Accrual of arguments in defeasible argumentation

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

ABSTRACT

In this paper we address an often overlooked problem in defeasible argumentation: how do we deal with arguments that are on their own *defeated, but* together *remain undefeated? Pollock (1991) finds this accrual of arguments a natural supposition, but then surprisingly denies its existence. We think that arguments do accrue. To handle the accrual of arguments, we introduce* compound defeat *of arguments. We call the defeat of arguments compound, if groups of arguments can be defeated by other groups of arguments. The formalism presented in this paper is based on this notion of compound defeat. It adequately handles the accrual of arguments.*

1 INTRODUCTION

Recently the formal study of defeasible argumentation has got much attention. Formalisms of defeasible argumentation are often distinguished from nonmonotonic logics in general. The principal distinction is that a notion of argument is central. In defeasible argumentation arguments can become defeated by other arguments.

The main new element of the formalism in this paper is the *compound defeat* of arguments. By compound defeat we mean that groups of arguments can be defeated by other groups of arguments, instead of only single arguments by single arguments. We think that the interaction of arguments is not modeled adequately by a binary defeat relation on arguments. For instance, it can be the case that two arguments α_1 and α_2 are both on their own defeated by β, but together remain undefeated, and even defeat β. So, we can 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.[1]

This defeat information essentially involves *three* arguments. It is an example of what Pollock (1991) calls the *accrual* of arguments.[2] Even though Pollock 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. In section 5, we come back to Pollock's argument against this principle of accrual.

Defeasible argumentation has been formally studied by Pollock (1987-1994), Lin and Shoham (1989), Vreeswijk (1991, 1993), Simari and Loui (1992), Prakken (1993a, b), Dung (1993), and Bondarenko *et al.* (1993). These formalisms cannot adequately deal with the accrual of arguments. We present a formalism based on compound defeat that can.

The idea to incorporate compound defeat in a formalism for defeasible argumentation is inspired by the research on Reason-Based Logic (Hage, 1993; Hage and Verheij, 1994; Verheij, 1994). In Reason-Based Logic, a restricted form of compound defeat occurs. Rules can lead to

---

[1] The following natural language example is taken from Verheij (1994). 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).

[2] Pollock (1991) speaks of the accrual of *reasons*.

reasons for and against a conclusion. Explicit information on how the reasons are weighed leads to a conclusion. In this way, the group of reasons for a conclusion can defeat the group of reasons against a conclusion (and the other way round).

In the next two sections we describe the basic elements of our formalism: arguments and defeaters. Then we define defeasible argumentation theories and their extensions. In the last section we come back to the main points of the formalism: compound defeat and accrual of arguments.

2 ARGUMENTS

We start with the formal definition of an *argument*. Our notion of an argument is related to that of Lin and Shoham (1989) and Vreeswijk (1991, 1993), 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 2.1 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 2.2 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 $A \to s$ is an argument in L. In this case we define

   $\mathrm{Conc}(A \to s) = s$

   $\mathrm{Conds}(A \to s) = \mathrm{Conds}[A]$

   $\mathrm{Sents}(A \to s) = \{s\} \cup \mathrm{Sents}[A]$

   $\mathrm{Initials}(A \to s) = \{A \to s\} \cup \mathrm{Initials}[A]$

   $\mathrm{Finals}(A \to s) = \{s\} \cup \{B \to s \mid \exists f\text{: f is a surjective function from A onto B, such that } \forall\alpha\text{: } f(\alpha) \in \mathrm{Finals}(\alpha)\}$[5]

Conc(α) is the *conclusion* of α. An element of Conds(α), Sents(α), Initials(α), and Finals(α) 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 $S \to s$, 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:

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

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.

---

[3] Lin and Shoham (1989), 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.

[4] If $f: V \to W$ 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.

Notation 2.3 If A is finite, i.e. $A = \{\alpha_1, \alpha_2, ..., \alpha_n\}$, we write $\alpha_1, \alpha_2, ..., \alpha_n \to s$ for an argument $A \to s = \{\alpha_1, \alpha_2, ..., \alpha_n\} \to s$, if no confusion can arise.

Intuitively, if $A \to s$ 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: c, $\varnothing \to a$, $\{a\} \to b$, $\{\{a\} \to b\} \to c$, $\{\{\to a\} \to b, \varnothing \to c\} \to d$, $\{\{a\} \to c, \{b\} \to c\} \to d$. They are graphically represented in figure 1.

