Dialectical proof theory for defeasible argumentation with defeasible priorities (preliminary report)

Henry Prakken*

Computer/Law Institute, Free University, De Boelelaan 1105 Amsterdam
email: henry@rechten.vu.nl

Abstract. In this paper a dialectical proof theory is proposed for logical systems for defeasible argumentation that fit a certain format This format is the abstract theory developed by Dung, Kowalski and others. A main feature of the proof theory is that it also applies to systems in which reasoning about the standards for comparing arguments is possible.

1 Introduction

One form which nonmonotonic logics can take is systems for defeasible argumentation [Pollock 87, Simari & Loui 92, Vreeswijk 93a, Dung 95, Prakken & Sartor 95]. In such systems nonmonotonic reasoning is analyzed in terms of the interactions between arguments for alternative conclusions. Nonmonotonicity, or defeasibility, arises from the fact that arguments can be defeated by stronger counter-arguments. In this paper a dialectical proof theory is proposed for systems of this kind that fit a certain abstract format, viz. the one defined by [Dung 95]. The use of dialectical proof theories for defeasible reasoning was earlier studied by [Dung94] and, inspired by [Rescher 1977], by [Loui 93, Vreeswijk 93b, Brewka 94], while also [Royakkers & Dignum 1996] contains ideas that can be regarded as a dialectical proof theory. The general idea is based on game–theoretic notions of logical consequence developed in dialogue logic (for an overview see [Barth & Krabbe 82]). Here a proof of a formula from certain premises takes the form of a dialogue game between a proponent and an opponent of the formula. Both players have certain ways available of attacking and defending a statement. A formula is implied by the premises iff it can be succesfully defended against every possible attack.

Below first the general framework of [Dung 95] will be described (Section 2), after which in section 3 the dialectical proof theory is presented. Then in Section 4 Dung's framework and the proof theory will be adapted in such a way that the standards used for comparing conflicting arguments are themselves (defeasible) consequences of the premises.

The proof–theoretical ideas described in this paper were originally developed in [Prakken & Sartor 96b], for a system presented in [Prakken & Sartor 96a],

* Henry Prakken was supported by a research fellowship of the Royal Netherlands Academy of Arts and Sciences, and by Esprit WG 8319 'Modelage'.

which in turn extends and revises [Dung 93b]'s application of his semantics to extended logic programming. The main purpose of the present paper is to show that these ideas apply to any system of the format defined by [Dung 95]. For this reason the present paper does not express arguments in a formal language; it just assumes that this can be done.

2 An abstract framework for defeasible argumentation

Inspired by earlier work of Bondarenko, Kakas, Kowalski and Toni, [Dung 95] has proposed a very abstract and general argument–based framework. An up–to–date technical survey of this approach is [Bondarenko et al. 95]. The two basic notions of the framework are a set of arguments, and a binary relation of defeat among arguments. In terms of these notions, various notions of argument extensions are defined, which aim to capture various types of defeasible consequence. Then it is shown that many existing nonmonotonic logics can be reformulated as instances of the abstract framework.

The following version of this framework is kept in the abstract style of [Dung 95], with some adjustments proposed in [Prakken & Sartor 96a]. Important differences will be indicated when relevant.

Definition 1. An argument–based theory (AT) is a pair $(Args, defeat)$, where $Args$ is a set of arguments, and $defeat$ a binary relation on $Args$.

- An AT is *finitary* iff each argument in $Args$ is defeated by at most a finite number of arguments in $Args$.
- An argument A *strictly defeats* an argument B iff A defeats B and B does not defeat A.
- A set of arguments is *conflict–free* iff no argument in the set is defeated by another argument in the set.

The idea is that an AT is defined by some nonmonotonic logic or system for defeasible argumentation. Usually the set $Args$ will be all arguments that can be constructed in these logics from a given set of premises. In this paper I will (almost) completely abstract from the source of an AT. Moreover, unless stated otherwise, I will below implicitly assume an arbitrary but fixed AT.

The relation of $defeat$ is intended to be a weak notion: intuitively 'A defeats B' means that A and B are in conflict and that A is not worse than B. This means that two arguments can defeat each other. A typical example is the Nixon Diamond, with two defaults 'Quakers are pacifists' and 'Republicans are not pacifists', and with the facts that Nixon was a quaker and a republican. If there are no grounds for preferring one default over the other, the two arguments that Nixon was, and was not a pacifist, defeat each other.

