a_1\ \alpha_2)$ the defeater $\beta\ (\alpha_2)$ is activated. In the stage $a_1\ \alpha_1\ a_2\ \alpha_2\ (\beta)$ 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

- The defeat of each of the defeated arguments is justified.

  The default is namely that an argument is *not* defeated. Defeat is justified by activated defeaters.
- If the defeat of an argument is justified, it must actually be defeated.

  Otherwise put, defeaters that are activated in the stage must be obeyed.
- No relevant defeater is unjustly ignored.

  This requirement needs 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 justifies the defeat of one of its activating arguments. (It is of course even sufficient that the defeat of a subargument or a strengthening of one of the activating arguments is justified.) However, $\alpha_1\ \alpha_2\ (\beta)$ and $\beta\ (\alpha_1)$ 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 $\alpha_1\ \alpha_2\ (\beta)$ overrules $\beta\ (\alpha_1)$, 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 5.5 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. A is not a subset of Δ, or B is not a proper subset of Γ.

We can finally define when an argumentation stage is *acceptable* with respect to an argumentation theory.[10] The requirements in the definition have already been briefly explained just after definition **5.4**.

Definition 5.6 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.

An *extension* of an argumentation theory is an acceptable stage of argumentation that cannot be continued. It must therefore be maximal with respect to set inclusion.[11]

Definition 5.7 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 $\Sigma' \cup$ T'.

The unique[12] extension of our example theory is a_1 α_1 a_2 α_2 (β). Even though the theory contains the defeaters β (α_1), β (α_2) that can defeat α_1 and α_2 separately, they remain undefeated by supporting each other. This 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 reinforced each other.

6 Other formalisms

In order to compare our formalism with others, we end with an overview of some of its characteristics.

---

[10] The way we define acceptable defeasible argumentation stages is related to the way Dung defines his admissible sets of arguments [Du93], and Pollock his partial status assignments [Po94]. However, these do not represent *stages* in the process of argumentation.

[11] Dung's preferred extensions [Du93] and Pollock's status assignments [Po94, p. 393] are defined similarly.

[12] A theory can have any number of extensions: zero, one, or several.

- The *structure of arguments* can influence which arguments are defeated and which are not (section 1).
- *Other arguments* can influence which arguments are defeated and which are not (section 1).
- The formalism is *independent* of the choice of a *language* (section 3).
- Defeat is the result of *explicit defeat information* (section 4).
- Defeat information represents *direct defeat* (section 4). Defeat is not triggered by a conflict.
- *Stages* in the process of argumentation are represented (section 5).
- Arguments can *accrue* (section 2, 3, 4, 5).
- *Defeated arguments influence* the process of argumentation (section 2, 4, 5).

Table 1 relates these characteristics to some other formalisms. An entry in the table is crossed if the formalism in that column has the characteristic indicated in the row.

	BoToKo93	Du93	Pr93a, Pr93b	HaVe94	LiSh89, Li93	Po87-Po94	Vr91, Vr93
Structure of argument	×		×		×		×
Other arguments	×	×	×	×		×	
Language independent		×			×		×
Explicit defeat information		×	×			×	×
Direct defeat information		×				×	
Argumentation stages							×
Accrual of arguments				×			
Influence of defeated arguments				×			

Table 1: Formalisms of defeasible argumentation and their characteristics

7 Conclusion

In this paper we have argued that the process of argumentation is better modeled if the arguments that are defeated at a certain stage of argumentation are not ignored. This is especially clear if several arguments for a conclusion accrue: They can at early stages of argumentation be taken into account, but be defeated by other arguments, and later exert their influence, and become undefeated. We have provided a formalism that captures these ideas. Key definitions are those of argumentation stages (definition **5.2**), in which the defeated arguments are explicitly taken into account, and of relevance of defeaters (definition 5.3), which assures that only defeaters are used that contain arguments actually 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

[BoToKo93] A. Bondarenko, F. Toni and R. A. Kowalski, An assumption-based framework for non-monotonic reasoning, in *Logic programming and non-monotonic reasoning. Proceedings of the second international workshop*, L. M. Pereira and A. Nerode (eds.), MIT Press, Cambridge (Massachusetts), 1993, pp. 171-189.

[Du93] P. M. Dung, On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming, and human's social and economical affairs, manuscript, 1993.

[HaVe94] J. Hage and B. Verheij, Reason-Based Logic: a logic for reasoning with rules and reasons, to appear in *Law, Computers and Artificial Intelligence*, Vol. 3 (1994), No. 2, report SKBS/B3.A/94-10.

[LiSh89] F. Lin and Y. Shoham, Argument Systems: a uniform basis for nonmonotonic reasoning, in *Proceedings of the First International Conference on Principles of Knowledge Representation and Reasoning*, R. J. Brachman, H. J. Levesque and R. Reiter (eds.), Morgan Kaufmann Publishers, San Mateo (California), 1989, pp. 245-255.

[Li93] F. Lin, An argument-based approach to nonmonotonic reasoning. *Computational Intelligence*, Vol. 9 (1993), No. 3, pp. 254-267.

[Po87] J. L. Pollock, Defeasible reasoning, *Cognitive Science* 11 (1987), pp. 481-518.

[Po90] J. L. Pollock, A theory of defeasible reasoning, *International Journal of Intelligent Systems*, Vol. 6 (1990), pp. 33-54.

[Po91] J. L. Pollock, Self-defeating arguments, *Minds and Machines* 1 (1991), pp. 367-392.

[Po92] J. L. Pollock, How to reason defeasibly, *Artificial Intelligence* 57 (1992), pp. 1-42.

[Po94] J. L. Pollock, Justification and defeat, *Artificial Intelligence* 67 (1994), pp. 377-407.

[Pr93a] H. Prakken, *Logical tools for modelling legal argument*, H. Prakken, Amsterdam, 1993.

[Pr93b] H. Prakken, A logical framework for modelling legal argument, in *The Fourth International Conference on Artificial Intelligence and Law. Proceedings of the Conference*, ACM, New York, 1993, pp. 1-9.

[SiLo92] G. R. Simari and R. P. Loui, A mathematical treatment of defeasible reasoning and its applications, *Artificial Intelligence* 53 (1992), pp. 125-157.

[Ve94] H. B. Verheij, Reason Based Logic and legal knowledge representation, in *Proceedings of the Fourth National Conference on Law, Computers and Artificial Intelligence*, I. Carr and A. Narayanan (eds.), University of Exeter, 1994, pp. 154-165, report SKBS/B3.A/94-5.

[Ve95] B. Verheij, Accrual of arguments in defeasible argumentation, in *Dutch/German Workshop on Nonmonotonic Reasoning. Proceedings of the Second Workshop*, Delft University of Technology, Utrecht University, 1995, pp. 217-224, report SKBS/B3.A/95-01.

[Vr91] G. Vreeswijk, Abstract argumentation systems: preliminary report, in *Proceedings of the First World Conference on the Fundamentals of Artificial Intelligence*, D. M. Gabbay and M. De Glas (eds.), Angkor, Paris, 1991, pp. 501-510.

[Vr93] G. Vreeswijk, *Studies in defeasible argumentation*, G. A. W. Vreeswijk, Amsterdam, 1993.