A logical framework for modelling legal argument

Henry Prakken
Computer/Law Institute
*Free University Amsterdam**
The Netherlands

Abstract

This paper investigates the relevance of the logical study of argumentation systems for AI-and-law research, in particular for modelling the adversarial aspect of legal reasoning. It does so in applying the argumentation framework of Prakken (1993a/b) to the legal domain. Three elements of the framework are particularly illustrated: firstly, its generality, in that it leaves room for any standard for comparing pairs of arguments; secondly, its ability to model the combined use of these standards; and finally, its relevance for modelling metalevel reasoning. These three features make the framework suitable as a logical framework for any theory of legal argument.

1 Introduction

This paper investigates the relevance for AI-and-law research of a new development in logic, the formal study of argumentation as pursued by e.g. Pollock (1987), Prakken (1991a/b, 1993a/b), Vreeswijk (1991) and Simari and Loui (1992). The main feature of this development is the modelling of inconsistency-tolerant and defeasible reasoning as constructing and comparing arguments for incompatible conclusions. This emphasis on the notion of an argument makes the resulting theories particularly relevant for one of the main issues in AI-and-law research, the modelling of the adversarial aspect of legal reasoning, i.e. of legal reasoning as a rule-guided rather than a rule-governed activity (Gardner, 1987; Rissland and Ashley, 1989; Skalak and Rissland, 1991; Gordon, 1991). The claim of the present paper is that the logical argumentation theories can provide formal foundations for these developments, since they deal with questions like: how is an argument formally defined, what is the role of premise-orderings in comparing arguments, how does the status of an argument affect the status of other arguments, and so on. This claim will be illustrated by applying the argumentation framework of Prakken (1993a/b) to legal reasoning.

A first version of my framework was presented in Prakken (1991a). However, in that paper the focus was on one particular way of comparing conflicting arguments, viz. with Poole's (1985) specificity principle, in an attempt to model reasoning with implicit exceptions; the general ideas behind the theory as being an argumentation framework were underemphasized. Since then I have become aware that the general structure of the framework is more

* The current address of the author is Imperial College, Dept. of Computing, 180 Queen's Gate, London SW7 2BZ, UK. E-mail: hp3@doc.ic.ac.uk.
The research reported in this paper was supported by a grant from the Legal Research Foundation, which is part of the Netherlands Foundation for Scientific Research.

important than the particular 'kind of defeat' which is used; therefore the present paper focuses on the flexibility of that general structure. In particular, I will investigate the combined use of several criteria for comparing arguments, and the relevance of the framework for metalevel reasoning.

After outlining the overall structure of my argumentation framework (Section 2) I will discuss some possible legal standards for comparing arguments and how their use can be formalized within my framework (Section 3). Then I study the combined use of these standards (Section 4) and the relevance of the framework for metalevel reasoning (Section 5), after which I end with a discussion of related research (Section 6).

2 Constructing and comparing arguments in default logic

2.1 Main ideas

The investigations of Prakken (1993a/b) are a further development of earlier investigations of reasoning with inconsistent information by, among others, Alchourrón and Makinson (1981), Poole (1985,1988) and Brewka (1989). The general idea of all these approaches, perhaps expressed in their as yet most optimal form by Sartor (1992), is to deal with inconsistency and disagreement by focusing on *consistent subsets* of the available information and comparing these sets with respect to some appropriate ordering relation. Although as a general idea this is feasible, the claim of Prakken (1993a/b) is that in defining such a comparison a further notion is indispensable, that of an *argument*. Most importantly, the fact should be reflected that arguments are compared as they are constructed, viz. step-by-step, proceeding from intermediate to final conclusions. A second claim is that classical logic is less suitable as the language which underlies inconsistency-tolerant reasoning than is often claimed, since the material implication is logically too strong: particularly the validity of contraposition makes arguments possible which intuitively cannot be constructed. I have shown that Reiter's default logic is a better knowledge representation language, but I do not exclude that other logics with a defeasible conditional are even more suitable.

