TELEOLOGICAL REASONING IN REASON-BASED LOGIC

Jaap Hage
University of Limburg
Department of Metajuridica
P.O. Box 616
6200 MD Maastricht
+31 43 88 3053 / 3020
email: jaap.hage@metajur.rulimburg.nl

ABSTRACT

This paper describes a version of Reason-Based Logic that has special facilities for teleological reasoning. The logic is specified by model-theoretic means, and is illustrated by means of an extended example. This example illustrates how RBL can be used to model the logical difference between legal rules on the one hand and legal principles and goals on the other hand. The appendices give more details on the ontological presuppositions of RBL and on the phenomenon of 'deontic collapse'.

Keywords

Reason-Based Logic, reasoning with goals, logical characterisation of rules vs. goals and principles

1. INTRODUCTION

Legal arguments are not only based on cases, legal rules or principles. Teleological reasoning, in which goals, policies, rights and interests play a role, is also very important [Summers 1978]. In the field of Law and AI, teleological reasoning in one form or another is a topic in [McCarty 1980, Koers and Kracht 1991, and Sanders 1991]. This paper describes a version of Reason-Based Logic (RBL) that not only deals with legal rules and principles, but also with goals. Moreover, this paper contains a model-theoretic characterisation of the conclusions that can validly be drawn from an RBL-theory. After a short introduction of RBL I will illustrate the possibilities of RBL concerning teleological reasoning with an extended example that also deals with the logical difference between legal rules on the one hand, and legal principles and goals on the other hand.

A more elaborate description, an informal characterisation, and more examples of the applications of RBL can be found in [Hage and Verheij 1994] and (especially) in [Hage and Verheij *to appear*]. These papers do not deal with goals, however. Moreover, next to some minor

changes, they define valid conclusions in RBL by means of extensions of RBL-theories, rather than by means of model theory. An early description which emphasises the interconnection with case-based reasoning is [Hage 1993], while [Verheij and Hage 1994] shows how RBL can be used to model analogous rule application.

2. AN INFORMAL INTRODUCTION TO REASON-BASED LOGIC

RBL is a nonmonotonic logic which was especially developed for reasoning with rules and in particular for legal reasoning. Its main sources of influence are the work of Raz on practical reasoning [e.g. Raz 1975], Toulmin's distinction between statements and inference rules (warrants) [Toulmin 1958], the work of Naess on argumentation theory [Naess 1966] and Dworkin's analysis of the distinction between legal rules and legal principles [Dworkin 1978]. Moreover, RBL was inspired by typical legal forms of reasoning such as analogous rule application, weighing reasons, and reasoning about the validity and the application of rules.

The basic idea of RBL is that the application of a rule[1] only leads to a reason which pleads for the rule's conclusion. Similarly the 'application' of a goal leads to a reason which pleads for the presence of a state that contributes to that goal, or for performing an action that contributes to the goal. The actual derivation of a conclusion is based on weighing all the reasons that plead either for or against this conclusion. As a consequence, derivation is a two step-procedure. The first step consists of the determination of all reasons that plead for or against the possible conclusion; the second step consists of weighing these reasons.

2.1 Weighing reasons

Suppose we have the following rule and goal:

Rule *theft*: If somebody is a thief, he ought to be imprisoned.
Goal *youth-protection*: Young persons are to be protected from harmful environments.

If we want to infer whether John should be imprisoned, we should collect all the reasons that plead for and against his imprisonment. If John is a thief, application of the rule

---

[1] The notion of a 'rule' is used here in a broad sense, that also includes legal principles.

theft makes this fact into a reason why John ought to be imprisoned. If John is still young, the goal to protect the youth makes the fact that imprisonment of John places him in a harmful environment into a reason against imprisoning John. These reasons have to be weighed to determine whether John ought to be imprisoned.

To determine which reason wins, we need more information. Such information is supplied by an additional premise, for instance that the reason that imprisonment of a young persons places him in a harmful environment outweighs the reason that he is a thief where the issue of imprisonment is involved. Given this 'weighing knowledge' it is possible to derive that John ought not to be imprisoned.

If we do not have any information about the relative weight of the reasons for and against a conclusion, it is not possible to infer anything. There is, however, one important case in which we need no explicit weighing knowledge, namely when there are only reasons for a conclusion, or only reasons against a conclusion. If, for instance, only the rule *theft* would be applied, we only have a reason for imprisoning John, and no reasons against imprisoning him. In such a case it is possible to infer that John ought to be imprisoned, without the need for weighing knowledge.

