On Resolving Conflicts Between Arguments
On Resolving Conflicts Between Arguments

Nico Roos
ON RESOLVING CONFLICTS BETWEEN ARGUMENTS

NICO ROOS

Maastricht University, Department of Computer Science,
P.O. Box 616, 6200 MD Maastricht, The Netherlands, e-mail: roos@cs.unimaas.nl

Argument systems originate from philosophy. They are based on the idea that one can construct arguments for propositions. These arguments can be viewed as structured reasons that justify the belief in a proposition. A justification need not be valid in all circumstances when arguments are constructed using defeasible rules. Using defeasible rules, it might be possible to construct an argument for a proposition as well as its negation. When arguments support conflicting propositions, one of the arguments, or a sub-argument of these arguments must be defeated in order to resolve the conflict. This raises the question concerning the arguments or sub-arguments that can be subject to defeat.

The paper shows that in case defeasible rules possess no contraposition, no proper sub-argument of an argument can be subject to defeat. This result makes it possible to resolve conflicts by constructing an undercutting argument for the last rule of a defeated argument. So, rebutting defeat is replaced by undercutting defeat. This has two consequences. Firstly, instead of determining the sets of valid arguments, it suffices to determine sets of valid rules. A linear time algorithm for determining a set of valid rules will be given. Secondly, the developed argument system can be related to Default Logic.

Key words: argumentation, defeasible rules, non-monotonic reasoning.

1. INTRODUCTION

Argument systems originate from philosophy (Toulmin 1958). More recently they have also been studied in AI (Bondarenko et al. 1997; Cayrol 1995; Dung 1993; Fox et al. 1992; Geffner 1994; Hunter 1994; Kraus et al. 1995; Lin & Shoham 1989; Loui 1987; Pollock 1987; Pollock 1992; Pollock 1994; Poole 1988; Prakken 1993; Simari & Loui 1992; Vreeswijk 1991; Vreeswijk 1997). When such argument systems are used for reasoning with defeasible rules (Fox et al. 1992; Geffner 1994; Hunter 1994; Kraus et al. 1995; Loui 1987; Pollock 1987; Prakken 1993; Simari & Loui 1992; Vreeswijk 1991; Vreeswijk 1997), a rule is viewed as a justification for believing the consequent of the rule whenever we have a justification for believing its antecedent (Toulmin 1958). A justification for believing the antecedent can consist of facts about the world, denoted as evidence or premises, and of propositions that are justified by other defeasible rules. So, we can construct a tree of defeasible rules that justifies the belief in some proposition with respect to some evidence. This tree is called an argument for the proposition.

A defeasible rule used in the construction of an argument, can be interpreted in different ways. It can be interpreted as describing that a proposition is preferred to hold in some context, ignoring sub-contexts that represent exceptions. Examples of such rules are legal norms, instructions of operation, and so on. A defeasible rule can also be interpreted as a conditional probability that is almost 1. Some authors interpret a probability that is almost 1 as a probability that is infinitely close to 1 (Geffner & Pearl 1992; Pearl 1989). Clearly, this is an unrealistic assumption that we should try to avoid.

Both interpretations of defeasible rules have in common that they ignore exceptional situations in which the rule describes an invalid relation. Ignoring possible exceptions offers three advantages.

• We can establish a relation without the need to specify a possibly endless list unlikely exceptions. By doing so, simpler rules can be formulated. Furthermore, we can still apply a rule when lacking information about the exceptional but unlikely cases.
The following picture denotes the situations in which a rule can be applied.

- We can handle unforeseen cases in which a rule is applicable though it should not be. For example, legal rules can be subject to exceptions because a legislator cannot foresee all possible cases on which the rule can be applied, but which do not correspond with the underlying motivation of the rule (Prakken 1993). In such specific cases a judge can make an exception.

- We can handle unforeseen interactions between rules. Rules supporting conflicting conclusions can be applicable in unforeseen cases. Clearly, in such cases, one of the rules cannot be a valid justification for its consequent.

Since defeasible rules need not be valid in all circumstances, it is possible to construct arguments for two or more propositions that are in conflict. A number of these conflicts can be resolved by comparing the arguments involved and subsequently decide which (sub-) arguments are valid. In most argument systems proposed in the literature, one of the arguments supporting the conflicting propositions is defeated (Loui 1987; Pollock 1987; Pollock 1992; Pollock 1994; Prakken 1993; Simari & Loui 1992; Vreeswijk 1991; Vreeswijk 1997). Since a sub-argument is defeated only if it contains the last defeasible rule of the defeated argument, this approach can be interpreted as defeating the last defeasible rule of the argument. Yet others have argued that one should not only consider the last rule as a candidate for defeat (Geffner 1994). Geffner (1994), for example, allows the chain of argumentation to be broken at any rule of an argument if the argument conflicts with an observed fact. Unlike other authors, his defeasible rules represent causal relations.

Although most authors agree on the fact that the last rule of one of the conflicting arguments must be defeated, they do not show the necessity of this fact. In this paper, we will investigate whether this is necessary. Furthermore, we will use the result of this investigation to resolve conflicts by constructing arguments for undercutting defeat of rules. Constructing arguments for the undercutting defeat of rules will enable us to present a new way of defining the set of valid arguments. The valid arguments are those arguments that do not contain invalid (defeated) rules. So if all defeasible steps of an argument are valid, then so is the argument, and the proposition that it supports should be believed. This approach is more natural than using a defeat relation between arguments (Pollock 1987; Pollock 1994;
Simari & Loui 1992; Vreeswijk 1997). Furthermore, it will enable us to calculate a set of valid rules in linear time.

The next section formalizes the arguments that can be constructed using defeasible rules. 
Sections 3 discusses the resolution of conflicts between arguments. Based on the results of Section 3, Section 4 proposes a preference relation on the set of rules to evaluate conflicting arguments. Section 5 uses the result of Section 3 to construct arguments for the undercutting defeat of rules, and to defined the extensions and the belief set of a defeasible theory. Section 6 discusses how to compute an extension and Section 7 discusses some properties of the here proposed approach. Section 8 discusses related work and Section 9 concludes the paper.

2. THE ARGUMENT SYSTEM

We will derive arguments using a defeasible theory \( \langle \Sigma, D \rangle \). Here, \( \Sigma \) represents a set of premises and \( D \) represents a set of defeasible rules. The set of premises \( \Sigma \) is a subset of the propositional logic \( L \). \( L \) is recursively defined from a set of atomic propositions \( At \) and the operators \( \neg, \land \) and \( \lor \).

For every defeasible rule \( \varphi \leadsto \psi \in D \) there holds that \( \varphi \) is a proposition in \( L \) and that \( \psi \) is either a proposition in \( L \) or the negation of a defeasible rule in \( D \); i.e. \( \psi = \neg(\alpha \leadsto \beta) \) and \( \alpha \leadsto \beta \in D \). The negation of a defeasible rule \( \neg(\alpha \leadsto \beta) \) will be interpreted as: ‘\( \alpha \) may no longer justify \( \beta \)’. So the negation of a rule explicitly blocks the conclusive force of the defeasible rule. It will be used to describe the undercutting defeat of an argument. If we have a valid argument for \( \neg(\alpha \leadsto \beta) \), then no argument containing the rule \( \alpha \leadsto \beta \) can be valid.

In an argument system, a defeasible rule is viewed as a justification for believing the consequent of the rule whenever we have a justification for believing its antecedent (Toulmin 1958). A justification for believing the antecedent can consist of facts about the world, denoted as evidence or premises, and of propositions that are justified by other defeasible rules. So, we can construct a tree of defeasible rules that justifies the belief in some proposition with respect to some evidence. This tree is called an argument for the proposition.

Logically sound deductions need not be represented in an argument. None of these deduction steps can be subject to defeat. Only the relations described by defeasible rules need not be valid in all circumstances.

**Definition 1.** Let \( \langle \Sigma, D \rangle \) be a defeasible theory where \( \Sigma \) is the set of premises and \( D \) is the set of rules.

Then an argument\(^1\) \( A \) for a proposition \( \psi \) is recursively defined in the following way:

- For each \( \psi \in \Sigma \): \( A = \{ \langle \emptyset, \psi \rangle \} \) is an argument for \( \psi \).
- Let \( A_1, ..., A_n \) be arguments for respectively \( \varphi_1, ..., \varphi_n \). If \( \varphi_1, ..., \varphi_n \vdash \psi \), then \( A = A_1 \cup ... \cup A_n \) is an argument for \( \psi \).
- For each \( \varphi \leadsto \psi \in D \) if \( A' \) is an argument for \( \varphi \), then \( A = \{ \langle A', \varphi \leadsto \psi \rangle \} \) is an argument for \( \psi \).