To give a brief outline of default logic, it is based on a set F of *facts*, expressed in the language of first-order logic and assumed consistent, and a set Δ of *defaults*, which are inference rules of the form $\phi:\psi/\chi$, in which ϕ is the *prerequisite*, ψ the *justification*, and χ the *consequent*. Informally, this reads as 'If ϕ holds and ψ may be consistently assumed, χ may be inferred'. New beliefs can be derived by using ground instances of any default of Δ one wishes, as long as consistency is preserved. If as many defaults as possible are thus used, i.e. if applying any new default would cause an inconsistency, sets result which are called *extensions* of (F,Δ). Since defaults can conflict, a default theory may have several, mutually inconsistent, extensions.

Now in my framework the idea is to represent the facts of the case at hand as elements of F, together with necessary truths such as 'a man is a person' or 'a lease contract is a contract', and to formalize defeasible rules as normal defaults $\phi:\psi/\psi$, in this paper written as $\phi \Rightarrow \psi$. Unconditional defeasible rules can be represented as defaults of the form $\Rightarrow \phi$, which is shorthand for $\top \Rightarrow \phi$, where $\top$ stands for any valid formula.

2.2 Formalization

In this subsection I will present a simplified version of my argumentation framework, also introducing new terms for some of the notions of Prakken (1993a/b). For the present purposes it is particularly important that the framework is of a general nature, in that it allows for any standard for comparing pairs of arguments. It consists of four parts: the first concerns the notion of an argument, the second says when arguments are in conflict, the third is about ways of comparing arguments and the final part defines what it means that an argument is justified. It is important to note that all definitions below are given at the background of a fixed default theory (F,Δ).

To start with arguments, they are divided into a set of facts and a set of defaults, i.e. they are default theories. The facts are the set F of the background default theory (F,Δ), while the defaults are a subset of the set of ground instances of Δ. Not every subset of Δ will do, but only sets which result in a 'coherent' argument, i.e. one with a unique extension. Furthermore, all defaults of an argument are required to be applicable. Finally, if a formula ϕ is in the extension of a certain argument A, A is said to explain ϕ. Since all defaults of an argument are applicable, this is the case if and only if ϕ is deductively implied by the facts and/or the joint consequents of the defaults of the argument.

Definition 2.1: **a.** A = (F,D) (where D is a finite subset of ground instances of Δ) is an *argument* iff it has a unique extension E(A) such that of all elements of D both the prerequisites and the consequents are in E(A).

b. A *explains* a formula ϕ iff ϕ is in E(A). ϕ is then the *conclusion* of A.

c. An argument A' = (F,D') is a *subargument* of an argument A = (F,D) iff D' ⊂ D.

The second main notion of the framework is that of a conflict between arguments. This is defined in terms of the *final conclusions* of an argument, which are its conclusions not also explained by one of its subarguments.

Definition 2.2: An argument A' *attacks* an argument A iff A and A' explain contradictory final conclusions.

Note that Definition 2.2 implies that A attacks A' iff A' attacks A. Arguments attacking each other are also said to be *counterarguments* of each other, and if A' attacks A then A' is called an *attack* of A.

The third main element of the framework is a way of comparing arguments, which is defined as a comparison of *pairs* of arguments; the alternative is a comparison of sets of arguments, as in Vreeswijk (1991). An important feature of the definition is that, to capture the intended generality of the framework, it assumes the existence of some *unspecified* standard 'R-defeat' for comparing pairs of arguments, provided by the user of the framework; the only assumptions which will be made about R-defeat are that it is an asymmetric relation, i.e. no argument can R-defeat itself, and that there are no infinite sequences of arguments defeating each other. It is this part of the framework where, for example, Poole's specificity definition or a standard on the basis of premise-orderings can be applied (if there is no danger of confusion, the R will in the notation be omitted).

The final main element of the framework is the definition of a justified argument, i.e. of an argument which is better than any counterargument. In order to reflect the step-by-step nature of argumentation this notion is defined inductively: the idea is that in each inductive step arguments attacking each other are only compared with respect to their final conclusions, which is captured by clause (2) of the definition; intermediate conclusions should already have been justified at earlier steps in the induction, which is expressed by clause (1). A further idea captured by clause (2) is that an argument which is not itself better than a counterargument can still be saved ('reinstated') by another argument which *is* better than this counterargument.

