Computational Intelligence, Volume 16, Number 3, 2000

ON RESOLVING CONFLICTS BETWEEN ARGUMENTS

Nico Roos

Institute for Knowledge and Agent Technology, Maastricht University,
P.O. Box 616, 6200 MD Maastricht, The Netherlands, e-mail: roos@cs.unimaas.nl

Argument systems are based on the idea that one can construct arguments for propositions; i.e., structured reasons justifying the belief in a proposition. Using defeasible rules, arguments need not be valid in all circumstances, therefore, 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 must be defeated, which raises the question of *which (sub-)arguments can be subject to defeat*?

In legal argumentation, meta-rules determine the valid arguments by considering the last defeasible rule of each argument involved in a conflict. Since it is easier to evaluate arguments using their last rules, *can a conflict be resolved by considering only the last defeasible rules of the arguments involved*?

We propose a new argument system where, instead of deriving a defeat relation between arguments, *undercutting-arguments* for the defeat of defeasible rules are constructed. This system allows us, (*i*) to resolve conflicts (a generalization of rebutting arguments) using only the last rules of the arguments for inconsistencies, (*ii*) to determine a set of valid (undefeated) arguments in linear time using an algorithm based on a JTMS, (*iii*) to establish a relation with Default Logic, and (*iv*) to prove closure properties such as *cumulativity*. We also propose an extension of the argument system that enables *reasoning by cases*.

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

This revised version of the paper provides a more clear explanation of reasoning by cases (Section 8), fixes an issue in the proof of the closure property Cautious Monotony mentioned in Theorem 3, and corrects several small errors.

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; Dung 1995; Fox et al. 1992; Geffner 1994; Hunter 1994; Kraus et al. 1995; Lin & Shoham 1989; Loui 1987; Loui 1998; Pollock 1987, 1992, 1994; Poole 1988; Prakken 1993; Prakken & Vreeswijk 1999; Simari & Loui 1992; Vreeswijk 1991, 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; Prakken & Vreeswijk 1999; Simari & Loui 1992; Vreeswijk 1991, 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.

Since the rules used in the construction of arguments are defeasible, it might be possible to construct an argument for a proposition as well as its negation. Clearly, only one of these arguments can give a valid justification for the proposition it supports. In most argument systems proposed in the literature, one of the arguments supporting the conflicting propositions; i.e. for a proposition and its negation, is defeated (Loui 1987; Pollock 1987, 1992,

[0]This paper extends (Roos 1997b).

1994; Prakken 1993; Prakken & Vreeswijk 1999; Simari & Loui 1992; Vreeswijk 1991, 1997). Generally, if an argument is defeated, there is a defeated sub-argument (not necessarily a proper sub-argument) that has a single last defeasible rule, and no sub-argument of this sub-argument is defeated. An exception are arguments that are based on defeasible causal relations. Geffner (1994), for example, allows the chain of causal arguments to be broken at any rule of an argument if the proposition supported by the argument conflicts with an observed fact.

Defeasible rules representing causal relations have one property, not found in non causal defeasible rules. Causal defeasible rules can be used in contraposition. This raises the following question. *If rules cannot be used in contraposition, can a conflict be resolved by defeating a last defeasible rule of the argument supporting one of the conflicting propositions*?

In legal argumentation, meta rules are used to resolve conflicts (Prakken 1993). These meta rules determine the valid arguments by considering the last defeasible rule, with respect to the chain of argumentation, of each proposition involved in the conflict. Hence, an argument should not only reflect the information used in an argumentation, but also the structure of the argumentation.

Some argument system represent this structure explicitly (Vreeswijk 1991, 1997), while others represent it implicitly (Pollock 1987, 1992). Vreeswijk (1991, 1997) only considers defeasible rules that are definite horn clauses and a special symbol $\perp$ to denote an inconsistency. In this language, each argument for a conflict; i.e. for $\perp$, is unique. If we use, however, full propositional or predicate logic, there can be more than one way of argumentation for deriving 'the same conflict'.

Suppose for example that we have the following three defeasible rules: $\{a \rightsquigarrow \neg d, b \rightsquigarrow \neg e, c \rightsquigarrow (d \vee e)\}$ and the facts: $\{a, b, c\}$. Then we can construct arguments for conflicting propositions in at least three different ways.

$$\left.\begin{array}{r}\left.\begin{array}{r} a \\ a \rightsquigarrow \neg d\end{array}\right\} \neg d \\ \left.\begin{array}{r} c \\ c \rightsquigarrow (d \vee e)\end{array}\right\} d \vee e\end{array}\right\} e \text{ and } \quad \left.\begin{array}{r} b \\ b \rightsquigarrow \neg e\end{array}\right\} \neg e$$

$$\left.\begin{array}{r}\left.\begin{array}{r} b \\ b \rightsquigarrow \neg e\end{array}\right\} \neg e \\ \left.\begin{array}{r} c \\ c \rightsquigarrow (d \vee e)\end{array}\right\} d \vee e\end{array}\right\} d \text{ and } \quad \left.\begin{array}{r} a \\ a \rightsquigarrow \neg d\end{array}\right\} \neg d$$

$$\left.\begin{array}{r}\left.\begin{array}{r} b \\ b \rightsquigarrow \neg e\end{array}\right\} \neg e \\ \left.\begin{array}{r} a \\ a \rightsquigarrow \neg d\end{array}\right\} \neg d\end{array}\right\} \neg d \wedge \neg e\}\neg(d \vee e) \text{ and } \quad \left.\begin{array}{r} c \\ c \rightsquigarrow d \vee e\end{array}\right\} d \vee e$$

Each of these three couples of arguments supporting conflicting propositions, uses exactly the same information to derive a conflict. We would therefore expect that we have only one argument for the conflict instead of three.

The heart of the problem is that the consequences of the three defeasible rules used in the above presented arguments are logically inconsistent. How these three rules are used to derive an inconsistency, should not matter. The derivation of the inconsistency takes place through logically sound deductions which cannot be subject to defeat. The line of reasoning that is followed using the logical sound deductions should not matter. Therefore, in the definition of an argument given in the next section, we abstract away from the actual logical sound deductions. For the same reason, we consider arguments for inconsistencies $\perp$, instead of arguments of the conflicting propositions.

As mentioned above, in legal argumentation only the last rules for an inconsistency are considered in order to resolve the inconsistency. Using a preference relations, one of the last rules is identified as the culprit. There are $\frac{1}{2}n(n-1)$ pairs of rules between which there can be a preference, where n is the number of defeasible rules. If we use a preference on the arguments, ignoring the structure of the arguments, there can be at most $\frac{1}{2}2^n(2^n-1)$ pairs of arguments between which there can be a preference. Clearly, it is easier to evaluate arguments using their last rules, than using the whole argument. We will therefore investigate the following question. *Is there a need for considering more than only the last rules of an argument for an inconsistency*?

The next section formalizes the arguments that can be constructed using defeasible rules. The here defined arguments do not only represent the defeasible rules that are used, but also the line of reasoning. Section 3 discusses whether we can resolve an inconsistency just by defeating one of the last defeasible rules of the argument for the inconsistency. Section 4 investigates whether we can select the rule (or argument) to be defeated just be looking at the last rules of the argument for the inconsistency. Based on the results of Sections 3 and 4, Section 5 proposes a new argument system. What is essentially new is that inconsistencies are resolved by constructing an argument for the *undercutting* defeat of one of the defeasible rules of the argument for the inconsistency. Section 6 discusses how to compute an extension and presents a linear time algorithm for doing so. Section 7 discusses closure properties of the argument system and the relation with default logic. Section 8 presents an extension of the presented argument system that enables reasoning by cases. Finally, Section 9 discusses related work and Section 10 concludes the paper.

2. THE ARGUMENT SYSTEM

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

For every defeasible rule $\varphi \rightsquigarrow \psi \in D$ there holds that φ is a proposition in L and that ψ is either a proposition in L or the negation of a defeasible rule in D; i.e. $\psi = \neg(\alpha \rightsquigarrow \beta)$ and $\alpha \rightsquigarrow \beta \in D$. The negation of a defeasible rule $\neg(\alpha \rightsquigarrow \beta)$ will be interpreted as: 'α may no longer justify β'. 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 rule. If we have a valid argument for $\neg(\alpha \rightsquigarrow \beta)$, then no argument containing the rule $\alpha \rightsquigarrow \beta$ can be valid. For example, we can undercut the rule: 'something that looks red is red' by the rule: 'something that stands below a red light need not be red if it looks red'.

Notice the correspondence of the defeasible rules $\alpha \rightsquigarrow \beta$ and $\varphi \rightsquigarrow \neg(\alpha \rightsquigarrow \beta)$ with respectively the semi- and non-normal default rules $\frac{\alpha:\beta,\omega_1}{\beta}$ and $\frac{\varphi:\omega_2}{\neg\omega_1}$ where ω_1 and ω_2 summarize the exceptions on the default rules. Also notice the difference with Nute's (1988, 1994) defeater rule $\varphi\ ?\!\!\rightarrow \neg\beta$. If we have a valid argument for φ, Nute's defeater rule $\varphi\ ?\!\!\rightarrow \neg\beta$ defeats *any* argument containing a rule of the form $\alpha \rightsquigarrow \beta$. We can, however, use the defeater rule $\varphi \wedge \alpha\ ?\!\!\rightarrow \neg\beta$ to describe $\varphi \rightsquigarrow \neg(\alpha \rightsquigarrow \beta)$.

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 deduction steps need not be represented explicitly 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 Σ is the set of premises and D is the set of rules.

Then an argument[1] A for a proposition ψ is recursively defined in the following way:

- For each $\psi \in \Sigma$: $A = \{\langle \emptyset, \psi \rangle\}$ is an argument for ψ.
- 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 ψ.
- For each $\varphi \rightsquigarrow \psi \in D$ if A' is an argument for φ, then $A = \{\langle A', \varphi \rightsquigarrow \psi \rangle\}$ is an argument for ψ.

Let $A = \{\langle A'_1, \alpha_1 \rangle, ..., \langle A'_n, \alpha_n, \rangle\}$. Then:

$$\begin{aligned}
&\vec{A} = \{\alpha_1, ..., \alpha_n\} \cap D; \\
&\hat{A} = \{c(\alpha_1), ..., c(\alpha_n)\} \text{ where } c(\alpha) = \alpha \text{ if } \alpha \in L \text{ and, } c(\alpha \rightsquigarrow \beta) = \beta; \\
&\tilde{A} = \{\alpha_i \mid 1 \leq i \leq n, \alpha_i \in D\} \cup \textstyle\bigcup_{i=1}^n \tilde{A}'_i; \\
&\bar{A} = \{\alpha_i \mid 1 \leq i \leq n, \alpha_i \in \Sigma\} \cup \textstyle\bigcup_{i=1}^n \bar{A}'_i
\end{aligned}$$

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

$$\left.\begin{array}{r} \alpha \\ \beta \vdash \beta \rightsquigarrow \gamma \vdash \gamma \rightsquigarrow \delta \end{array}\right| \vdash \varphi$$

Then $\vec{A} = \{\gamma \rightsquigarrow \delta\}$ denotes the last rules used in the argument A. Furthermore, $\hat{A} = \{\alpha, \delta\}$ denotes the propositions that represent the beliefs $Th(\{\alpha, \delta\})$ supported by the argument A. Clearly, A is an argument for every proposition $\varphi \in Th(\{\alpha, \delta\})$. $\tilde{A} = \{\gamma \rightsquigarrow \delta, \beta \rightsquigarrow \gamma\}$ denotes the set of all rules in A, and $\bar{A} = \{\alpha, \beta\}$ 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 is interpreted as describing 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.