\(^1\)We will sometimes add the index \( \psi \) to an argument \( (A_\psi) \) to denote that it is an argument for \( \psi \). Of course there can be more than one argument for \( \psi \).
Let $A = \{ \langle A'_1, \alpha_1 \rangle, ..., \langle A'_n, \alpha_n \rangle \}$. Then:

$\bar{A} = \{ \alpha_1, ..., \alpha_n \} \cap D$;

$A = \{ c(\alpha_1), ..., c(\alpha_n) \} \text{ where } c(\alpha) = \alpha \text{ if } \alpha \notin D \text{ and } c(\alpha \leadsto \beta) = \beta$;

$\bar{A} = \{ \alpha_i | 1 \leq i \leq n, \alpha_i \in D \} \cup \bigcup_{i=1}^{n} A'_i$;

$\check{A} = \{ \alpha_i | 1 \leq i \leq n, \alpha_i \in \Sigma \} \cup \bigcup_{i=1}^{n} A'_i$

**Example 1.** Let $A = \{ \langle \emptyset, \alpha \rangle, \{ \{ \langle \emptyset, \beta \rangle, \gamma \leadsto 1\} \}, \gamma \leadsto \delta \} \}$ be an argument for $\varphi$.

\[
\beta \vdash \beta \leadsto \gamma \vdash \gamma \leadsto \delta \vdash \varphi
\]

Then $\bar{A} = \{ \gamma \leadsto \delta \}$ denotes the last rules used in the argument $A$. Furthermore, $\check{A} = \{ \alpha, \delta \}$ denotes the propositions that represent the base belief set of the argument $A$. The base belief set is equal to $Th(\{ \alpha, \delta \})$. Clearly, $A$ is an argument for every proposition $\varphi \in Th(\{ \alpha, \delta \})$.

\[
\bar{A} = \{ \gamma \leadsto \delta, \beta \leadsto \gamma \}$ denotes the set of all rules in $A$, and $\check{A} = \{ \alpha, \beta \}$ denotes the premises used in the argument $A$.

In the above definition of an argument, we do not apply the contraposition of a defeasible rule in the construction of an argument. In general, the contraposition of a defeasible rule is invalid. A rule describes that its consequent should hold or probably holds in context described by its antecedent. By no means this implies that the antecedent does not hold if the consequent does not hold. If the defeasible rule represents a conditional probability, $Pr(\psi | \varphi) > t$ does not imply that $Pr(\neg \varphi | \neg \psi) > t$. In fact, if $Pr(\psi | \varphi) < 1$, $Pr(\neg \varphi | \neg \psi)$ can have any value in the interval $[0,1]$. Only in the event that we also know the a priori probabilities of $Pr(\varphi)$ and of $Pr(\psi)$, we can verify whether $Pr(\neg \varphi | \neg \psi) > t$ holds.

Also if the defeasible describes a preference, the negation of the consequent does not imply that the negation the antecedent should hold. A rule describes what should hold in the context described by its antecedent. The converse need not hold. So, knowing that John may not drive a car, we may not conclude that he does not own a driving license. It may just be the case that we have an exceptional situation, e.g. John is drunk, John has collected too many speeding tickets, John may not drive a car on doctors orders, and so. Especially if most people own a driving license, an exceptional situation need not be unlikely.

**Causal rules** are a special kind of defeasible rules that do possess a contraposition (Geffner, 1994). If, ‘normally, $\varphi$ causes $\psi’$, then $\neg \psi$ implies $\neg \varphi$, unless we have an exceptional situation. Such a rule can be described by a conditional probability, as is done in Bayesian Belief Networks. This description is incomplete unless we know or we can calculate the a priori probabilities of the antecedent and the consequent. Bayesian Belief Networks guarantee the latter. Here, however, we do not have this information. Therefore, to guarantee that the contraposition is applied correctly, we need a specialized approach. Geffner (1994) discusses the properties of such an approach. In the remainder of this paper, however, we will not consider causal rules.

Two arguments can be related to each other. The relation that is of interest for us is whether one argument uses the same inference steps as another argument. If so, the former is called a sub-argument of the latter. Though an argument can be viewed as a tree, a sub-argument is not exactly a sub-tree.

**Definition 2.** An argument $A$ is a sub-argument of $B$, $A \subseteq B$, if and only if every $\langle A', \alpha \rangle \in A$ is a sub-structure of the argument $B$.\footnote{Notice that we reach the base of the recursion if $A$ is an empty set. If $A$ is an empty set, it is trivial that every $\langle A', \alpha \rangle \in A$ is a sub-structure of the argument $B$.}
\( (A', \alpha) \) is a sub-structure of an argument \( B \) if and only if

- either there exists a \( (B', \alpha) \in B \) such that \( A' \) is a sub-argument of \( B' \);
- or there exists a \( (B', \beta) \in B \) such that \( (A', \alpha) \) is a sub-structure of \( B' \).

**Example 2.** Let \( A = \{ (\emptyset, \alpha), \{ \{ (\emptyset, \beta) \}, \beta \leadsto \gamma \}, \gamma \leadsto \delta \} \) be an argument. Then

\[
\begin{align*}
    \{ (\emptyset, \alpha), \{ \{ (\emptyset, \beta) \}, \beta \leadsto \gamma \} \}, \gamma \leadsto \delta \\
    \{ (\emptyset, \alpha), \{ (\emptyset, \beta) \}, \beta \leadsto \gamma \} \\
    \{ (\emptyset, \alpha) \} \\
    \{ \{ \{ (\emptyset, \beta) \}, \beta \leadsto \gamma \}, \gamma \leadsto \delta \} \\
    \{ \{ (\emptyset, \beta) \}, \beta \leadsto \gamma \} \\
    \{ (\emptyset, \beta) \}
\end{align*}
\]

are sub-arguments of \( A \).

An argument represents a derivation tree of defeasible rules. Since a rule in an argument \( A \) gives a justification for its consequent, the argument can be viewed as a global justification for a proposition \( \varphi \), \( \hat{A} \vdash \varphi \), that is grounded in the premises \( \bar{A} \). Whether an argument is valid depends on whether the argument or one of its sub-arguments is defeated. When an argument \( A \) for some proposition \( \varphi \) is valid we say that \( \varphi \) follows from the premises \( \bar{A} \) using the rules \( \tilde{A} \).

### 3. CONFLICTING ARGUMENTS

A defeasible rule \( \varphi \leadsto \psi \) describes a either preferred or a probabilistic relation. Therefore, there may exist situations in which the relation it represents is invalid. In these exceptional situations, either \( \neg \psi \) must holds or both \( \psi \) and \( \neg \psi \) must be unknown. Following Pollock (1987), the former situation is called rebutting defeat and the latter is called undercutting defeat. Brewka (1991a) uses the terms hard and weak exceptions.

In case of rebutting defeat, we can construct an argument for an inconsistency (\( \bot \)). Arguments supporting propositions that are inconsistent are said to disagree. If \( A_\bot = \{ (A'_1, \mu_1), ..., (A'_n, \mu_n) \} \) is an argument for \( \bot \), \( \hat{A}_\bot \vdash \bot \), then the arguments \( A_1 = \{ (A'_1, \mu_1) \}, ..., A_n = \{ (A'_n, \mu_n) \} \) are said to disagree. Disagreeing arguments must be compared in order to resolve the derived inconsistency. After comparing the arguments, we may decide that one of the arguments is defeated by the others. By this we mean that the argument may no longer justify a proposition because of the other arguments.

What should it exactly mean: an argument is defeated by other arguments? For example, does it mean that only the last rule in the argument may no longer justify its consequent or can it also mean that one of its proper sub-arguments is also defeated. Let us assume that latter case is allowed for. Suppose that \( \{ \psi, \mu \} \) is an inconsistent set, that \( \psi \) is justified by the argument

\[
A_\psi = \{ \{ \{ (\emptyset, \alpha) \}, \alpha \leadsto \varphi \}, \varphi \leadsto \psi \}
\]

\[
\alpha \vdash \alpha \leadsto \varphi \vdash \varphi \leadsto \psi \vdash \psi
\]

and that \( \mu \) is justified by

\[
A_\mu = \{ \{ (\emptyset, \eta) \}, \eta \leadsto \mu \}
\]

\[
\eta \vdash \eta \leadsto \mu \vdash \mu
\]
If $A_\mu$ defeats the sub-argument

$$A_\varphi = \{ (\emptyset, \alpha), \alpha \models \varphi \}$$

of $A_\varphi$, then the situation in which $\alpha$ and $\eta$ hold ($\alpha \land \eta$) represents an exception on the rule $\alpha \models \varphi$. In this exceptional situation either $\lnot \varphi$ holds or $\varphi$ is unknown.

