Legal decision making as dialectical theory construction with argumentation schemes

Bart Verheij
Department of Metajuridica, Universiteit Maastricht
P.O. Box 616, 6200 MD Maastricht, The Netherlands
bart.verheij@metajur.unimaas.nl, http://www.metajur.unimaas.nl/~bart/

Abstract
In the present paper, an approach to the formal modeling of legal decision making is proposed. It is presented in reply to a critique of two major existing approaches. The critique reveals topics that these approaches treat confusingly or fail to address at all. The topics concern their philosophical and technical underpinnings, their empirical adequacy and their heuristic value. The approach of this paper is shown to address the topics neglected by the criticized approaches.

1. Introduction
Substantial progress has been made in the logic of law. Abstract topics like defeasibility and the process of argumentation have been formally analyzed with greater theoretical satisfactoriness, and concrete topics like precedent-based and statute-based reasoning have been formally analyzed in more detail and with greater empirical adequacy. Still it is hard to get a coherent view on the different opinions and approaches that have been proposed. Though recently some attempts have been made to bridge the unwarranted gap between precedent-based and statute-based reasoning models (cf. the work of Hage 1993 and Prakken and Sartor 1998), this integrative task has not yet been completed to full satisfaction. Recent interesting work by Bench-Capon and Sartor (2000) adds to the diversity of approaches. A reason can be that there still is no common set of primitive notions that has turned out sufficiently flexible and expressive to accommodate the different approaches. In the present paper, a model of dialectical argumentation with argumentation schemes is presented in an attempt to fill this gap. It can provide an approach to the empirically adequate formal modeling of legal decision making.

The approach presented here combines three ideas. First, in legal decision making any statement can be just as well supported (by a reason for it) as attacked (by a reason against it). It turns out that this seemingly trivial point becomes a powerful tool when it is noted that it applies also to statements expressing support or attack. Second, legal decision making uses dedicated argumentation schemes (like rule application, precedent distinction and analogy) and these schemes tend to be defeasible or even contingent. The modeling of argumentation schemes results in a blurred border between the logical object and meta-level, which is unproblematic when done sensibly. Third, legal decision making is a kind of dialectical theory construction. Initial assumptions and conclusions are critically addressed by adducing reasons for and against them. The result is a gradually changing dialectical theory in which statements can at one time be justified and at others defeated or unsettled. This process is heuristically guided by the data and argumentation schemes available.

As a start, two existing approaches to the formal modeling of legal decision making are critically discussed in order to position the approach of the present paper.

2. A critique of two existing approaches
Many approaches to the formal modeling of legal decision making have been proposed. In order to position the one of the present paper, it is presented in reply to a critique of two of the major existing approaches, viz. Prakken's (1997) and Hage's (1997).[1] (Below, page numbers refer to these sources.) Prakken has provided logical tools for modeling legal argument, in the form of a theory of defeasible reasoning, presented as a dialogue game. Hage presents a theory of reasoning with rules and the reasons resulting from them, formalized in the logical system Reason-Based Logic. Both give applications to legally relevant issues, such as rule exceptions, conflicts and validity.

Both approaches have deepened the understanding of the formal modeling of legal decision making, the former especially with regards to the formalization of its defeasible aspects, the latter especially with regards to its philosophical and legal-theoretical underpinning. Here five themes are discussed that are relevant for the formal modeling of legal decision making, and it is discussed to what extent and with how much success these are covered by Prakken's and Hage's approach. (The following assumes acquaintance with the two approaches.) The themes are the logic of argument defeat, the logic of rules, the logic of law, the empirical adequacy of the approaches, and their heuristic value.[2]

2.1 The logic of argument defeat
It is by now common knowledge that legal decision making involves defeasible arguments, i.e., otherwise valid arguments that can lose their justifying power in the light of defeating counter-arguments. (By convention, I will speak of defeat when referring to the *property* of arguments (or of statements or of one's favorite objects of defeat) that they are defeated, and of attack when referring to the *relation* between arguments (or statements or ...) that results in defeat.) A key insight is Pollock's (1987) distinction between kinds of defeat, viz. defeat by a rebutter or by an undercutter. Briefly, a rebutter is a reason attacking a reason for an opposite conclusion, and an undercutter a statement that attacks that another reason supports its conclusion. Both Prakken and Hage give an account of undercutters and rebutters, but in

[1] Prakken's approach has been developed in cooperation with Sartor, see e.g. Prakken and Sartor (1996). The present author has contributed to Hage's approach, see e.g. Verheij (1996).

[2] Verheij (2000a) gives a 30 odd page critique of Prakken's (1997) approach, some of which points are recapitulated here.

ICAIL-01 St. Louis, Missouri

different logical forms (both differing from Pollock's form as well). While Prakken (p. 174) considers them as types of argument attack (where arguments are a kind of derivations as they occur in standard logic), Hage (p. 166f.) treats them as different applications of the exclusion of rules as they underlie reasons.[3]

How are undercutters formalized in the two approaches? In Hage's approach, Excluded(thief(x) $\Rightarrow$ punishable(x)) has the effect that Punishable(john) cannot be derived from Valid(thief(x) $\Rightarrow$ punishable(x)) and Thief(john). As a result, Excluded(thief(x) $\Rightarrow$ punishable(x)) (or a reason for it) can be seen as an undercutter for the reason R(thief(x), punishable(x)).[4] In Prakken's approach, an argument of the form [P, $P \Rightarrow Q$, Q] undercuts an argument of the form [P', $P' \wedge {\sim}Q \Rightarrow R$, R]. Here a sentence like ${\sim}Q$ stands for 'Assuming that not Q', which is not counted as a premise of an argument in which it occurs, and thus needs no support. Only arguments containing a weakly negated statement like ${\sim}Q$ in the antecedent of a rule can be undercut, viz. by an argument for Q.

An unsatisfactory aspect of Hage's approach is that it gives a rather complicated elaboration of Pollock's simple and straightforward notion of undercutting. Hage's example of undercutting uses the full apparatus of Reason-Based Logic. As a result, undercutting involves a rule that is valid and has a satisfied condition, but is excluded, and therefore not applicable; provided there are no other reasons concerning the rule's application, the rule then ought not to be applied, and the rule's conclusion does not follow. Notwithstanding the value of Reason-Based Logic's richness, when modeling undercutting, it seems to give an unnecessarily involved picture. We return to this point in section 2.4.

Prakken's model of undercutting has a wholly different shortcoming: it cannot model the undercutting of reasons to its full generality. An example is given by the argument [P, $P \Rightarrow Q$, Q] in which P is a reason for Q. This reason cannot be undercut. (The argument *is* defeasible: it can be rebutted). If R were a possible undercutter of P as a reason for Q, that would have to be encoded in the conditional underlying the reason: $P \Rightarrow Q$ would have to be replaced with $P \wedge {\sim}R \Rightarrow Q$. Now the adapted argument [P, $P \wedge {\sim}R \Rightarrow Q$, Q] can be undercut, e.g., by the mini-argument [R]. This limited representability of undercutters is a severe drawback, since it precludes that a reason is *unexpectedly* undercut. In Prakken's approach, reasons can only be undercut when that is foreseen at the time of representation. This is especially unfortunate since dealing with sources of defeat that by their nature cannot be expected (like new legislation), is one of the principle aims of modeling defeasible reasoning in the first place. (We briefly return to this point in section 2.4.)

In the logical system DEFLOG it is straightforward and simple to express undercutters and other kinds of defeat (see section 3.3).