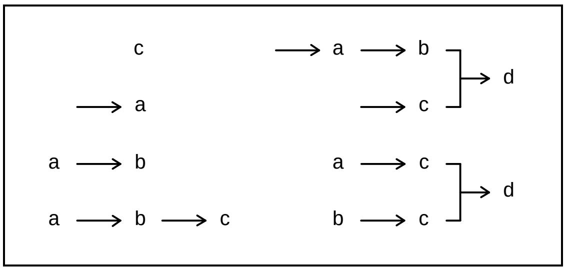

Figure 1. Examples of arguments

The conditions of the argument $\{\{a\} \to c, \{b\} \to c\} \to d$ are a and b. It has d as its conclusion. The argument c, a sentence, is its own conclusion and condition. The argument $\varnothing \to a$ has a as its conclusion, and no condition. Some initial arguments of $\{\{\to a\} \to b, \varnothing \to c\} \to d$ are $\varnothing \to c$, $\{\to a\} \to b$ and the argument itself. Some of its final arguments are d, $\{b, c\} \to d$, and $\{b, \varnothing \to c\} \to d$. Some subarguments of the argument $\{\{a\} \to c, \{b\} \to c\} \to d$ are c and $\{c, \{b\} \to c\} \to d$.

The formal structure of our arguments differs from those of Lin and Shoham (1989), Vreeswijk (1991, 1993). 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 $\{\{b\} \to c\} \to d$ is a *weakening* of the argument $\{\{a\} \to c, \{b\} \to c\} \to d$. The latter contains $\{a\} \to c$ and $\{b\} \to c$ to support the intermediate conclusion c, while the former only contains $\{b\} \to c$.

Definition 2.4 Let L be a language. For any argument α in the language L we recursively define a set of arguments $\mathrm{Weaks}(\alpha)$:

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

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

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

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

An element of $\mathrm{Weaks}(\alpha)$ 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.

The weakenings of the argument $\{\{a\} \to c, \{b\} \to c\} \to d$ are $\{\{a\} \to c\} \to d$, $\{\{b\} \to c\} \to d$, and the argument itself. Weakenings are in general not subarguments. For instance, $\{\{a\} \to c\} \to d$ is not a subargument of $\{\{a\} \to c, \{b\} \to c\} \to d$.

If an argument α is a weakening of an argument β, the subrules of α are also subrules of β. For instance, the subrules $\{a\} \to c$ and $\{c\} \to d$ of $\{\{a\} \to c\} \to d$ are also subrules of $\{\{a\} \to c,$

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

{b} → c} → d. This argument should not be confused with {{a, b} → c} → d. The latter has the rule {a, b} → c as a subrule, the former hasn't. The rule {a, b} → c is *not* a strengthening of the rules a → c and b → c.

Different arguments can have the same subrules. The arguments {{{a} → c} → d, {{b} → c} → d} → e, {{{a} → c, {b} → c} → d} → e and {{{a} → c} → d, {{b} →c} → d, {{a} → c, {b} → c} → d} → e (see figure 2) have the same subrules, namely {a} → c, {b} → c, {c} → d and {d} → e. The first argument is a proper weakening of the second. The second is a proper weakening of the third. The third argument has no proper strengthening with the same subrules. It uses its subrules as effectively as possible. There is some redundancy in the argument, because the arguments {{a} → c} → d and {{b} → c} → d are weakenings of the argument {{{a} → c, {b} → c} → d} → e.

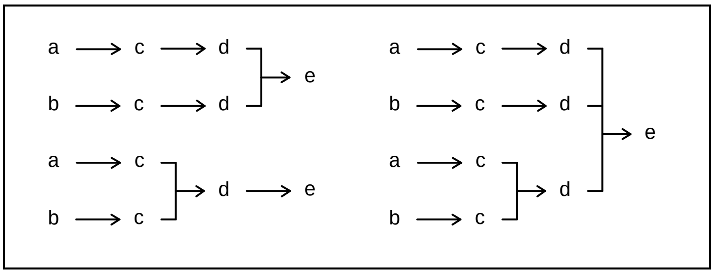

Figure 2. Different arguments with the same subrules

In other formalisms different arguments with the same subrules are not distinguished. For instance, the formal arguments of Pollock (1987-1994), Simari and Loui (1992), Prakken (1993a, b), and Bondarenko *et al.* (1993) are (more or less) sets of rules, so that the distinction is concealed. In the formalisms of Lin and Shoham (1989) and Vreeswijk (1991, 1993), the arguments in figure 2 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.

3 DEFEATERS

As said before, in defeasible argumentation, arguments are *defeasible*. In our formalism, *all* arguments are defeasible. Except for Dung (1993), 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 (1987-1994) defeaters, Prakken's (1993a, b) 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 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 (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*. In contrast with Pollock's (1987-1994) defeaters, and Dung's (1993) attacks, our defeaters can explicitly represent compound defeat, because they consist of *sets* of arguments for a conclusion. A defeater represents that a set of arguments is defeated if some other set of arguments is undefeated.