Definition 2.3: *(justified arguments)*. An argument A = (F,D) is an R-*justified argument* iff

(1) All subarguments of A are R-justified arguments;

(2) A R-defeats all attacks A' for which it holds that neither A' nor one of its subarguments is R-defeated by another R-justified argument.

In fact, this definition is not the one of Prakken (1993a/b), but the more intuitive definition of Prakken (1991a/b), which, as already explained there, is circular. In Prakken (1993a/b) the circularity is avoided by rewriting the definition as conditions on *sets* of arguments and defining an argument to be R-justified iff it is in the *smallest* set satisfying these conditions. However, in practical applications the new definition can without harm be read as its simpler version Definition 2.3, which is what I will do in the present paper.

A very important aspect of Definition 2.3 is that it divides arguments into three classes. The first class is, of course, that of *justified* arguments. Furthermore, if there are arguments which defeat other arguments, there are, of course, also arguments which are *defeated*; formally, they are defined as the arguments which are attacked by a justified argument. Finally, Definition 2.3 leaves room for an interesting, non-empty class of arguments which are neither justified, nor defeated, but merely *defensible*. Consider two arguments A and B attacking each other and not attacked by any other argument, and assume that neither of the arguments R-defeats the other. Then neither of them is justified since they do not satisfy condition (2), but then also neither of them is defeated, since the only argument attacking them is not justified. The significance of such defensible arguments is that according to Definition 2.3 an argument needs not itself be justified in order to prevent a counterargument from being justified; it needs merely be defensible.

Finally, Definition 2.3 makes it possible to define the *defeasible consequences* of a default theory (F,Δ) plus a kind of defeat R: these are the conclusions for which there is an argument which according to R is better than any competing argument. This set, denoted by DC(F,Δ,R), can simply be defined as the set of all formulas explained by an R-justified argument (F,D) such that D ⊆ Δ. A natural requirement for a theory of comparing arguments is that if a formula is deductively implied by justified formulas, it is also itself justified. In my framework this depends on the kind of defeat which is used.

3 Kinds of defeat

In this section I discuss how the 'R' of Definition 2.3 can be given 'flesh and blood'. First, in 3.1, I give a brief overview of some possible legal standards for comparing arguments and then, in 3.2, I give a technique of formalizing most of these standards within my framework.

3.1 Legal standards for comparing arguments

In the legal domain many standards for comparing arguments can be found. First of all, there are, at least in continental legal systems, the three general conflict resolution metaprinciples based on hierarchy (Lex Superior), specificity (Lex Specialis) and the time of enactment of a provision (Lex Posterior). In addition, many regulations contain special rules about conflicts between particular classes of rules. For example, Section 1624 of the Dutch civil code (BW) states that if a contract has features of both lease of business accommodation and another contract, and a regulation concerning the other contract type conflicts with a regulation concerning lease of business accommodation, the latter prevails. 1637c BW is a similar provision for labour contracts. Finally, all kinds of domain specific standards can be used, based, for example, on legal principles or on the purpose of a rule.

Most of these standards result in orderings on individual premises; the only exception which I know of is specificity, but in order not to complicate the story too much, I will in this paper adhere to the often-used method of expressing specificity as an ordering on individual premises, which in most applications gives satisfactory results. Now, in the next subsection I will present a formalization in my framework of reasoning with ordered premises.