A deep question concerning Prakken's work on argument defeat is related to his layered view on argumentation (p. 270f.), especially with respect to his logic and dialectical layer. At the logic layer, it is specified what contradiction is and which arguments support a conclusion at all. The dialectical layer deals with attack and defeat, and determines whether an argument justifies its conclusion. The logic layer can for instance be a standard logic, like first-order predicate logic, or a nonmonotonic logic. In the first case, something very strange happens: first-order predicate logic is about *truth-preserving* arguments, so how can they be *defeasible*? As a result, for a logic based on truth-preservation the dialectical layer must remain empty. In the second case, it is very unclear why the nonmonotonic logic is further 'nonmonotonized' at the dialectical layer. My way out of this is to *integrate* the logical and the dialectical layer into a dialectical logic, such as my system DEFLOG below. Arguments become dialectical in the sense of incorporating both support and attack, and can be *dialectically justifying* (i.e., withstand counterarguments in a certain specific sense), instead of only merely supportive.

2.2 The logic of rules

In the context of legal decision making, rules play a central role. The typical example is the rule that thieves are punishable. Prakken expresses rules as a kind of conditional sentences like $P \Rightarrow Q$. They are a kind of rules of inference (i.e., not a part of the ordinary logical object language) in order to make them 'unidirectional' to use a term by Prakken, in the sense of not allowing contraposition (from not-Q and $P \Rightarrow Q$, infer not-P). The conditional $\Rightarrow$ cannot be nested, so that $P \Rightarrow (Q \Rightarrow R)$ is not a rule. In order to express rule properties (like a rule's validity or applicability), Prakken allows that rules are named. For instance, that d is the name of $P \Rightarrow Q$ is denoted as d: $P \Rightarrow Q$. Hage on the other hand does not want to formalize rules as conditional sentences, and emphasizes that they are individuals (and not states of affairs). He therefore denotes rules as logical terms with a conditional form, like $p \Rightarrow q$. Properties of rules are now straightforwardly expressible using predicates, like in Valid($p \Rightarrow q$). It should be stressed that both Prakken and Hage do not derive the behavior of their rules from existing formalizations (like the material implication of standard logic) but define it from scratch.

Both approaches can be made to work, but not in all respects equally satisfactory. The problem of Prakken's approach is an inherent limitation of his naming method: though it can be used to express properties of rules, these properties cannot by themselves affect the logical behavior of the named rule. An example is the rule d: $P \Rightarrow Q$. It is now for instance possible to express that the rule has an exception (e.g. as exc(d)) and that R gives rise to an exception ($R \Rightarrow$ exc(d)). However, as it stands the behavior of the rule d: $P \Rightarrow Q$ is *unaffected* by such expressions. For instance, assuming or concluding that exc(d) does as yet not lead to new ways of blocking the rule d: $P \Rightarrow Q$. Only when the rule is somehow suitably *adapted*, the naming technique can affect the rule's logical behavior. An example is the change of the rule d: $P \Rightarrow Q$ to d: $P \wedge {\sim}$exc(d) $\Rightarrow Q$. Now exc(d) can have a new effect, for it can undercut an argument that contains d: $P \wedge {\sim}$exc(d) $\Rightarrow Q$.

The limitation of Prakken's naming method also implies that reasoning about rules cannot be modeled to full satisfaction. For instance, it is hard to formalize the derivation of a rule from a fact (e.g., conclude $P \Rightarrow Q$ from R, as in the argument that if John is a thief he is punishable since thieves are punishable) or the derivation of a rule from other rules (e.g., conclude $P \Rightarrow R$ from $P \Rightarrow Q$ and $Q \Rightarrow R$). The main causes are that Prakken's rules form a

[3] It seems that Hage's (p. 167-8) example is not really one of Pollock's rebutters since it does not recognize that the rebutting statement should itself actually be a reason to have its rebutting effect. Hage's notion of a pro outweighing a con seems to be closer to Pollock's notion of a rebutter.

[4] Hage (1997) uses a somewhat peculiar mixture of formal and semi-formal symbols: negation is denoted by the symbol ~, the validity of a rule by a recognizable predicate Valid, and the central notion of a reason by the abbreviation R. Prakken (1997) also uses a mixture, but more systematic. The logical core uses formal symbols like $\Rightarrow$ and $\prec$, whereas the legal applications use recognizable predicates like appl.

fixed part of the domain theory and are not part of the ordinary object language. Prakken suggests a way to repair this (p. 176-7): to any rule, add valid(d) to the antecedent when d is the rule's name. Thus we get d: $P \wedge$ valid(d) $\Rightarrow Q$.[5] In this way, the rule can only be used when valid(d) can otherwise be argued for. Technically, this works. For instance, the derivation of $P \Rightarrow Q$ from R now can be formalized by two rules, viz. d: $P \wedge$ valid(d) $\Rightarrow$ Q and $R \Rightarrow$ valid(d), and the derivation of $P \Rightarrow R$ from $P \Rightarrow Q$ and $Q \Rightarrow R$ can be formalized in four rules, viz. d: $P \wedge$ valid(d) $\Rightarrow Q$, e: $Q \wedge$ valid(e) $\Rightarrow R$, f: $P \wedge$ valid(f) $\Rightarrow R$ and valid(e) $\wedge$ valid(f) $\Rightarrow$ valid(d). The solution is unnecessarily complicated: Prakken's method comes down to mimicking a conditional that can be reasoned about using a delicate method involving fixed rules and their names. Effectively, d: $P \wedge$ valid(d) $\Rightarrow Q$ makes valid(d) into an imitation of $P \Rightarrow Q$, but one that can be argued about.

Hage's approach does not have this problem. His formalization approach is as expressive as one could want. Two questions spring to mind however. The first is: why does Hage model rules as logical terms and not as sentences? The second is: are simple arguments, e.g., those based on elementary rule application, not unnecessarily complicated in Reason-Based Logic? Answering the second question is postponed to section 2.4 on empirical adequacy. I know of two reasons that might answer the first question. One is that rules are to be formalized as logical terms since their properties can be at issue in reasoning. The other is that rules are not a kind of facts, but constitute factual relations (and do not describe them) (cf. p. 76).[6] I agree with the first reason, but not when amplified to the claim that the occurrence of rule properties implies that rules are *only* to be formalized as logical terms. They can well be modeled as sentences *and* as terms. Moreover, if rules were facts (next to individuals, as I will argue below), the occurrence of rule properties in reasoning would not distinguish them from other facts: fact properties occur in reasoning too. For instance, an argument can be about the issue whether *it is just that* it has been decided that someone is acquited, or about the issue whether *it has been approved by parliament* that the prince gets married. The second reason is related to the observation that in modeling legal decision making one should pay attention to constitutive rules (that abound in the law), possibly more than to descriptive rules (cf. note 6). It should be noted however that the reason only works when rules are not a kind of facts *since* they constitute factual relations (for saying that rules are not facts is equivalent to saying that they should not be modeled as sentences). I do not see why there cannot be a fact of the matter that one fact constitutes another fact (or that one kind of fact typically constitutes another kind of facts). What else could it be? That someone is a thief can constitute that someone is punishable, but not that someone is punished (at least not in the actual world). Clearly certain constitutive relations between (kinds of) facts obtain, and others don't. As a result, the most straightforward way to model the constitutive relations is as a kind of states of affairs, that by the structure of these relations is expressible by a conditional sentence. Hage shows exactly how: the fact that Valid($p \Rightarrow q$) (and *not* the rule-individual $p \Rightarrow q$) has the effect that the fact that P can constitute the fact that Q. So in Hage's approach the sentence Valid($p \Rightarrow q$) can express (or imply) the state of affairs that a fact that P can constitute a fact that Q. But formally Valid($p \Rightarrow q$) is just another way of denoting a conditional sentence of the form $P \Rightarrow Q$. When the sentences Valid($p \Rightarrow q$) are recognized as a kind of conditional sentences, the existence of Hage's Reason-Based Logic shows that there is moreover no *logical* reason for treating constitutive rules (only) as terms.[7]

I draw two conclusions. First, rules (in the sense of constitutors of factual relations) can well be regarded as a kind of facts. Second, (returning to the first reason for treating rules as terms) rule properties need to be expressed, but not just rule properties: properties of all facts must be expressible.