Also if the defeasible rule is interpreted as describing a conditional probability, $Pr(\psi \mid \varphi) > t$ does not imply that $Pr(\neg\varphi \mid \neg\psi) > t$. In fact, if $Pr(\psi \mid \varphi) < 1$, $Pr(\neg\varphi \mid \neg\psi)$ 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 $Pr(\neg\varphi \mid \neg\psi) > t$ holds.

[1] We will sometimes add the index ψ to an argument (A_ψ) to denote that it is an argument for ψ. Of course there can be more than one argument for ψ.

Causal rules are a special kind of defeasible rules that do possess a contraposition (Geffner 1994). If, '*normally*, φ **causes** ψ', 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, $A \leq B$, if and only if every $\langle A', \alpha \rangle \in A$ is a sub-structure of the argument B.[2]

$\langle A', \alpha \rangle$ is a sub-structure of an argument B if and only if

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

Example 2. let $A = \{\langle \emptyset, \alpha \rangle, \langle \{\langle \{\langle \emptyset, \beta \rangle\}, \beta \rightsquigarrow \gamma \rangle\}, \gamma \rightsquigarrow \delta \rangle\}$ be an argument for φ.

$$\left.\begin{array}{r}\alpha \\ \beta \vdash \beta \rightsquigarrow \gamma \vdash \gamma \rightsquigarrow \delta\end{array}\right| \vdash \varphi$$

Then

$$A_1 = \{\langle \emptyset, \alpha \rangle, \langle \{\langle \{\langle \emptyset, \beta \rangle\}, \beta \rightsquigarrow \gamma \rangle\}, \gamma \rightsquigarrow \delta \rangle\}$$

$$\left.\begin{array}{r}\alpha \\ \beta \vdash \beta \rightsquigarrow \gamma \vdash \gamma \rightsquigarrow \delta\end{array}\right|$$

$$A_2 = \{\langle \emptyset, \alpha \rangle, \langle \{\langle \emptyset, \beta \rangle\}, \beta \rightsquigarrow \gamma \rangle\}$$

$$\left.\begin{array}{r}\alpha \\ \beta \vdash \beta \rightsquigarrow \gamma\end{array}\right|$$

$$A_3 = \{\langle \emptyset, \alpha \rangle, \langle \emptyset, \beta \rangle\}$$

$$\left.\begin{array}{r}\alpha \\ \beta\end{array}\right|$$

$$A_4 = \{\langle \emptyset, \alpha \rangle\}$$

$$\alpha \mid$$

$$A_5 = \{\langle \{\langle \{\langle \emptyset, \beta \rangle\}, \beta \rightsquigarrow \gamma \rangle\}, \gamma \rightsquigarrow \delta \rangle\}$$

$$\beta \vdash \beta \rightsquigarrow \gamma \vdash \gamma \rightsquigarrow \delta \mid$$

$$A_6 = \{\langle \{\langle \emptyset, \beta \rangle\}, \beta \rightsquigarrow \gamma \rangle\}$$

$$\beta \vdash \beta \rightsquigarrow \gamma \mid$$

$$A_7 = \{\langle \emptyset, \beta \rangle\}$$

$$\beta \mid$$

are sub-arguments of A.

[2] 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.

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 φ, $\hat{A} \vdash \varphi$, 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 φ is valid we say that φ follows from the premises $\bar{A}$ using the rules $\tilde{A}$.

3. DEFEATING A LAST RULE OF AN ARGUMENT

A defeasible rule $\varphi \rightsquigarrow \psi$ describes either a 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 ψ and $\neg\psi$ must be unknown. Since an argument is basically a tree constructed using defeasible rules, an argument containing a rule that is not valid in the current context, can neither be valid.

There are two reasons for an argument to become invalid. Either the argument contains a rule $\alpha \rightsquigarrow \beta$ while we have a valid argument for $\neg(\alpha \rightsquigarrow \beta)$, or the argument is a sub-argument of an argument for an inconsistency. In the latter situation the question is, which sub-argument(s) of the argument for an inconsistency, can no longer be valid? In the discussion of this question, we will use the term *disagreeing arguments* which is defined in the following way.

Definition 3. Let $A_\perp = \{\langle A'_1, \mu_1\rangle, ..., \langle A'_n, \mu_n\rangle\}$ be an argument for and inconsistency ($\hat{A}_\perp \vdash \perp$). Then, the arguments $A_1 = \{\langle A'_1, \mu_1\rangle\}, ..., A_n = \{\langle A'_n, \mu_n\rangle\}$ are said to *disagree*.

Clearly, in order to restore consistency, some of the disagreeing arguments can no longer be valid. These arguments are said to be defeated because of the other arguments. It is also clear that it is sufficient to defeat only one of the disagreeing arguments in order to restore consistency if the argument for the inconsistency is a (subset) minimal argument. Without lost of generality, we may assume that the argument for the inconsistency is a minimal argument. Resolving inconsistencies using the minimal arguments will also resolve inconsistencies based on non minimal arguments. We can therefore reformulate the above raised question. Is it sufficient to defeat a disagreeing argument, but no proper sub-argument of this disagreeing argument, to resolve an inconsistency? We will see that a set of defeasible rules can always be extended such that indeed no proper sub-argument of a disagreeing argument needs to be defeated.

Let

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

be a set of disagreeing arguments; i.e. $A = \bigcup_{i=1}^n A_i$ is an argument for $\perp$. Suppose that some proper sub-argument $A_\varphi = \{\langle A', \alpha \rightsquigarrow \varphi\rangle\}$ of the disagreeing argument $A_k = \{\langle A'_k, \mu_k\rangle\}$ is defeated because of the inconsistency and that no proper sub-argument of A_φ is defeated. Then, $\bigcup_{i=1}^n \bar{A}_i$ represents an exceptional situation in which either $\neg\varphi$ holds or φ is unknown.

Suppose that $\neg\varphi$ holds. We cannot use the contraposition of the rules to derive $\neg\varphi$. We can, however, introduce rules that enable us to construct an argument $A_{\neg\varphi}$ for $\neg\varphi$ such that $\bar{A}_{\neg\varphi} \subseteq \bigcup_{i=1}^n \bar{A}_i$. In that case A_φ is a disagreeing argument in another inconsistency. This inconsistency can be used to defeat A_φ. Hence, there is no need for defeating A_φ because of the argument A for $\perp$. For example, let

$$A_\perp = \{\langle\{\langle\{\langle\{\langle\emptyset, \alpha\rangle\}, \alpha \rightsquigarrow \varphi\rangle\}, \varphi \rightsquigarrow \eta\rangle, \langle\{\langle\{\langle\emptyset, \beta\rangle\}, \beta \rightsquigarrow \psi\rangle\}, \psi \rightsquigarrow \neg\eta\rangle\}$$

$$\left.\begin{array}{r} \alpha \vdash \alpha \rightsquigarrow \varphi \vdash \varphi \rightsquigarrow \eta \\ \beta \vdash \beta \rightsquigarrow \psi \vdash \psi \rightsquigarrow \neg\eta \end{array}\right| \vdash \perp$$

be an argument for $\bot$. Then we can defeat $\alpha \leadsto \varphi$ by introducing the rule $\alpha \wedge \beta \leadsto \neg\varphi$.

Now suppose that φ is unknown. We cannot introduces rules that enable us to construct an argument $A_{\neg\varphi}$ since φ is unknown. We can, however, introduce rules that enable us to construct an argument $A_{\neg(\alpha\leadsto\varphi)} = \{\langle A', \xi \leadsto \neg(\alpha \leadsto \varphi)\rangle\}$ for $\neg(\alpha \leadsto \varphi)$ such that $\bar{A}_{\neg(\alpha\leadsto\varphi)} \subseteq \bigcup_{i=1}^{n} \bar{A}_i$. Since $A_\varphi = \{\langle A', \alpha \leadsto \varphi\rangle\}$ is defeated if $A_{\neg(\alpha\leadsto\varphi)}$ is valid, there is no need for defeating A_φ because of the argument A for $\bot$. For example, let

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

$$\left.\begin{array}{r} \alpha \vdash \alpha \leadsto \varphi \vdash \varphi \leadsto \eta \\ \beta \vdash \beta \leadsto \psi \vdash \psi \leadsto \neg\eta \end{array}\right| \vdash \bot$$

be an argument for $\bot$. Then we can defeat $\alpha \leadsto \varphi$ by introducing the rule $\alpha \wedge \beta \leadsto \neg(\alpha \leadsto \varphi)$.

Hence, we can avoid the need for defeating a proper sub-argument of a disagreeing argument, if necessary by introducing additional rules. Since a disagreeing argument has a unique last rule, defeating a disagreeing argument implies defeating its *last rule*. Hence, it suffice to defeat one of the last rules $\vec{A}$ of an argument A for an inconsistency.

4. A PREFERENCE RELATION ON RULES

In the previous section we have seen that no proper sub-argument of one of the disagreeing arguments needs to be subject to defeat. This makes it possible to defeat a disagreeing argument by defeating its *last rule*. We will now investigate whether we can determine the rule to be defeated by considering only the last rule of each of the disagreeing arguments; i.e. the last rules of the argument for the inconsistency.

Defeating the last rule of one of the disagreeing arguments in case of an inconsistency offers three advantages. Firstly, we on longer have to consider a defeat relation between arguments as is done in: (Pollock 1987, 1994; Simari & Loui 1992; Vreeswijk 1997). This significantly simplifies the preference relation that we must consider. If we use a preference relation on rules, then there are $\frac{1}{2}n(n-1)$ pairs of rules between which there can be a preference, where n is the number of defeasible rules. If we use a preference on the arguments, ignoring the structure of the arguments, there can be $\frac{1}{2}2^n(2^n-1)$ pairs of arguments between which there can be a preference.

Secondly, an argument looses its conclusive force (is defeated) if it contains defeated rules. This simplifies the handling of arguments. And, as we will see in Section 6, it enables us to determine a set of valid arguments in linear time.

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

To show that there is no need for a preference relation on arguments, we will show that a dependence on sub-arguments can be removed by reformulating the set of rules. Suppose that we have two different arguments for an inconsistency where both arguments have the same set of last rules.

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

$$\left.\begin{array}{r} \alpha \vdash \alpha \leadsto \varphi \vdash \varphi \leadsto \eta \\ \beta \vdash \beta \leadsto \psi \vdash \psi \leadsto \neg\eta \end{array}\right| \vdash \bot$$

and

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

$$\left.\begin{array}{r}\varphi \vdash \varphi \leadsto \eta \\ \psi \vdash \psi \leadsto \neg\eta\end{array}\right| \vdash \perp$$

Also suppose that $\varphi \leadsto \eta$ must be defeated given $A_\perp$ and $\psi \leadsto \neg\eta$ must be defeated given $A'_\perp$. In the former case, the situation described by α and β represents an exception on the rule $\varphi \leadsto \eta$. We can, for example, describe this exception by introducing the rule $\alpha \wedge \beta \leadsto \neg\eta$ with preference $\alpha \wedge \beta \leadsto \neg\eta \succ \varphi \leadsto \eta$, or the rule $\alpha \wedge \beta \leadsto \neg(\varphi \leadsto \eta)$. 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.

Another possibly problematic situation arises when a set of arguments supporting a proposition, is stronger than each individual argument. This is known as *accrual of reasons*. Such a situation suggest that we need to consider preferences between sets of rules. We can, however, handle such situations 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 γ and either α or β hold. By introducing a rule $\alpha \wedge \beta \leadsto \psi$ and by preferring it to $\gamma \leadsto \neg\psi$, we can assure that ψ holds whenever α, β and γ 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 for such situations. Fortunately, a set of arguments that weakens the support for a proposition, seems to be counter-intuitive.