Suppose that $\lnot \varphi$ holds. Since we cannot use contraposition of $\varphi \models \psi$, there is only way we can ensure that $\lnot \varphi$ holds. It must be possible to construct an argument $A_{\lnot \varphi}$ for $\lnot \varphi$ such that $A_{\lnot \varphi} \subseteq \{ \alpha, \eta \}$. So, there must be a set of defeasible rules $\{ \xi_1 \models \nu_1, \ldots, \xi_k \models \nu_k \}$ such that $\xi_1, \ldots, \xi_k$ are derivable from $\{ \alpha, \eta \}$, and such that $\{ \nu_1, \ldots, \nu_k, \alpha, \eta \}$ implies $\lnot \varphi$. But then, there is no longer a need for allowing that $A_\mu$ defeats $A_\varphi$ since

$$A_{\lnot (\alpha \models \varphi)} = \{ \langle A, \xi \models \lnot (\alpha \models \varphi) \rangle \}$$

already defeats $A_\varphi$.

If, on the other hand, $\varphi$ is unknown, either a proper sub-argument of $A_\varphi$ is defeated by $A_\mu$ or there must be a rule $\xi \models \lnot (\alpha \models \varphi)$ where $\xi$ is derivable from $\{ \alpha, \eta \}$. In the former case, we have the fact $\alpha$ which cannot be defeated. In the latter case, there is no need for allowing that $A_\mu$ defeats $A_\varphi$ since

$$A_{\lnot (\alpha \models \varphi)} = \{ \langle A, \xi \models \lnot (\alpha \models \varphi) \rangle \}$$

already defeats $A_\varphi$.

The line of reasoning presented above also applies to a set of disagreeing arguments.

**Theorem 1.** A conflict described by a set of disagreeing arguments can be resolved by defeating one of the disagreeing arguments.

**Proof.** Let

$$A_1 = \{ \langle A'_1, \mu_1 \rangle, \ldots, A_n = \{ \langle A'_n, \mu_n \rangle \}$$

be the set of disagreeing arguments; i.e. $\bigcup_{i=1}^n A_i$ is an argument for $\bot$. Suppose that some proper sub-argument $A_\varphi = \{ \langle A', \alpha \models \varphi \rangle \}$ of one of the disagreeing arguments $A_k = \{ \langle A'_k, \mu_k \rangle \}$ is defeated by $A_\mu = A_1 \cup \ldots \cup A_{k-1} \cup A_{k+1} \cup \ldots \cup A_n$ and that no proper sub-argument of $A_\varphi$ is defeated by $A_\mu$. Then, $\bigcup_{i=1}^n A_i$ represents an exceptional situation in which either $\lnot \varphi$ holds or $\varphi$ is unknown.

Suppose that $\lnot \varphi$ holds. Since we cannot use the contraposition of a rule, it must be possible to construct an argument $A_{\lnot \varphi}$ for $\lnot \varphi$ such that $A_{\lnot \varphi} \subseteq \bigcup_{i=1}^n A_i$. So, there must be a set of defeasible rules $\{ \xi_1 \models \nu_1, \ldots, \xi_k \models \nu_k \}$ such that $\xi_1, \ldots, \xi_k$ are derivable from $\bigcup_{i=1}^n A_i$, and such that $\{ \nu_1, \ldots, \nu_k \} \cup \bigcup_{i=1}^n A_i$ implies $\lnot \varphi$. But then, there is no longer a need for allowing that $A_\mu$ defeats $A_\varphi$ since $A_{\lnot \varphi}$ can and should defeat $A_\varphi$.

Now suppose that $\varphi$ is unknown. Then either a proper sub-argument of $A_\varphi$ is defeated by $A_\mu$ or there must be a rule $\xi \models \lnot (\alpha \models \varphi)$ where $\xi$ is derivable from $\bigcup_{i=1}^n A_i$. The former situation is not possible since we assumed that $A_\varphi$ is a smallest ($\subseteq$) sub-argument that is defeated by $A_\mu$. In the latter situation, $A_{\lnot (\alpha \models \varphi)}$ with $A_{\lnot (\alpha \models \varphi)} \subseteq \bigcup_{i=1}^n A_i$ already defeats $A_\varphi$. 


Notice that the absence of the contraposition of a defeasible rule is essential for this result. If defeasible rules possess a contraposition, the line of reasoning given above is no longer valid.

4. EVALUATING ARGUMENTS

How do we determine which argument of a set of disagreeing arguments must be defeated? We will assume that it suffices to consider only the last rules of each of the disagreeing arguments for this purpose. This assumption offers two advantages. Firstly, it becomes possible to defeat rules instead of arguments. So, an argument looses its conclusive force (is defeated) if it contains defeated rules. For the resolution of conflicts this has the following implication.

A conflict described by a set of disagreeing arguments must be resolved by defeating the last rule of one of the disagreeing arguments.

Secondly, the resolution of inconsistencies is cumulative. It does not matter whether the antecedent of a last rule is an observed fact or derived through reasoning. Furthermore, an observed fact may be based on some hidden reasoning of which we are not aware.

Considering the whole argument for an inconsistency to determine the rule to be defeated, is not necessary either. This is illustrated by following two arguments for an inconsistency;

\[ A_\bot = \{\langle\{\emptyset, \alpha\}\rangle, \alpha \leadsto \varphi\}, \{\langle\{\emptyset, \beta\}\rangle, \beta \leadsto \psi\}, \psi \leadsto \neg\eta\} \]

\[ \alpha \vdash \alpha \leadsto \varphi \vdash \varphi \leadsto \eta \mid \bot \]

and

\[ A'_\bot = \{\langle\emptyset, \varphi\rangle\}, \varphi \leadsto \eta\}, \{\emptyset, \psi\}, \psi \leadsto \neg\eta\} \]

\[ \varphi \vdash \varphi \leadsto \eta \mid \bot \]

\[ \psi \vdash \psi \leadsto \neg\eta \mid \bot \]

Suppose that \(\varphi \leadsto \eta\) must be defeated given \(A_\bot\) and \(\psi \leadsto \neg\eta\) must be defeated given \(A'_\bot\). In the former case, the situation described by \(\alpha\) and \(\beta\) represents an exception on the rule \(\varphi \leadsto \eta\). We can, for example, describe this exception by introducing the rule \(\alpha \land \beta \leadsto \neg\eta\) or the rule \(\alpha \land \beta \leadsto \neg(\varphi \leadsto \eta)\). Since each of these rules defeats \(\varphi \leadsto \eta\), \(\varphi \leadsto \eta\) can no longer defeat \(\psi \leadsto \neg\eta\). Hence, we only have to consider the last rules of an argument for an inconsistency.

An additional motivation for considering only the last rules of an argument for an inconsistency, comes from Prakken’s investigation of legal argumentation (Prakken 1993). He observes that in legal argumentation, only the last rules of an argument for an inconsistency are used to resolve the inconsistency.

To determine the rule to be defeated by other rules after deriving an inconsistency, usually, some preference relation on the rules is used. The most commonly used preference relation is a specificity order. In legal argumentation other preference relations are applied as well (Prakken 1993). These preference relations are specified by meta-rules. This makes them subject to defeat in situations where the meta-rules specify incompatible relations between rules (Brewka 1994).

Using a preference relation on the set of defeasible rules, it seems obvious that the least preferred rule of the set of last rules of an argument for an inconsistency must be defeated.
There is one possible objection against this approach. A set of arguments supporting the
same proposition may be stronger than each individual argument. This is known as *accrual of reasons.* We can, however, handle this situation by using a rule that combines the last
rules of each argument for that proposition. To illustrate this, suppose that we have the
following defeasible rules: \( \alpha \leadsto \psi, \beta \leadsto \psi \) and \( \gamma \leadsto \neg \psi. \) Let the last rule be preferred to
the first two rules. Then \( \neg \psi \) must hold if \( \gamma \) and either \( \alpha \) or \( \beta \) hold. By introducing a rule
\( \alpha \land \beta \leadsto \psi \) and by preferring it to \( \gamma \leadsto \neg \psi, \) we can assure that \( \psi \) holds whenever \( \alpha, \beta \) and \( \gamma \) hold.

Another problem arises when a set of arguments for a proposition weakens the support for
the proposition. The approach presented here offers no solution such situations. Fortunately,
a set of arguments that weakens the support for a proposition, seems to be counter intuitive.

**Definition 3.** Let \( \langle \Sigma, D \rangle \) be a defeasible theory and let \( \succ \) be a partial preference relation on
\( D. \) Furthermore, let

\[
A_1 = \{ \langle A'_1, \eta_1 \leadsto \psi_1 \rangle \}, \ldots, A_k = \{ \langle A'_k, \eta_k \leadsto \psi_k \rangle \},
A_{k+1} = \{ \emptyset, \sigma_1 \}, \ldots, A_n = \{ \emptyset, \sigma_i \}
\]

be disagreeing arguments.