A related topic is that of rules of inference. These are discussed in the context of the logic of law.

2.3 The logic of law

The law uses dedicated argumentation schemes, some of which are typical for the law. Examples are rule application, analogy and precedent distinction. Elsewhere (Verheij 1999b) I have argued that this observation naturally leads to a *context-dependent conception of logical validity*, i.e., a contingent logical validity that pertains to the kind of reasoning in a specific domain, such as legal decision making.[8] The simple point is that the argumentation schemes that are dedicated to a specific domain can be regarded as the rules of inference of a context-dependent logic.

Interestingly, though Prakken and Hage both deal with the formal modeling of legal reasoning, they do this in rather different ways. Whereas Hage designs a logic with a rich set of primitives (like the dedicated predicates and terms expressing or denoting rules, principles, validity, exclusion etc.), Prakken uses a system with relatively few, more abstract primitives (like argument defeat and priority handling) which is subsequently adapted to legal reasoning by adding more concrete, domain-dependent elements (like predicates expressing rule validity and applicability). It can be said that Hage's system is a concrete logic of law (with applications outside the law)[9], while Prakken's is an abstract

[5] Prakken suggests a second way, viz. assuming that all rules are valid by default. I see no situation in which that is a reasonable assumption.

[6] That thieves are punishable is a constitutive rule since it *makes* that thieves are punishable; that about 1% of all thieves is punished is descriptive since it *describes* a state of affairs. Constitutive rules are indeed central in the law, and perhaps more so than descriptive rules. The former shape the world, whereas the latter are shaped by the world. And indeed the law, when seen as at least in part a human-made institution, shapes the world. Note however that in evidential reasoning (i.e., when 'establishing' facts on the basis of evidence) descriptive rules (such as statistical correlations) are very relevant.

[7] I actually think that both constitutive and descriptive rules can be logically formalized by the *same* conditional. It is not a coincidence that the rule sentence 'Thieves are punishable' can be read both constitutively (i.e., as the rule that makes that thieves are punishable) and descriptively (i.e., as the rule that it is actually, maybe even accidentally, the case that thieves are punishable). In my opinion, the difference comes simply from different *relations with other facts*. For instance, DEFLOG's conditional $\leadsto$ is in itself constitutive (since it just validates *Modus ponens*), but can be made descriptive when facts like $\psi \leadsto (\varphi \leadsto \psi)$ and not-$\varphi \leadsto (\varphi \leadsto \psi)$ obtain.

[8] This is conform my use of the term 'logic', viz. as the task or result of formalizing good reasoning, or some aspects thereof.

[9] Here we will consider Reason-Based Logic as a logic of law, though Hage (1997) may prefer to describe it as a logic of rules, principles, goals and reasons. What matters here is that Reason-

logical system that can be adapted to the legal context (cf. also Verheij, 1999b).[10]

Hage's approach is apparently based on a brave assumption, viz. that there is a logic of law (or of rules, principles and goals) in some absolute sense, or at least that it is possible to ever better approximate such an absolute logic. Reason-Based Logic is intended to be the *right* logic of rules and reasons, in some sense or another. Still to me Reason-Based Logic seems to be *an* expressive logic of rules and reasons, modeling a *particular* view. In fact, the version history of Reason-Based Logic can be taken as evidence for the latter: there are many versions, possibly gradually improving and surely incorporating more topics. I think that a logic of rules and reasons (that is empirically inclined) can never be really right, since the topic is itself fundamentally messy: the terms 'rule' and 'reason' are used in many ways, and intendedly so. People use the terms in the way that is appropriate for the topic at hand. It suffices that they use the terms on each particular occasion in a clear and unambiguous way. As long as it is transparent how rules lead to reasons in today's argument, it does not matter much how they do in tomorrow's. Interestingly, defeasible logic provides just the tools to allow such 'modular' uses of relevant terms: using a term in one way should block using it in another way.

Moreover, it remains obscure which topics of reasoning are to be modeled in the 'logical core' (i.e., in Reason-Based Logic itself) and which must be a part of the domain theory. The question of delimiting the fixed logical core from the logical aspects modeled in the contingent domain theory is relevant since the concrete argumentation schemes of legal reasoning tend to be defeasible and sometimes even contingent. But how can it be determined which argumentation schemes are defeasible and which contingent as soon as one recognizes that a new situation can shed new light on an earlier answer? Again, Reason-Based Logic's version history is an example of the hazardous nature of this task: when new topics are added or difficulties are discovered, the formalism is adapted. Though I actually like the adaptive approach (since it reflects a decent critical stance that is the basis of good science), it calls for the recognition that *there is no clear border between a fixed logical core and contingent logical aspects*. Every time and again, there will come a moment in which it becomes appropriate to take an element out of the logical core and make it contingent, or the other way around. An example is the prohibition of analogical rule application in criminal law. This could be included in the logical core of a dedicated 'logic of criminal law', but it is conceivable that some day the dogma of the prohibition of analogy is loosened, or turns out to be less of a dogma than at first thought. In principle, a single Supreme Court case can require the change of our hypothesized logic of criminal law.

There is a way out: start with an abstract logical core of sufficiently expressive primitives, and define the concrete, domain-dependent logic in terms thereof by means of argumentation schemes. Now at all 'ordinary' occasions the concrete logic constrains our view of reasoning: it determines what counts as good reasoning and what not. For all normal purposes, it forms our fixed logical core, and the abstract logic underlying it remains hidden. However, at the odd times that the concrete logic needs to be changed for some reason or other, the curtain hiding the underlying abstract logic is raised, and the relevant portion of the definition of the concrete logic is again seen to be a contingent statement in the abstract logic instead of a fixed part of the concrete logic. This approach reminds of an analogical phenomenon in computer science: the virtual machines that an actual machine can become by suitable programming. The analogy is that normal communication with a computer does not go beyond the virtual machine level as provided by the application programs on it. As such, normally a computer is just a word processing or internet access machine. Only sometimes it is opportune to go to a deeper level, such as that of the operation system, or further down, the machine language.

One further point with regards to the topic of a logic of law needs to be addressed: where do the argumentation schemes come from and how are they justified? Is there an analogue of the validity of rules of inference in standard logic for the defeasible, contingent argumentation schemes encountered here? Both Prakken and Hage find inspiration for the topics of legal reasoning they formalize in *actual* legal reasoning. As a result, there is a praiseworthy empirical slant to their approaches. Unlike Prakken, Hage explicitly discusses the topic of the validity of argumentation schemes. According to Hage, rules of inference are social rules, and thus are pragmatically valid (p. 248f.). One aspect of pragmatic validity is that rules of inference should withstand criticism: even though they depend on a particular social group, within that group there can be arguments about the validity of a rule of inference. Hage does not elaborate on this point, and leaves a question unanswered: if the rules of inference of Reason-Based Logic are to be argued about (and they *are*, as the version history shows), what logic can then be used for arguing? Again Reason-Based Logic? That would not get us very far. I would say that such arguing should take place in a more abstract logic than Reason-Based Logic, one that has less primitives, but which is sufficiently expressive to model Reason-Based Logic in terms of it. Moreover, I have a good candidate for a 'bottom line' dialectical logic, i.e., a logic for which it makes little sense to ask for further specification, since its principles are only just enough to allow dialectical argument in the first place. (Obviously what one counts as a bottom line logic depends on one's goals.) That logic is DEFLOG, to be presented below.