A stonger notion is captured by strict defeat (not used in Dung's work), which by definition is asymmetric. A standard example is the Tweety Triangle, where (if arguments are compared with specificity) the argument that Tweety flies because it is a bird is strictly defeated by the argument that Tweety doesn't fly since it is a penguin.

A central notion of Dung's framework is acceptability. Intuitively, it defines how an argument that can possibly not defend itself, can be protected from attacks by a set of arguments. Since [Prakken & Sartor 96a, Prakken & Sartor 96c], on which this paper's proof theory is based, use a slightly different notion of acceptability, I will tag Dung's version with a d.

Definition 2. An argument A is *d–acceptable* with respect to a set S of arguments iff each argument defeating A is defeated by some argument in S.

The other variant will just be called 'acceptability'.

Definition 3. An argument A is *acceptable* with respect to a set S of arguments iff each argument defeating A is strictly defeated by some argument in S.

So the only difference is that Dung uses 'defeat' where we use 'strict defeat'. In Section 4.1 I will comment on the significance of this difference.

Another of Dung's notions is that of an admissible set.

Definition 4. A conflict–free set of arguments S is *admissible* iff each argument in S is d–acceptable with respect to S.

On the basis of these definitions several alternative notions of 'argument extensions' can be defined. For instance, Dung defines the following credulous notions.

Definition 5. A conflict–free set S is a *stable extension* iff every argument that is not in S, is defeated by some argument in S.

The Nixon Diamond has two stable extensions, one with the argument that Nixon was a pacifist, and one with the argument that he was not a pacifist. On the other hand, the Tweety Triangle has only one stable extension, with the argument that Tweety doesn't fly.

Since a stable extension is conflict–free, it reflects in some sense a coherent point of view. Moreover, it is a maximal point of view, in the sense that every possible argument is either accepted or rejected. The maximality requirement makes that not all AT's have stable extensions. Consider, for example, an AT with three arguments A, B and C, and such that A defeats B, B defeats C and C defeats A (such circular defeat relations can occur, for instance, in logic programming because of negation as failure, and in default logic because of the justification part of defaults.)

To give also such AT's a credulous semantics, Dung defines the notion of a preferred extension.

Definition 6. A conflict–free set is a *preferred extension* iff it is a maximal (with respect to set inclusion) admissible set.

Clearly all stable extensions are preferred extensions, so in the Nixon Diamond and the Tweety Triangle the two semantics coincide. However, not all preferred extensions are stable: in the above example with circular defeat relations the empty set is a (unique) preferred extension, which is not stable.

Preferred and stable semantics clearly capture a credulous notion of defeasible consequence: in cases of an irresolvable conflict as in the Nixon diamond, two, mutually conflicting extensions are obtained. Dung also defines a notion of skeptical consequence, and this is for which I will define the dialectical proof theory. Application of the proof theory to the credulous semantics will be briefly discussed in Section 5. Dung defines the skeptical semantics with a monotonic operator, which for each set S of arguments returns the set of all arguments d–acceptable to S. Its least fixpoint captures the smallest set which contains every argument that is acceptable to it. I will use the variant with plain acceptability.

Definition 7. Let $AT = (Args, defeat)$ be an argument–based theory and S any subset of $Args$. The *characteristic function* of AT is:

- $F_{AT} : Pow(Args) \longrightarrow Pow(Args)$
- $F_{AT}(S) = \{A \in Args | A \text{ is acceptable with respect to } S\}$

I now give the, perhaps more intuitive, definition of [Prakken & Sartor 96a], which for finitary AT's is equivalent to the fixpoint version (which is also used in [Prakken & Sartor 96c]). The formal results on the proof theory hold for both formulations, although for the fixpoint formulation completeness holds under the condition that the AT is finitary.

Definition 8. For any AT $= (Args, defeat)$ we define the following sequence of subsets of $Args$.

- $F^0_{AT} = \emptyset$
- $F^{i+1}_{AT} = \{A \in Args \mid A \text{ is acceptable with respect to } F^i_{AT}\}$.

Then the set $JustArgs_{AT}$ of arguments that are justified on the basis of AT is $\cup^{\infty}_{i=0}(F^i_{AT})$.