2.2 Teleological reasoning

The basic idea of teleological reasoning is that if performing some action contributes to the achievement of a goal, this is a reason for performing that action, or - in logical terminology - for the truth of the sentence that this action ought to be performed. For instance, if continuation of a rent contract with the new owner of a house contributes to the protection of the lessees, this is a reason why the rent contract is to be continued.

This same example can also be described in terms of a state that contributes to another state. The state that the rent contract is continued contributes to the state where the lessees are protected.

There are intricate relations between these two ways of describing the situation, the discussion of which falls outside the scope of this paper. My purpose here is to demonstrate the facilities which RBL offers to deal with teleological reasoning, and for that purpose I will describe a mechanism that can deal with both actions and states that contribute to goals.

Teleological reasoning does not only deal with means to achieve a goal, but also with avoiding situations that have a negative influence on the achievement of the goal. Actions and states that detract from a goal are to be avoided, and consequently there is a reason against performing such actions and having those states. If a prohibition of racist propaganda infringes the right of free speech, this is a (non-decisive) reason against having such a prohibition.

2.3 Exclusionary reasons and the relevancy of goals

Until now we have assumed that the rule *theft* was applied and that the goal *youth-protection* was relevant. Normally a rule is applied if its conditions are satisfied. This means that the rule *theft* normally is applied if John is a thief. Similarly, the goal *youth-protection* is relevant if some act protects John from a harmful environment.

However, the application of a rule can be excluded. Suppose that we have the rule

Rule *prescription*: Nobody is to be imprisoned if his crime is prescribed.

Since the rule *prescription* prevails over rule *theft* (this should be explicitly stated in an additional premise), the application of rule *theft* is excluded by the applicability of the rule *prescription*. As a consequence, the fact that John is a thief does not even become a reason why John ought to be imprisoned.

Similarly, the goal *youth-protection* is not relevant if there is a state of emergency. This can be stated in

Rule *emergency*: In cases of emergency, the goal *youth-protection* is not relevant.

Goals which are not relevant for a particular case do not generate reasons concerning that case. So if there is an emergency, the goal *youth-protection* does not generate a reason against John's being imprisoned.

Clearly, if a reason is not generated, the issue of weighing reasons does not even arise concerning that reason.

3. SPECIAL FUNCTION AND PREDICATE SYMBOLS OF REASON-BASED LOGIC

RBL is based on first order predicate logic (FOPL) in the sense that the language of RBL is the language of FOPL, and that the theorems of FOPL are all true in RBL.

RBL uses a rich ontology, that includes actions[2], states of affairs, facts, and reasons.[3] A *state of affairs* is that part of reality that is expressed by a closed sentence. For instance, the sentence Thief(john) expresses the state of affairs that John is a thief. A *fact* is a state of affairs that is denoted by a *true* sentence. False sentences express states of affairs that do not actually obtain; facts are states of affairs that actually obtain. A *reason* is a fact that has a particular significance for a conclusion, in that it either pleads for or against it. From a logical point of view, all states of affairs, including facts and reasons, are individuals.

In RBL-theories it is both necessary to use sentences and to refer to states of affairs (facts, reasons) that are expressed

---

2   Actions are action-types, such as imprisoning somebody, or - more concretely - imprisoning John. A more comprehensive theory of teleological and deontic reasoning should also deal with action tokens such as this particular case of imprisoning John.

3   Some ontological presuppositions of RBL are exposed in Appendix 1.

by these same sentences. Moreover, the logical connection between these sentences and the states of affairs that they express should be maintained. For instance, we want both to argue that John is a thief, and to refer to the fact that John is a thief as a reason for imprisoning him. Since predicates can only have terms as arguments, we need a simple means to translate sentences into terms that denote the states of affairs expressed by these sentences, and back. For this purpose, predicate symbols are written as strings of characters that begin with an uppercase letter, and function symbols (including terms) as strings of characters beginning with a lowercase letter. To obtain the term that corresponds to a sentence, the first (uppercase) letter of each predicate symbol in the sentence is replaced by the same letter in lowercase.[4] For example, the states of affairs expressed by the sentences

Thief(john)
Guilty(john) & Ought(imprison(john))

are respectively referred to by the terms
thief(john)
guilty(john) & ought(imprison(john)).

Actions are denoted by functions, such as steal, or imprison(john).

The language of RBL is that of FOPL, but contains a number of special function and predicate symbols, that is rule/3[5], goal/2, { ., ., ..., . }/n (for n = 1, 2, ...), reasons_pro/1, reasons_con/1, do/1, ~do/1, O/1, Ought/1, Valid/1, Accepted/1, Excluded/3, Applicable/3, Contributes_to/2, Relevant/2, Applies/3, Reason/3, and Outweighs/3.