Then, if \( \eta_i \leadsto \psi_i \) is the least preferred last rule in \( \tilde{A} \) where \( A = A_1 \cup \ldots \cup A_n, \eta_i \leadsto \psi_i \)
must be defeated.

Since we are using a preference relation on the set of defeasible rules in order to solve
conflicts, we should extend the definition of a defeasible theory \( \langle \Sigma, D \rangle \) with the preference
relation \( \succ, \) i.e. \( \langle \Sigma, D, \succ \rangle. \) Certainly, to describe legal argumentation, this extension is
necessary. If we restrict ourselves to one specific preference relation, namely *specificity,* there
is no need to extend the definition of a defeasible theory. The *specificity* preference relation
can be derived from the set of defeasible rules of a defeasible theory.\(^3\)

Specificity is the principle by which rules applying to a more specific situation override
those applying to more general ones. The most specific situation to which a rule can be
applied is the situation in which only its antecedent is known to hold. In that situation, its
consequent must hold. The following preference relation is based on the fact that the most
specific situation to which a rule can be applied is the situation in which only its antecedent
is known to hold.

**Definition 4.** Let \( D \) be a set of defeasible rules and let \( \varphi \leadsto \psi, \eta \leadsto \mu \) be two rules in \( D.
\)

\( \varphi \leadsto \psi \) is more specific than \( \eta \leadsto \mu \) if and only if, given the premises \( \{ \varphi \}, \) there is an
argument \( A_\eta \) for \( \eta \) such that \( A_\eta \subseteq \{ \varphi \}. \)

\( \varphi \leadsto \psi \) is strictly more specific than \( \eta \leadsto \mu, \varphi \leadsto \psi \succ spec \eta \leadsto \mu, \) if and only if \( \varphi \leadsto \psi \)
is more specific than \( \eta \leadsto \mu \) and \( \eta \leadsto \mu \) in not more specific than \( \varphi \leadsto \psi. \)

**Example 3.** Let \( \varphi \leadsto \psi, \varphi \leadsto \eta \) and \( \eta \leadsto \neg \psi \) be three defeasible rules.

Given the premises \( \{ \varphi \}, \) we can derive the argument \( A_\eta = \{ \langle \emptyset, \varphi \rangle, \varphi \leadsto \eta \}. \) Since
\( A_\eta \subseteq \{ \varphi \}, \varphi \leadsto \psi \) is more specific than \( \eta \leadsto \neg \psi. \) Furthermore, since, given the premises \( \{ \eta \}, \)
there is no argument for \( \varphi, \varphi \leadsto \psi \) is strictly more specific than \( \eta \leadsto \neg \psi. \)

The above defined specificity preference relation is corresponds with definition of specificity
implied by the axioms of conditional logics (Geffner & Pearl 1992). This definition of
specificity is relatively weak. Vreeswijk (1991) presents an example showing that a slightly
stronger definition can result in counter intuitive conclusions.

\(^3\)Since the set of rules \( D \) is usually considered as background knowledge, we can determine the specificity
preference relation in advance.
5. THE BELIEF SET

An inconsistency can be resolved considering the last rules of the argument for the inconsistency. This implies that in case the inconsistency is resolved, one of these last rules may no longer justify the belief in its consequent; i.e. the rule is defeated. For this rule we can construct an argument supporting the undercutting defeat of this rule.

Definition 5. Let $A_\bot$ be an argument for an inconsistency and let $\varphi \leadsto \psi \in \min_\varphi (\bar{A}_\bot)$ be a least preferred last rule for the inconsistency.

If $(A_\varphi, \varphi \leadsto \psi) \in A_\bot$, then $A_{\neg(\varphi \leadsto \psi)} = (A - \{(A_\varphi, \varphi \leadsto \psi)\}) \cup A_\varphi$ is an argument for the defeat of $\varphi \leadsto \psi$.

Example 4. Let

$$A_\bot = \{\{\{\emptyset, \alpha\}\}, \alpha \leadsto \varphi\}, \varphi \leadsto \psi\}, \{\{\emptyset, \eta\}\}, \eta \leadsto \mu\}$$

If $\eta \leadsto \mu$ is preferred to $\varphi \leadsto \psi$, then

$$A_{\neg(\varphi \leadsto \psi)} = \{\{\{\emptyset, \alpha\}\}, \alpha \leadsto \varphi\}, \{\{\emptyset, \eta\}\}, \eta \leadsto \mu\}$$

Given these arguments for the defeat of rules, we can define an extension. Here an extension is a set of propositions for which we have valid arguments. A valid argument is an argument of which the rules are not defeated. This also holds for the arguments for the defeat of rules. A rule is defeated if the argument for its defeat is valid; i.e. the argument does not contain defeated rules.

Definition 6. Let $\mathcal{A}$ be a set of all derivable arguments and let

$$\text{Defeat}(\Gamma) = \{\alpha \leadsto \beta | A_{\neg(\alpha \leadsto \beta)} \in \mathcal{A}, \bar{A}_{\neg(\alpha \leadsto \beta)} \cap \Gamma = \emptyset\}.$$  

Then the set of defeated rules $\Omega$ is defined as:

$$\Omega = \text{Defeat}(\Omega).$$

Theorem 2. The set of defeated rules $\Omega$ is well founded. I.e. for each $\Lambda \neq \Omega$ such that $\Lambda = \text{Defeat}(\Lambda)$, neither $\Lambda \subset \Omega$ nor $\Lambda \supset \Omega$ holds.

Proof. Suppose $\Lambda \subset \Omega$. Then, by the definition of Defeat: $\text{Defeat}(\Lambda) \supset \text{Defeat}(\Omega)$. Hence, $\Lambda \supset \Omega$. Contradiction.

Suppose $\Omega \subset \Lambda$. Then, by the definition of Defeat: $\text{Defeat}(\Omega) \supset \text{Defeat}(\Lambda)$. Hence, $\Omega \supset \Lambda$. Contradiction.  

All argument systems agree on the fact that an argument is defeated if one of its sub-arguments is defeated. So, no rule of an argument is defeated if and only if no sub-argument is defeated.
Notice that the set of defeated rules need not be unique. Even if every inconsistency has a unique least preferred last rule, the set of defeated rules need not be unique. Consider for example the facts \( \alpha \) and \( \beta \) and rules \( \alpha \Rightarrow \gamma \), \( \beta \Rightarrow \delta \), \( \gamma \Rightarrow \neg \delta \) and \( \delta \Rightarrow \gamma \), where the last two rules are preferred to the first two. Here there are two sets of defeated rules \( \Omega \): \( \{ \alpha \Rightarrow \gamma \} \) and \( \{ \beta \Rightarrow \delta \} \).

Given a set of defeated rules, the extensions and the belief set can be defined. An extension consists of all propositions for which we have a valid argument. Following Pollock (1987), these propositions are said to be warranted.

**Definition 7.** Let \( \Omega \) be a set of defeated rules and let \( A \) be a set of all derivable arguments.

Then an extension \( E \) is defined as:

\[
E = \{ \varphi \mid A_\varphi \in A, \hat{A}_\varphi \cap \Omega = \emptyset \}.
\]

**Definition 8.** Let \( (\Sigma, D, \succ) \) be a defeasible theory. Furthermore, let \( E_i, \ldots, E_n \) be the corresponding extensions.

Then the belief set \( B \) is defined as:

\[
B = \bigcap_{i=1}^{n} E_i.
\]

### 6. Determination of the Fixed Point of Defeat

The determination of the fixed points of Defeat can be viewed as a labeling problem of a JTMS (Doyle 1979).\(^5\) Such a JTMS must contain a node \( N \) for every proposition of the form \( \neg(\alpha \Rightarrow \beta) \) for which we have an argument in \( A \). Furthermore, for each node \( N_{\neg(\varphi \Rightarrow \psi)} \) representing \( \neg(\alpha \Rightarrow \beta) \) and for each argument in \( A \) supporting \( \neg(\alpha \Rightarrow \beta) \), the JTMS contains a justification \( \langle \text{in-nodes}, \text{out-nodes}, N_{\neg(\varphi \Rightarrow \psi)} \rangle \). Such a justification consists of an empty set of in-nodes and a set of out-nodes. If \( A \) is an argument for \( \neg(\alpha \Rightarrow \beta) \), then

\[
\langle \emptyset, \{ N_{\neg(\varphi \Rightarrow \psi)} \mid \varphi \Rightarrow \psi \in \hat{A} \}, N_{\neg(\varphi \Rightarrow \psi)} \rangle
\]

is a justification for \( N_{\neg(\varphi \Rightarrow \psi)} \).

Each valid labeling of the nodes corresponds with a fixed point of Defeat. The fixed point consists of the those negated rules that are label ‘IN’.