In this definition the notion of acceptability captures reinstatement of arguments: if all arguments that defeat A are themselves defeated by an argument in F^i, then A is in F^{i+1}.

In [Prakken & Sartor 96a, Prakken & Sartor 96c] it is shown that each set of justified arguments is conflict–free. These papers also contain examples illustrating the definition.

3 A dialectical theorem prover

3.1 General idea and illustrations

In this section a dialectical theorem prover will be defined for the just–presented keptical semantics. Essentially it is a notational variant of [Dung94]'s dialogue game version for his skeptical semantics as applied to extended logic programs. A proof of a formula will take the form of a dialogue tree, where each branch of the tree is a dialogue, and the root of the tree is an argument for the formula. The idea is that every move in a dialogue consists of an argument based

on an implicitly assumed AT, and that each stated argument attacks the last move of the opponent in a way that meets the player's burden of proof. That a move consists of a complete argument, means that the search for an individual argument is conducted in a 'monological' fashion, determined by the nature of the underlying logic; only the process of considering counterarguments is modelled dialectically. The required force of a move depends on who states it, and is motivated by the definition of acceptability. Since the proponent wants a conclusion to be justified, a proponent's move has to be strictly defeating, while since the opponent only wants to prevent the conclusion from being justified, an opponent's move may be just defeating.

Let us illustrate this with an informal example of a dialogue (recall that it implicitly assumes a given AT). Let us denote the arguments stated by the proponent by P_i and those of the opponent by O_i. The proponent starts the dispute by asserting that P_1 is a justified argument.

P_1: Assuming the evidence concerning the glove was not forged, it proves guilt of OJ.

Now the opponent has to defeat this argument. Supppose it can do so in only one way.

O_1: I know that the evidence concerning the glove was forged, so your assumption is not warranted.

The proponent now has to counterattack with an argument that strictly defeats O_1. Consider the following argument

P_2: The evidence concerning the glove was not forged, since it was found by a police officer, and police officers don't forge evidence.

and suppose (for the sake of illustration) that defeat is determined by specificity considerations. Then P_2 strictly defeats O_1, so P_2 is a possible move. If the opponent has no new moves available from $Args_{AT}$, s/he loses, and the conclusion that OJ is guilty has been proved.

In dialectical proof systems a 'loop checker' can be implemented in a very natural way: no two moves of the proponent in the same branch of the dialogue may have the same content. It is easy to see that this rule will not harm P; if O had a move the first time P stated the argument, it will also have a move the second time, so no repetition by P can make P win a dialogue.

Assume for illustration that the arguments in $Args$ are those that can be made by chaining one or more of the following premises:

(1) Mr. F forged the glove–evidence
(2) Someone who forges evidence is not honest
(3) Mr. F is a police officer
(4) Police officers are honest

(5) Someone who is honest, does not forge evidence.

Assume again that defeat is determined by specificity, in the obvious way. Now the proponent argues that Mr. F did not forge the glove–evidence.

P_1: Mr. F is a police officer, so he is honest and
therefore does not forge evidence.

O attacks this argument on its 'subconclusion' that Mr. F is honest; and since the counterargument is more specific, this is a defeating argument.

P_1: I know that F forged evidence, and this shows that he is not honest.

P now wants to attack O's argument in the same way as O attacked P's argument: by launching a more specific attack on O's 'subconclusion' that F forged the glove–evidence. However, P has already stated that argument at the beginning of the dispute, so the move is not allowed. And no other strictly defeating argument is available. So it is not provable that Mr. F did not forge the glove–evidence, not even that he is honest. However, by a completely symmetric line of reasoning we obtain that also the contrary conclusions are not provable. So no conclusion about whether Mr. F is honest or not, and forged evidence or not, is provably justified.

3.2 The proof theory

Now the dialectical proof theory will be formally defined. Again the definitions assume an arbitrary but fixed AT. Although thus the parties in a dispute are restricted to using rules from a given 'pool' of premises, this is just a theoretical restriction; the definitions equally apply if it is assumed that $Args_{AT}$ consists of every argument put forward by the players in a dialogue.