- rule/3 (and rule/1)

In RBL rules (and principles) are denoted by *terms* of the language. In this way it is possible to refer to them and to reason about them. A term denoting a rule or principle has the form:[6]

rule(*rid, condition, conclusion*)

Here *condition* is a sentence of RBL and *conclusion* a literal of RBL. We assume that *condition* is a disjunction of conjunctions of one or more literals. Each disjunct of *condition* is a possible reason for *conclusion*.

The first argument of a rule, *rid*, is called the identifier of the rule. It is assumed that each rule has a unique identifier.

The term rule(*rid*) is used as an abbreviation of the term rule(*rid, condition, conclusion*). Because the identifier of a rule is unique, this does not lead to confusion.

- goal/2 (and goal/1)

Just like rules, goals are denoted by terms. The parameters of a goal specify respectively the unique identifier of the goal and the state (of affairs) that is to be obtained: goal(*gid, state*)

state is described by means of an RBL-term that denotes a state of affairs, for instance goal(protection_john, protected_from_harmful_environment(john)).

The term goal(*gid*) is used as an abbreviation of the term goal(*gid, state*).

- { ., ., ..., . }/n (for n = 0, 1, 2, ...)

These symbols are used to refer to sets of states of affairs, most often reasons.

- reasons_pro(atom) = {r | Reason(r, atom, pro)
  reasons_con(atom) = {r | Reason(r, atom, con)

The function symbols reasons_pro(*atom*) and reasons_con(*atom*) have as their values the sets of all reasons that plead for respectively against *Atom*.

- do/1 and ~do/1

do/1 and ~do/1 are functions that operate on action types. do(*action*) denotes the doing of *action*. ~do(*action*) denotes the refraining from *action*. do/1 and ~do/1 are normally used in combination with the Ought/1-predicate.

- O/1 is a deontic predicate for ought-to-be. It can be applied iteratedly.[7]

O(*state*) means that the state of affairs *state* ought to be the case.

- Ought/1

Ought is a deontic predicate for ought-to-do. Ought(do(*action*)) means that an *action* ought to be performed. Ought(~do(*action*)) means that an *action* ought not to be performed.

- Valid/1

The sentence Valid(rule(*rid*)) means that the rule with identifier *id* is valid.

- Accepted/1

The sentence Accepted(goal(*gid*)) means that the goal *gid* is accepted. Acceptance is for goals what validity is for rules.

- Excluded/3

The sentence Excluded(rule(*rid*), *Iconds, Iconcl*) means that the rule with identifier *id* is excluded, where *Iconds* and *Iconcl* are respectively the instantiated conditions and conclusion of the rule.

- Applicable/3

The sentence Applicable(rule(*rid*), *facts, conclusion*) means

4   The connectives of FOPL, e.g. → and &, are treated as if they are also function symbols. By overloading the notation, the translation of sentences to terms is as simple as described in the text.

5   The number following / indicates the arity of the function or predicate symbol.

6   Metavariables for sentences will be denoted by strings of italic characters beginning with an upper case character, e.g., *Atom*. Metavariables for terms will be denoted as strings of italic lower case characters, e.g., *atom*. I use the convention that matching metavariables, such as *Atom* and *atom*, represent a sentence and its corresponding term.

7   Appendix 1 contains some additional remarks on the meaning of the deontic predicates.

that the rule with identifier *rid* is made applicable by the facts denoted by the term *facts* and may generate a reason for the conclusion denoted by the term *conclusion*.

- Contributes_to/2

The sentence Contributes_to(*action*, *state*) means that the an action of the type *action* contributes to the goal *state*. If the goal *state* is accepted and relevant, this means that there is a reason to perform an *action*, or a reason for the truth of the sentence which expresses that an action ought to be performed: Ought(do(*action*)).

The sentence Contributes_to(*state1*, *state2*) means that the state *state1* contributes to the goal *state2*. If the goal *state2* is accepted and relevant, this means that there is a reason for the presence of the state of affairs *state1*, or for the truth of the sentence that expresses that it ought to obtain: O(*state1*).

The sentence Contributes_to(*action*, ~*state*) means that the action *action* detracts from the goal *state*. If the goal to achieve *state* is accepted and relevant, this means that there is a reason not to perform an *action*, or for the truth of the sentence which expresses that no action ought to be performed: Ought(~do(*action*)).

The sentence Contributes_to(*state1*, ~*state2*) means that the presence of *state1* detracts from the goal *state2*. If the goal to achieve *state2* is accepted and relevant, this means that there is a reason against the presence of state of affairs *state1*, or for the truth of the sentence that expresses that it ought not to obtain: O(~*state1*).

- Relevant/2

The sentence Relevant(goal(*gid*), *action*) means that the goal *gid* is relevant for the performance of an *action*.