**Theorem 3.** A set of rules \( \Omega \) is a fixed point of Defeat if and only if there is a labeling of the JTMS such that for each \( \varphi \Rightarrow \psi \in \Omega \), \( N_{\neg(\varphi \Rightarrow \psi)} \) is labeled IN.

**Proof.** Let \( \Omega = \{ \varphi \Rightarrow \psi \mid N_{\neg(\varphi \Rightarrow \psi)} \text{ is labeled IN} \}. \)

Suppose that \( \Omega \) is not a fixed point. Then, for some \( \varphi \Rightarrow \psi \), \( \varphi \Rightarrow \psi \in \text{Defeat}(\Omega) \) and \( \varphi \Rightarrow \psi \notin \Omega \) or \( \varphi \Rightarrow \psi \notin \text{Defeat}(\Omega) \) and \( \varphi \Rightarrow \psi \in \Omega \).

Suppose that \( \varphi \Rightarrow \psi \in \text{Defeat}(\Omega) \) and \( \varphi \Rightarrow \psi \notin \Omega \). Since \( \varphi \Rightarrow \psi \in \text{Defeat}(\Omega) \), there is an argument \( A_{\neg(\varphi \Rightarrow \psi)} \) such that \( A_{\neg(\varphi \Rightarrow \psi)} \cap \Omega = \emptyset \). But then, by the construction of the JTMS, \( N_{\neg(\varphi \Rightarrow \psi)} \) must be labeled IN. Contradiction.

\(^5\) A JTMS consists of nodes representing propositions, and of justifications. A node is either labeled IN or OUT, which corresponds with respectively ‘is believed’ and ‘is not believed’. A justification is a triple consisting of a set of in-nodes, a set of out-nodes and a consequent node. The consequent node is labeled IN if all in-nodes are labeled IN and no out-node is labeled IN.
Suppose that \( \varphi \triangleright \psi \notin \text{Defeat}(\Omega) \) and \( \varphi \sim \psi \in \Omega \). Since \( \varphi \sim \psi \notin \text{Defeat}(\Omega) \), there is no argument \( A_{\neg(\varphi \sim \psi)} \) such that \( A_{\neg(\varphi \sim \psi)} \cap \Omega = \emptyset \). But then, by the construction of the JTMS, \( N_{\neg(\varphi \sim \psi)} \) must be labeled \text{OUT}. Contradiction.

Hence, \( \Omega \) is a fixed point.

Now let \( \Omega \) be a fixed point of \text{Defeat}.

Suppose that it is impossible to label all nodes corresponding with \( \Omega \)\text{IN}. If we nevertheless label the nodes that correspond with \( \Omega \)\text{IN} and all remaining nodes \text{OUT}, then there must be a justification that is invalidated. Let \( \langle \emptyset, X, N_{\neg(\varphi \sim \psi)} \rangle \) be this justification. Then a node \( N_{\neg(\alpha \sim \beta)} \in X \) is labeled \text{IN}. By the construction of the JTMS, there is an argument \( A_{\neg(\varphi \sim \psi)} \) such that \( X = \{ N_{\neg(\eta \sim \mu)} \mid \eta \sim \mu \in A_{\neg(\varphi \sim \psi)} \} \). Therefore, since \( N_{\neg(\alpha \sim \beta)} \in X \) is labeled \text{IN}, \( \varphi \sim \psi \notin \text{Defeat}(\Omega) \). Contradiction.

Hence, a valid labeling corresponds with a fixed point of \( \Omega \).

Much research has been done on algorithms for labeling nodes in a JTMS network (Doyle 1979; Goodwin 1987; Reinfrank 1989; Witteveen & Brewka 1993). Some also deal with situations in which no admissible labeling exists (Witteveen & Brewka 1993). Partial labeling has been proposed for these situations.

When no admissible labeling exists, the set of arguments \( A \) contains forms of self defeating arguments. In its most simple form, self defeat is related to one argument \( \alpha \sim \beta \in \tilde{A}_{\neg(\varphi \sim \psi)} \). In general, self defeat is represented by odd loops in the corresponding JTMS.

No extension exists if a set of arguments contains self-defeating arguments. This is a problematic case that seems to present a defect in our knowledge. We have a set of rules \( D \) that in some circumstances represents nonsense. In that case, the set of rules \( D \) must be revised. I.e. some rules must be removed or reformulated. Though this is an important topic, it falls outside the scope of this paper.

Odd loops in the network can be determined in linear time with respect to \( n \cdot d \) where \( n \) is the number of nodes and \( d \) is the maximum number of outgoing justifications of any node. After detecting an odd loop we can mark the nodes involved as being ‘UNDETERMINED’, as well as the nodes that necessarily depend on nodes in an odd loop. This labeling of some of the nodes can subsequently be replaced by \text{IN} or \text{OUT} if the labeling of the remaining nodes enforces this. Hopefully, after labeling all nodes, no UNDETERMINED nodes remain. By doing so, we handle odd loops in an pragmatic way.

Finding a grounded labeling of a JTMS network is, in general, an NP-Hard problem. Fortunately, for the above presented JTMS networks without odd loops, we can find a labeling in linear time with respect to \( n \cdot d + j \) where \( n \) is the number of nodes, \( d \) is the maximum number of outgoing justifications of any node and \( j \) is the total number of justifications. An algorithm for finding a grounded labeling will be given in Appendix B.

7. PROPERTIES

Default logic. In Section 3, we have seen that it suffices to consider only the last rules of an argument for an inconsistency. This property enables us to define a default logic that is equivalent with respect to the belief set. This default logic will be based on Brewka’s prioritized default logic (Brewka 1994). Brewka argues that it is sufficient to use only normal default rules in combination with a preference relation on these rules. Here, we will follow a similar approach, but we will use the preference relation in a different way as Brewka proposes.

Since we only consider normal default rules, it suffices to verify whether a rule is applicable –its antecedents hold–, and whether it is not defeated by other rules –its consequent
The consequences of a set of applicable rules, together with the premises, may form an inconsistent set of propositions. Since defeasible rules are viewed as normal default rules, one of these rules must be defeated. The partial preference relations on the rules will be used to determine the rule that must be defeated. If an applicable rule is defeated, there must be a set of non-defeated applicable rules that implies, together with the premises, the negation of its consequent. Furthermore, the defeated rule may not be preferred to any of rules that cause its defeat.

**Definition 9.** Let $(\Sigma, D, \succ)$ be a defeasible theory.

Let $\Gamma(S)$ be a smallest set, with respect to the inclusion relation $(\subseteq)$, for which the following conditions hold:

1. $\Sigma \subseteq \Gamma(S)$;
2. $\Gamma(S) = Th(\Gamma(S))$;
3. if there is a $\Delta \subset D$ that defeats $\varphi \leadsto \psi$ with respect to $\succ$, then $\neg(\varphi \leadsto \psi) \in \Gamma(S)$;
4. if $\varphi \in \Gamma(S)$, $\varphi \leadsto \psi \in D$ and $\neg(\varphi \leadsto \psi) \not\in S$, then $\psi \in \Gamma(S)$.

$\Delta$ defeats $\varphi \leadsto \psi$ with respect to $\succ$ if and only if

- $\varphi \in \Gamma(S)$,
- $\Delta \subseteq \{ \eta \leadsto \mu \in D \mid \{ \eta, \mu \} \subseteq \Gamma(S) \}$,
- $\{ \mu \mid \eta \leadsto \mu \in \Delta \} \cup \Sigma \vdash \neg \psi$ and
- for no $\eta \leadsto \mu \in \Delta$ there holds: $\varphi \leadsto \psi \succ \eta \leadsto \mu$.

$E$ is an extension of the default theory if and only if $E = \Gamma(E)$.

**Theorem 4.** Let $(\Sigma, D, \succ)$ be a defeasible theory.

Then, the set of extensions determined by the argument system is equal to the set of extensions determined by the default logic.

The proof of this theorem can be found in Appendix A.

**Example 5.** Let $(\Sigma, D, \succ)$ be a defeasible theory where $\Sigma = \{ \alpha, \beta \}$, $D = \{ \alpha \leadsto \delta, \beta \leadsto \neg \delta \}$ and $\succ = \{(\beta \leadsto \neg \delta, \alpha \leadsto \delta)\}$.

Then we can construct the following arguments.

- $A_\delta = \{ \{\emptyset, \alpha\}, \alpha \leadsto \delta \}$;
- $A_\neg \delta = \{ \{\emptyset, \beta\}, \beta \leadsto \neg \delta \}$;
- $A_{\neg(\alpha \leadsto \delta)} = \{ \{\emptyset, \beta\}, \beta \leadsto \neg \delta, \{\emptyset, \alpha\} \}$

This set of arguments result in one fixed point, $\Omega = \{ \alpha \leadsto \delta \}$ and for the function Defeat. So, we have an extension

$E = Th(\{ \alpha, \beta, \neg \gamma, \neg \delta, \neg(\alpha \leadsto \delta) \})$.