Definition 9. A *dialogue* is a finite nonempty sequence of moves $move_i = (Player_i, Arg_i)$ $(i > 0)$, such that

1. $Player_i = P$ iff i is odd; and $Player_i = O$ iff i is even;
2. If $Player_i = Player_j = P$ and $i \neq j$, then $Arg_i \neq Arg_j$;
3. If $Player_i = P$, then Arg_i strictly defeats Arg_{i-1};
4. If $Player_i = O$, then Arg_i defeats Arg_{i-1}.

The first condition says that the proponent begins and then the players take turns, while the second condition prevents the proponent from repeating its attacks. The last two conditions form the heart of the definition: they state the burdens of proof for P and O.

Definition 10. A *dialogue tree* is a finite tree of moves such that

1. Each branch is a dialogue;

2. If $Player_i = P$ then the children of $move_i$ are all defeaters of Arg_i.

The second condition of this definition makes dialogue trees candidates for being proofs: it says that the tree should consider all possible ways in which O can attack an argument of P.

Definition 11. A player wins a dialogue if the other player cannot move. And a player wins a dialogue tree iff it wins all branches of the tree.

The idea of this definition is that if P's last argument is undefeated, it reinstates all previous arguments of P that occur in the same branch of a tree, in particular the root of the tree.

Definition 12. An argument A is *provably justified* iff there is a dialogue tree with A as its root, and won by the proponent.

In [Prakken & Sartor 96c] it is shown that this proof theory is sound and for finitary AT's also complete with respect to the skeptical fixpoint semantics. This is not surprising, since what the proof theory does is, basically, traversing the sequence defined by Definition 8 in the reverse direction. Note that it implies that an argument A is justified iff there is a sequence $F^1, \ldots, F^n$ such that A occurs for the first time in F^n (in the explicit fixpoint definition of [Dung 95, Prakken & Sartor 96c] this only holds for finitary AT's; in the general case only the 'if' part holds). We start with A, and then for any argument B defeating A we find an argument C in F^{n-1} that strictly defeats B and so indirectly supports A. Then any argument defeating C is met with a strict defeater from F^{n-2}, and so on. Since the sequence is finite, we end with an argument indirectly supporting A that cannot be defeated.

4 Defeasible priorities

In several argumentation frameworks, as in many other nonmonotonic logics, the defeat relation is partly defined with the help of priorities on the premises. In most systems these priorities are undisputable and assumed consistent. However, as discussed in e.g. [Gordon 95, Prakken & Sartor 96b], these features are often unrealistic. In several domains of practical reasoning, such as legal reasoning, the standards for conflict resolution are themselves subject to debate and disagreement, and therefore a full theory of defeasible argumentation should also be able to formalise arguments about priorities, and to adjudicate between such arguments.

This section presents a formalisation of this feature, which forms the main technical addition to [Dung 93b, Dung94]. In [Prakken & Sartor 96a] the semantics of [Dung 93b] is revised, and the same is done in [Prakken & Sartor 96b, Prakken & Sartor 96c] with the proof theory of [Dung94], also presented in the previous section. Here these revisions are generalised to any system fitting the format of [Dung 95].

However the generalisation is only well–defined if the logic generating an AT satisfies some additional assumptions. Firstly, for each AT I assume as given an ordering $<$ on the premises from which the arguments of the AT can be constructed; properties of $<$ can be assumed if necessary. Then I assume that the defeat relation is determined with the help of these priorities, i.e. that a notion *A defeats B on the basis of* $<$ has been defined. Finally, I assume that the language in which arguments can be expressed contains a distinguished twoplace predicate symbol $\prec$, intended to denote the relation $<$.

4.1 Changing the semantics

Now how can we make the priorities that are needed to determine defeat, defeasible consequences of the AT, according to Definition 8? The idea is that in determining whether an argument is acceptable with respect to F^i_{AT}, we look at those priority statements that are conclusions of arguments in F^i_{AT}. Formally (for any set of arguments):

Definition 13. For any set S of arguments

$$<_S = \{r < r' \mid name\_of\_r \prec name\_of\_r' \text{ is a conclusion of some } A \in S\}$$

I will abbreviate 'A defeats B on the basis of S' as 'A S–defeats B'. For singleton sets $\{C\}$ I will write '$\{C\}$–defeats' as 'C–defeats'.