3.2 Orderings of premises

In order to apply the framework to reasoning with hierarchically ordered defaults I have developed in Prakken (1991b) a definition of 'hierarchical defeat'. The set of defaults Δ is assumed to be ordered by a partial preorder, i.e. a transitive and reflexive relation. 'x ≤ y' stands for 'x is at least as low as y'; 'x ≈ y' is shorthand for 'x ≤ y and y ≤ x'; and 'x < y' abbreviates 'x ≤ y and not y ≤ x'. The definitions should take care of two things: they should identify the defaults which are *relevant* to a conflict, and they should tell how to *compare* the relevant sets of conflicting arguments. The relevant defaults are identified according to the following idea. Informally, since intermediate conclusions are irrelevant to a conflict between arguments, we would like to identify exactly the defaults which are 'at the end of the argument chain' for the conclusion at stake; intermediate conclusions are then dealt with by the inductive part of Definition 2.3. Consider any argument A = (F,D). Then formally the relevant defaults, which I will call the *top rules* of A, are all defaults d ∈ D such that (F,D-{d}) explains the prerequisite of d. The set of top-rules of an argument A for ϕ will be denoted by $[\phi]$A. Consider by way of illustration F = {a}, d1 = a ⇒ b, d2 = a ⇒ c and the arguments A = (F,{d1}) and B = (F,{d1,d2}): then [b]A = {d1} and [b ∧ c]B = {d1,d2}.

Now an argument A hierarchically defeats an argument A' iff for all formulas ϕ about which they are in conflict it holds that all members of $[\phi]$A are higher than *all* members of $[\neg\phi]$A'.

Definition 3.1: An argument A *H-defeats* an argument A' iff for all final conclusions ϕ of A and $\neg\phi$ of A' it holds that d < d' for every d ∈ $[\phi]$A and d' ∈ $[\neg\phi]$A'.

In Prakken (1993a/b) it is shown that for H-defeat the set DC(F,Δ,H) of defeasible consequences is deductively closed.

3.3 Applications

Example 3.2 Consider by way of illustration an example about a provision (5 GW) of the Dutch constitution declaring every person to have the right to submit a written request to the proper authority. Imagine two case law decisions, one of a lower court, saying that a request which is not sent by ordinary mail is not a written request and one of a higher court, stating that a request by fax is a written request. Imagine further an act stating that prisoners do not have the right to submit requests to any authority. I will leave the ordering relation between case law decisions and legislation undefined. Finally, I will, in order to make the example closer to most legal systems, assume that the Dutch constitution is higher than statutes; in fact in Dutch law their relation is more complicated.

1 ¬By-mail ⇒ ¬Written
2 By-fax ⇒ Written
3 Prisoner ⇒ ¬May-request-authority
4 Written ⇒ May-request-authority
F = {By-fax, ¬By-mail, Prisoner}; H: 1 < 2, 3 < 4.

Let A1 = (F,{1,3}), A1' = (F,{1}), A2 = (F,{2,4}) and A2' = (F,{2}). Then if we check clause (2) of Definition 2.3, we see with Definition 3.1 that A2 H-defeats A1, since [¬May-request-authority]A1 = {3}, [May-request-authority]A2 = {4} and 3 < 4. Furthermore, according to clause (1) of Definition 2.3 we have to check whether all subarguments of A2 are justified; A2' is the only subargument of A2 which is attacked, viz. by A1', and A2' H-defeats A1' since [Written]A2' = {2}, [¬Written]A1' = {1} and 1 < 2. In conclusion, A2 is a justified argument.

A nice property of this formalization is that it avoids a problem of many other approaches to reasoning with priorities, viz. the need to express priorities between defaults which intuitively have nothing to do with each other, i.e. between (1) and (2) on the one hand and (3) and (4) on the other. Particularly if there are no general criteria for assigning the priorities, as holds for case law versus statutory rules, the formalization process may become complicated. In my formalization this problem does not occur: the reason is that the comparison of A1 and A2 is made step-by-step, i.e. the conflict about their intermediate conclusions is dealt with before the conflict about their final conclusions, and what is essential is that in doing so at each step only the defaults relevant to the conclusion *of that step* are taken into account.