A last motivation for using a preference relation on rules, comes from Prakken's (1993) investigation of legal argumentation. He points out that in legal argumentation, meta rules, such as 'lex superior' and 'lex posterior', are used to resolve the inconsistency. These meta rules define a preference relation on legal norms (the defeasible rules). When arguments disagree, the meta rules are applied to the last rules of the disagreeing arguments in order to determine the argument to be defeated. Prakken illustrates this with legal examples. Also notice that meta rules can also be subject to defeat in situations where they specify incompatible relations between rules (Brewka 1994).

From the above discussion, we can draw the following conclusion.

> Let $\langle \Sigma, D\rangle$ be a defeasible theory and let $\succ$ be a partial preference relation on D. Furthermore, let
>
> $$A_\perp = \{\langle A'_1, \eta_1 \leadsto \psi_1\rangle, \dots, \langle A'_k, \eta_k \leadsto \psi_k\rangle\langle \emptyset, \sigma_1\rangle, \dots, \langle \emptyset, \sigma_j\rangle\}$$
>
> be an argument for an inconsistency. So,
>
> $$\begin{array}{l} A_1 = \{\langle A'_1, \eta_1 \leadsto \psi_1\rangle\}, ..., A_k = \{\langle A'_k, \eta_k \leadsto \psi_k\rangle\}, \\ A_{k+1} = \{\langle \emptyset, \sigma_1\rangle\}, \dots, A_n = \{\langle \emptyset, \sigma_j\rangle\}\end{array}$$
>
> are disagreeing arguments.
> Then, if $\eta_i \leadsto \psi_i$ is the least preferred last rule in $\vec{A}_\perp$, $\eta_i \leadsto \psi_i$ must be defeated.

Since we are using a preference relation on the set of defeasible rules in order to resolve 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 also 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]

[3]Since the set of rules D is usually considered as background knowledge, we can determine the specificity preference relation in advance.

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 $K \subseteq L$ be some general background knowledge, let D be a set of defeasible rules, and let $\varphi \rightsquigarrow \psi, \eta \rightsquigarrow \mu$ be two rules in D.

$\varphi \rightsquigarrow \psi$ is more specific than $\eta \rightsquigarrow \mu$ if and only if, given the premises $\{\varphi\}$, there is an argument A_η for η such that $\bar{A}_\eta \subseteq \{\varphi\} \cup K$.

$\varphi \rightsquigarrow \psi$ is strictly more specific than $\eta \rightsquigarrow \mu$, $\varphi \rightsquigarrow \psi \succ_{\text{spec}} \eta \rightsquigarrow \mu$, if and only if $\varphi \rightsquigarrow \psi$ is more specific than $\eta \rightsquigarrow \mu$ and $\eta \rightsquigarrow \mu$ in not more specific than $\varphi \rightsquigarrow \psi$.

Example 3. Let $\varphi \rightsquigarrow \psi$, $\varphi \rightsquigarrow \eta$ and $\eta \rightsquigarrow \neg\psi$ be three defeasible rules.

Given the premises $\{\varphi\}$, we can derive the argument $A_\eta = \{\langle\{\langle\emptyset, \varphi\rangle\}, \varphi \rightsquigarrow \eta\rangle\}$. Since $\bar{A}_\eta \subseteq \{\varphi\} \cup K$, $\varphi \rightsquigarrow \psi$ is more specific than $\eta \rightsquigarrow \neg\psi$. Furthermore, since, given the premises $\{\eta\}$, there is no argument for φ, $\varphi \rightsquigarrow \psi$ is strictly more specific than $\eta \rightsquigarrow \neg\psi$.

The above defined specificity preference relation 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.

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 $A_\perp$ be an argument for an inconsistency and let $\varphi \rightsquigarrow \psi \in min_\succ(\vec{A}_\perp)$ be a least preferred last rule for the inconsistency.

If $\langle A_\varphi, \varphi \rightsquigarrow \psi\rangle \in A_\perp$, then $A_{\neg(\varphi\rightsquigarrow\psi)} = (A_\perp \backslash \{\langle A_\varphi, \varphi \rightsquigarrow \psi\rangle\}) \cup A_\varphi$ is an argument for the defeat of $\varphi \rightsquigarrow \psi$.

Example 4. Let

$$A_\perp = \{\langle\{\langle\{\langle\emptyset, \alpha\rangle\}, \alpha \rightsquigarrow \varphi\rangle\}, \varphi \rightsquigarrow \psi\rangle, \langle\{\langle\emptyset, \eta\rangle\}, \eta \rightsquigarrow \mu\rangle\}$$

$$\left.\begin{array}{r} \alpha \vdash \alpha \rightsquigarrow \varphi \vdash \varphi \rightsquigarrow \psi \\ \eta \vdash \eta \rightsquigarrow \mu \end{array}\right| \vdash \perp$$

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

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

$$\left.\begin{array}{r} \alpha \vdash \alpha \rightsquigarrow \varphi \\ \eta \vdash \eta \rightsquigarrow \mu \end{array}\right| \!\circ\, \neg(\varphi \rightsquigarrow \psi)$$

We use the symbol $|\!\circ$ to denote that $\neg(\varphi \rightsquigarrow \psi)$ does not deductively follow from φ and μ. $\neg(\varphi \rightsquigarrow \psi)$ "follows" from $\varphi \rightsquigarrow \psi$, A_φ, A_μ and $\succ$.

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, let Γ be a set of defeasible rules and let

$$Defeat(\Gamma) = \{\alpha \rightsquigarrow \beta \mid A_{\neg(\alpha \rightsquigarrow \beta)} \in \mathcal{A}, \tilde{A}_{\neg(\alpha \rightsquigarrow \beta)} \cap \Gamma = \emptyset\}.$$

Then the set of defeated rules Ω is defined as:

$$\Omega = Defeat(\Omega).$$

Proposition 1. The set of defeated rules Ω are incomparable. I.e. for each $\Lambda \neq \Omega$ such that $\Lambda = Defeat(\Lambda)$, neither $\Lambda \subset \Omega$ nor $\Lambda \supset \Omega$ holds.

Proofs can be found in Appendix A.

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 α and β and rules $\alpha \rightsquigarrow \gamma$, $\beta \rightsquigarrow \delta$, $\gamma \rightsquigarrow \neg\delta$ and $\delta \rightsquigarrow \neg\gamma$, where the last two rules are preferred to the first two. Here there are two sets of defeated rules Ω; $\{\alpha \rightsquigarrow \gamma\}$ and $\{\beta \rightsquigarrow \delta\}$.

Given the sets 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*.[4]

Definition 7. Let Ω be a set of defeated rules and let $\mathcal{A}$ be a set of all derivable arguments.

Then an extension $\mathcal{E}$ is defined as:

$$\mathcal{E} = \{\varphi \mid A_\varphi \in \mathcal{A}, \tilde{A}_\varphi \cap \Omega = \emptyset\}.$$

The belief set contains the propositions that hold in every extension. This is the skeptical view in which the belief set consists of every proposition for which we have a valid argument in every extension.

Definition 8. Let $\langle \Sigma, D, \succ \rangle$ be a defeasible theory. Furthermore, let $\mathcal{E}_1, ..., \mathcal{E}_n$ be the corresponding extensions.

Then the belief set B is defined as:

$$B = \bigcap_{i=1}^{n} \mathcal{E}_i.$$

It is possible to have a set of arguments for which no extension exists. Such a situation can arise when the set of arguments contain self-defeating arguments. In its most simple form, self-defeat is related to one argument $\alpha \rightsquigarrow \beta \in \tilde{A}_{\neg(\alpha \rightsquigarrow \beta)}$.[5] Since the set of defeasible

[4] Some proposals made in the literature do not consider multiple extensions. Instead, they consider provisionally defeated arguments. These are arguments that are valid in some extensions but not in all extensions. For a discussion see (Vreeswijk 1997; Prakken & Vreeswijk 1999).

[5] Prakken & Vreeswijk (1999) present an instance of the liar's paradox as an example of self-defeat. With a different formulation of the example, however, we can solve the paradox by using defeasible rules without introducing self-defeat.

rules belong to the background knowledge, self-defeat seems to present a defect in our knowledge. Hence, a revision of the set of rules D is needed. I.e. some rules must be removed or reformulated. Though this is an important topic, it does not help us much in practical situations. Hence, we need a way to draw useful conclusions even if self-defeat occurs. Pollock (1994) introduces partial status assignments to deal with the problem. Here, we can do something similar. Firstly, we will reformulate Definition 6 in terms of a status assignment.

> A status assignment is an assignment of *defeated* and *undefeated* to rules in D based on the following condition.
> A rule $\varphi \rightsquigarrow \psi \in D$ is assigned *defeated* if and only if there is an argument $A_{\neg(\varphi \rightsquigarrow \psi)}$ such that every $\alpha \rightsquigarrow \beta \in \tilde{A}_{\neg(\varphi \rightsquigarrow \psi)}$ is assigned *undefeated*. Otherwise, the rule $\varphi \rightsquigarrow \psi \in D$ is assigned *undefeated*

Ω is the set of rules that are assigned the status *defeated*.

Proposition 2. A set of rules Ω is a fixed point of *Defeat* if and only if there is a status assignment such that Ω is the set of rules that are assigned the status *defeated*.

To deal with self-defeat, following Pollock (1994), we can use a partial status assignment.

> A *partial* status assignment is an assignment of *defeated* and *undefeated* to a largest subset of the rules in D based on the following conditions.
>
> - A rule $\varphi \rightsquigarrow \psi \in D$ is assigned *defeated* if and only if there is an argument $A_{\neg(\varphi \rightsquigarrow \psi)}$ such that every $\alpha \rightsquigarrow \beta \in \tilde{A}_{\neg(\varphi \rightsquigarrow \psi)}$ is assigned *undefeated*.
> - A rule $\varphi \rightsquigarrow \psi \in D$ is assigned *undefeated* if and only if for every argument $A_{\neg(\varphi \rightsquigarrow \psi)}$ there is a rule $\alpha \rightsquigarrow \beta \in \tilde{A}_{\neg(\varphi \rightsquigarrow \psi)}$ that is assigned *defeated*.
>
> A rule $\varphi \rightsquigarrow \psi \in D$ that is not is assigned *defeated* or *undefeated* are denoted as being *undetermined*

Since we should only consider conclusions based on arguments containing *undefeated* rules, Ω is the set of rules that are *not* assigned the status *undefeated*.

In the remainder of this paper, with exception of the next section, we will assume that status assignments are complete.

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). 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.

Such a JTMS must contain a node N for every proposition of the form $\neg(\alpha \rightsquigarrow \beta)$ for which we have an argument in $\mathcal{A}$. Furthermore, for each node $N_{\neg(\alpha \rightsquigarrow \beta)}$ representing $\neg(\alpha \rightsquigarrow \beta)$ and for each argument in $\mathcal{A}$ supporting $\neg(\alpha \rightsquigarrow \beta)$, the JTMS contains a justification $\langle$*in*-nodes, *out*-nodes, $N_{\neg(\alpha \rightsquigarrow \beta)}\rangle$. Such a justification consists of an *empty set* of *in*-nodes and a set of *out*-nodes. If A is an argument for $\neg(\alpha \rightsquigarrow \beta)$, then

$$\langle \emptyset, \{N_{\neg(\varphi \rightsquigarrow \psi)} \mid \varphi \rightsquigarrow \psi \in \tilde{A}\}, N_{\neg(\alpha \rightsquigarrow \beta)}\rangle$$

is a justification for $N_{\neg(\alpha \rightsquigarrow \beta)}$.