The sentence Relevant(goal(*gid*), *state*) means that the goal *gid* is relevant for the presence of the state of affairs *state*.

Relevancy is to goals what non-exclusion is to rules. Goals only generate reasons if they are relevant.

- Applies/3

The sentence Applies(rule(*rid*), *facts*, *conclusion*) means that the rule with identifier *rid* applies on the basis of the facts denoted by the term *facts* and generates a reason for the conclusion denoted by the term *conclusion*. E.g. Applies(rule(theft), thief(john), o(punished(john))) means that the rule theft applies with its conditions instantiated to thief(john), and its conclusion instantiated to o(punished(john)).

- Reason/3

The sentence Reason(*facts*, *atom*, pro) means that the facts denoted by the term *facts* are a reason for the conclusion denoted by the term *atom*. The sentence Reason(*facts*, *atom*, con) means that *facts* are a reason against *atom*.

- Outweighs/3

The sentence Outweighs(*reasons1*, *reasons2*, *atom*) means that the reasons in the set denoted by the term *reasons1* outweigh the reasons in the set denoted by the term

reasons2 (as reasons concerning *atom*). The terms *reasons1* and *reasons2* must both have the form { *facts1*, *facts2*, ..., *factsn* }, with n > 0.

4. VALID CONCLUSIONS

A conclusion can validly be drawn from an RBL-theory if and only if it is true in all *preferred* RBL-models of that theory. An RBL-model of an RBL-theory T is an RBL-possible world in which all sentences of T come out true. An RBL-possible world is subject to the usual constraints that hold for model theoretical semantics of FOPL[8] and the following ones which especially hold for RBL[9]:

1. Assume that the sentence *Facts* is an instance of one of the disjuncts of the sentence *Condition* under some substitution σ, and that the term *iconcl* is the instance of the term *conclusion* under σ.

   If Valid(rule(*rid*, *condition*, *conclusion*)), and *Facts* are true, and if Excluded(rule(*rid*), *facts*, *iconcl*) is false, then Applicable(rule(*rid*), *facts*, *iconcl*) is true.

   If Applicable(rule(*rid*), *facts*, *atom*) is true, then Reason(applicable(rule(*rid*), *facts*, *atom*), applies(rule(*rid*), *facts*, *atom*), pro) is also true.[10]

2. a) If Accepted(goal(*gid*, *state*)), Contributes_to(*action*, *state*), and Relevant(goal(*gid*), *action*) are true, then Reason(contributes_to(*action*, *state*), ought(do(*action*)), pro)) is true.

   b) If Accepted(goal(*gid*, *state2*)), Contributes_to(*state1*, *state2*), and Relevant(goal(*gid*), *state1*) are true, then Reason(contributes_to(*state1*, *state2*), o(*state1*), pro) is true.

   c) If Accepted(goal(*gid*, *state*)), Contributes_to(*action*, ~*state*), and Relevant(goal(*gid*), *action*) are true, then Reason(contributes_to(*action*, ~*state*), ought(~do(*action*)), pro) is true.

   d) If Accepted(goal(*gid*, *state2*)), Contributes_to(*state1*, ~*state2*), and Relevant(goal(gid), *state1*) are true, then Reason(contributes_to(*state1*, ~*state2*), o(~*state1*), pro) is true.

---

[8] Such constraints on mere (predicate) logically possible worlds can e.g. be found in Lukaszewicz 1990, definition 1.25.

I consider the way in which the truth value of a compound sentence depends on the truth values of its more elementary parts as a constraint on logically possible worlds. E.g. it is a constraint on logically possible worlds that the sentence A&B is true, if and only if the sentences A and B are both true. The constraints discussed in this section are constraints on RBL-possible worlds, a subset of logically possible worlds. Readers who prefer to work with an accessibility-relation might want to say that the constraints specify which worlds are accessible. Valid conclusions are then to be defined as sentences that are true in all preferred accessible logically possible worlds.

[9] In the following, it is assumed that *Atom* denotes an RBL-atom. Moreover, free variables are assumed to be bound under universal quantification.

[10] This definition contains a minor deviation from earlier versions of RBL as described in [Hage and Verheij 1994, and *to appear*]

3 a) If Applies(rule(*rid*), *facts*, *atom*) is true, then
   Reason(*facts*, *atom*, pro) is also true.

   b) If Applies(rule(*rid*), *facts*, ~*atom*) is true, then
   Reason(*facts*, *atom*, con) is also true.

4 a) If Outweighs(reasons_pro(*atom*),
   reasons_con(*atom*), *atom*) is true, then *Atom* is
   true.

   b) If Outweighs(reasons_con(*atom*),
   reasons_pro(*atom*), *atom*) is true, then *Atom* is
   false.