2.4 The empirical adequacy of the approaches

An important aim of Hage's and Prakken's approaches to the formal modeling of legal reasoning is to do better than standard logic (like first-order predicate logic) can. One evaluation criterion is then the empirical adequacy of the approaches.[11] In section 2.1, it was already discussed that Prakken's approach to undercutters leaves no room for unexpected undercutting: a rule $P \Rightarrow Q$ has to be changed to $P \wedge \sim R \Rightarrow Q$ to allow undercutting. Of course it is possible to let R take the form of an open-ended exception or applicability clause (like in d: $P \wedge \sim \mathrm{exc}(d) \Rightarrow Q$), but that does not take away the problem that actual rules (like the rule

Based Logic contains special-purpose elements inspired by phenomena that typically occur in legal reasoning.

[10] Some doubt that it makes sense to speak of a logic of law at all. For instance, Soeteman (1989, especially p. 20-22) states that a system of formal logic would never be intrinsically legal. Verheij (1999b) shows why it makes sense to speak of a logic of law (and why there is no clear border between Soeteman's 'formal' and 'material' logic).

[11] Prakken (p. 103f.) pays explicit attention to formalization criteria, some of which have to do with empirical adequacy.

that thieves are punishable) do *not* require change and do *not* have exception clauses built into them. Hage's approach shows that this problem of Prakken's approach is not a formal necessity: Reason-Based Logic formalizes rules in a way that they need no change or exception clauses, in order to be unexpectedly undercut. A minor point concerning the empirical adequacy of Prakken's approach is that the weakly negated conditions like ~*R* in the antecedent of a rule seem to have no empirical counterpart.

Hage's approach raises a different kind of question with respect to empirical adequacy, however, one that was already asked in section 2.2: are simple arguments, e.g., those based on elementary rule application, not unnecessarily complicated in Reason-Based Logic? A simple case of rule application involves in Reason-Based Logic a valid rule with a satisfied conclusion, that is not excluded and therefore applicable; the rule then ought to be applied provided there are no other relevant reasons concerning the application of the rule, and thus the rule's conclusion follows. This echoes the discussion in section 2.1 of Hage's involved modeling of undercutters. If one takes empirical adequacy seriously, one must agree that simple arguments become too complicated in Reason-Based Logic. In actual legal reasoning, simple arguments of the type 'John is a thief. Therefore John is punishable' can occur just as naturally as an elaborate argument on the reasons for and against the obligation to apply the rule that thieves are punishable. Apparently, there is both room for arguments with simple primitives and for arguments with complex primitives. An empirically adequate formal model of legal reasoning should therefore allow both kinds of arguments. It may be said that the simple argument is *actually* a kind of abbreviation of the complex one, that must be regarded as the 'real' one, the 'deep' structure, its 'rational reconstruction', or whatever, but then one also admits that empirical adequacy is not one's goal (or one commits the fallacy of seeing what one wants to see). In fact, it occurs in practice (though perhaps not very often) that the argument is about the question *which* argumentation scheme has been or should be used.[12] As a result, the *Hinein-interpretierung* of an argument (e.g., by claiming that an argument is 'actually' another argument) should be replaced by a discussion about the interpretation of the argument.

The way out suggested here is to allow both the formalization of simple and of complex arguments, and to show how they relate to each other. The idea is then that the complex argument is one of several possible *elaborations* of the simple argument, e.g., that result from making explicit which argumentation scheme is used. A simple argument like 'John is a thief. Therefore John is punishable' would for example be elaborated as an instance of an argument based on the complex notion of rule application as it occurs in Reason-Based Logic.

2.5 The heuristic value of the approaches

Do the approaches by Hage and Prakken give clues for *doing* argumentation? In other words, do the approaches have heuristic value? With respect to Hage's approach, one can say that due to its richness it can provide inspiration for types of arguments that one can perform. An actual reasoner can use Reason-Based Logic as a kind of catalog of kinds of arguments, and thus find that for instance reasoning about the application of a rule can serve his argumentative purposes. Since Hage's work (1997) is mainly semantically and ontologically styled, he pays here little attention to the actual process of argumentation (but see e.g. Hage 2000 for some of his views).

Prakken discusses two ways in which argumentation can proceed, viz. by justifying decisions and by suggesting premises (p. 26-27). Prakken considers analogical reasoning as an example of the latter, deductive reasoning of the former. One could say that Prakken here is looking for heuristics for argumentation: deduction is a means to justify decisions, analogy a means to suggest premises. When I reviewed Prakken's book (Verheij 2000a) I had trouble to understand exactly what Prakken meant, though, and I still have. Prakken's discussion did provoke me to think further about these matters however. Here are my current views.

To me, a distinction between reasoning to conclusions and premise selection is fine, but is wholly independent of the *kind* of reasoning involved. Both rule application and analogy can be used to draw conclusions and to find premises (cf. also Verheij, 2000a), simply since both can be expressed as argumentation schemes (the latter looking for instance something like this: From 1. By the application of rule *R*, *A* follows from *C*, 2. *B*, and 3. *B* is relevantly similar to *A*, conclude: *C*). My claim is supported by the fact that the end product of reasoning is both after rule application and after reasoning by analogy an argument that shows how certain premises lead to certain conclusions (possibly in the face of attacks), the only relevant difference being the argumentation schemes used. Obviously, an argumentation scheme for reasoning by analogy has premises that are harder to support (like the premise that *B* is relevantly similar to *A*) and may ask for further backing (i.e., an answer to the question why reasoning by analogy is allowed in the first place).

2.6 Summary of the critique

With respect to the *logic of argument defeat*, it can be said that Hage uses a rather complicated elaboration of Pollock's undercutters, while Prakken cannot model them in full generality. Prakken's view on the logic of argument defeat with a logical and a dialectical layer is unclear.

With respect to the *logic of rules*, it was found that Prakken's naming method for rules is inherently limited and requires the adaptation of the rules involved in order to have effects. Hage's claim that rules are individuals is unwarranted; they can just as well be regarded as facts, provided that rule properties (just like other fact properties) are expressible.

With respect to the *logic of law*, it was noticed that Hage's approach seems to be based on the assumption that it is (or can become) the right approach to rules and reasons, which seems to be overly brave. With respect to Hage's approach it is unclear which logical elements are to be a part of the logical core and which of the domain theory. Both Prakken and Hage underspecify where to find argumentation schemes and how to justify or argue about them.

With respect to the *empirical adequacy* of the approaches, it was noted that Prakken's approach to undercutters does not allow unexpected undercutting, unless rules are not modeled as they actually occur. Hage's elaborate logical system was argued to be not fully appropriate for modeling simple arguments, since their modeling would require their considerable, and thus arguable, elaboration. Neither Prakken nor Hage answer the problem that in practice different uses of primitive terms like 'rule' and 'reason'

---

[12] In a recent Dutch Supreme Court case (HR January 7, 2000, NJ 2000, 496), the argument is in part about whether extensive interpretation or analogical rule application was (or could be) used in a particular case of civil procedural law.

are used sensibly, simply by separating different uses in a modular way.

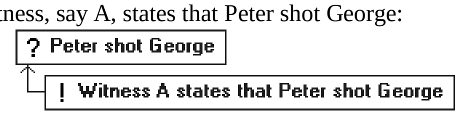
With respect to the *heuristic value* of the approaches, it was observed that Prakken gave a confusing view on the difference between deductive and analogical reasoning in terms of justifying decisions and suggesting premises, while Hage (in his 1997) did not pay attention to actually doing argumentation.

3. DEFLOG - a logic of dialectical interpretation

Recently I have developed a theory of dialectical argumentation and a corresponding logic of dialectical interpretation, called DEFLOG. It is related to my work on automated argument assistance (e.g., Verheij 1999a). Below the theory is summarized. For a more extensive account, the reader may want to consult Verheij (2000b).