It is not difficult to verify that each valid labeling of the nodes corresponds one to one with status assignment to the corresponding rules. A rule is assigned the status *defeated* if and only if the corresponding node of the JTMS is labeled IN.

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 $\mathcal{A}$ contains self-defeating arguments. In its most simple form, self-defeat is related to one argument $\alpha \rightsquigarrow \beta \in \tilde{A}_{\neg(\alpha \rightsquigarrow \beta)}$. In general, self-defeat is represented by *odd loops* in the corresponding JTMS.

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 IN or 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 a pragmatic way.

Finding a 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 labeling will be given in Appendix B. Although a labeling can be found in linear time, the number of possible labelings can be exponential in number of minimal arguments for inconsistencies if no preference relation over the defeasible rules is specified; i.e. $\succ = \emptyset$.

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. Semi-normal and non-normal default rules are used to realize undercutting defeat or to define preferences between default rules. Using semi-normal and non-normal default rules for the encoding of preferences is not very elegant. A more important problem is, however, that we cannot specify preferences between default rules that cause an inconsistency because of contingent information. Prioritized default logic does not have this drawback. The prioritized default logic proposed below is similar to Brewka's prioritized default logic. We will, however, 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 holds*–. 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 causes its defeat.

Definition 9. Let $\langle \Sigma, D, \succ \rangle$ 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 $\varphi \rightsquigarrow \psi$ with respect to $\succ$, then $\neg(\varphi \rightsquigarrow \psi) \in \Gamma(S)$;
4. if $\varphi \in \Gamma(S)$, $\varphi \rightsquigarrow \psi \in D$ and $\neg(\varphi \rightsquigarrow \psi) \notin S$, then $\psi \in \Gamma(S)$.

Δ defeats $\varphi \rightsquigarrow \psi$ with respect to $\succ$ if and only if

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

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

Theorem 1. Let $\langle \Sigma, D, \succ \rangle$ be a defeasible theory. The set of extensions determined by the argument system is equal to the set of extensions determined by the default logic.

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

Then we can construct the following arguments.

$$\begin{aligned} &A_\delta = \{\langle\{\langle\emptyset, \alpha\rangle\}, \alpha \rightsquigarrow \delta\rangle\}; \\ &A_{\neg\delta} = \{\langle\{\langle\emptyset, \beta\rangle\}, \beta \rightsquigarrow \neg\delta\rangle\}; \\ &A_{\neg(\alpha\rightsquigarrow\delta)} = \{\langle\{\langle\emptyset, \beta\rangle\}, \beta \rightsquigarrow \neg\delta\rangle, \langle\emptyset, \alpha\rangle\} \end{aligned}$$

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

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

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 $\alpha \rightsquigarrow \delta$. Then δ must belong to the extension. Furthermore, since $\beta \rightsquigarrow \neg\delta \succ \alpha \rightsquigarrow \delta$, $\beta \rightsquigarrow \neg\delta$ will not be defeated either. Therefore, $\neg\delta$ will belong to the extension. But then $\alpha \rightsquigarrow \delta$ will be defeated. Contradiction.

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

So, we have one extension

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

We can establish a relation between this new prioritized default logic and Reiter's default logic. Firstly, we can translate defeasible rules to default rules. Since we must be able to denote that the application of a default rule is no longer valid, $\neg(\alpha \rightsquigarrow \beta)$, 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 $n_{\alpha\rightsquigarrow\beta}$ is the name of the translation of $\alpha \rightsquigarrow \beta$, then $\neg n_{\alpha\rightsquigarrow\beta}$ will be the translation of $\neg(\alpha \rightsquigarrow \beta)$. To ensure that a default rule named $n_{\alpha\rightsquigarrow\beta}$ will not be applied if

$\neg n_{\alpha \rightsquigarrow \beta}$ 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 $\alpha \rightsquigarrow \beta$ to the default rule

$$\frac{\alpha : \beta, n_{\alpha \rightsquigarrow \beta}}{\beta}.$$

The translations 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 preference relation on the defeasible rules is not taken into account, some Reiter-extensions need not be extensions according to Definition 9. To eliminate these extensions, we must encode the preference relation using default rules. To do this properly, we must also change the translation of $\alpha \rightsquigarrow \beta$.

For every rule $\alpha \rightsquigarrow \beta \in D$, introduce a non-normal default rule:

$$\frac{\alpha : n_{\alpha \rightsquigarrow \beta}}{\beta}.$$

For every set of rules $\{\eta_1 \rightsquigarrow \mu_1, ..., \eta_k \rightsquigarrow \mu_k\} \subseteq D$ and for every $\varphi \rightsquigarrow \psi \in D$ such that $\Sigma \cup \{\mu_1, ..., \mu_k\} \vdash \neg\psi$ and for no $\eta_i \rightsquigarrow \mu_i$ there holds: $\varphi \rightsquigarrow \psi \succ \eta_i \rightsquigarrow \mu_i$, introduce a default rule:

$$\frac{\eta_1 \wedge ... \wedge \eta_k : n_{\eta_1 \rightsquigarrow \mu_1}, ..., n_{\eta_k \rightsquigarrow \mu_k}, \neg n_{\varphi \rightsquigarrow \psi}}{\neg n_{\varphi \rightsquigarrow \psi}}.$$

A disadvantage of this translation is that it depends on the premises Σ.

We can also translate default rules to defeasible rules, with the exception of non-normal default rules. Consider a normal or semi-normal default rule of the form:

$$\frac{\alpha : \beta_1, ..., \beta_k, \gamma}{\gamma}.$$

We can translate this default rule to the following defeasible rules:

$$\alpha \rightsquigarrow \gamma, \neg\beta_1 \rightsquigarrow \neg(\alpha \rightsquigarrow \gamma), ..., \neg\beta_k \rightsquigarrow \neg(\alpha \rightsquigarrow \gamma).$$

Specificity. Poole (1985) gives a semantic 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 α describes a set of rules D needed to derive α; $F_c \cup D \cup F_n \models \alpha$. Here, the defeasible rules D are represented by implications. Furthermore, F_c and F_n denote the contingent and the necessary facts respectively.

According to Poole (1985), $\langle D_1, \psi \rangle$ is more specific than $\langle D_2, \mu \rangle$, if for every set of possible fact F_p:

if $F_p \cup D_1 \cup F_n \models \psi$ and $F_p \cup D_2 \cup F_n \not\models \psi$, then $F_p \cup D_2 \cup F_n \models \mu$.

In this paper, another definition has been given. This definition can be related to Poole's definition of specificity.

Theorem 2. Let $\varphi \rightsquigarrow \psi$ and $\eta \rightsquigarrow \mu$ be two rules.

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

The converse of this theorem need not hold. The reason why the converse need not hold is because a set of rules (D_1 or D_2) need not uniquely determine a single argument. Furthermore, $D_1 = \{\varphi \rightsquigarrow \psi\}$ is more specific than $D_2 = \{\eta \rightsquigarrow \mu\}$ according to Poole's definition while it is only more specific according to the definition given in this paper if there exists an argument A_η such that $\tilde{A}_\eta \subseteq \{\varphi\}$.

Closure properties. Gabbay (1985) has initiated the study of the closure properties of the non-monotonic derivability relation: '$\mid\!\sim$' (Gabbay 1985; Kraus et al. 1990; Makinson 1988). Here, the non-monotonic derivability relation is defined as:

$\Sigma \mid\!\sim \varphi$ if and only if B is the belief set of $\langle \Sigma, D, \succ \rangle$ and $\varphi \in B$.

Gabbay (1985) argues that there are three axioms that must be satisfied by all non-monotonic logics.

Reflexivity
if $\varphi \in \Sigma$, then $\Sigma \mid\!\sim \varphi$;
Cut
if $\Sigma \mid\!\sim \varphi$ and $\Sigma \cup \{\varphi\} \mid\!\sim \psi$, then $\Sigma \mid\!\sim \psi$;
Cautious Monotonicity
if $\Sigma \mid\!\sim \varphi$ and $\Sigma \mid\!\sim \psi$, then $\Sigma \cup \{\varphi\} \mid\!\sim \psi$;

These axioms characterize the property called *cumulativity.*

We wish, of course, that all logical consequences of the set of premises are also derivable.

Deduction
if $\Sigma \vdash \varphi$, then $\Sigma \mid\!\sim \varphi$;

This axiom implies *Reflexivity*, it implies together with *Cut* the axiom *Right Weakening*, and it implies together with *Cautious Monotonicity* and *Cut* the axiom *Left Logical Equivalence.* The latter two axioms have been proposed by Kraus, Lehmann and Magidor (1990). They also proposed an axiom characterizing reasoning by cases.

Or
if $\Sigma \cup \{\varphi\} \mid\!\sim \eta$ and $\Sigma \cup \{\psi\} \mid\!\sim \eta$, then $\Sigma \cup \{\varphi \vee \psi\} \mid\!\sim \eta$;

Non-monotonic logics satisfying *Deduction*, *Cautious Monotonicity*, *Cut* and *Or* are said to belong to system **P**.

Theorem 3. The defeasible theory $\langle \Sigma, D, \succ \rangle$ satisfies:
Reflexivity, Deduction, Cut and, in the absence of odd loops, *Cautious Monotony.*
An *odd loop* is an odd number of arguments $A_1, \ldots, A_n$ where every A_{i+1} defeats a rule in $\tilde{A}_i$, and A_1 defeats a rule in $\tilde{A}_n$.

A defeasible theory does not satisfy the closure property *Or*, and therefore does not allow for *reasoning by cases.*

8. REASONING BY CASES

To enable reasoning by cases in an argument system, the usual approach is to use indirect argumentation. Indirect argumentation involves subsidiary arguments that justify a conclusion with respect to the premises and some assumptions. If we have an argument for φ under the assumption α, an argument for φ under the assumption β and an argument for $\alpha \vee \beta$, then we can construct an argument for φ without the assumptions α and β using reasoning by cases. Most argument systems, however, do not allow for subsidiary argumentation and therefore are not able to reason by cases. This also holds for the argument system proposed in the previous sections.

We can of course extend the argument system by (i) allowing for subsidiary arguments and (ii) introducing a rule for combining arguments through reasoning by cases. If arguments were not defeasible, such a simple extension would suffice. Unfortunately, arguments are defeasible, so we must also address the defeasibility of an argument when reasoning by cases. The question that we have to address is whether an argument can be defeated by a subsidiary argument when reasoning by cases. If an argument is defeated by a subsidiary argument in every case described by a disjunction, then the answer is clearly Yes. Roos (1997a, 1998) has argued that the answer must also be Yes when an argument is defeated by only one subsidiary argument corresponding with a case of a disjunction. He illustrates the necessity for this with the following example.

Suppose that we have the following rules:

- A person who injures another person must be punished.
- A person who injures another person in self-defense, should not be punished.
- A person who is dragged into a fight against his/her will, is acting in self-defense.

Now suppose that John has injured Peter and that a reliable witness testifies that either John or Paul has been dragged into the fight against his will. If the argument for not punishing John in case he acted in self-defense, would not defeat the argument for punishing John, we will conclude that John must be punished. This would be most unfortunate for John if he was dragged into the fight against his will.

When reasoning by cases, we should be able to apply defeasible rules in a case. The above example suggests that we should also resolve conflicts within the context of a case. There is one issue with resolving conflicts within the context of a case, as is illustrated by the following example.