5 Outweighs({*facts₁*, ..., *factsₙ* }, ∅, *atom*) with $n > 0$, is
   true.

Preferred RBL-models are those RBL-models that are not
less preferred than any other RBL-model. An RBL-model
M1 is less preferred than an RBL-model M2 if either one
of the following is the case:

a. any sentence of the form Relevant(goal(*gid*),
   action/state) is false in M1 and true in M2.

b. any sentence of the form Excluded(rule(*rid*), *iconds*,
   iconcl)) is true in M1 and false in M2.

c. any sentence of the form Reason(*facts*, *atom*, pro/con)
   is true in M1 and false in M2.[11]

This preference-relation maximises the number of relevant
goals, and minimises the number of reasons and excluded
rules.

5. SOME CHARACTERISTICS OF RBL

Since RBL is an extension of FOPL; it allows all
inferences that are possible in FOPL. In addition it has
special facilities to reason defeasibly with both rules and
goals.

As a nonmonotonic logic for legal reasoning, RBL is
comparable to, amongst others, the work described in
[Prakken 1993; Sartor 1994; Gordon 1994].

The possibility that rules are excluded and that goals are
irrelevant corresponds to the use of undercutters [Pollock
1987]. Weighing reasons can be compared to the effect of
rebutters, with the understanding that weighing reasons is
more powerful because it cannot only deal with a conflict
between a pair of reasons (defaults or arguments in other
theories), but also between sets of them.

Since rules and goals are individuals from a logical point
of view, it is possible to argue about their validity
(acceptability), their exclusion and applicability
(relevancy), and their application. These arguments can all
use the full potential of RBL for dealing with reasons.

Because it is possible to argue about rule exclusion on the
basis of the same logical mechanism that holds for other
arguments, conflict rules and their (recursive) conflicts can

be dealt with easily. Moreover, since the information about
the relative weights of sets of reasons is contained in
ordinary sentences. it is possible to argue about this
knowledge in the same way too. Such arguments are
comparable to arguments about the ordering of defaults or
arguments in other logics [Brewka 1994; Prakken 1995].
Since it is possible to have reasons for the application of a
rule other than that the rule conditions are satisfied, RBL
can deal with analogous rule application [Verheij and
Hage 1994].

RBL-conclusions obey the rules designated as a default
reasoning core in Geffner and Pearl [1992].

RBL has no built-in priorities, nor does it allow backward
reasoning with rules. or the derivation of rules in the
fashion of a Hypothetical Syllogism.[12]

Finally, RBL has a model theoretic semantics. However, it
presently lacks constructive rules of inference.

6. AN APPLICATION OF RBL

It is not possible to illustrate all possibilities of RBL in a
conference-paper. Therefore I will only give one extended
example that combines reasoning with rules and with
goals. and that illustrates how the phenomenon of
exclusionary reasons makes the logical difference between
legal rules and principles understandable. The example I
have in mind is a polished, but realistic representation of a
part of Dutch law that deals with the transfer of property
by non-owners.

6.1 Colliding reasons

There is a legal principle that deals with the transfer of
property by non-owners, namely the principle that nobody
can transfer a right that he does not have. Let us translate
this principle into the language of RBL as follows[13]:

Valid(rule(nemo_plus,
   transfers(*x*, *y*, *z*) & ~owner(*x*, *z*),
   ~o(owner(*y*, *z*))))

---

11 Notice that it is possible that both M1 is less preferred than M2, and M2
is less preferred than M1. In such a case, neither M1 nor M2 is a
preferred model.

12 For instance. it is not possible to use the rule that thieves are punishable
to derive that somebody is not a thief from that he is not punishable.

   It is possible to add rules to a domain theory, which would make one or
   more of the mentioned kinds of inferences possible. Cf. Hage and
   Verheij *to appear*.

13 Here we encounter a difficulty that has not yet been solved. The actual
   consequence of the principle is that the third party *is not* the new owner,
   rather than that he *ought not to be* the new owner. Yet, the goal, which
   is assumed to collide with the principle, has as its consequence that the
   third party ought to be the new owner. In other words, the questions
   whether the third party is the new owner and whether he ought to be the
   new owner seem to collapse in Dutch law. I discuss this phenomenon in
   Appendix 2.

   To overcome the resulting logical difficulties, I have given the legal
   principle in its formalised version the conclusion that it ought not be the
   case that the third party becomes the new owner. In this way the logical
   problems are 'solved' by cheating a bit in the formalisation of the
   principle. This same manoeuver is later on also used in the formalisation
   of the rule *transfer_of_mutables*.