Definition 3.1 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 3.2 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 $a[b \to c]$ defeats the rule $b \to c$, 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 $a \to b \to c$ is activating in a defeater, it cannot defeat the argument $b \to c$. If the argument $a \to c \to d$ is activating in a defeater, it cannot defeat the argument $\{a \to c, b \to c\} \to d$.

Defeaters can represent compound defeat: a set of arguments for a conclusion defeats another set of arguments. For instance, the example in the introduction requires not only regular defeaters, namely $\beta[\alpha_1]$ and $\beta[\alpha_2]$, but also a defeater that represents compound defeat, namely $\alpha_1, \alpha_2[\beta]$.

4 THEORIES AND EXTENSIONS

We are about to define a *defeasible argumentation theory*. It formally represents which arguments can be made, and when arguments become defeated. Our notion of a defeasible 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 4.1 A *(defeasible argumentation) theory* is a triple (L, Args, D), where

1. L is a language,
2. Args is a set of arguments in L, closed under initial arguments,[8] and
3. D is a set of defeaters of L, with their ranges in Args.[9]

For instance, a theory that represents the example in the introduction is defined as follows:[10]

$L = \{a_1, a_2, a, b\}$,
$Args = \{a_1, a_1 \to a, a_2, a_2 \to a, b\}$,
$D = \{\beta[\alpha_1], \beta[\alpha_2], \alpha_1, \alpha_2[\beta]\}$, where $\alpha_1 = a_1 \to a$, $\alpha_2 = a_2 \to a$, $\beta = b$.

---

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

[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 $\alpha[\beta]$ and $\beta[\alpha]$ can be defeaters of a theory. A classic example of such a situation is the Nixon diamond.

[10] A natural language interpretation can be found in note 1.

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 they 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. It is however too simple to illustrate all aspects of the definitions.

The main question concerning a theory is to determine the status of its arguments: which arguments remain undefeated and which become defeated? An intuitive first requirement for any reasonable set of undefeated (grounded) arguments is that it is closed under initial arguments. Any reasonable set of undefeated arguments is therefore a *defeasible argumentation structure*.[11]

Definition 4.2 Let (L, Args, D) be a defeasible argumentation theory. A *(defeasible argumentation) structure* of (L, Args, D) is a subset Σ of arguments of (L, Args, D), such that all initial arguments of an argument in Σ are also in Σ.

Some defeasible argumentation structures of the example theory are $\{\beta\}$ and $\{a_1, \alpha_1, a_2, \alpha_2, \beta\}$.

Which arguments of a theory are defeated and which undefeated is determined by its defeaters. We assume that arguments are normally undefeated, but can become defeated because of defeaters. A defeater only justifies the defeat of its defeated arguments, if its activating arguments are used in undefeated arguments, i.e., if they are subarguments of undefeated arguments. The defeater is then *activated*.

Definition 4.3 Let (L, Args, D) be a defeasible argumentation theory, A[B] a defeater in D, and Σ a defeasible argumentation structure of (L, Args, D). A[B] is *activated* in Σ, if $A \subseteq \text{Subs}[\Sigma]$.

In the defeasible argumentation structure $\{\beta\}$ the defeaters $\beta[\alpha_1]$ and $\beta[\alpha_2]$ are activated. In the structure $\{a_1, \alpha_1, a_2, \alpha_2, \beta\}$ all three defeaters are activated.

An acceptable set of undefeated arguments must have two intuitive properties:

- Arguments cannot be undefeated if their defeat is justified.
- Defeaters cannot be ignored unjustly.

The first of these properties becomes formally: Arguments of which the defeat is justified by an activated defeater cannot be contained in an acceptable defeasible argumentation structure. The second becomes: If a defeater is not activated in an acceptable defeasible argumentation structure, it must be *deactivated*.

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 even sufficient that the defeat of a subargument or a strengthening of one of the activating arguments is justified. So, it seems that $\alpha_1, \alpha_2[\beta]$ and $\beta[\alpha_1]$ deactivate each other. This is not the case, because of the accrual of the arguments α_1 and α_2. The defeater $\alpha_1, \alpha_2[\beta]$ *overrules* $\beta[\alpha_1]$. Therefore, the defeater $\beta[\alpha_1]$ cannot deactivate $\alpha_1, \alpha_2[\beta]$. This leads to the second condition: a defeater can only be deactivated by a defeater it does not overrule.

Definition 4.4 Let (L, Args, D) be a defeasible argumentation theory, A[B] and $\Gamma[\Delta]$ defeaters in D, and Σ a defeasible argumentation structure of (L, Args, D). A[B] *overrules* $\Gamma[\Delta]$ in Σ, if $A \supseteq \Delta$ and $B \supseteq \Gamma$.