> John normally attends a party when he is invited: $ji \rightsquigarrow jp$. Bob and John never attend the same party: $\neg(jp \wedge bp)$. John is invited to a party: ji.

The proposition $\neg(jp \wedge bp)$ implies two cases, one in which John does not attend the party, and one in which Bob does not. Clearly the former case conflicts with the conclusion of the rule $ji \rightsquigarrow jp$. Since facts defeat defeasible rules, we would conclude that John will not attend the party, in the former case. This conclusion is not valid because the conclusion of the rule $ji \rightsquigarrow jp$ is consistent with the other case described by the proposition $\neg(jp \wedge bp)$.

To address the above described problem, we will use the following principles for reasoning by cases:

- Conclusions drawn in a case may not change when other cases are eliminated because of additional information. Of course the overall conclusions may change because they depend on all cases.
- Conflicts must be evaluated using the initial information and the conclusions of applied defeasible rules. The defeasible rules may be applied in a case implied by one or more propositions.

A possible way to enable reasoning by case proposed in (Roos 1997a) is by introducing special defeasible rules, called *hypotheses*, that generate cases. To avoid that we consider cases α and β that follow from the disjunction $\alpha \vee \beta$, simultaneously, we ensure that the cases are mutually exclusive. Considering cases α and β simultaneously corresponds to the case $\alpha \wedge \beta$, which is only one of the three possibilities implied by the disjunction $\alpha \vee \beta$. The following set of *hypotheses* can be used to introduce the mutually exclusive cases.

$$H = \{\alpha \vee \beta \rightsquigarrow \alpha \wedge \neg\beta, \alpha \vee \beta \rightsquigarrow \alpha \wedge \beta, \alpha \vee \beta \rightsquigarrow \neg\alpha \wedge \beta \mid \alpha \vee \beta \in L\ \}$$

For every proposition ψ that is unknown with respect to the partial models, we can derive $\psi \vee \neg\psi$ and $\varphi \vee \psi$ where φ is a known proposition. These disjunctions should not be considered for reasoning by cases. If we would, we could defeat the rule 'birds fly' using the disjunction $\mathit{penguins} \vee \neg\mathit{penguins}$ and the rule 'penguins do not fly'. Clearly, we do not want this.

The reason why we should not consider these disjunctions is because of the difference between *unknown* and *uncertain*. Uncertainty is expressed by multiple cases, while unknown is expressed by a single case of which we do not (yet) know the truth-value some atomic propositions. Some of the rules may fill in the yet unknown information.

We may apply a disjunction $\alpha \vee \beta$ for reasoning by cases if $\alpha \vee \beta$ is not a derived tautology and if $\alpha \vee \beta$ has not been derived from α or β. A characteristic of these requirements is that $\alpha \vee \beta$ may not contain more atomic propositions than the proposition from which it is derived. This requirement blocks the possibility of introducing irrelevant cases. Furthermore, the only tautologies that are allowed according to this requirement, cannot do any harm as we will see below. Hence, we can formulate the following modified definition of an argument.

Definition 10. *Definition 1 revised.* Let $\langle \Sigma, D, \succ \rangle$ be a defeasible theory where Σ is the set of premises and D is the set of rules.

Then an argument A for a proposition ψ is recursively defined in the following way:

- For each $\psi \in \Sigma$: $A = \{\langle \emptyset, \psi \rangle\}$ is an argument for ψ.
- 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 ψ.
- For each $\varphi \rightsquigarrow \psi \in D$ if A' is an argument for φ, then $A = \{\langle A', \varphi \rightsquigarrow \psi \rangle\}$ is an argument for ψ.
- For each $\varphi \rightsquigarrow \psi \in H$ if A' is an argument such that $\hat{A}' \vdash \varphi$ and $At(\varphi) \subseteq At(\hat{A}')$, then $A = \{\langle A', \varphi \rightsquigarrow \psi \rangle\}$ is an argument for ψ.

The function $At(\cdot)$ denotes the set of atomic propositions used in a propositions or a set of propositions.

If we have, for example, and argument for $\alpha \wedge \beta$, we can derive an argument for $\alpha \vee \neg\alpha$ and for $\alpha \vee \neg\beta$. It is, however, invalid to apply $\alpha \vee \neg\alpha$ and $\alpha \vee \neg\beta$ for reasoning by cases. Fortunately, since any argument for the cases $\neg\alpha$ and $\neg\beta$ will be inconsistent with the argument for $\alpha \wedge \beta$, we can resolve the problem using preferences. By preferring any defeasible rule in D to any hypothesis in H, we guarantee that a case described by a disjunction will not be considered if another case described by the disjunction is derivable.

Definition 11. Let D be a set of defeasible rules and let H be a set of hypotheses.

For each $\alpha \rightsquigarrow \beta \in D$ and for each $\gamma \rightsquigarrow \delta \in (H \backslash D)$: $\alpha \rightsquigarrow \beta \succ \gamma \rightsquigarrow \delta$.

The above definitions allows us to construct an argument for a case described by a disjunction. The preference relation ensures that we can apply reasoning by cases if no constituent of a disjunction is derivable. Furthermore, since the cases introduced by the hypotheses are mutually exclusive, each case will be represented by a separate extension. So, disjunctions can be viewed as describing *possible extensions*.

Viewing a disjunction as describing possible extensions is an important deviation from the 'normal' interpretation of a disjunction. In argument systems multiple extensions arise because there is no preference between two or more conflicting arguments; e.g. the Nixon diamond. This can be interpreted as a disjunction stating that one of the arguments is valid. For each case described by this disjunction, we create an extension describing that case. For real disjunctions we can do the same. We can introduce an extension for each case described by a disjunction. Above we have realized this by using hypotheses. These hypotheses create an extension for each case described by a disjunction.

To illustrate reasoning by cases using defeasible rules, we will apply the above presented results to the example of John who might be dragged into a fight.

$$\begin{array}{l}
\mathit{John_injures_Peter} \\
\mathit{John_dragged_into_fight} \vee \mathit{Paul_dragged_into_fight} \\
\neg(\mathit{John_dragged_into_fight} \wedge \mathit{Paul_dragged_into_fight}) \\
\mathit{John_dragged_into_fight} \rightsquigarrow \mathit{self_defense_John} \\
\mathit{John_injures_Peter} \rightsquigarrow \mathit{John_must_be_punished} \\
\mathit{self_defense_John} \rightsquigarrow \neg \mathit{John_must_be_punished}
\end{array}$$

$$(\mathit{self\_defense\_John} \rightsquigarrow \neg \mathit{John\_must\_be\_punished}) \succ (\mathit{John\_injures\_Peter} \rightsquigarrow \mathit{John\_must\_be\_punished})$$

Using these facts and rules, we can construct arguments. Two of these argument are:

$$\begin{array}{l}
A_{\mathit{John_must_be_punished}} = \\
\quad \mathit{John_injures_Peter} \vdash \\
\quad\quad \mathit{John_injures_Peter} \rightsquigarrow \mathit{John_must_be_punished} \vdash \\
\quad\quad\quad \mathit{John_must_be_punished}
\end{array}$$

$$\begin{array}{l}
A_{\neg \mathit{John_must_be_punished}} = \\
\quad \mathit{John_dragged_into_fight} \vee \mathit{Paul_dragged_into_fight} \vdash \\
\quad\quad \mathit{John_dragged_into_fight} \vee \mathit{Paul_dragged_into_fight} \rightsquigarrow \\
\quad\quad\quad \mathit{John_dragged_into_fight} \wedge \neg \mathit{Paul_dragged_into_fight} \vdash \\
\quad\quad\quad\quad \mathit{John_dragged_into_fight} \rightsquigarrow \mathit{self_defense_John} \vdash \\
\quad\quad\quad\quad\quad \mathit{self_defense_John} \rightsquigarrow \neg \mathit{John_must_be_punished} \vdash \\
\quad\quad\quad\quad\quad\quad \neg \mathit{John_must_be_punished}
\end{array}$$

Using all derivable arguments, we can determine the following two extensions.

$$\begin{array}{rl}
E_1 = Th(\{ & \mathit{John_injures_Peter}, \\
& \mathit{John_dragged_into_fight}, \\
& \neg \mathit{Paul_dragged_into_fight}, \\
& \mathit{self_defense_John}, \\
& \neg \mathit{John_must_be_punished}, \\
& \neg(\mathit{John_injures_Peter} \rightsquigarrow \mathit{John_must_be_punished}) \})
\end{array}$$

$$\begin{array}{rl}
E_2 = Th(\{ & \mathit{John_injures_Peter}, \\
& \mathit{Paul_dragged_into_fight}, \\
& \neg \mathit{John_dragged_into_fight}, \\
& \mathit{John_must_be_punished} \})
\end{array}$$

Since in only one of the two situations John must be punished, we do not know whether John must be punished. Additional information should be collected to enable us to make a choice between the two situations that are represented by the two extensions.

Reasoning by cases does not guarantee that the closure property *Or* holds because cases are mutually exclusive. We can, however, proof an *Exclusive Or* property.

Theorem 4. The defeasible theory $\langle \Sigma, D, \succ \rangle$ satisfies *Exclusive Or*:
if $\Sigma \cup \{\varphi \wedge \neg\psi\} \mid\sim \eta$, $\Sigma \cup \{\neg\varphi \wedge \psi\} \mid\sim \eta$, then $\Sigma \cup \{\varphi \veebar \psi\} \mid\sim \eta$;

9. RELATED WORK

In the literature, several argument systems that apply defeasible rules have been proposed (Fox et al. 1992; Geffner 1994; Krause et al. 1995; Pollock 1987; Prakken 1993; Prakken & Vreeswijk 1999; 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; Krause 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; Prakken & Vreeswijk 1999; Simari & Loui 1992; Vreeswijk 1997).

Krause, Ambler, Elvang-Gøransson and Fox (Fox et al. 1992; Krause 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[6]. Furthermore, the argument system enables them to evaluate the strength of arguments for an inconsistent set of propositions symbolically. Krause 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. In case of specificity, however, he uses 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 principles 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 φ 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 from 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 arguments 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 and study the mathematical properties of the resulting system.

Pollock (1990) observes that his theory of inductive warrant is not without problems.

[6]A rule is not interpreted as representing a conditional probability.

Therefore, he introduces a new theory of warrant based on the idea of multiple 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.

Dung (1995) discusses the theory of warrant on an abstract level. He presents several notions of acceptable arguments based on a set of arguments and a binary attack relation on the set of arguments. Here, arguments are considered as atomic entities. No relation with an argument system is specified. The different notions of acceptability correspond with different ways of dealing with self-defeat and with multiple extensions. The extensions based on Definition 6 correspond with Dung's *stable extensions* and the extensions based on the partial status assignment correspond with Dung's *preferred extensions*.

Prakken and Vreeswijk (1999) give an overview of argument systems proposed in the literature. In their overview they discuss the strong and weak points of many theories of warrant that have been proposed in the literature. One of the aspect they look at is the handling of self-defeat. Furthermore, they discuss several arguments systems in detail.

The theories of warrant presented by Pollock (1987, 1994), Dung (1995), Vreeswijk (1997), and Prakken and Vreeswijk (1999), 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 these authors is concerned with selecting a set of valid arguments. It seems more natural, however, 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 above mentioned proposals.

Nute (1988, 1994) proposes a *defeasible logic* that is closely related to argument systems for reasoning with defeasible rules. Nute's logic, which seems to be inspired by logic programming, does not derive arguments that are subsequently evaluated to determine the valid conclusions. Instead, Nute introduces a proof system that guarantees that only valid conclusions are derived. The proof system consists of four rules for deriving formulas that hold and three rules for formulas that cannot hold (Nute 1994). Since the preference relation 'specificity' is an integral part of these rules, the formulation of the rules is rather complex. Furthermore, the approach is less flexible. Adding other preference relations requires a reformulation of the rules.