15

(If x transfers z to y, and x was not the owner of z, then y should not become the owner of z.)

Moreover, in Dutch civil law, the goal is accepted that third parties in good faith are to be protected:

Accepted(goal(protection_good_faith,
    in_good_faith(y) → protected(y)))

Suppose we have the following facts:

Transfers(casey, joyce, book)
~Owner(casey, book)
In_good_faith(joyce))
Contributes_to(owner(joyce),
    in_good_faith(joyce)) → protected(joyce))

Since there is no reason to conclude that the nemo-plus principle is excluded, the principle is to be applied, and we obtain the reason:

Reason(transfers(casey, joyce, book) &
    ~owner(casey, book), o(owner(joyce, book)), con)

(The facts that Casey transfers the book to Joyce and that Casey was not the owner of the book are together a reason why it ought not be the case that Joyce becomes owner of the book.)

Since there is also no reason to conclude that the accepted goal that parties in good faith are to be protected is irrelevant, we also have

Reason(contributes_to(owner(joyce),
    in_good_faith(joyce)) → protected(joyce)),
    o(owner(joyce, book)), pro)

(The fact that if Joyce becomes the owner of the book, this contributes to the protection of parties in good faith, is a reason why it ought to be the case that Joyce becomes the owner of the book.)

To decide whether Joyce becomes the owner of the book, we need additional weighing knowledge, such as:

Outweighs(
    {contributes_to(owner(joyce),
        in_good_faith(joyce)) → protected(joyce))},
    {transfers(casey, joyce, book) & ~owner(casey, book)},
    o(owner(joyce, book)))

(As far as the question is concerned whether it ought to be that Joyce becomes the owner of the book, the reason that if Joyce becomes the owner of the book, this contributes to the protection of parties in good faith, outweighs the reason that Casey transfers the book to Joyce and that Casey was not the owner of the book.)

If the reason based on the protection of parties in good faith actually outweighs the reason based on the nemo-plus principle, it can be derived that Joyce should become the new owner of the book. However, whether judges will use this weighing knowledge in particular cases, is uncertain. Therefore, legal security asks for the introduction of rules that govern this situation.

6.2 Rules and principles

To increase legal security, the legislator might adopt the following rule:

Valid(rule(transfer_of_mutables,
    mutable(z) & transfers(x, y, z) & ~owner(x, z) &
        in_good_faith(y),
    o(owner(y, z))))

(If somebody, who is not the owner, transfers a mutable good to somebody else who acts in good faith, it ought to be the case that the latter party becomes the owner of the good.)

Moreover, since this rule is meant to replace the nemo-plus principle and the goal to protect parties in good faith, the following will also hold:

Replaces(rule(transfer_of_mutables), rule(nemo_plus))
Replaces(rule(transfer_of_mutables),
    goal(protection_good_faith))

(The rule *transfer_of_mutables* replaces both the nemo-plus principle and the goal *protection_good_faith*.)

If a principle or goal is replaced by a rule, it should not be applied to a case if the rule is applicable to that case. The following RBL-rules take care for that[14]:

Valid(rule(rule_replacement,
    applicable(rule(*rid1*),*conds1*, *concl1*) &
        replaces(rule(*rid1*), rule(*rid2*),
        excluded(rule(*rid2*),*conds2*, *concl2*)))

(If a rule that is applicable to a case, replaces some other rule, the latter rule is excluded in that case.)

Valid(rule(goal_replacement,
    applicable(rule(*rid*),*conds*, *concl*) &
        replaces(rule(*rid*), goal(*gid*),
    ~relevant(goal(*gid*), *concl*)))

(If a rule that is applicable to a case, replaces a goal, this goal is not relevant for that case.)

By excluding the nemo-plus principle and making the goal to protect parties in good faith irrelevant, the rule *transfer_of_mutables* monopolises the case to which it is applicable. That is, it generates a reason why the party in good faith becomes the owner and prevents the generation of other reasons that deal with the same issue. As a consequence the one reason needs not to be weighed against other reasons. In this way, although the rule only generates a reason for its conclusion, it seems as if the rule applies in an all-or-nothing fashion. This is more or less in accordance with the logical difference between legal rules and principles as pointed out by Dworkin. [Dworkin 1978, p. 24]. The accordance is only more or less, because to which extent a rule applies in an all-or-nothing fashion, depends on how many principles and goals are discarded by an applicable rule. For example, it can be argued that a rule only discards those principles and goals which were

---

taken into account in making that rule [Hage *forthcoming*].