Example 3.3: *spotting issues*. The relevance of the framework for AI-and-law research can also be illustrated by giving, in the same spirit as Gordon (1991), a logical analysis of Gardner's (1987) program for spotting legal issues. To give a very brief summary of the program, its task is to distinguish hard from easy questions. Besides the input facts, Gardner distinguishes three kinds of information: legal rules, common-sense rules and previously decided cases. Conflicts are dealt with in the following way. If two cases conflict then the problem is reported as hard, but if the conflict is between a case and a common-sense rule, then it is regarded as easy, in that the case prevails. Formalizing this in my framework is rather simple: all three kinds of legal information are defaults, and cases and common sense-rules are ordered by a simple ordering saying that the first are strictly higher than the latter. Since in Gardner's program cases and common-sense rules only serve to provide antecedents of legal rules, in my framework their relation to the latter needs not be defined (cf. Example 3.2). Now in terms of my framework the second kind of a hard question is the existence of two H-defensible arguments for opposite conclusions, where H refers to the ordering on cases and common-sense rules.

4 Combining kinds of defeat

Once various individual kinds of defeat have been investigated, the question naturally arises as to how they are combined in comparing arguments. This is the topic of the present section. Again I will first discuss some issues from legal theory, before I propose a possible formalization within my framework.

4.1 Requirements for formalizing multiple defeat

An important observation from legal theory is that standards for comparing arguments are applied in their order of importance. For example, the general metaprinciples seem to be used in the following way: first it is checked whether an argument is justified according to Lex Superior and only if this gives no solution, another principle is applied. Which principle exactly comes next, is not entirely undisputed: most legal scholars seem to give Lex Posterior precedence over Lex Specialis, but sometimes the reverse is argued for, e.g. by Sartor (1992). It might even be defended that there is no precedence relation at all between these two principles. For these reasons a formal theory of combined defeat should ideally leave room for equal and undefined relations between kinds of defeat. Nevertheless, because of space limitations this paper will only discuss the case of a linear order; for the full theory the reader is referred to Prakken (1993a).

4.2 Formalization

I will now sketch a formalization of multiple defeat.

I will assume a set $\mathfrak{R}$ of kinds of defeat, which are denoted by capitals R, R',... or by a more specific name, like H for hierarchy, T for Temporality and S for specificity. $\mathfrak{R}$ is assumed to be ordered by a linear order $\prec$, i.e. for all distinct elements R and R' of $\mathfrak{R}$ it holds that $R \prec R'$ or $R' \prec R$. Finally, individual default-orderings will sometimes be indexed by their name.

The next simplified version of the theory developed in Prakken (1993a) suffices to illustrate the relevance to legal reasoning. The idea of this version is to combine individual premise-orderings into one 'overall' ordering on the premises, which can then be applied in the way defined in Section 3. This is similar to the proposal of Sartor (1992). One important aspect of the full theory which will be left undiscussed is the inclusion of specificity as an ordering on arguments instead of on premises, which inclusion avoids some problems concerning transitivity (cf. Prakken, 1993a:170-1).

The construction of a combined default-ordering, which will be denoted by O, is based on the following idea. Consider two defaults d1 and d2. If in the highest kind of defeat R it holds that $d1 <_R d2$ or that R is for d1 and d2 undefined, then this also holds in the overall ordering; if instead $d1 \approx_R d2$, then the next highest kind of defeat is inspected to look for a '<' or undefined relation between d1 and d2, and so on; only if the relation of equality between d1 and d2 holds in all kinds of defeat, then this relation also holds in O.

These considerations are summarized in the following definition, in which in expressions of the form $d \circledast_R d'$ the symbol $\circledast$ is a variable which can have the values $<$, $\leq$ and $\approx$.

Definition 4.1: *(overall ordering on defaults O)*. Let (F,Δ) be a default theory (F,Δ) and $\mathfrak{R}$ a finite set of kinds of defeat which are based on preorders of Δ. Let $\mathfrak{R}$ be ordered by a linear order $\prec$. Then for all $d,d' \in \Delta$: $d \circledast_O d'$ iff there is an $R \in \mathfrak{R}$ such that

- $d \circledast_R d'$ and
- for all $R' \in \mathfrak{R}$ such that $R \prec R'$: $d \approx_{R'} d'$.