Conclusions that follow form the defeasible logic can be weaker than the conclusions one would expect. Ideally, in case of multiple extensions, conclusions should be based on those arguments that are valid in all extensions. Prakken & Vreeswijk (1999) point out that in Nute's defeasible logic, this is not always the case. Nute's approach sometimes allows for a smaller number of conclusions than necessary.

An advantage of Nute's approach is its suitability for realizing an implementation. His approach gives us a recursive procedure for the determination of validity of a conclusion. This is in contrast with the procedure proposed by Loui (1998). Loui views the procedure of determining the validity of a conclusion as a dialectics satisfying some protocol.

10. CONCLUSION

A defeasible rule describes a 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 conclude that an inconsistency can be resolved by defeating one of the last rules of the argument supporting the inconsistency. Furthermore, we conclude 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.

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 *undercuts* the application of the rule. Hence, *rebutting* defeat is 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.

A relation between default logic and the proposed argument system has been established and closure properties have been studied. Finally, an extension of the argument system enabling reasoning by cases has been proposed.

APPENDIX A

Proposition 1. The set of defeated rules Ω are incomparable. i.e. for each $\Lambda \neq \Omega$ such that $\Lambda = \mathit{Defeat}(\Lambda)$, neither $\Lambda \subset \Omega$ nor $\Lambda \supset \Omega$ holds.

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

Suppose $\Omega \subset \Lambda$. Then, by the definition of *Defeat*: $\mathit{Defeat}(\Omega) \supseteq \mathit{Defeat}(\Lambda)$. Hence, $\Omega \supseteq \Lambda$. Contradiction. ■

Proposition 2. A set of rules Ω is a fixed point of *Defeat* if and only if there is a status assignment such that Ω is the set of rules that is assigned the status *defeated*.

Proof. Let $\Omega = \{\varphi \rightsquigarrow \psi \mid \varphi \rightsquigarrow \psi$ is assigned the status *defeated*$\}$.

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

Suppose that $\varphi \rightsquigarrow \psi \in \mathit{Defeat}(\Omega)$ and $\varphi \rightsquigarrow \psi \notin \Omega$. Since $\varphi \rightsquigarrow \psi \in \mathit{Defeat}(\Omega)$, there is an argument $A_{\neg(\varphi \rightsquigarrow \psi)}$ such that $A_{\neg(\varphi \rightsquigarrow \psi)} \cap \Omega = \emptyset$. But then, by the definition of a status assignment, $\varphi \rightsquigarrow \psi$ is assigned the status *defeated*. Contradiction.

Suppose that $\varphi \rightsquigarrow \psi \notin \mathit{Defeat}(\Omega)$ and $\varphi \rightsquigarrow \psi \in \Omega$. Since $\varphi \rightsquigarrow \psi \notin \mathit{Defeat}(\Omega)$, there is no argument $A_{\neg(\varphi \rightsquigarrow \psi)}$ such that $A_{\neg(\varphi \rightsquigarrow \psi)} \cap \Omega = \emptyset$. But then, by the definition of a status assignment, $\varphi \rightsquigarrow \psi$ is assigned the status *undefeated*. Contradiction.

Hence, Ω is a fixed point.

Now let Ω be a fixed point of *Defeat*. We assign the status *defeated* to all rules in Ω and *undefeated* to all rules not in Ω.

Suppose that this is not a valid status assignment. Then there is a rule $\varphi \rightsquigarrow \psi$ that is assigned the status *defeated* while there is no argument $A_{\neg(\varphi \rightsquigarrow \psi)}$ such that every $\alpha \rightsquigarrow \beta \in \tilde{A}_{\neg(\varphi \rightsquigarrow \psi)}$ is assigned the status *undefeated*, or $\varphi \rightsquigarrow \psi$ that is assigned the status *undefeated* while there in an argument $A_{\neg(\varphi \rightsquigarrow \psi)}$ such that every $\alpha \rightsquigarrow \beta \in \tilde{A}_{\neg(\varphi \rightsquigarrow \psi)}$ is assigned the status *undefeated*.

In the former case, for every argument $A_{\neg(\varphi \rightsquigarrow \psi)}$ there is a rule $\alpha \rightsquigarrow \beta \in \tilde{A}_{\neg(\varphi \rightsquigarrow \psi)}$ that is assigned the status *defeated*. Therefore, for every argument $A_{\neg(\varphi \rightsquigarrow \psi)}$, $\tilde{A}_{\neg(\varphi \rightsquigarrow \psi)} \cap \Omega \neq \emptyset$. Hence $\varphi \rightsquigarrow \psi \notin \Omega$ and therefore $\varphi \rightsquigarrow \psi$ is assigned the status *undefeated*. Contradiction.

In the latter case, there is an argument $A_{\neg(\varphi\rightsquigarrow\psi)}$ such that every $\alpha \rightsquigarrow \beta \in \tilde{A}_{\neg(\varphi\rightsquigarrow\psi)}$ is assigned the status *undefeated*. Therefore, $\tilde{A}_{\neg(\varphi\rightsquigarrow\psi)} \cap \Omega = \emptyset$. Hence $\varphi \rightsquigarrow \psi \in \Omega$ and therefore $\varphi \rightsquigarrow \psi$ is assigned the status *defeated*. Contradiction. ■

To prove Theorem 1, the following lemmas will be used.

Lemma 1. Let Γ be a set of propositions and let $\mathcal{E}$ be the deductive closure of Γ. Furthermore, let there be an argument for each proposition in Γ.

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

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

Lemma 2. Let $\mathcal{E}$ be an extension according to Definition 9 and let $\Omega = \{\alpha \rightsquigarrow \beta \mid \neg(\alpha \rightsquigarrow \beta) \in \mathcal{E}\}$. Furthermore, let there be an argument A for each proposition in $\mathcal{E}$ and let $\tilde{A} \cap \Omega = \emptyset$.

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

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

Since $\alpha \rightsquigarrow \beta \in \Omega$, either there exists a $\gamma \rightsquigarrow \neg(\alpha \rightsquigarrow \beta)$ and $\gamma \in \mathcal{E}$, or there exists a Δ 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 $\{\mu \mid \eta \rightsquigarrow \mu \in \Delta\} \cup \Sigma \vdash \neg\beta$ 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 $A_{\neg(\alpha\rightsquigarrow\beta)} = \{\langle A_\gamma, \gamma \rightsquigarrow \neg(\alpha \rightsquigarrow \beta)\rangle\}$ and $\tilde{A}_{\neg(\alpha\rightsquigarrow\beta)} \cap \Omega = \emptyset$. Hence, $\alpha \rightsquigarrow \beta \in \mathit{Defeat}(\Omega)$. Contradiction.

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

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

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

In the former case, $\alpha \in \mathcal{E}$ and $\neg\beta \in \mathcal{E}$. But then $\neg(\alpha \rightsquigarrow \beta) \in \mathcal{E}$. Contradiction.

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

Hence, $\Omega = \mathit{Defeat}(\Omega)$. ■

Theorem 1. Let $\langle\Sigma, D, \succ\rangle$ be a defeasible theory. The set of extensions determined by the argument system is equal to the set of extensions determined by the default logic.

Proof. 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; i.e., we assume that $\mathcal{E} = \Gamma(\mathcal{E})$.

1. Clearly for each $\alpha \in \Sigma$ we have an argument $\{\langle\emptyset, \alpha\rangle\}$. Since it contains no rules, it cannot be defeated. Therefore, $\Sigma \subseteq \mathcal{E}$.
2. According to the definition of an argument, $\mathcal{E}$ is deductively closed.
3. Let $\Delta = \{\eta_1 \rightsquigarrow \mu_1, ..., \eta_n \rightsquigarrow \mu_n\}$ defeat $\alpha \rightsquigarrow \beta$ given $\mathcal{E}$. Then $\{\mu_1, ..., \mu_n, \alpha\} \subseteq \Gamma(\mathcal{E}) = \mathcal{E}$ and for no $\eta_i \rightsquigarrow \mu_i \in \Delta$: $\alpha \rightsquigarrow \beta \succ \eta_i \rightsquigarrow \mu_i$. Since $\{\mu_1, ..., \mu_n, \alpha\} \subseteq \mathcal{E}$, we have valid arguments $A_{\mu_1}, ..., A_{\mu_n}, A_\alpha$. Hence, we have a valid argument $A_{\neg(\alpha\rightsquigarrow\beta)}$ for $\neg(\alpha \rightsquigarrow \beta)$. Therefore, $\neg(\alpha \rightsquigarrow \beta) \in \mathcal{E}$.

4. Let $\alpha \in \Gamma(\mathcal{E}) = \mathcal{E}$ and $\neg(\alpha \rightsquigarrow \beta) \notin \mathcal{E}$. Then there exists a valid argument $A_\beta = \{\langle A_\alpha, \alpha \rightsquigarrow \beta\rangle\}$ for β. Hence, $\beta \in \mathcal{E} = \Gamma(\mathcal{E})$.

Hence, $\Gamma(\mathcal{E}) \subseteq \mathcal{E}$.

Suppose that $\mathcal{E}$ is not a minimal set satisfying the requirements of $\Gamma(\mathcal{E})$. Then there is a $\varphi \in \mathcal{E}\backslash\Gamma(\mathcal{E})$ and a corresponding valid argument A_φ. Let A_ψ be the smallest sub-argument such that $\psi \notin \Gamma(\mathcal{E})$.

Suppose that $A_\psi = \{\langle\emptyset, \psi\rangle\}$. Since $\psi \in \Sigma$, $\psi \in \Gamma(\mathcal{E})$. Contradiction.

Suppose that $\varphi_1, ..., \varphi_n \vdash \psi$ and $\{A_{\varphi_1}, ..., A_{\varphi_n}\} \subseteq \mathcal{A}$. Then, since A_ψ is the smallest sub-argument, $\{\varphi_1, ..., \varphi_n\} \subseteq \Gamma(\mathcal{E})$. Therefore $\psi \in \Gamma(\mathcal{E})$. Contradiction.

Suppose that A_ψ with $\psi = \neg(\eta_i \rightsquigarrow \mu_i)$ is the result of an argument $A_\perp = \{\langle A'_1, \eta_1 \rightsquigarrow \mu_1\rangle, \ldots, \langle A'_k, \eta_k \rightsquigarrow \mu_k\rangle\langle\emptyset, \sigma_1\rangle, \ldots, \langle\emptyset, \sigma_\ell\rangle\}$ for an inconsistency. Clearly for no $j \neq i$: $\eta_j \rightsquigarrow \mu_j \succ \eta_i \rightsquigarrow \mu_i$. Since A_ψ is the smallest sub-argument, $\{\mu_1, ..., \mu_{i-1}, \mu_{i+1}, ..., \mu_k, \sigma_1, ..., \sigma_\ell\} \subseteq \Gamma(\mathcal{E})$. Therefore $\psi = \neg(\eta_i \rightsquigarrow \mu_i) \in \Gamma(\mathcal{E})$. Contradiction.

Suppose that $A_\psi = \{\langle A_\mu, \eta \rightsquigarrow \psi\rangle\}$. Since $\mathcal{E}$ is an extension according to Definition 7, there is no valid argument for $\neg(\eta \rightsquigarrow \psi)$. Therefore, $\neg(\eta \rightsquigarrow \psi) \notin \mathcal{E}$. Hence, $\psi \in \Gamma(\mathcal{E})$. Contradiction.