Definition 4.5 Let (L, Args, D) be a defeasible argumentation theory, A[B] and $\Gamma[\Delta]$ defeaters in D, and Σ a defeasible argumentation structure of (L, Args, D). A[B] *deactivates* $\Gamma[\Delta]$, if the following hold:

1. There is an element of B that is a subargument or a strengthening of an element of Γ.
2. $\Gamma[\Delta]$ does not overrule A[B].

[11] Our definition of a defeasible argumentation structure is related to those 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 3).

We can finally define acceptable sets of undefeated arguments.[12] The requirements in the definition correspond to the two intuitive properties explained above.

Definition 4.6 Let (L, Args, D) be a defeasible argumentation theory, and Σ a defeasible argumentation structure of (L, Args, D). Σ is *acceptable* with respect to (L, Args, D), if the following hold:

1. If $A[B] \in D$ is activated, then $\Sigma \cap B = \emptyset$.
2. If $A[B] \in D$ is not activated, then there is an activated $\Gamma[A'] \in D$ that deactivates $A[B]$.

An acceptable defeasible argumentation structure of our example theory is $\{\alpha_1, \alpha_2\}$. The structure $\{\beta\}$ is not acceptable, because $\alpha_1, \alpha_2[\beta]$ is not deactivated.

We can now define an extension of a theory as an acceptable defeasible argumentation structure that is maximal with respect to set inclusion.[13]

Definition 4.7 Let (L, Args, D) be a defeasible argumentation theory. An acceptable defeasible argumentation structure Σ of (L, Args, D) is an *extension* of the theory, if

1. Σ is acceptable.
2. There is no acceptable defeasible argumentation structure Σ', such that $\Sigma \subset \Sigma'$.

The unique[14] extension of our example theory is the one we wanted: $\{\alpha_1, \alpha_2\}$. Even though the theory contains the defeaters $\beta[\alpha_1]$, $\beta[\alpha_2]$ 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 the following restricted theory that lacks α_2:

$L = \{a_1, a_2, a, b\}$,
$Args = \{a_1, a_1 \rightarrow a, b\}$,
$D = \{\beta[\alpha_1]\}$, where $\alpha_1 = a_1 \rightarrow a$, and $\beta = b$.

Now α_1 is defeated by β, and the extension of this theory is $\{\beta\}$. The argument α_1 does only remain undefeated if it is reinforced by α_2.

5 CONCLUSION

We conclude with a discussion of the main points of this paper: Defeat can be compound and arguments can accrue.

- Defeat can be compound.

We know of no other formalism for defeasible argumentation that models compound defeat. Restricted compound defeat occurs in Vreeswijk's (1991, 1993) formalism.[15] It might be argued that compound defeat can be modeled by any formalism that has a language with conjunction. In our opinion that is a wrong approach, because it obscures what is going on. Accruing arguments are *separate* arguments for a conclusion instead of one *composite* argument for that conclusion. Combining the arguments $a \rightarrow c$ and $b \rightarrow c$ must be distinguished from the argument $a, b \rightarrow c$.

- Arguments can accrue.

The reason that we have incorporated compound defeat in our formalism is the accrual of arguments. Accrual of arguments occurs in our formalism in two closely related ways: In strengthenings of an argument, and in defeaters that are activated by groups of arguments for a conclusion. We think it is obvious that arguments can accrue. Pollock's (1991) main point against the accrual of arguments is the following thought experiment. He asks to imagine a

[12] Acceptable defeasible argumentation structures are related to Dung's (1993) admissible sets of arguments and Pollock's (1994) partial status assignments.

[13] Dung's (1993) preferred extensions and Pollock's (1994, p. 393) status assignments are defined similarly.

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

[15] This is not obvious, because defeat is indirect in Vreeswijk's formalism. 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.

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. *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. Defeaters of the form $[\alpha_1, \alpha_2]$, where α_1 and α_2 represent different testimonies, can model Pollock's example.

ACKNOWLEDGEMENTS

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 Gerard Vreeswijk for his valuable suggestions to improve this paper.

REFERENCES

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.

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.

Hage, J. C. (1993). Monological reason based logic. A low level integration of rule-based reasoning and case-based reasoning. *The Fourth International Conference on Artificial Intelligence and Law. Proceedings of the Conference*, pp. 30-39. ACM, New York.

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*. Published as report SKBS/B3.A/94-10.

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

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

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

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

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

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

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

Prakken, H. (1993a). *Logical tools for modelling legal argument.* H. Prakken, Amsterdam.

Prakken, H. (1993b). 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.

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

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. Also published as report SKBS/B3.A/94-5.

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.

Vreeswijk, G. (1993). *Studies in defeasible argumentation.* G. A. W. Vreeswijk, Amsterdam.