In Prakken (1993a/b) it is shown that the resulting ordering is well-behaved, in that it preserves the properties of reflexivity and transitivity of the individual orderings. Because of this result the definition of O-defeat is extremely simple: since what is constructed is an individual preorder on defaults, we can simply use the definition of H-defeat of individual hierarchies.

4.4 Applications

Example 4.2: The first example serves two aims. Firstly, it illustrates some technical aspects of the definition of combined defeat, in showing that even if a '<' relation between defaults does not result in defeat at a higher level, it can still influence the outcome at a lower level. Secondly, it suggests a plausible extension of Gardner's program, viz. reducing the number of hard questions by comparing cases with respect to specificity, as is also done in HYPO (Rissland and Ashley, 1989). First the relation between the two kinds of defeat must be defined. Assume that the hierarchy is regarded as higher than specificity and consider then the next variation on Example 3.2 ('csk' stands for 'common-sense knowledge').

case1: $\neg$By-snailmail $\Rightarrow$ $\neg$Written
case2: $\neg$By-snailmail $\wedge$ By-fax $\Rightarrow$ Written
csk: Electronic $\Rightarrow$ $\neg$Written
F: {$\neg$By-snailmail, By-fax, Electronic};

H: case1 $\approx$ case2 csk < case1 csk < case2
S: case1 < case2
$\mathfrak{R}$ = {H,S}; $S \prec H$.

According to Definition 4.1 this results in O being

csk < case1 csk < case2 case1 < case2

Let A1 = (F,{case1}), A2 = (F,{case2}) and A3 = (F,{csk}). Then according to Definition 3.1 (with O = H) O results in A2 O-defeating both A1 and A3, for which reason the question 'Written?' can be regarded as clear with a positive answer, while according to Gardner's original program it is hard. The technical point illustrated by this example is that if H would be applied individually, A3 would not be H-defeated, since it would be 'saved' by the conflict at H between A1 and A2. However, according to Definition 4.1 the '<' relations of H still have effect in the lower kind of defeat S.

Example 4.3 *The scope of kinds of defeat.* It might be argued that the Lex Posterior and Lex Specialis principle have a limited scope in that they only apply to conflicts between rules of the same authority or regulation; in this view it is impossible that, if two distinct authorities or regulations are of equal level, a rule of one of them is set aside by a later or more specific rule of the other one. Consider by way of illustration the Dutch civil regulations on contracts and on tort. It might be held that, although these are both statute regulations, they are so independent from each other that if they are in conflict, Lex Posterior and Lex Specialis are inapplicable. How can this view be formalized? My solution is to say that if for two particular distinct regulations these principles do not apply to such conflicts, these regulations are not of equal level but hierarchically incomparable. Technically, this will have the effect that the lower kinds of defeat become inapplicable.

I will illustrate this with three imaginary sections of the Dutch civil code, the first two about certain types of compensation in case of breach of contract, and the third one about compensation in case of tort. 'Comp' stands for 'Compensation' and $\veebar$ for exclusive disjunction.

d1 Breach-of-labour-contract ⇒ Comp1
d2 Breach-of-lease-contract ⇒ Comp2
d3 Tort ⇒ Comp3
F = {Breach-of-labour-contract, Breach-of-rent-contract,Tort},
(Comp1 $\veebar$ Comp2 $\veebar$ Comp3)}
$\mathfrak{R}$ = {H,T,S}; S ≺ T ≺ H.

With H = {d1 ≈ d2 ≈ d3} T and S can be applied to all conflicts, but in order to make T and S inapplicable to conflicts between d1 and d2 on the one hand and d3 on the other, H should only be defined for (d1,d2), i.e. as H = {d1 ≈ d2}: then S and T are only applicable to conflicts between d1 and d2.