For arbitrary sets S it is not guaranteed that $<_S$ has the desired properties, for instance, in [Prakken & Sartor 96a] those of a strict partial order. However, it is sufficient that the properties hold for each $<_{F^i_{AT}}$. In virtually any nonmonotonic logic this can be assured by including the axioms of a strict partial order for $\prec$ in the undebatable part of the premises.

I now redefine the notion of acceptability as follows (d–acceptability can be changed in the same way).

Definition 14. An argument A is *acceptable* with respect to a set S of arguments iff all arguments S–defeating A are strictly S–defeated by some argument in S.

Note that this definition not only changes Dung's definition (by using strict defeat), but also refines it: Dung does not consider defeasible priorities and therefore does not make defeat relative to sets of arguments.

Definition 8 can now be applied with Definition 14. The following two properties are crucial in proving that each F^i is contained in F^{i+1}, which guarantees that each set of justified arguments is conflict–free. They are also crucial in proving that the explicit-fixpoint definition of [Prakken & Sartor 96c] is monotonic.

Property 4.1 *For any two conflict–free sets of arguments S and S' such that $S \subseteq S'$, and any two arguments A and B we have that*

1. *If A S'–defeats B, then A S–defeats B.*

2. If A strictly S–defeats B, then A strictly S'–defeats B.

Property 4.2 *For any conflict–free set of arguments S and arguments A and B: A strictly S– defeats B, if and only if there is a minimal (w.r.t. set inclusion) $X \subseteq S$ such that A strictly X–defeats and strictly $A + X$–defeats B.*

Given our weak interpretation of *defeat* these properties seem very reasonable and easy to obtain: the idea is to define *A defeats B on the basis of* $<$ in terms of the absence of priorities in $<$ that would make A worse than B; then adding more priorities to $<$ can only make defeat relations go away. Therefore I will below assume that these properties hold for any argument–based input theory.

I can now comment on the use of strict defeat in Definitions 3 and 14: Property 4.1(2) does not hold for defeat, yet it is essential to make Definition 8 well–behaved in case of defeasible priorities.

4.2 Changing the proof theory

Now I change the proof theory. The main problem here is on the basis of which priorities the defeating force of the moves should be determined. What is to be avoided is that we have to generate all priority arguments before we can determine the defeating force of a move. The pleasant surprise is that, to achieve this, a few very simple conditions suffice. For O it is sufficient that its move $\emptyset$–defeats P's previous move. This is so since Property 4.1 implies that if A is for some S an S–defeater of P's previous move, it is also an $\emptyset$–defeater of that move. So O does not have to take priorities into account. Let us illustrate this by modifying our glove dialogue as follows (we again leave it to the readers to formalise the arguments in their favourite formalism). Again the proponent starts with

P_1: Assuming the evidence concerning the glove was not forged,
it proves guilt of OJ.

Suppose the opponent now replies with

O_1: I know that the evidence concerning the glove was forged,
since I was told so, so your assumption is not warranted.

In agreement with most nonmonotonic logics, I assume that an attack on an assumption succeeds if no priority relations hold: i.e. O_1 $\emptyset$–defeats P_1.

P, on the other hand, should take some priorities into account, since strict defeat usually requires 'better than' relations between rules. However, it suffices to apply only those priorities that are stated by P's move; more priorities are not needed, since Property 4.1 also implies that if P's argument Arg_i strictly $Args_i$– defeats O's previous move, it will also do so whatever more priorities will be derived. So P can reply to O_1 with

P_2: The evidence concerning the glove was not forged, since it was found by a police officer, and as a general rule police officers don't forge evidence. This rule is more reliable than your rule that what what you are told is true.

Because of the priority statement at the end, P_2 strictly P_2–defeats O_1. However, this is not the only type of move that the proponent should be allowed to make. To see this, note that O can respond with repeating O_1 as O_2, at least assuming that O_1 $\emptyset$–defeats P_2, which in many systems it will do (e.g. in [Prakken & Sartor 96a]).

$O_2 = O_1$

And because of the nonrepetition rule P cannot respond to O_2 with $P_3 = P_2$. So P must be allowed to state a priority argument that neutralises the defeating force of O_2, i.e. it is OK if P_3 is such that O_2 does not P_3–defeat P_1. If P is allowed to make such a move, it can repeat the priority part of P_2:

P_3: The rule that police officers don't forge evidence is more reliable than your rule that what you are told is true.