3.1 Dialectical argumentation

In dialectical argumentation, statements can not only support other statements, but also attack them. For instance, as a reason to support that Peter shot George, the statement can be made that some witness, say A, states that Peter shot George:

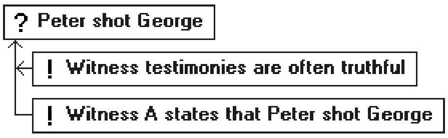


The exclamation mark indicates an assumed statement, the question mark a statement that is at issue. Here the issue that Peter shot George is settled (the statement is justified, as is indicated by the dark, bold font) since there is a justifying reason for it, namely A's testimony.

As a reason against the issue that Peter shot George, the statement can be made that some other witness, say B, states that the shooting did not take place:

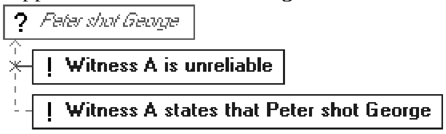
Assuming only B's testimony (not A's), the issue that Peter shot George is again settled, but this time the statement is defeated, as is indicated by the struck-through font.

That some statement supports or attacks another statement can itself be at issue. For instance, it can be argued that A's testimony supports that Peter shot George since witness testimonies are often truthful:



Likewise, a reason can be given to support that some statement *attacks* another statement.

A's unreliability can be adduced in order to attack that A's testimony supports that Peter shot George:



Here the issue that Peter shot George is unsettled, as is indicated by the light italic font, since it is not justified (e.g., by a justifying reason for it) nor defeated (e.g., by a defeating reason against it). Similarly, a reason can be given to attack that some statement attacks another statement.

Dialectical argumentation can be regarded as the gradual construction of dialectical arguments as above (cf. Verheij 2000c).

3.2 The dialectical interpretation of theories

Dialectical argumentation can be made formally precise in terms of the logical system DEFLOG. Its starting point is a simple logical language with two connectives $\times$ and $\rightsquigarrow$. The first is a unary connective that is used to express the defeat of a statement, the latter is a binary connective that is used to express that one statement supports another. When φ and ψ are sentences, then $\times\varphi$ (φ's so-called *dialectical negation*) expresses that the statement that φ is defeated, and $(\varphi \rightsquigarrow \psi)$ that the statement that φ supports the statement that ψ. Attack, occasionally denoted as $\rtimes$, is defined in terms of these two connectives: $\varphi \rtimes \psi$ is defined as $\varphi \rightsquigarrow \times\psi$, and expresses that the statement that φ attacks the statement that ψ, or equivalently that the statement that φ supports that the statement that ψ is defeated. When p, q, r and s are elementary sentences, then $p \rightsquigarrow (q \rightsquigarrow r)$, $p \rightsquigarrow \times(q \rightsquigarrow \times r)$ and $(p \rightsquigarrow q) \rightsquigarrow (p \rightsquigarrow \times(r \rightsquigarrow s))$ are some examples of sentences. (For convenience, outer brackets are omitted.)

The central definition of DEFLOG is its notion of the *dialectical interpretation* (or *extension*) of a theory. A theory is any set of sentences, and when it is dialectically interpreted, all sentences in the theory are evaluated, either as justified or as defeated. (This is in contrast with the interpretation of theories in standard logic, where all sentences in an interpreted theory are assigned the same positive value, namely true, e.g., by giving a model of the theory.)

An assignment of the values justified or defeated to the sentences in a theory gives rise to a dialectical interpretation of the theory, when two properties obtain. First, the justified part of the theory must be conflict-free. Second, the justified part of the theory must attack all sentences in the defeated part. Formally the definitions are as follows.

(i) Let T be a set of sentences and φ a sentence. Then T *supports* φ when φ follows from T by the repeated application of $\rightsquigarrow$-Modus ponens (i.e., from $\varphi \rightsquigarrow \psi$ and φ, conclude ψ). T *attacks* φ when T supports $\times\varphi$.

(ii) Let T be a set of sentences. Then T is *conflict-free* when there is no sentence φ that is both supported and attacked by T.

(iii) Let Δ be a set of sentences, and let J and D be subsets of Δ that have no elements in common and that have Δ as their union. Then (J, D) *dialectically interprets* the theory Δ when J is conflict-free and attacks all sentences in D. The sentences in J are the *justified statements* of the theory Δ, the sentences in D the *defeated statements*.

(iv) Let Δ be a set of sentences and let (J, D) dialectically interpret the theory Δ. Then (Supp(J), Att(J)) is a *dialectical interpretation* or *extension* of the theory Δ. Here Supp(J) denotes the set of sentences supported by J, and Att(J) the set of sentences attacked by J. The sentences in Supp(J) are the *justified statements* of the dialectical interpretation, the sentences in Att(J) the *defeated statements*.

Note that when (J, D) dialectically interprets Δ and (Supp(J), Att(J)) is the corresponding dialectical interpretation, J is equal to Supp(J) $\cap$ Δ, and D to Att(J) $\cap$ Δ.

The examples discussed in section 3.1 can be used to illustrate these definitions, since the evaluated dialectical arguments of that section naturally correspond to dialectically interpreted theories. Let s express Peter's shooting of George, a A's testimony, b B's testimony, t the truthfulness of testimonies, and u A's unreliability. Then the first example corresponds to the two-sentence theory $\{a, a \leadsto s\}$. The arrow in the figure corresponds to the sentence $a \leadsto s$. The theory has a unique extension in which all statements of the theory are justified. In the extension, one other statement is justified, viz. s. The second example corresponds to the theory $\{b, b \leadsto \times s\}$. The arrow ending in a cross in the figure corresponds to the sentence $b \leadsto \times s$. The theory has a unique extension in which again all sentences of the theory are justified. In the extension, there are two other interpreted statements, viz. $\times s$, which is justified, and s, which is defeated. (The reader may wish to check that the theory $\{b, b \leadsto \times s, s\}$, which is not conflict-free, has the same unique extension, but that one of the statements in the theory is defeated.) The third example corresponds to the theory $\{a, t, t \leadsto (a \leadsto s)\}$. In its unique extension, all statements of the theory are justified, and in addition $a \leadsto s$ and s. The fourth example corresponds to the theory $\{a, u, u \leadsto \times(a \leadsto s)\}$. In its unique extension, $a \leadsto s$ is defeated and s is not interpreted (i.e., neither justified nor defeated). (The theory $\{a, u, u \leadsto \times(a \leadsto s), a \leadsto s\}$ is not conflict-free, but has the same unique extension.)

There is a lot to say about the dialectical interpretation of theories. An important notion is that of *dialectical justification*, defined as follows. A conflict-free set C is said to *dialectically justify* a statement φ with respect to the theory Δ when C attacks any conflict-free set C' that is incompatible with C. (Conflict-free sets are incompatible when their union is not conflict-free.) The reader is referred to Verheij (2000b).

The above gives rise to a correspondence between dialectical argumentation and dialectical theory construction since the evaluated dialectical arguments of section 3.1 correspond to dialectically interpreted theories (cf. also Verheij 2000c).

3.3 Expressiveness of the logical language

DEFLOG'S logical language only uses two connectives, viz. $\leadsto$ and $\times$. Notwithstanding its simple structure, many central notions of dialectical argumentation can be analyzed in terms of it. Here it is briefly shown how to express conjunctive support, undercutting attack, rebutting attack and priority attack (see Verheij 2000b for more).