5 Metalevel reasoning

As was said in Section 3, an important part of the present framework is the 'grey' area of defensible arguments: after all, all 'hard cases' are cases in which neither of the conflicting arguments is defeated according to the general or statutory legal metaprinciples. What lawyers should do in such conflicts is convince their opponents and the judiciary of their solution of the conflict. Such a discussion can be a very complicated argumentation process, involving all kinds of domain specific standards and probably also the use of precedents as studied by e.g. Rissland and Ashley (1989) and Skalak and Rissland (1991). A natural way to model this is as reasoning with information expressed declaratively at a metalevel (cf. Hamfelt and Barklund, 1989). Now this gives rise to an interesting application of the argumentation framework: the outcome of such a metalevel discussion can often be translated into an ordering on premises or arguments and then the framework can be used to calculate the resulting status of the various arguments.

The next example gives an illustration and also points at an issue for further research. Assume that the specific conflict resolution rules 1624 and 1637 BW, discussed in Section 3.1, are declaratively represented at a metalevel as

1624 BW: Buss.acc.contr-reg(x) ∧ Contr-reg(y) ∧ In-conflict(x,y) ⇒ y < x
1637c BW: Labour.contr-reg(x) ∧ Contr-reg(y) ∧ In-conflict(x,y) ⇒ y < x
F includes $\forall x \neg (x < y \wedge y < x)$

x and y are variables for defaults of Δ. The idea of this is that first with these metarules the ordering on the object level defaults is derived, after which the conflicts at the object level are dealt with as described in this paper.

An interesting problem is the treatment of conflicts at the metalevel: for example, do the general legal metaprinciples also apply to conflicts between specific statutory metarules? This question is not merely hypothetical, as the present example shows. Both metarules are about conflicts between regulations about contracts but, interestingly, they themselves are also regulations of this kind. Now assume there is a conflict between a rule r1 about labour contracts and a rule r2 about lease of business accommodation: then also 1624 and 1637c are in conflict, since their solutions of the conflict between r1 and r2 contradict each other: 1624 says r1 < r2 while 1637c says r2 < r1. To solve the conflict between 1624 and 1637c other metanorms might be invoked but the question is which ones apply. To make it even more complicated, in the present formalization both rules *themselves* apply to the conflict between them, since both are of the type required by their antecedents while, moreover, In-conflict(1624,1637c) is true! What is the best way of representing and reasoning with such rules or, more generally, what is the best way of modelling reasoning *about* priorities? This is an interesting topic for further research.

6 Related research

In recent years, the idea of regarding defeasible reasoning as constructing and comparing alternative arguments has also been developed by others, notably by Pollock (1987), Vreeswijk (1991) and Simari and Loui (1992). All these systems have the same general structure of the present framework, in that they define the notions of arguments, conflicts between arguments, standards for comparing them and justified arguments; the systems differ in their ways of formalizing these notions. For a detailed comparison the reader is referred to Prakken (1993a); it now suffices to say that all of these systems, including my own, both have their merits and their drawbacks, which calls for further logical research on combining the good points while avoiding the weak points. Desirable features of the present framework seem to be that it is defined for arbitrary standards for comparing arguments (as also Vreeswijk, 1991), that it is the only system formalizing the combined use of such standards, and that the assessment of arguments is three-valued, in that it leaves room for arguments which are neither justified nor defeated, but merely defensible. However, a better way to obtain threevaluedness seems to be the one of Vreeswijk's (1991): whereas Definition 2.3 always results in a *unique* set of justified arguments, Vreeswijk's definition allows in case of unsolvable conflicts for *alternative* sets of arguments so-called "in force" and their corresponding DC's, and he then defines the defensible conclusions as the formulas in some but not all DC's and the justified conclusions as the ones in all DC's: thus he can account for the justification of 'floating conclusions', i.e. of conclusions common to defensible arguments attacking each other (cf. Prakken, 1993a:156).

A theory which is not really a logical argumentation framework, but which is very relevant if it comes to *applying* such a system to the legal domain, is the "abductive theory of legal issues" of Gordon (1991). This theory is formally very similar to Poole's (1988) logical framework for default reasoning. Like Poole, Gordon is not interested in defining a new notion of logical consequence, but in demonstrating how logic can be *used* for analyzing various forms of reasoning, in Gordon's case "spotting legal issues". To this end he defines a number of notions like 'argument', 'rebuttal', 'counterargument', and 'issue', resulting in a definition of what a clear and a hard case is. At the logical level Gordon's theory lacks some essential features of my framework, particularly an inductive notion of constructing and comparing arguments and a defeasible conditional operator. However, more important is that his definitions can rather easily be adapted to an argumentation framework, which makes his work very relevant to any theory of modelling legal argument. Prakken (1993a) gives some examples of profiting from Gordon's investigations in this way.