According to the default logic given in this section, an extension must at least contain the premises $\Sigma = \{ \alpha, \beta \}$.

Suppose now that we cannot defeat $\alpha \leadsto \delta$. Then $\delta$ must belong to the extension. Furthermore, since $\beta \leadsto \neg \delta \succ \alpha \leadsto \delta$, $\beta \leadsto \neg \delta$ will not be defeated either. Therefore, $\neg \delta$ will belong to the extension. But then $\alpha \leadsto \delta$ will be defeated. Contradiction.

Hence, $\alpha \leadsto \delta$ must be defeated. Since we cannot defeat $\beta \leadsto \neg \delta$, $\neg \delta$ will belong to the extension. Therefore we can derive $\neg(\alpha \leadsto \delta)$.

So, we have one extension

$E' = Th(\{ \alpha, \beta, \neg \gamma, \neg \delta, \neg(\alpha \leadsto \delta) \})$. 

On resolving conflicts between arguments

We can translate this new default logic into Reiter’s default logic by translating the defeasible rules into default rules. Since we must be able to denote that the application of a default rule is no longer valid, \( \neg(\alpha \leadsto \beta) \), we will associate a name with each default rule. This name will be used to denote that the rule may no longer be applied. So if \( n_{\alpha \leadsto \beta} \) is the name of the translation of \( \alpha \leadsto \beta \), then \( \neg n_{\alpha \leadsto \beta} \) will be the translation of \( \neg(\alpha \leadsto \beta) \). To ensure that a default rule named \( n_{\alpha \leadsto \beta} \) will not be applied if \( \neg n_{\alpha \leadsto \beta} \) holds, we must use the name of the default rule as one of the justifications of this default rule. Hence, we translate a defeasible rule \( \alpha \leadsto \beta \) into the default rule \( \alpha : \beta, n_{\alpha \leadsto \beta} \). The translation of the defeasible rules are all semi-normal default rules.

It is not difficult to verify that every extension according to Definition 9 is also a Reiter-extension. Since the preferences relation on the default rules is not taken into account, some Reiter-extensions need not be extensions according to Definition 9. To eliminate these extensions, we must verify for each extension whether it is compatible with the preference relation on the rules. An extension is compatible with the preference relation if for each blocked\(^6\) default rule \( \alpha : \beta_1, \ldots, \beta_m \) there is a set of active\(^7\) default rules \( \Delta \subseteq R \) such that the consequences of the rules in \( \Delta \) together with the premises \( \Sigma \) imply \( \neg \beta \), and for no rule \( \psi : \eta_1, \ldots, \eta_n \) in \( \Delta \), \( \alpha : \beta_1, \ldots, \beta_m \not\psi \) holds.

So, \( E \) is an extension of a defeasible theory \( \langle \Sigma, D \rangle \) if and only if \( E \) is an extension of the default logic \( \langle R, \Sigma \rangle \) and \( E \) is compatible with the preference relation on \( R \).

**Specificity.** Poole (1985) gives a semantical definition of specificity based on the comparison of arguments (theories). His definition does not use the last rules of an argument as a starting point. Instead, Poole compares sets of rules. A Poole-argument \( \langle D, \alpha \rangle \) for a proposition \( \alpha \) describes a set of rules \( D \) needed to derive \( \alpha \); \( F_c \cup D \cup F_n \models \alpha \). Here \( F_c \) and \( F_n \) respectively denote the contingent and the necessary facts.

In this paper, a new definition has been given. This new definition can be related to Poole’s definition of specificity.

**Theorem 5.** Let \( \varphi \leadsto \psi \) and \( \eta \leadsto \mu \) be two rules.

If \( \varphi \leadsto \psi \) is more specific than \( \eta \leadsto \mu \) according to Definition 4, then there are two Poole-arguments \( \langle D_1, \psi \rangle \) and \( \langle D_2, \mu \rangle \) with \( \varphi \leadsto \psi \in D_1 \) and \( \eta \leadsto \mu \in D_2 \) for which there hold that \( \langle D_1, \psi \rangle \) is more specific than \( \langle D_2, \mu \rangle \).

**Proof.** We must prove that for every set of possible facts \( F_p \) there must hold: if \( F_p \cup D_1 \cup F_n \models \psi \) and \( F_p \cup D_2 \cup F_n \not\models \psi \), then \( F_p \cup D_2 \cup F_n \models \mu \). Let \( D_2 = \tilde{A}_\eta \cup \{ \eta \leadsto \mu \} \) and \( D_1 = \{ \varphi \leadsto \psi \} \).

Since \( \varphi \leadsto \psi \) is more specific than \( \eta \leadsto \mu \), given the premise \( \{ \varphi \} \) there must exist an argument \( \tilde{A}_\eta \) for \( \eta \).

Suppose that \( \tilde{A}_\eta = \emptyset \). Then \( \langle D_1, \psi \rangle \) is more specific than \( \langle D_2, \mu \rangle \).

Suppose that \( \tilde{A}_\eta = \{ \varphi \} \). Then, any possible fact \( F_p \) for which the antecedent of Poole’s definition holds, must imply \( \varphi \). Hence, \( \langle D_1, \psi \rangle \) is more specific than \( \langle D_2, \mu \rangle \). \( \blacksquare \)

The converse of this theorem need not hold. The reason why the converse need not hold is because a set of rules \( \langle D_1 \rangle \) or \( \langle D_2 \rangle \) need not uniquely determine a single argument.

---

\(^6\)The application of a default rule is blocked in an extension if its prerequisite belongs to the extension and one of its justifications is inconsistent with the extension.

\(^7\)A default rule is active in an extension if its prerequisite belongs to the extension and its justifications are consistent with the extension.
Furthermore, Poole’s definition of specificity is not violated if for some possible fact $F_p \cup D_1 \cup F_n \models \psi$ and $F_p \cup D_2 \cup F_n \models \psi$, but $F_p \cup D_2 \cup F_n \not\models \mu$.

8. RELATED WORK

In the literature, several argument systems that apply defeasible rules have been proposed (Fox et al. 1992; Geffner 1994; Kraus et al. 1995; Pollock 1987; Prakken 1993; Simari & Loui 1992; Vreeswijk 1991; Vreeswijk 1997). These related papers can roughly be divided in three groups; those that discuss the strength of an argument (Fox et al. 1992; Kraus et al. 1995), those that discuss the evaluation of arguments supporting conflicting propositions (Geffner 1994; Prakken 1993; Simari & Loui 1992; Vreeswijk 1991) and those that discuss validity of arguments (Pollock 1987; Pollock 1994; Simari & Loui 1992; Vreeswijk 1997).

Kraus, Ambler, Elvang-Gransson and Fox (Fox et al. 1992; Kraus et al. 1995) present argument systems that enable us to determine the strength of an argument for a proposition. They use a simple logic consisting of atoms, including $\perp$, and Horn clauses. For this logic they develop an argument system that enables them to evaluate the strength of arguments for a consistent set of propositions probabilistically. Furthermore, the argument system enables them to evaluate the strength of arguments for an inconsistent set of propositions symbolically. This evaluation partially corresponds with those valid arguments that are present in all extensions, and those valid arguments that are present in any extension. Kraus et al. do not, however, discuss how to defeat one of the disagreeing arguments.

Closely related to the strength of an argument is the evaluation of disagreeing arguments that support an inconsistency. Simari and Loui (1992) have proposed to apply Poole’s definition of specificity for this purpose. In this definition it is necessary to consider all the rules of the disagreeing arguments. The same approach is taken by Prakken (1993). Prakken argues that in legal argumentation only the last rules of an argument for an inconsistency are considered. Since he does not know how to do this in case of specificity, he decides to use Poole definition.

Vreeswijk (1991) discusses some general principles to evaluate disagreeing arguments. He proposes a scheme for evaluating disagreeing arguments based on the types of these arguments. He derives these types from the structure of the arguments. Furthermore, he argues that beside these weak general principle there are no general guild-lines to evaluate arguments. The definition of specificity given in Section 4 correspond with the general principles of Vreeswijk. However, applying it as a preference relation does not.

Geffner (1994) argues that any rule of an argument for a proposition $\varphi$ can be defeated if $\neg \varphi$ is a known fact. As we have seen in Section 2, Geffner uses causal rules which are a special kind of defeasible rules. These kind of rules have been excluded form this paper. Many defeasible rules are not causal rules. Furthermore, a discussion of causal rules would also require a study of causality.

The theory of warrant is concerned with the validity of arguments. These are the argument that are not defeated by other arguments. In (Pollock 1987), Pollock introduces the theory of inductive warrant. Simari and Loui (1992) combine the theory of inductive warrant with Poole’s definition of specificity. They study the mathematical properties of the resulting system.