Conjunctive support occurs when a statement is supported by a conjunctive combination of statements. For instance, in order to support the statement that someone has committed murder a conjunction of three statements can be adduced as a reason, namely of the statement that someone's life has been taken, the statement that this was done intentionally, and the statement that this was done on premeditated plan. Conjunctive support requires a conditional with a conjunction as antecedent, as in the sentence $(\varphi \wedge \psi) \leadsto \chi$ that can express that the conjunction of the two statements expressed by the sentences φ and ψ supports the statement expressed by the sentence χ. In DEFLOG's language, which lacks the connective $\wedge$ for the expression of conjunctions, the nested conditional $\varphi \leadsto (\psi \leadsto \chi)$ can be used for the expression of support by a conjunction of two statements. Similar nested conditionals can be used to express conjunctive support in general. In the rest of the paper, $(\varphi \wedge \psi) \leadsto \chi$ (or simply $\varphi \wedge \psi \leadsto \chi$) will be used as an abbreviation of $\varphi \leadsto (\psi \leadsto \chi)$ to express support by a conjunction of two statements, and similarly for conjunctive support in general.

Undercutting attack occurs when a statement attacks that another statement supports a third statement (cf. Pollock 1987). An example is the attack by the statement that there is a ground of justification, of someone's crime being a reason for punishment. In DEFLOG's language, undercutting attack can be expressed straightforwardly as $\varphi \leadsto \times(\psi \leadsto \chi)$. The sentence expresses that the statement that φ attacks that the statement that ψ supports the statement that χ.

Rebutting attack occurs when a reason for some conclusion attacks that a statement supports an opposite conclusion.[13] An example is the attack of someone's crime as support for punishment by someone's being a first offender as a reason against punishment. Rebutting attack can be expressed as $(\varphi \leadsto \psi) \leadsto (\varphi \leadsto \times(\chi \leadsto \text{not-}\psi))$. Here ψ and not-ψ express opposites. The sentence is slightly easier to read when it is taken as a case of conjunctive support, as follows: $((\varphi \leadsto \psi) \wedge \varphi) \leadsto \times(\chi \leadsto \text{not-}\psi)$. The sentence reads that the conjunction of the statement that φ supports that ψ and the statement that φ attack the statement that χ supports that not-ψ.

Priority attack is only a slight generalization of rebutting attack. It occurs when a reason for some conclusion attacks that another statement supports some other conclusion. The attack is the result of the priority of one reason over another. The difference between priority attack and rebutting attack is that the latter always involves opposite conclusions, while the former can involve any pair of conclusions. Priority attack occurs for instance when the conflict rule *Lex superior derogat legi inferiori* (which says that in case of conflict a rule made by a higher authority has priority) is applicable. Formally, priority attack can be expressed as $(\varphi \leadsto \psi) \leadsto (\varphi \leadsto \times(\chi \leadsto \omega))$, the *only* difference with rebutting attack being that ψ and ω are not necessarily opposites. The sentence expresses that the reason φ for ψ attacks that χ supports ω. If one prefers to have a dedicated notation for the priority relation, the sentence can be abbreviated as $(\varphi \leadsto \psi) > (\chi \leadsto \omega)$.

To conclude, it is noted that an important insight of DEFLOG's dialectically interpreted logical language is that there is a strong parallel between support and attack, and more specifically between Toulmin's (1958) warrants and Pollock's (1987) undercutters. Both support and attack are expressed using DEFLOG's conditional $\leadsto$, viz. as $\varphi \leadsto \psi$ and as $\varphi \leadsto \times\psi$, respectively, and have their effect by the simple application of $\leadsto$-Modus ponens. Toulmin's warrants (in the pragmatic optimum sense of the term[14]) correspond to a nested conditional $\varphi \leadsto (\psi \leadsto \chi)$, where φ expresses the warrant, while Pollock's undercutters correspond to $\varphi \leadsto \times(\psi \leadsto \chi)$, where φ expresses the undercutter.

4. Argumentation schemes

Legal reasoning does not only involve reasons for and against conclusions, as analyzed in the previous section, but also makes use of dedicated, sometimes typically legal argumentation

[13] The following convention is useful to distinguish the case that a statement (say expressed by the sentence φ) *supports* another statement (expressed by ψ) from the case that a statement *is a reason for* another statement. The former is expressed by the conditional $\varphi \leadsto \psi$, the latter by the conjunction of $\varphi \leadsto \psi$ and φ. As a result, being a reason involves not only giving support, but also being true/justified/believed etc.

[14] Toulmin (1958) sometimes uses the term 'warrant' for the particular conditional relation between the statements involved (like 'If witness A states that Peter shot George, then Peter shot George'), and sometimes for a corresponding generic conditional relation (like 'If some witness states that some fact obtains, then that fact obtains'). The former is sometimes called the *logical minimum*, the latter the *pragmatic optimum*.

schemes. In the following, it is shown how argumentation schemes can be incorporated in the abstract model of dialectical argumentation of section 3.

4.1 Concrete logics specified by defeasible argumentation schemes

Argumentation schemes can be seen as a generalization of the rules of inference as they are familiar from classical logic. Examples of rules of inference include *Modus ponens* (From 1. If *P*, then *Q*, and 2. *Q*, conclude *Q*) and the *Disjunctive syllogism* (From 1. Either *P* or *Q*, and 2. It is not the case that *P*, conclude *Q*). Argumentation schemes address a broader scope of types of reasoning than the rules of inference of classical logic. Whereas for instance the rules of inference of classical propositional logic are purely an analysis of the types of valid reasoning in a language with truth-functional connectives, argumentation schemes can be about any kind of reasoning as it occurs in practice. There is a similar goal: both rules of inference and argumentation schemes intend to model *valid* kinds of reasoning (in some sense of the term 'valid').

Other examples of argumentation schemes are what might be called the *Practical syllogism* (From 1. Doing *A* contributes to goal *G*, and 2. *P* has goal *G*, conclude: *P* should do *A*) and *Analogous rule application* (From 1. By the application of rule *R*, *A* follows from *C*, 2. *B*, and 3. *B* is relevantly similar to *A*, conclude *C*). Lists of argumentation schemes occur in argumentation theory (cf. e.g. Walton 1996). The analogizing and distinguishing of cases can also provide examples of argumentation schemes. Bench-Capon and Sartor's (2000) theory constructors seem to be closely related to argumentation schemes.

Argumentation schemes are used to construct arguments: instances of the schemes can be chained to form arguments, just as derivations are constructed from the rules of inference of standard logic. (The term 'argument' is used somewhat differently in section 3.)

The properties of argumentation schemes can differ strongly from those of the rules of inference of classical logic. First, argumentation schemes can be *defeasible*, in the sense that though generally valid and giving rise to good arguments, there can be exceptional circumstances under which that is not the case (see below). Second, argumentation schemes can be *contingent*, in the sense that it can depend on the particular circumstances whether they can be used to form good arguments (see section 4.3).

A basic example of a defeasible argumentation scheme is *Modus ponens* for an inconclusive rule, e.g.:

(1) From 1. As a rule, if *P*, then *Q*, and 2. *P*, conclude *Q*.

When the rule is inconclusive, i.e., when there can be circumstances that even though it is the case that *P*, still it is not to be concluded that *Q*, then the argumentation scheme is defeasible: an argument constructed using it can be defeated under such exceptional circumstances. Let's use a dedicated expression for such circumstances:

(2) There is an exception to the rule that if *P*, then *Q*.

What we want is that an argument supporting this conclusion, attacks an argument that is constructed using the scheme *Modus ponens*, in the sense that the former can make the latter non-justifying. For instance, the derivation

(3) As a rule, if Peter has violated a property right,
then Peter has committed a tort.
Peter has violated a property right.
Therefore: Peter has committed a tort.

should be attacked by the (mini-)derivation

> There is an exception to the rule that if Peter has violated a property right, then Peter has committed a tort.