Of course, O might challenge P's priority argument, for instance, by saying that instead the 'what I am told is true' rule is more reliable since O only listens to very reliable people. However, I will end the discussion of our example and describe the changes of the proof theory. All we have to change is the burdens of proof in Definition 9:

(3) If $Player_i = P$ *then*

- Arg_i *strictly* Arg_i*–defeats* Arg_{i-1}*; or*
- Arg_{i-1} *does not* Arg_i*–defeat* A_{i-2}*.*

(4) If $Player_i = O$ *then* Arg_i $\emptyset$*–defeats* Arg_{i-1}*.*

The other definitions stay the same.

In [Prakken & Sartor 96c] it is shown that the proof theory is, with respect to the fixpoint semantics, sound in the general case and complete for finitary AT's. The corresponding results for the system with fixed priorities are proven as a special case. Although these results are proven for a particular system, the proofs use no more than the properties assumed above.

5 Proof theory for credulous semantics

In this section I sketch how a dialectical proof theory can be developed for the credulous semantics discussed in Section 2. I will first focus on the case with fixed priorities. Defining a proof theory for stable semantics will not be easy, since we

always have to prove that a stable extension exists. Therefore I concentrate on preferred semantics. This is also relevant for stable semantics, since [Dung 93b] identifies conditions under which preferred and stable semantics coincide.

Note first that the existence of a proof means that the argument is in *some* preferred extension. Now the idea is to reverse the burden of proof of P and O. P now only has to defeat O's arguments, while O now must strictly defeat P's moves. Moreover, the non–repetition rule now holds for O instead of for P, while the children of P's moves are now all its *strict* defeaters. Finally, since preferred extensions are conflict–free, we must require that in each dialogue the set of all moves of the proponent is conflict-free.

With respect to soundness and completeness, it is relevant that every admissible set is contained in some preferred extension. Then soundness follows since it is easy to see that the union of all P's arguments in a dialogue tree is an admissible set. Completeness can be proven for the finite case, by showing that each finite admissible set corresponds to a proof for each of its members. For the infinite case there are obvious counterexamples. Consider e.g. an infinite set of arguments $\{A_1, \ldots, A_n, \ldots\}$, where each $A_i (i > 1)$ strictly defeats A_{i-1}: both the set of all 'odd', and that of all 'even' arguments are preferred extensions, but any 'proof' has to be infinite.

Extending these ideas to the case with defeasible priorities is still to be investigated.

Finally, we could also turn things around: instead of starting with the semantics and finding a corresponding proof theory, we might start with the proof theory, i.e. state some reasonable properties on the dialectical protocol, and then ask which semantics they generate. For instance, in the proof theory of Section 3 we might require of O's moves that that they are not strictly defeated by any previous move of P in the same dialogue. Assume by way of illustration an AT with $Args = \{A, B, C, D\}$, A strictly defeats C, B strictly defeats D, and both A and D and B and C defeat each other. Then $JustArgs_{AT}$ is empty, but might be argued that it should consist of A and B, as has been done by [Horty et al. 90] for similar examples in the context of defeasible inheritance.

The change in the proof theory gives this result. Perhaps the corresponding semantics is what [Dung 95, Bondarenko et al. 95] call a *complete extension*, which is any conflict–free fixpoint of F_{AT}, not just its least one. Such extensions capture the arguments that are justified if first some arguments are assumed to be justified.

6 Concluding remarks

This paper has discussed three contributions to the formalisation of defeasible argumentation. Firstly, I have discussed how the abstract framework of [Dung 95, Bondarenko et al. 95] can be extended with defeasible priorities. Secondly, I have, by generalising work of [Dung94], discussed how dialectical proof theories can be defined for this framework and its extension. Finally, I have given

an impression of the research questions that arise in the dialectical approach to the proof theory of defeasible argumentation.

As for future research, first of all the preliminary contributions of this paper should, of course, be further developed. Moreover, it would be interesting to investigate the relation between dialectical proof theories and dialectical protocols for disputation as defined by e.g. [Loui & Norman 95, Gordon 95]. The leading idea of such protocols is that rationality has a procedural side: an argument is acceptable if it has been successfully defended against criticism in a properly conducted dispute. The main aim of this line of research is to find out what makes a dispute proper, i.e. what makes it fair and effective.