7 Conclusion

This paper has presented and applied to legal reasoning a formal theory of constructing and comparing arguments. Three features have particularly been illustrated: its generality in that it allows for any standard of comparing pairs of arguments, its ability to model the combined use of these standards, and its relevance for metalevel reasoning. The claim of this paper is that these features support more sophisticated models of legal reasoning as rule-guided instead of rule-governed reasoning.

Future logical research on argumentation should aim at combining the technical merits of the various existing argumentation systems and on formalizing the combination of argumentation systems with systems for metalevel reasoning. Future AI-and-law research should clarify which standards for comparing arguments are used in legal reasoning, and in what way they are used. For example, how exactly are cases, legal principles or the purposes of rules used, and how are the various standards combined? Particularly a study of the use of precedents may contribute to an integrated theory of rule-based and case-based reasoning in law.

References

Alchourrón and Makinson 1981: C.E. Alchourrón and D. Makinson, Hierarchies of regulations and their logic. In R.Hilpinen (ed.): *New studies in deontic logic*, Reidel, Dordrecht 1981, 125-148.

Brewka 1989: G. Brewka, Preferred subtheories: an extended logical framework for default reasoning. *Proceedings IJCAI-1991*, 1043-1048.

Gardner 1987: A. von der L. Gardner, *An Artificial Intelligence approach to legal reasoning*. MIT press, 1987.

Gordon 1991: T.F. Gordon, An abductive theory of legal issues. *International Journal of Man-Machine Studies* 35 (1991), 95-118.

Hamfelt and Barklund 1989: A. Hamfelt and J. Barklund, Metalevels in legal knowledge and their runnable representation in logic. In *Preproceedings* *of the III International Conference on "Logica, Informatica, Diritto"*, Florence 1989, Vol. II, 557-576.

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

Poole 1985: D.L. Poole, On the comparison of theories: Preferring the most specific explanation. *Proceedings IJCAI-1985*, 144-147.

Poole 1988: D.L. Poole, A logical framework for default reasoning. *Artificial Intelligence* 36 (1988), 27-47.

Prakken 1991a: H. Prakken, A tool in modelling disagreement in law: preferring the most specific argument. *Proceedings ICAIL-1991*, Oxford. ACM Press, 1991, 165-174.

Prakken 1991b: H. Prakken, Reasoning with normative hierarchies (extended abstract). *Proceedings of the first International Workshop on Deontic Logic and Computer Science*, Amsterdam 1991, 315-334.

Prakken 1993a: H. Prakken, *Logical tools for modelling legal argument*. Doctoral Dissertation Free University Amsterdam, 1993.

Prakken 1993b: H. Prakken, An argumentation framework in default logic. To appear in *Annals of Mathematics and Artificial Intelligence*.

Rissland and Ashley 1989: E.L. Rissland and K.D. Ashley, HYPO: A precedent-based legal reasoner. In G.P.V. Vandenberghe (ed.): *Advanced topics in law and information technology.* Kluwer, Deventer, 1989, 213-234.

Sartor 1992: G. Sartor, Normative conflicts in legal reasoning. *Artificial Intelligence and Law*, Vol. 1, Nos. 2-3, 1992, 209-235.

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

Skalak and Rissland 1991: D.B. Skalak and E.L. Rissland, Argument moves in a rule-guided domain. *Proceedings ICAIL-1991*, Oxford. ACM Press, 1991, 1-11.

Vreeswijk 1991: G.A.W. Vreeswijk, Abstract argumentation systems. *Proceedings of the First World Conference on the Fundamentals of AI WOCFAI '91*, Paris, 1991, pp. 501-510. Also to appear in *Studia Logica*.