Hence, $\mathcal{E}$ is a fixed point of Γ.

Let $\mathcal{E}$ be an extension according to Definition 9 and let $\Omega = \{\alpha \rightsquigarrow \beta \mid \neg(\alpha \rightsquigarrow \beta) \in \mathcal{E}\}$. So, $\mathcal{E} = \Gamma(\mathcal{E})$. We will proof 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 $\tilde{A} \cap \Omega = \emptyset$. If we have an argument for each $\varphi \in \mathcal{E}$, then, by Lemma 2, Ω satisfies Definition 6, i.e. $\Omega = \mathit{Defeat}(\Omega)$. Since for each $\varphi \in \mathcal{E}$, we have an argument A such that $\tilde{A} \cap \Omega = \emptyset$, A must be a valid argument for φ

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 $\varphi \in \Gamma_0$ we can construct an argument $A_\varphi = \{\langle\emptyset, \varphi\rangle\}$. Furthermore, by Lemma 1, we can construct an argument for each $\varphi \in \mathcal{E}_0$. Clearly, $\tilde{A}_\varphi \cap \Omega = \emptyset$.

Proceeding inductively, let $\mathcal{E}_i \subseteq \mathcal{E}$ be a smallest deductively closed subset such that $\Gamma_i \subseteq \mathcal{E}_i$. Suppose that $\mathcal{E}_i \subset \mathcal{E}$. Then there is a $\varphi \in (\mathcal{E}\backslash\mathcal{E}_i)$ such that either $\varphi = \neg(\alpha \rightsquigarrow \beta)$ and $\alpha \rightsquigarrow \beta$ is defeated given $\mathcal{E}_i$, or $\alpha \in \mathcal{E}_i$, $\alpha \rightsquigarrow \varphi \in D$ and $\neg(\alpha \rightsquigarrow \varphi) \notin \mathcal{E}$, or neither of these two possibilities.

In the third case, $\Gamma(\mathcal{E})$ is not a minimal set. Hence, this case is impossible.

In the first case there is a $\Delta \subseteq \{\eta \rightsquigarrow \mu \in D \mid \{\eta, \mu\} \subseteq \mathcal{E}_i\}$ that defeats $\alpha \rightsquigarrow \beta$. Hence we can construct an argument $A_{\neg(\alpha\rightsquigarrow\beta)}$ for φ such that $\tilde{A}_{\neg(\alpha\rightsquigarrow\beta)} \cap \Omega = \emptyset$.

In the second case $A_\varphi = \{\langle A_\alpha, \alpha \rightsquigarrow \varphi\rangle\}$ is an argument for φ such that $\tilde{A}_\varphi \cap \Omega = \emptyset$.

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

Hence, for every proposition in $\mathcal{E}$ we can construct an argument A such that $\tilde{A} \cap \Omega = \emptyset$. Given these arguments, there holds according to Lemma 2 that $\Omega = \{\alpha \rightsquigarrow \beta \mid \neg(\alpha \rightsquigarrow \beta) \in \mathcal{E}\}$ satisfies Definition 6, i.e. $\Omega = \mathit{Defeat}(\Omega)$. Hence, the arguments for the propositions in $\mathcal{E}$ are valid arguments.

Now suppose that we can construct a valid argument A_φ for a proposition $\varphi \notin \mathcal{E}$, i.e., $\tilde{A}_\varphi \cap \Omega = \emptyset$. Since $\bar{A}_\varphi \subseteq \Sigma$, either for some rule $\alpha \rightsquigarrow \beta \in \tilde{A}_\varphi$ there holds: $\alpha \in \mathcal{E}$ and $\beta \notin \mathcal{E}$. So, $\alpha \rightsquigarrow \beta \in \Omega$. Contradiction.

Hence, $\mathcal{E}$ is an extension according to Definition 7. ■

Theorem 2. Let $\varphi \rightsquigarrow \psi$ and $\eta \rightsquigarrow \mu$ be two rules.

If $\varphi \rightsquigarrow \psi$ is more specific than $\eta \rightsquigarrow \mu$ according to Definition 4, then there are two Poole-arguments $\langle D_1, \psi\rangle$ and $\langle D_2, \mu\rangle$ with $\varphi \rightsquigarrow \psi \in D_1$ and $\eta \rightsquigarrow \mu \in D_2$ 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 $D_2 = \tilde{A}_\eta \cup \{\eta \rightsquigarrow \mu\}$ and $D_1 = \{\varphi \rightsquigarrow \psi\}$.

Since $\varphi \rightsquigarrow \psi$ is more specific than $\eta \rightsquigarrow \mu$, given the premise $\{\varphi\}$ there must exist an argument A_η for η.

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

Suppose that $\bar{A}_\eta = \{\varphi\}$. Then, any possible fact F_p for which the antecedent of Poole's definition holds, must imply φ. Hence, $\langle D_1, \psi\rangle$ is more specific than $\langle D_2, \mu\rangle$. ■

Theorem 3. The defeasible theory $\langle \Sigma, D, \succ\rangle$ satisfies:
Reflexivity, Deduction, Cut and, in the absence of odd loops, *Cautious Monotony.*
An *odd loop* is an odd number of arguments $A_1, \ldots, A_n$ where every A_{i+1} defeats a rule in $\tilde{A}_i$, and A_1 defeats a rule in $\tilde{A}_n$.

Proof. *Reflexivity.* For each $\varphi \in \Sigma$, $A = \{\langle \emptyset, \varphi\rangle\}$ is an argument for φ. Since A contains no rule, it cannot be defeated. Therefore, $\varphi \in B$.

Deduction. For each φ such that $\Sigma \vdash \varphi$, $A = \{\langle \emptyset, \psi\rangle \mid \psi \in \Sigma\}$ is an argument for φ. Since A contains no rules, it cannot be defeated. Therefore, $\varphi \in B$.

Cut. Let $\mathcal{E}$ be an extension of the defeasible theory $\langle \Sigma, D, \succ\rangle$, let B be the belief set of $\langle \Sigma, D, \succ\rangle$, and let $\varphi \in B$

Suppose that $\mathcal{E}$ is no longer an extension after adding some $\varphi \in B$ to Σ. Let Ω be the set of defeated rules that correspond with the extension $\mathcal{E}$. Then after adding φ there must be a new argument $A_{\neg(\alpha \rightsquigarrow \beta)}$ such that $\tilde{A}_{\neg(\alpha \rightsquigarrow \beta)} \cap \Omega = \emptyset$ and $\alpha \rightsquigarrow \beta \notin \Omega$. Since $\tilde{A}_{\neg(\alpha \rightsquigarrow \beta)} \cap \Omega = \emptyset$ and $\alpha \rightsquigarrow \beta \notin \Omega$, $\{\langle \emptyset, \varphi\rangle\}$ must be a sub-argument of $A_{\neg(\alpha \rightsquigarrow \beta)}$

Now three situations are possible.

- $A_{\neg(\alpha \rightsquigarrow \beta)} = \{\langle A_\xi, \xi \rightsquigarrow \neg(\alpha \rightsquigarrow \beta)\rangle\}$. Then there is an $A^*_{\neg(\alpha \rightsquigarrow \beta)}$ in which $\{\langle \emptyset, \varphi\rangle\}$ is replaced by A_φ. Since A_φ is valid; i.e. $\tilde{A}_\varphi \cap \Omega = \emptyset$, there holds that $\alpha \rightsquigarrow \beta \in \Omega$. Contradiction.
- $A_{\neg(\alpha \rightsquigarrow \beta)}$ is derived from $A_\perp$ and $\{\langle \emptyset, \varphi\rangle\}$ is not a disagreeing argument. Since $\{\langle \emptyset, \varphi\rangle\}$ is a sub-argument of $A_\perp$, there is an $A^*_\perp$ in which $\{\langle \emptyset, \varphi\rangle\}$ is replaced by A_φ. Clearly, $\vec{A}_\perp = \vec{A}^*_\perp$. Hence, since A_φ is valid, $\alpha \rightsquigarrow \beta \in \Omega$. Contradiction.
- $A_{\neg(\alpha \rightsquigarrow \beta)}$ is derived from $A_\perp$ and $\{\langle \emptyset, \varphi\rangle\}$ is a disagreeing argument. Then there is an $A^*_\perp$ in which $\{\langle \emptyset, \varphi\rangle\}$ is replaced by A_φ. Hence, $A_\varphi \subseteq A^*_\perp$. Since A_φ is valid, for no $\eta \rightsquigarrow \mu \in \vec{A}^*_\perp$: $\alpha \rightsquigarrow \beta \succ \eta \rightsquigarrow \mu$. Therefore, there is an $A^*_{\neg(\alpha \rightsquigarrow \beta)} = (A^*_\perp \backslash \{\langle A_\alpha, \alpha \rightsquigarrow \beta\rangle\}) \cup A_\alpha$ and $A^*_{\neg(\alpha \rightsquigarrow \beta)} \cap \Omega = \emptyset$. Hence, $\alpha \rightsquigarrow \beta \in \Omega$. Contradiction.

Cautious Monotonicity. Let $\langle \Sigma, D, \succ\rangle$ be a defeasible theory, and let B be the belief set of $\langle \Sigma, D, \succ\rangle$.

Suppose that $\mathcal{E}$ is an extension of the defeasible theory $\langle \Sigma \cup \{\varphi\}, D, \succ\rangle$ for some $\varphi \in B$, but not of $\langle \Sigma, D, \succ\rangle$. Let Ω be the set of defeasible rules determining the extension $\mathcal{E}$. Every extension $\mathcal{E}'$ of $\langle \Sigma, D, \succ\rangle$ determined by the defeasible rules Λ, is also an extension of the defeasible theory $\langle \Sigma \cup \{\varphi\}, D, \succ\rangle$ according to the property *Cut.* Therefore, $\Lambda \not\subseteq \Omega$ and $\Omega \not\subseteq \Lambda$ according to Proposition 1.

Consider an extension $\mathcal{E}'$ of the defeasible theory $\langle \Sigma, D, \succ\rangle$ determined by the defeasible rules Λ. Since $\varphi \in B$, there is an argument A_φ generated by $\langle \Sigma, D, \succ\rangle$ such that $\tilde{A}_\varphi \cap \Lambda = \emptyset$.

Every argument determined by the defeasible theory $\langle \Sigma, D, \succ \rangle$ is also an argument of the defeasible theory $\langle \Sigma \cup \{\varphi\}, D, \succ \rangle$. Moreover, every argument A_ψ determined by the defeasible theory $\langle \Sigma \cup \{\varphi\}, D, \succ \rangle$ is either an argument of the defeasible theory $\langle \Sigma, D, \succ \rangle$, or contains $\{\langle \emptyset, \varphi \rangle\}$ as a sub-argument. If we replace every sub-argument $\{\langle \emptyset, \varphi \rangle\}$ in A_ψ by A_φ, denoted by A^*_ψ, then we get an argument of the defeasible theory $\langle \Sigma, D, \succ \rangle$.