The exception statement can of course itself be supported by a non-trivial derivation, e.g., based on a rule that a case of *force majeure* is an exception to the rule that if Peter has violated a property right, then Peter has committed a tort.

The exception-statement (2) can be explicitly integrated into the scheme (1), e.g., as follows:

> From 1. As a rule, if *P*, then *Q*, and 2. *P*, conclude *Q*, unless there is an exception to the rule that if *P*, then *Q*.

This argumentation scheme, that might be dubbed *Modus ponens remittens*, is explicitly defeasible since it specifies how an instance (of its ordinary premises-conclusion part) can be attacked. In general, such defeasible argumentation schemes not only consist of *premises* (here 'As a rule, if *P*, then *Q*' and '*P*') and a *conclusion* ('*Q*'), but also of *attacks* ('There is an exception to the rule that if *P*, then *Q*').[15]

This conception of defeasible argumentation schemes (i.e., argumentation schemes with attacks) can be used to specify a 'concrete dialectical logic'. Such a specification consists of a *language* that constrains the sentence types and their composition, and a set of *defeasible argumentation schemes* as above that determine the possible derivations and the ways to attack them.[16] In the example, the language consists of two sentence types, viz. 'As a rule, if *P*, then *Q*' and 'There is an exception to the rule that if *P*, then *Q*', both containing sentence variables *P* and *Q*. When 'John is a thief', 'John is punishable' and 'John acted under force majeure' are elementary sentences, then 'As a rule, if John acted under force majeure, then there is an exception to the rule that if John is a thief, then John is punishable' is an example of a composite sentence that nests the two sentence types. In general, it is convenient when sentence types can contain not only sentence variables, but also term variables. With respect to the language it is noted that our rather liberal (but formally simple) setting has the effect that the use of sentence variables makes the expression of fact properties unproblematic. For instance, we could have the sentence type 'It is probable that *P*', where *P* is a sentence, or 'The rule that *R* is valid', where *R* is a sentence of a rule-like form, e.g., 'Thieves are punishable'.

It has not been elaborated how exactly it is to be determined which arguments of a particular concrete dialectical logic justify their conclusions and which not, given a set of premises. This will be achieved by embedding the specification of a concrete dialectical logic into DEFLOG.

4.2 Embedding a concrete logic into DEFLOG

It is not hard to embed concrete dialectical logics as they were introduced in section 4.1 into DEFLOG. Recall that a concrete dialectical logic was specified as a language and a set of defeasible argumentation schemes. In order to embed the concrete logic into DEFLOG, the first step is to consider the sentences of the concrete logic as elementary sentences of DEFLOG. As a result, DEFLOG does not 'see' the structure of these sentences. New

---

[15] Of course the analogy with Reiter's (1980) defaults is strong.

[16] If axioms are convenient for the specification of some concrete logic, they can be regarded as zero-premise argumentation schemes.

sentences become expressible, that were not part of the language of the concrete logic. For instance, the sentence 'Peter has violated a property right $\leadsto$ Peter has committed a tort' is expressible in the DEFLOG embedding of a language that itself contains the sentences 'Peter has violated a property right' and 'Peter has committed a tort'. These new sentences can be used to embed the defeasible argumentation schemes of the concrete logic into DEFLOG. The idea is straightforward. Let

> From 1. φ_1, 2. φ_2, ..., and *n*. φ_n, conclude ψ, unless
> χ_1, χ_2, ..., or χ_m.

be an instance of a defeasible argument scheme of the concrete logic. Then the DEFLOG embedding of the logic must contain the following sentences (recall the notational convention on conjunctive antecedents of section 3.3):

$\varphi_1 \wedge \varphi_2 \wedge ... \wedge \varphi_n \leadsto \psi$
$\chi_1 \leadsto \times(\varphi_1 \wedge \varphi_2 \wedge ... \wedge \varphi_n \leadsto \psi)$
$\chi_2 \leadsto \times(\varphi_1 \wedge \varphi_2 \wedge ... \wedge \varphi_n \leadsto \psi)$
...
$\chi_m \leadsto \times(\varphi_1 \wedge \varphi_2 \wedge ... \wedge \varphi_n \leadsto \psi)$

For instance, the embedding of the scheme *Modus ponens remittens* consists of all instances of the following two schemes:

> As a rule, if *P*, then $Q \wedge P \leadsto Q$
> There is an exception to the rule that if *P*, then *Q* $\leadsto \times$(As a rule, if *P*, then $Q \wedge P \leadsto Q$)

The effect is as expected. Normally, the conclusion that ψ is justified in a dialectical interpretation of a theory (using DEFLOG's terminology) when the scheme's premises φ_1, φ_2, and φ_n are justified. However, when one of the attacks χ_i is justified, then ψ will not be justified on the basis of the justified premises φ_1, φ_2, and φ_n, since then the conditional with that effect, viz. $\varphi_1 \wedge \varphi_2 \wedge ... \wedge \varphi_n \leadsto \psi$, is defeated. Note that by embedding the language and argumentation schemes of a concrete dialectical logic, also its arguments have a natural embedding: the embedding of an argument is the set of sentences that results from the embedding of the argument's premises and of the instances of the argumentation schemes used in it. For instance, the embedding of derivation (3) consists of the following three sentences:

> As a rule, if Peter has violated a property right, then Peter has committed a tort.
> Peter has violated a property right.
> As a rule, if Peter has violated a property right, then Peter has committed a tort $\wedge$ Peter has violated a property right $\leadsto$ Peter has committed a tort.

By embedding our concrete dialectical logics into DEFLOG, we have achieved two things. First, the embedding formally elaborates when the arguments of a concrete dialectical logic justify their conclusion, and when they don't. The answer is simply that an argument justifies its conclusion when all sentences of its DEFLOG embedding are justified, and otherwise it does not. Second, a side effect of the embedding is that it becomes clear how to argue about a concrete logic, for instance about the validity of its argumentation schemes, as will be seen next.

4.3 Arguing about argumentation schemes

There can be circumstances when it turns out that the argumentation schemes of a particular logic are it issue (cf. the end of section 2.4, and McBurney and Parsons 2000). Continuing our small example of the *Modus ponens remittens* logic, it can become clear that a new kind of attack has to be added, for instance that there are outweighing reasons against the rule's conclusion. Arguably, this would require a change of the concrete logic, resulting in an adaptation of *Modus ponens remittens*:

> From 1. As a rule, if *P*, then *Q*, and 2. *P*, conclude *Q*, unless 1. There is an exception to the rule that if *P*, then *Q*, or 2. There are reasons against *Q* outweighing *P*.

An argument can also be about the whole scheme. The typical example is reasoning by analogy, e.g., molded into an argumentation scheme as follows:

(4) From 1. By the application of rule *R*, *A* follows from *C*, 2. *B*, and 3. *B* is relevantly similar to *A*, conclude *C*.

In criminal law, such a scheme is considered to be unacceptable, in civil or administrative law, it can be acceptable, but not under all circumstances. For instance, it can be argued that when a rule burdens citizens, it should not be applied analogically. As a result, whether an argumentation scheme is acceptable can depend on the circumstances.

These kinds of arguing about argumentation schemes are formally unproblematic in our setting of concrete logics embedded in DEFLOG. For instance, that in criminal law analogy is not acceptable would take a logical form like $\varphi \leadsto \times(\psi_1 \wedge \psi_2 \wedge \psi_3 \leadsto \chi)$, where φ stands e.g. for 'The case at hand is a case of criminal law' and the ψ_i are (instances of) the three premises of the argumentation scheme (4) and χ its (instantiated) conclusion.

By the embedding of concrete logics in DEFLOG, it has become possible to consider the concrete logic as a *temporarily fixed filter* that keeps DEFLOG's particulars out of view, but that can be taken away (perhaps in part), when appropriate in order to argue about that part of the logic. It can well be the case that after an adaptation of the concrete logic, the adapted version is used as the new filter, for as long as is convenient.