Pollock observes that his theory of inductive warrant is not without problems. Therefore, in (Pollock 1994), he introduces a new theory of warrant based on the idea of multiple

8A rule is not interpreted as representing a conditional probability.
extensions. Vreeswijk (1991) has made a similar proposal.

In (Vreeswijk 1997), Vreeswijk relates the theory of inductive warrant to a theory of warrant based on extensions. He discusses several ways of defining a theory of warrant and discusses the mutual relation.

The theories of warrant proposed by Pollock (1987, 1994) and Vreeswijk (1997), start from a defeat relation on the set of derived arguments. This relation is the result of resolving conflicts between the propositions supported by the arguments. The theory of warrant as described by Pollock and by Vreeswijk is concerned with selecting a set of arguments that do not defeat each other. What complicates their theories of warrant is that we may not use any set meeting this requirement. An argument that is not in the set must be defeated by an argument that is.

It seems more natural to express the validity of an argument in terms of the validity of the defeasible steps that are used in the argument. In this respect, the theory of warrant proposed in this paper differs from the proposals of Pollock (1987, 1994) and Vreeswijk (1997).

9. CONCLUSIONS

A defeasible rule describes an preferred or a probabilistic relations between propositions. Such defeasible rules can be used to construct arguments for propositions. For both interpretations of a defeasible rule, we have shown that an inconsistency must be resolved by defeating one of the last rules of the argument supporting the inconsistency. This result has been used to show that it suffices to consider only the rules are candidates for defeat, to select the rule to be defeated. For this purpose, a preference relation on the set of rules has been proposed. A definition of specificity that generates such a preference relation on the set of rules has been given.

In legal argumentation, an inconsistency is resolved by evaluating the last rules of the argument for the inconsistency and subsequently defeating one of these rules. This confirms the presented results.

Since one of the last rules of the argument for an inconsistency must be defeated, we can formulate an argument for the defeat of this rule. Such an argument undercut the application of the rule. Hence, rebutting defeat can be reformulated as undercutting defeat after determining the rule to be defeated. Although this approach does not lead to new results, it is more intuitive. An argument gives a valid justification for a conclusion, if all step (the rules) of the justification are valid. Furthermore, it enables us to determine the extensions of valid beliefs using a Reason Maintenance System.

Finally, a relation between default logic and reasoning with arguments has been established.

APPENDIX A

To prove Theorem 4, the following lemmas will be used.

Lemma 1. Let $A$ be a set of arguments and let $E$ be a corresponding extension.

Then for each rule $\alpha \succ \beta$ that is defeated according to Definition 6, there holds that $\neg (\alpha \succ \beta) \in \Gamma(E)$ according to Definition 9.

Proof. Let $\alpha \succ \beta$ is defeated according to Definition 6. Then there either exists a valid argument $A_{\neg (\alpha \succ \beta)} = \{ \langle A, \gamma \succ \neg (\alpha \succ \beta) \rangle \}$ or there exists a valid argument $A_{\neg (\alpha \succ \beta)} = \{ \langle A_1, \eta_1 \succ \mu_1 \rangle, \ldots, \langle A_n, \eta_n \succ \mu_n \rangle, \langle \emptyset, \nu_1 \rangle, \ldots, \langle \emptyset, \nu_m \rangle \}$. 
In the former case, $-(\alpha \rightsquigarrow \beta) \in \Gamma(\mathcal{E})$.

In the latter case, $\{\eta_1, \ldots, \eta_n, \mu_1, \ldots, \mu_n\} \subseteq \mathcal{E}$. This implies that the set $\Delta = \{\eta_1 \rightsquigarrow \mu_1, \ldots, \eta_n \rightsquigarrow \mu_n\}$ defeats $\alpha \rightsquigarrow \beta$ according to Definition 9. Hence, $-(\alpha \rightsquigarrow \beta) \in \Gamma(\mathcal{E})$. □

**Lemma 2.** Let $\Gamma$ be a set of propositions and let $\mathcal{E}$ be the deductive closure of $\Gamma$. Furthermore, let there be an argument for each proposition in $\Gamma$.

Then for each proposition in $\mathcal{E}$ we can construct an argument $A$.

Proof. For each $\varphi \in \mathcal{E} - \Gamma$ there holds that $\Gamma \vdash \varphi$. Hence, $\bigcup \{A_\varphi \mid \varphi \in \Gamma\}$ is an argument for $\varphi$. □

**Lemma 3.** Let $\mathcal{E}$ be an extension according to Definition 9 and let $\Omega = \{\alpha \rightsquigarrow \beta \mid -(\alpha \rightsquigarrow \beta) \in \mathcal{E}\}$. Furthermore, let there be an argument $A$ for each proposition in $\mathcal{E}$ as well as for each proposition that is derivable from these arguments according to the Definitions 1 and 5, and let $A \cap \Omega = \emptyset$.

Then $\Omega$ satisfies Definition 6, $\Omega = \text{Defeat}(\Omega)$.

Proof. Suppose that $\alpha \rightsquigarrow \beta \in \Omega$ and $\alpha \rightsquigarrow \beta \notin \text{Defeat}(\Omega)$.

Since $\alpha \rightsquigarrow \beta \in \Omega$, either there exists a $\gamma \rightsquigarrow -(\alpha \rightsquigarrow \beta)$ and $\gamma \in \mathcal{E}$, or there exists a $\Delta$ that defeats $\alpha \rightsquigarrow \beta$, $\alpha \in \mathcal{E}$ and $\Delta \subseteq \{\eta \rightsquigarrow \mu \in D \mid \{\eta, \mu\} \subseteq \mathcal{E}\}$ such that $\{\mu \mid \eta \rightsquigarrow \mu \in \Delta\} \cup \mathcal{E} \vdash \neg \beta$ and for no $\eta \rightsquigarrow \mu \in \Delta$ there holds: $\alpha \rightsquigarrow \beta \succ \eta \rightsquigarrow \mu$.

In the former case there exists an argument $A_{-(\alpha \rightsquigarrow \beta)} = \{(A_\gamma, \gamma \rightsquigarrow -(\alpha \rightsquigarrow \beta))\}$ and $A_{-(\alpha \rightsquigarrow \beta)} \cap \Omega = \emptyset$. Hence, $\alpha \rightsquigarrow \beta \in \text{Defeat}(\Omega)$. Contradiction.

In the latter case there exists an argument $A_{-(\alpha \rightsquigarrow \beta)} = \{A_\eta, \eta \rightsquigarrow \mu \mid \eta \rightsquigarrow \mu \in \Delta\} \cup \{(\emptyset, \varphi) \mid \varphi \in \Sigma\}$. Furthermore, $A_{-(\alpha \rightsquigarrow \beta)} \cap \Omega = \emptyset$. Hence, $\alpha \rightsquigarrow \beta \in \text{Defeat}(\Omega)$. Contradiction.

Hence, $\Omega \subseteq \text{Defeat}(\Omega)$.

Suppose that $\alpha \rightsquigarrow \beta \notin \Omega$ and $\alpha \rightsquigarrow \beta \in \text{Defeat}(\Omega)$. Then there exists an argument $A_{-(\alpha \rightsquigarrow \beta)}$ such that $A_{-(\alpha \rightsquigarrow \beta)} \cap \Omega = \emptyset$. This implies that either there exists an argument $A_{\gamma \rightsquigarrow \beta}$ for $\neg \beta$ such that $A_{\gamma \rightsquigarrow \beta} \cap \Omega = \emptyset$, or that $\gamma \rightsquigarrow -(\alpha \rightsquigarrow \beta) \in D$ and there exists an argument $A_\gamma$ for $\gamma$ such that $A_\gamma \cap \Omega = \emptyset$.

In the former case, $-\beta \in \mathcal{E}$. But then $\alpha \rightsquigarrow \beta \in \Omega$. Contradiction.

In the latter case, $\gamma \in \mathcal{E}$. But then $\alpha \rightsquigarrow \beta \in \Omega$. Contradiction.

Hence, $\Omega \subseteq \text{Defeat}(\Omega)$. □

**Proof. of Theorem 4** Let $\mathcal{E}$ be an extension according to Definition 7. We will proof that $\mathcal{E}$ is also an extension according to Definition 9 by showing that it is a fixed point satisfying the four requirements of Definition 9.

Clearly for each $\alpha \in \Sigma$ we have an argument $\{(\emptyset, \alpha)\}$. Since it contains no rules, it cannot be defeated. Therefore, $\Sigma \subseteq \mathcal{E}$.

According to the definition of an argument, $\mathcal{E}$ is deductively closed.

According to Lemma 1, $-(\alpha \rightsquigarrow \beta) \in \Gamma(\mathcal{E})$ if and only if $\alpha \rightsquigarrow \beta$ is defeated according to Definition 6. Hence, $-(\alpha \rightsquigarrow \beta) \in \mathcal{E}$.