7. CONCLUSION

RBL is a logic for defeasible legal reasoning. Despite its quite different conceptualisation of legal reasoning, it shares the important facilities of other such logics. In addition it can deal with weighing sets of reasons, and with reasoning about priorities. Moreover, it provides facilities for teleological reasoning, and a way to make the logical differences between rules and principles explicit. And finally, RBL can in a natural way be integrated with case-based reasoning [Hage 1993].

APPENDIX 1: THE ONTOLOGY BEHIND RBL

Introduction

Ontologies of domains are important for intelligent systems that operate in these domains [cf. Valente and Breuker 1994]. In this appendix I give a short exposition of the ontological assumptions behind RBL and the way in which they influenced the logic. This exposition must be short and is consequently rather apodictic, even where it is controversial. It cannot be expected to convince those who have strong different opinions, but it can, I hope, clarify my position to those who are interested in the logic described above. Moreover, it makes clear how RBL is not merely a technical device, thought up to deal with some peculiarities of legal reasoning, but is rather the result of a philosophical theory about the role of rules and reasons in (legal) thinking.

The world as mental construction

My starting point is that the ontological make up of the world we live in is to a large extent the work of the human mind. The world consists of objects and facts which fall into categories that are products of the mind. Not products which have been made up at will [Lorenz 1977], but nevertheless products that are not given with a mind-independent world [cf. Putnam 1976]. For example, without the human mind, there would not have been property rights, world championships, stocks, estuaries, or bad luck. This assumption that the constitutents of the world are a product of the human mind is the reason why the ontology of RBL shuns Occam's razor and liberally adopts a proliferation of entities.

The mind-dependent nature of facts is most conspicuous for those facts which I would call reason-based. Some facts seem to obtain almost independent of human influence, such as the fact that the seas are filled with water, that Earth turns around the Sun, etc. Other facts cannot exist independent of human culture, such as the facts that Michael Jordan is a basketball-player, the fact that I owe the grocer money for the potatoes I bought from him

[Anscombe 1956], the fact some particular behavior is rude [Foot 1958], and the fact that Jones is under an obligation to pay Smith five dollars [Searle 1969]. Some have connected these reason-based facts to social institutions, sometimes calling them *institutional facts* [MacCormick 1974 and MacCormick and Weinberger 1987]. However, unless the notion of an institution is taken very broadly, not all reason-based facts seem to be connected to institutions. Therefore I propose not to stick to that name.

Constitution

The presence of reason-based facts can only be established by means of other facts from which they are somehow 'derived'. For instance, a goal in a soccer match must be 'derived' from the fact that the ball passed the goal line. That my behavior was rude must be 'derived' from the nature of the behavior (slamming the door in someone else's face) and a given behavioral code which employs the notion of rudeness. And most notably, the existence of an obligation is established by the facts that generated the obligation (a promise, for instance), and - of course - a set of norms.

This phenomenon, that some facts can only exist thanks to other facts, their *supervenience* [Hare 1952, p. 80f.], manifests itself in connection with all value judgments, all modal sentences (anankastic, deontic, epistemic [White 1975]), and many classifications. E.g. a picture is only beautiful *thanks to* some of its other characteristics; an action is only forbidden *because* it belongs to a particular type or *because* it has certain consequences; the murderer of Smith must be insane, *given* the brutality of the murder; John is the owner of his house *because* he bought it from the builder.

In all of these cases, we can say that some facts *constitute* other facts. Constitution comes in two kinds. First, some facts can *amount to* some other fact. E.g. the specific nature of my behavior amounts to its being rude. Second, some facts constitute other facts because in some non-physical sense, they '*cause*' the existence of the other facts. E.g. my making a promise 'causes' the coming into existence of an obligation.

Constitution and derivation

I introduced the notion of constitution by showing that the presence of some facts had to be 'derived' from the presence of some other facts. The quotes around 'derived' indicate that in my opinion this use of the word is not completely satisfactory. In fact, I now want to distinguish between derivation and constitution. In the case of constitution, the very existence of the constituted facts is based on the constituting facts. For example, the goal in soccer only exists thanks to the fact that the ball passed the goal line. And I only have an obligation thanks to the

17

promise I made. The 'derivation' of the constituted facts from the constituting facts is the re-thinking of the relation of constitution (but cf. Hage *et al.* 1994 on the procedural nature of legal conclusions).

Often, however, we derive facts from other facts which do not constitute them, but which give evidence for their presence. This occurs, for instance, if we conclude that John is the murderer because he had a motive and the occasion. A motive and an occasion do not make John into a murderer; they are only indications that he might be the murderer. (However, these same facts constitute the epistemic modal fact that John is *probably* the murderer. The occasion and the motive *make it* probable that John committed the murder.)

The distinction between constitution and derivation can also be approached by means of a distinction between kinds of reasons. The reasons involved in constitution are reasons *why* something is the case. The reasons involved in derivation are reasons *to believe that* something is the case.