Consider the above mentioned extension $\mathcal{E}$ of $\langle \Sigma \cup \{\varphi\}, D, \succ \rangle$ determined by the defeated rules Ω. Clearly, $\tilde{A}_\varphi \cap \Omega \neq \emptyset$ otherwise $\mathcal{E}$ would als be an extension of $\langle \Sigma, D, \succ \rangle$. Therefore, there is a defeasible rule $\alpha \rightsquigarrow \beta \in (\tilde{A}_\varphi \cap \Omega)$ and a corresponding argument $A_{\neg(\alpha \rightsquigarrow \beta)}$ of the defeasible theory $\langle \Sigma \cup \{\varphi\}, D, \succ \rangle$. In no extension $\mathcal{E}'$, $A^*_{\neg(\alpha \rightsquigarrow \beta)}$ is a valid argument. This is only possible if the validity of $A_{\neg(\alpha \rightsquigarrow \beta)}$ given Ω depends, directly or indirectly through arguments defeating other arguments, on φ. So, $A_{\neg(\alpha \rightsquigarrow \beta)}$ depends, directly or indirectly, on an argument that has $\{\langle \emptyset, \varphi \rangle\}$ as a sub-argument. Hence, if we replace all arguments A on which $A_{\neg(\alpha \rightsquigarrow \beta)}$ depends by A^*, then $A^*_{\neg(\alpha \rightsquigarrow \beta)}$ is part of an odd loop. This contradicts the condition of the theorem. ■

Theorem 4. The defeasible theory $\langle \Sigma, D, \succ \rangle$ satisfies *Exclusive Or*:
if $\Sigma \cup \{\varphi \wedge \neg\psi\} \mid\!\sim \eta$, $\Sigma \cup \{\neg\varphi \wedge \psi\} \mid\!\sim \eta$, then $\Sigma \cup \{\varphi \veebar \psi\} \mid\!\sim \eta$;

Proof. Let $r_1 = \varphi \vee \psi \rightsquigarrow \varphi \wedge \neg\psi$, $r_2 = \varphi \vee \psi \rightsquigarrow \varphi \wedge \psi$ and $r_3 = \varphi \vee \psi \rightsquigarrow \neg\varphi \wedge \psi$.

To proof the theorem, we must prove that for every extension $\mathcal{E}$ of the defeasible theory $\langle \Sigma \cup \{\varphi \vee \psi\}, D, \succ \rangle$, either that $\mathcal{E}$ is an extension of $\langle \Sigma \cup \{\varphi \wedge \neg\psi\}, D, \succ \rangle$ or that $\mathcal{E}$ is an extension of $\langle \Sigma \cup \{\neg\varphi \wedge \psi\}, D, \succ \rangle$. Since for every extension $\mathcal{E}'$ of $\langle \Sigma \cup \{\varphi \wedge \neg\psi\}, D, \succ \rangle$ and of $\langle \Sigma \cup \{\neg\varphi \wedge \psi\}, D, \succ \rangle$, $\eta \in \mathcal{E}'$ holds, and since $B = \bigcap_i \mathcal{E}_i$, $\Sigma \cup \{\varphi \vee \psi\} \mid\!\sim \eta$.

Let $\mathcal{E}$ be an extension of $\langle \Sigma \cup \{\varphi \vee \psi\}, D, \succ \rangle$. Then because of the set of hypotheses H, $\varphi \wedge \neg\psi \in \mathcal{E}$ or $\neg\varphi \wedge \psi \in \mathcal{E}$. Notice that for no $\mathcal{E}$, $\varphi \wedge \psi \in \mathcal{E}$ unless Σ is inconsistent.

Suppose that $\varphi \wedge \neg\psi \in \mathcal{E}$. Then $\Omega = \{\alpha \rightsquigarrow \beta \mid \neg(\alpha \rightsquigarrow \beta) \in \mathcal{E}\}$. To prove that $\mathcal{E}$ is an extension of $\langle \Sigma \cup \{\varphi \wedge \neg\psi\}, D, \succ \rangle$, we have to prove that $\alpha \in \mathcal{E}$ if and only if there is an argument A^*_α such that $A^*_\alpha \cap \Omega = \emptyset$ given $\langle \Sigma \cup \{\varphi \wedge \neg\psi\}, D, \succ \rangle$.

Let $\alpha \in \mathcal{E}$. Then there is an A_α given $\langle \Sigma \cup \{\varphi \vee \psi\}, D, \succ \rangle$ with $\tilde{A}_\alpha \cap \Omega = \emptyset$. Therefore, we can construct an A^*_α given $\langle \Sigma \cup \{\varphi\}, D, \succ \rangle$ such that $A^*_\alpha \cap \Omega = \emptyset$ by first replacing each sub-argument $\{\langle \{\langle \emptyset, \varphi \vee \psi \rangle\}, r_1 \rangle\}$ in A_α by $\{\langle \emptyset, \varphi \rangle\}$ and subsequently by replacing each remaining sub-argument $\{\langle \emptyset, \varphi \vee \psi \rangle\}$ also by $\{\langle \emptyset, \varphi \rangle\}$. Hence, there is an argument A^*_α given $\langle \Sigma \cup \{\varphi \wedge \neg\psi\}, D, \succ \rangle$ with $\tilde{A}^*_\alpha \cap \Omega = \emptyset$.

Let A^*_α be an argument given $\langle \Sigma \cup \{\varphi \wedge \neg\psi\}, D, \succ \rangle$ with $\tilde{A}^*_\alpha \cap \Omega = \emptyset$. Then we can construct an A_α given $\langle \Sigma \cup \{\varphi \vee \psi\}, D, \succ \rangle$ by replacing each sub-argument $\{\langle \emptyset, \varphi \rangle\}$ in A^*_α by $\{\langle \{\langle \emptyset, \varphi \vee \psi \rangle\}, r_1 \rangle\}$. To make sure that $A_\alpha \cap \Omega = \emptyset$, we must make sure that r_1 is not defeated. If r_1 is defeated, there must be a valid argument for $\neg(\varphi \wedge \neg\psi)$. Since $\varphi \wedge \neg\psi \in \mathcal{E}$, there is no such argument.

In case $\neg\varphi \wedge \psi \in \mathcal{E}$, the proof is similar to the one given above. ■

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 of 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 its consequent 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
    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 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 IN or OUT. If we label the node 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, giving it a label and propagating the consequences, till all nodes are labeled. By backtracking on the choices that are made, we determine every labeling of the JTMS.

ACKNOWLEDGMENT

I thank the reviewers and Cees Witteveen for their comments which helped me to improve the paper.

REFERENCES

BONDARENKO, A., P. M. DUNG, R. A. KOWALSKI, and F. TONI. 1997. An abstract, argumentation-theoretic approach to default reasoning. Artificial Intelligence **93**:63–101.

BREWKA, G. 1994. Reasoning about priorities in default logic. Proceedings of the Twelfth National Conference on Artificial Intelligence, 940–945.

CAYROL, C. 1995. On the relation between argumentation and non-monotonic coherence-based entailment. Proceedings of the fourteenth International Joint Conference on Artificial Intelligence, 1443–1448.

DOYLE, J. 1979. A truth maintenance system. Artificial Intelligence, **12**:231–272.

Dung, P. M. 1993. The acceptability of arguments and its fundamental role in nonmonotonic reasoning and logic programming. Proceedings of the thirteenth International Joined Conference on Artificial Intelligence, 852–857.

Dung, P. M. 1995. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. Artificial Intelligence, **77**:321–357.

Fox, J., P. Krause and S. Ambler. 1992. Arguments, contradictions and practical reasoning. Proceedings of the tenth European Conference on Artificial Intelligence, 623–627.

Gabbay, D. M. 1985. Theoretical foundations of non-monotonic reasoning in expert systems. K. R. Apt (editor). Logic and models of concurrent systems, Springer Verlag, 439–457.

Geffner, H. and J. Pearl. 1992. Conditional entailment: bridging two approaches to default reasoning. Artificial Intelligence, **53**:209–244.

Geffner, H. 1994. Causal default reasoning, principles and algorithms, Proceedings of the twelfth National Conference on Artificial Intelligence, 245–250.

Goodwin, J. W. 1987. A theory and system for non-monotonic reasoning. Department of Computer and Information Science, Linköping University, Linköping, Sweden.

Hunter A. 1994. Defeasible reasoning with structured information. Proceedings of the fourth International Conference on Principles of Knowledge Representation and Reasoning, 281–292.

Krause, P., S. Ambler, M. Elvang-Gøransson, J. Fox. 1995. A logic of argumentation for reasoning under uncertainty. Computational Intelligence, **11**:113–131.

Kraus, S., D. Lehmann, M. Magidor. 1990. Nonmonotonic reasoning, preferential models and cumulative logics. Artificial Intelligence, **44**:167–207.

Lin, F. and Y. Shoham 1989. Argument Systems: a uniform basis for nonmonotonic reasoning. Proceedings of the first International Conference on Principles of Knowledge Representation and Reasoning, 245–255.

Loui, R. P. 1987. Defeat among arguments: a system of defeasible inference. Computational Intelligence, **3**:100–106.

Loui, R. P. 1998. Process and policy: resource-bounded nondemonstrative reasoning. Computational Intelligence, **14**:1–38.

Makinson, D. 1988. General Theory of cumulative inference. M. Reinfrank, J. de Kleer, M. L. Ginsberg and E. Sandewall (editors). Proceedings of the second International Workshop on Non-monotonic Reasoning, Springer-Verlag, 1–18.

Nute, D. 1988. Defeasible reasoning and decision support systems. Decision Support Systems, **4**:97–110.

Nute, D. 1994. A decidable quantified defeasible logic. D. Prawitz, B. Skyrms and D. Westerståhl (editors). Logic, Methodology And Philosophy of Science IX. Elsevier Science, 97–110.

Pollock, J. L. 1987. Defeasible reasoning. Cognitive Science, **11**:481–518.

Pollock, J. L. 1992. How to reason defeasibly. Artificial Intelligence, **57**:1–42.

Pollock, J. L. 1994. Justification and defeat. Artificial Intelligence, **67**:377–407.

Poole, D. 1985. On the comparison of theories: preferring the most specific explanation. Proceedings of the ninth International Joint Conference on Artificial Intelligence, 144–147.

Poole, D. 1988. A logical framework for default reasoning. Artificial Intelligence, **36**:27–47.

Prakken, H. 1993. Logical tools for modelling legal arguments. Ph.D. thesis, Vrije Universiteit, Amsterdam, The Netherlands.

Prakken, H., G. Vreeswijk. 2000. Logics for defeasible argumentation. To appear in: The Handbook of Philosophical Logic.

Reinfrank, M. 1989. Fundamentals and Logical Foundations of Truth Maintenance. Ph.D. thesis, Linköping University, Linköping, Sweden.

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

Roos, N. 1997a. Reasoning by cases using arguments. Dutch German Workshop on Non Monotonic Reasoning (DGNMR-97).

ROOS, N. 1997b. On Resolving Conflicts Between Arguments. Technical Report TR-CTIT-97-37 Centre for Telematics and Information Technology, University of Twente, Enschede. ISSN 1381-3625.

ROOS, N. 1998. Reasoning by cases in Default Logic. Artificial Intelligence **99**:165–183.

SIMARI, G. R., R. P. LOUI. 1992. A mathematical treatment of defeasible reasoning and its implementation. Artificial Intelligence, **53**:125–157.

TOULMIN, S. 1958. The uses of argument. Cambridge University Press.

VREESWIJK, G. 1991. The feasibility of defeat in defeasible reasoning. Proceedings of the second International Conference on Principles of Knowledge Representation and Reasoning, 526–534.

VREESWIJK, G. 1992. Nonmonotonicity and partiality in defeasible argumentation. W. van der Hoek, J.-J. Ch. Meyer, Y. H. Tan , C. Witteveen, (editors). Non-monotonic reasoning and partial semantics. Ellis Horwood, 107–155.

VREESWIJK, G. 1997. Abstract argumentation systems. Artificial Intelligence, **90**:225–279.

WITTEVEEN, C. and G. BREWKA. 1993. Skeptical reason maintenance and belief revision. Artificial Intelligence, **61**:1–36.