For each $\alpha \rightsquigarrow \beta \in D$ if $\alpha \in \Gamma(\mathcal{E})$ and $-(\alpha \rightsquigarrow \beta) \notin \mathcal{E}$, then $\beta \in \Gamma(\mathcal{E})$.

Suppose that $\alpha \in \mathcal{E}$. Since $-(\alpha \rightsquigarrow \beta) \notin \mathcal{E}$, there exists an valid argument $A_\beta = \{A_\gamma, \alpha \rightsquigarrow \beta\}$ for $\beta$. Hence, $\beta \in \mathcal{E}$.

Suppose that $\mathcal{E}$ is not a minimal set satisfying the requirements of $\Gamma(\mathcal{E})$. Then there exists an extension $\mathcal{E}' \subseteq \mathcal{E}$ satisfying the requirements of $\Gamma(\mathcal{E})$. This implies that there exists a valid argument $A_\varphi$ and $\varphi \notin \mathcal{E}'$. Since $\Sigma \subseteq \mathcal{E}'$, for some sub-argument $A_\beta \subseteq A_\varphi$ with $A_\beta = \{A_\alpha, \alpha \rightsquigarrow \beta\}$ there holds: $\alpha \in \mathcal{E}'$ and $\beta \notin \mathcal{E}'$. Hence, $-(\alpha \rightsquigarrow \beta) \in \mathcal{E}'$. Therefore
either \( \gamma \vdash \neg(\alpha \vdash \beta) \in D \), \( \gamma \vdash \neg(\alpha \vdash \beta) \not\in \Omega \) and \( \gamma \in \mathcal{E}' \), or \( \alpha \vdash \beta \) is defeated with respect to \( \mathcal{E}' \).

In the former case there must exist a valid argument for \( \gamma \) and \( \neg(\alpha \vdash \beta) \). But then \( A_{\beta} = \{ (\alpha, \alpha \vdash \beta) \} \) cannot be a valid argument. Contradiction.

In the latter case there exists a set of rules \( \Delta \subseteq \{ \eta \vdash \mu \mid \{ \eta, \mu \} \subseteq \mathcal{E}' \} \) that defeat \( \alpha \vdash \beta \). Since \( A_{\beta} \) is a valid argument for \( \beta \) in \( \mathcal{E}' \), by Lemma 1, \( \alpha \vdash \beta \) cannot be defeated. Contradiction.

Hence, \( \mathcal{E} \) is a fixed point of \( \Gamma \).

Let \( \mathcal{E} \) be an extension according to Definition 9 and let \( \Omega = \{ \alpha \vdash \beta \mid \neg(\alpha \vdash \beta) \in \mathcal{E} \} \). We will prove that \( \mathcal{E} \) is an extension according to Definition 7 by showing that for each proposition in \( \mathcal{E} \) there is a valid argument and for each proposition not in \( \mathcal{E} \) there is no such argument. We will show that there is a valid argument for each \( \varphi \in \mathcal{E} \) by showing that we can construct an argument \( A \) for each \( \varphi \in \mathcal{E} \) such that \( A \cap \Omega = \emptyset \). If we have an argument for each \( \varphi \in \mathcal{E}' \), then, by Lemma 3, \( \Omega \) satisfies Definition 6, i.e. \( \Omega = \text{Defeat}(\Omega) \). Since for each \( \varphi \in \mathcal{E} \), we have an argument \( A \) such that \( A \cap \Omega = \emptyset \), \( A \) must be a valid argument for \( \varphi \).

Let \( \Gamma_0 = \Sigma \) and let \( \mathcal{E}_0 \subseteq \mathcal{E} \) be a smallest deductively closed subset such that \( \Gamma_0 \subseteq \mathcal{E}_0 \). For each \( \varphi \in \mathcal{E} \), we can construct an argument \( A_{\varphi} = \{ \varphi \} \). Furthermore, by Lemma 2, we can construct an argument for each \( \varphi \in \mathcal{E}_0 \). Clearly, \( A_{\varphi} \cap \Omega = \emptyset \).

Proceeding inductively, let \( \mathcal{E}_i \subseteq \mathcal{E} \) be a minimal deductive closure subset such that \( \Gamma_i \subseteq \mathcal{E}_i \). Suppose that for no \( \varphi \in \mathcal{E} - \mathcal{E}_i \) an argument can be constructed. Then, according to Definition 9, \( \mathcal{E}_i = \Gamma(\mathcal{E}) \). Since \( \mathcal{E}_i \subseteq \mathcal{E} \), \( \mathcal{E} \) cannot be an extension according to Definition 9. Contradiction. Hence, there is a \( \varphi \in \mathcal{E} - \mathcal{E}_i \) for which we can construct an argument \( A_{\varphi} \).

Here, either \( \varphi = \neg(\alpha \vdash \beta) \) and \( \alpha \vdash \beta \) is defeated given \( \mathcal{E}_i \), or \( \alpha \in \mathcal{E}_i \), \( \alpha \vdash \varphi \in D \) and \( \neg(\alpha \vdash \varphi) \not\in \mathcal{E} \). Clearly, by the construction of \( A_{\varphi} \), \( A_{\varphi} \cap \Omega = \emptyset \).

Let \( \mathcal{E}_{i+1} \) be the deductive closure of \( \Gamma \cup \{ \varphi \} \). According to Lemma 2, for every proposition in \( \mathcal{E}_{i+1} \) we can construct an argument \( A \) such that \( A \cap \Omega = \emptyset \).

Hence, for every proposition in \( \mathcal{E} \), as well as for each proposition that is derivable from these arguments according to the Definitions 1 and 5, we can construct an argument \( A \) such that \( A \cap \Omega = \emptyset \). Given these arguments, there holds according to Lemma 3 that \( \Omega = \{ \alpha \vdash \beta \mid \neg(\alpha \vdash \beta) \in \mathcal{E} \} \) satisfies Definition 6, i.e. \( \Omega = \text{Defeat}(\Omega) \). Hence, the arguments for the propositions in \( \mathcal{E} \) are valid arguments.

Now suppose that we can construct a valid argument \( A_{\varphi} \) for a proposition \( \varphi \not\in \mathcal{E} \). Since \( A_{\varphi} \subseteq \Sigma \), for some rule \( \alpha \vdash \beta \in A_{\varphi} \) there holds: \( \alpha \in \mathcal{E} \) and \( \beta \notin \mathcal{E} \). Then, \( \alpha \vdash \beta \in \Omega \). Contradiction.

Hence, \( \mathcal{E} \) is an extension according to Definition 7.

\[ \square \]

**APPENDIX B**

Associate with each node and with each justification of the JTMS a counter. Initially, set the counter of a node equal to the number of incoming justifications and the counter of each justification equal to the number of the out-nodes of the justification. Determine all the nodes that have a justification with an empty set of out-nodes. Label these nodes IN, and place them on the in-list. Next execute *propagate*.

**propagate:**

for each node on the in-list

and for each out-going justification do

- decrement the counter of node;
- remove the justification;

...
if the counter of the node is equal to 0 then
    label the node OUT;
    place the node on the out-list;
end

delete the in-list;
for each node on the out-list
and for each out-going justification do
    decrement the counter of the justification;
    if the counter of the justification is equal to 0 then
        label its consequent node IN;
        place its consequent node on the in-list;
    end
end:
delete the out-list;
if the in-list is not empty then
    repeat propagate;
end
end

The above described procedure need not result in a complete labeling of the JTMS. When this is the case, more than one grounded labeling exist. To create a complete labeling, we must choose one of the unlabeled nodes, a node that is not labeled IN, OUT or UNDETERMINED, and label IN or OUT. If we label IN we place the node on the in-list, if we label it OUT we place it on the out-list. Subsequently, we must execute the procedure propagate.

We repeat the selection of a node and giving it a label and propagating the consequences till all nodes are labeled. By backtracking on the choices that are made, we determine every grounded labeling of the JTMS.

REFERENCES


On resolving conflicts between arguments


Centre for Telematics and Information Technology (CTIT)

Reports


Akkermans, H. and F. Ygge, "Smart Software as Customer Assistant in Large-Scale Distributed Load Management", TR-CTIT 97-26, November 1997, 16 pp.


Reports are freely available: University of Twente, Centre for Telematics and Information Technology, P.O. Box 217, NL-7500 AE Enschede, the Netherlands; Phone: +31 53 489 3779 / Fax: +31 53 489 3247/ email: castaned@ctit.utwente.nl


D'Argenio, P.R., Katoen, J.-P., Ruys, T.C. and Tretmans, J., "The Bounded Retransmission Protocol must be on time", TR-CTIT 97-03, January 1997,


Centre for Telematics and Information Technology

P.O. Box 217
7500 AE Enschede
The Netherlands