By the embedding of concrete logics in DEFLOG, the *validity of a concrete argumentation scheme* can now be identified with the dialectical justification of the statements that form the scheme's embedding. The required critical discussion of the scheme can take place in DEFLOG.

At the end of section 2.4, it was briefly mentioned that an argument can also be about *which* argumentation scheme is used. For instance, an argument 'John is a thief. Therefore John is punishable' could be said to be *actually* an elaborate case of rule application, perhaps in the style of Reason-Based Logic. When Reason-Based Logic is described as a concrete logic embedded in DEFLOG (see Verheij 2000b, for one way to do that) and is used as a filter hiding DEFLOG, the only way to formalize the simple argument is in the elaborate way of the concrete logic: the simple argument must be reshaped to fit the logic. When however the concrete logic is not used as a filter, but is seen as embedded in DEFLOG, then it becomes possible to consider the reshaping of the argument into its elaborate form as a *gradually constructed theory* about the argument. For instance, initially the theory might consist of two sentences, viz. φ and $\varphi \leadsto \psi$, where φ stands for 'John is a thief', and ψ for 'John is punishable'. Now the reshaping could for instance start by arguing that the argument is warranted by the rule that thieves are punishable. When χ stands for 'Thieves are punishable', then the changed theory about the argument would consist of φ, χ and $\chi \leadsto (\varphi \leadsto \psi)$. In this way, shaping an argument into a given concrete logic becomes the gradual construction of a theory about the argument, during which the argument is gradually elaborated, e.g., by specifying a possible backing.

5. Return to the critique

How does the approach to the formal modeling of legal decision making presented here answer the critical points observed in section 2? We summarize the answers given throughout the paper.

With respect to the *logic of argument defeat*, a simple and straightforward way of modeling important types of attack (like undercutters) has been proposed in the dialectical logic DEFLOG (cf. especially section 3.3). DEFLOG itself clarifies how Prakken's logical and dialectical layer can be sensibly related. The *Modus ponens* validating conditional $\leadsto$ corresponds to the former layer, dialectical negation $\times$ the latter. The view on dialectical arguments (section 3.1) and dialectical interpretation (section 3.2) shows the intertwining of the two layers.

With respect to the *logic of rules*, the treatment of DEFLOG's conditional $\leadsto$ showed that with respect to their defeasible aspects expressing rules as sentences works fine (even for constitutive rules). The liberal but formally simple approach to the language of concrete logics (as in section 4.1) showed how the need to express fact and rule properties can still be addressed.

With respect to the *logic of law*, a view of embedding concrete logics in an abstract logic like DEFLOG was presented (section 4.2) that showed how a concrete logic can be regarded as a temporarily fixed filter specifying a concrete kind of valid reasoning. By a partial release of the filter, elements of the concrete logic can be put at issue (section 4.3). Moreover, the validity of concrete argumentation schemes can be identified with the dialectical justification of the sentences that form the scheme's embedding. The actually occurring different uses of relevant terms like 'rule' and 'reason' suggest a modular approach to their modeling incorporating different, perhaps mutually exclusive, uses of the terms. Defeasible argumentation schemes are well-suited for that purpose.

With respect to *empirical adequacy*, the treatment of concrete argumentation schemes as suggested here allows the empirically adequate modeling of legal decision making. It has been discussed how a simple argument like 'John is a thief. Therefore John is punishable' can be modeled in relation with elaborations of it, in which for instance the argument's backing is made explicit (section 4.3).

With respect to the *heuristic value*, it can now be seen that both deductive reasoning and reasoning by analogy can be treated as reasoning on the basis of dedicated, concrete argumentation schemes (section 4.1). Since all argumentation schemes can be used both for suggesting premises and for drawing conclusions, no fundamental difference between deductive and analogical reasoning follows (from this perspective).

6. Conclusion

The present approach to the formal modeling of legal decision making involves two central elements. First, the abstract logic DEFLOG has been proposed as a 'bottom line dialectical logic', i.e., one that just allows the modeling of dialectical argument by means of dialectical theory construction. Second, contingent, defeasible argumentation schemes have been embedded in the bottom line logic for the flexible and empirically adequate modeling of concrete kinds of reasoning, such as statute-based and precedent-based reasoning in the law. In this way, a number of issues neglected by other approaches are addressed.

References

Bench-Capon, T., and Sartor, G. (2000). Using values and theories to resolve disagreement in law. *Legal knowledge and information systems. JURIX 2000: The Thirteenth Annual Conference* (eds. J. Breuker, R. Leenes and R. Winkels), pp. 73-84. IOS Press, Amsterdam.

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

Hage, J.C. (1997). *Reasoning with Rules. An Essay on Legal Reasoning and Its Underlying Logic.* Kluwer Academic Publishers, Dordrecht.

Hage, J.C. (2000). Dialectical Models in Artificial Intelligence and Law. *Artificial Intelligence and Law*, Vol. 8, p. 137-172.

McBurney, P., and Parsons, S. (2000). Tenacious Tortoises: A formalism for argument over rules of inference. *Computational Dialectics (ECAI 2000 Workshop).* Berlin, Germany.

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

Prakken, H. (1997). *Logical Tools for Modelling Legal Argument. A Study of Defeasible Reasoning in Law.* Kluwer Academic Publishers, Dordrecht.

Prakken, H., and Sartor, G. (1996). A Dialectical Model of Assessing Conflicting Arguments in Legal Reasoning. *Artificial Intelligence and Law*, Vol. 4, pp. 331-368.

Prakken, H., and Sartor, G. (1998). Modelling Reasoning with Precedents in a Formal Dialogue Game. *Artificial Intelligence and Law*, Vol. 6, pp. 231-287.

Reiter, R. (1980). A Logic for Default Reasoning. *Artificial Intelligence*, Vol. 13, pp. 81-132.

Soeteman, A. (1989). *Logic in Law.* Kluwer Academic Publishers, Dordrecht.

Toulmin, S.E. (1958). *The uses of argument.* University Press, Cambridge.

Verheij, B. (1996). *Rules, Reasons, Arguments. Formal studies of argumentation and defeat*. Dissertation Universiteit Maastricht.

Verheij, B. (1999a). Automated Argument Assistance for Lawyers. *Proceedings of the Seventh International Conference on Artificial Intelligence and Law*, pp. 43-52. ACM, New York.

Verheij, B. (1999b). Logic, context and valid inference. Or: Can there be a logic of law? *Legal Knowledge Based Systems. JURIX 1999: The Twelfth Conference* (eds. H.J. van den Herik, M.-F. Moens, J. Bing, B. van Buggenhout, J. Zeleznikow, and C.A.F.M. Grütters), pp. 109-121. Gerard Noodt Instituut, Nijmegen.

Verheij, B. (2000a). Book review: Formalism and interpretation in the logic of law. (Review of H. Prakken' s ' Logical Tools for Modelling Legal Argument' .) *Artificial Intelligence and Law*, Vol. 8, pp. 35-65.

Verheij, B. (2000b). DefLog - a logic of dialectical justification and defeat. *Draft*. See http://www.metajur.unimaas.nl/~bart/.

Verheij, B. (2000c). Dialectical Argumentation as a Heuristic for Courtroom Decision-Making. *Rationality, Information and Progress in Law and Psychology. Liber Amicorum Hans F. Crombag* (eds. P.J. van Koppen and N.H.M. Roos), pp. 203-226. Metajuridica Publications, Maastricht.

Walton, D.N. (1996). *Argumentation Schemes for Presumptive Reasoning.* Lawrence Erlbaum Associates, Mahwah.