Perhaps our soundness and completeness results are part of the criteria for fair and effective disputation. This is at least how [Vreeswijk 96] defines fairness and effectiveness: a protocol is fair iff every argument that can be successfully defended against every attack is justified, and it is effective iff every justified argument can be successfully defended against every attack.

References

[Barth & Krabbe 82] E.M. Barth and E.C.W. Krabbe. *From Axiom to Diaolgue: a Philosophical Study of Logic and Argumentation.* Walter de Gruyter, New York, 1982.

[Bondarenko et al. 95] A. Bondarenko, P.M. Dung, R.A. Kowalski and F. Toni. An abstract argumentation–theoretic approach to default reasoning. Available by anonymous ftp at laotzu.doc.ic.ac.uk/public/papers/Kowalski/arg-def.ps.Z

[Brewka 94] G. Brewka. A reconstruction of Rescher's theory of formal disputation based on default logic. *Proceedings of the 11th European Conference on Artificial Intelligence*, 366-370.

[Dung 93b] P.M. Dung. An argumentation semantics for logic programming with explicit negation. *Proceedings of the Tenth Logic Programming Conference*, MIT Press 1993, 616–630.

[Dung94] P.M. Dung. Logic programming as dialogue game. Unpublished paper.

[Dung 95] P.M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming, and n–person games. *Artificial Intelligence* 77 (1995), 321–357.

[Gordon 95] T.F. Gordon. *The Pleadings Game. An Artificial Intelligence Model of Procedural Justice.* Kluwer, 1995.

[Horty et al. 90] J.F. Horty, R.H. Thomasson and D.S. Touretzky. A sceptical theory of inheritance in nonmonotonic semantic networks. *Artificial Intelligence* 42 (1990), 311–348.

[Loui 93] R.P. Loui. Process and policy: resource–bounded non–demonstrative reasoning. Report WUCS–92–43, Washington–University–in–St–Louis, 1993. To appear in *Computational Intelligence.*

[Loui & Norman 95] R.P. Loui and J. Norman. Rationales and argument moves. *Artificial Intelligence and Law* 3: 159–189, 1995.

[Pollock 87] J.L. Pollock. Defeasible reasoning. *Cognitive Science* 11 (1987), 481–518.

[Prakken & Sartor 95] H. Prakken and G. Sartor. On the relation between legal language and legal argument: assumptions, applicability and dynamic priorities. *Proceedings of the Fifth International Conference on Artificial Intelligence and Law.* ACM Press 1995, 1–9.

[Prakken & Sartor 96a] H. Prakken and G. Sartor. A system for defeasible argumentation, with defeasible priorities. To appear in the *Proceedings of the International Conference on Formal Aspects of Practical Reasoning*, Bonn 1996. Springer Lecture Notes in AI, Springer Verlag, 1996.

[Prakken & Sartor 96b] H. Prakken and G. Sartor. Rules about rules: assessing conflicting arguments in legal reasoning. To appear in *Artificial Intelligence and Law*.

[Prakken & Sartor 96c] H. Prakken and G. Sartor. Argument–based extended logic programming with defeasible priorities. Submitted.

[Rescher 1977] N. Rescher. *Dialectics: a controversy–oriented approach to the theory of knowledge*. State University of New York Press, Albany, 1977.

[Royakkers & Dignum 1996] L. Royakkers and F. Dignum. Defeasible reasoning with legal rules. In M.A. Brown and J. Carmo (eds.) *Deontic Logic, Agency and Normative Systems*. Springer, Workshops in Computing, London etc. 1996, 174–193.

[Simari & Loui 92] G.R. Simari and R.P. Loui. A mathematical treatment of defeasible argumentation and its implementation. *Artificial Intelligence* 53 (1992), 125–157.

[Vreeswijk 93a] G. Vreeswijk. *Studies in defeasible argumentation*. Doctoral dissertation Free University Amsterdam, 1993.

[Vreeswijk 93b] G. Vreeswijk. Defeasible dialectics: a controversy–oriented approach towards defeasible argumentation. *Journal of Logic and Computation*, 1993, Vol.3, No.3., 317–334.

[Vreeswijk 96] G. Vreeswijk. Representation of formal dispute with a standing order. *Research Report MATRIX, University of Limburg, 1996.*

This article was processed using the LaTeX macro package with LLNCS style