In the law, there is a corresponding distinction between legal issues and issues of fact. The Dutch Supreme Court can only judge about the legal issues of a case. not about the factual issues.

Rules and principles

If a fact A is a reason for the presence of fact B. facts like A are reasons for facts like B. The reason-giving relation exists primarily between kinds of facts. and only secondarily between individual facts. The expression of the general relation between reason-giving facts and the facts for which they are reasons is a principle or a rule. Rules (from now on subsuming principles) do not describe this relation; they express it.

Acceptance (or validity) of a rule comes down to it that certain kinds of facts are considered to be reasons for other facts. Sometimes the rule can only be reconstructed with hindsight, when some facts are recognized as reasons. This is the case when case law is 'restated' in the form of rules. Some other times the rule is there before the reasons. such as in the case of statutory legal rules which only make facts into legal reasons after the rule has been created by means of legislation.

Rules as instruments

Reasons are facts that are significant for the existence of other facts. They derive this significance from assignment by humans. In using rules. humans assign significance to certain kinds of facts in relation to other facts. For instance, they consider promises as reasons why one ought to do what one has promised. This connection between types of facts has a dispositional nature. Concrete occurrences of promises are usually assigned significance for the existence of an obligation. These concrete

assignments are applications of the rule that promises create obligations.

It is, however, possible to refrain from the assignment of significance in particular cases. This happens, for instance, if the promise was forced. In such cases, there is a reason not to apply the rule that promises create obligations. Cases like this make it clear that rules need to be applied. The mere existence of rules (their acceptance or validity) does not suffice to transform brute facts into reasons. Next to existence, they need application. Only if they are actually applied, they generate reasons. In this respect, rules differ from statements. If a universal statement is true, it is true for all the cases to which it refers. The issue of application cannot even arise.

Because application is a non-issue in the case of statements, reasoning with statements cannot be defeasible. A statement is either false and cannot be used in arguments to infer true conclusions, or it is true and guarantees the truth of the conclusion (in case of a valid argument). If reasoning is considered defeasible, this goes to show that something like a rule (or a principle or goal) was involved in the argument. Seeming cases of defeasible arguments with statements are cases where material inference rules are disguised as statements [Toulmin 1958].

Directions of fit

Searle [1975] distinguished speech acts which aim to fit the world from speech acts that aim to make the world fit the act. The former have the word to world direction of fit, while the latter have the world to word direction of fit. For instance, the direction of fit of *statements* is from the speech act to the world. Statements aim to correspond to the world. *Imperatives*, on the other hand have a different direction of fit: they aim to make the world correspond to the content of the imperative.

A similar distinction can be made with respect to statements and rules. Rules aim to make the world correspond to the contents of the rule. For example, the rule that thieves ought to be punished makes it the case that in the world thieves ought usually to be punished. (This characteristic of rules may be the cause of the error to consider rules as a kind of imperatives [e.g. Ross 1968, p. 48f.]. Not everything that makes the world fit itself is an imperative. Baptising. for instance, also makes the world fit the speech act.)

It is not the case that the contents of the rule depend on what is the case in the world (although the existence of rules depends on the world). Therefore, rules are not true or false. If a reason-giving connection exists, this is a fact which can be described by saying that the rule which expresses this relation is *accepted* (in the case of social rules) or *valid* (in the case of institutional rules). Sentences saying that rules are accepted or valid are true or false; the

rules that occur in them have no truth value and are properly considered as logical individuals.

The truth conditions of deontic modalities

The distinction between rules and statements holds in the deontic as well as in the non-deontic sphere. Deontic statements, as opposed to formulations of deontic rules or norms, express deontic states of affairs. If these states of affairs obtain, the statements are true, otherwise not. This means that the truth values of deontic statements, just like the truth values of all other statements, depends on the states of affairs that obtain in the *actual* world, and not in some ideal world.

Which states of affairs correspond to deontic statements? What in the world, for instance, corresponds to the statement that Jones ought to pay Smith five dollars? The answer is: The state of affairs that Jones ought to pay Smith five dollars. Deontic statements are in this respect similar to all other statements. To the sentence 'The table is round' corresponds the state of affairs that the table is round [cf. Davidson 1967]. The semantics of ought-statements does in this respect not differ from the semantics of other statements.

There is of course another difference. The sentence that the table is round can directly be verified by inspecting the table, while the sentence that Jones ought to pay Smith five dollars is not amenable to direct empirical verification. It should be noticed, however, that this difference in the method of verification is not a difference in semantics. The connection between verification and meaning appears to me as an outdated remnant of the logical empirist tradition [cf. Hempel 1950].

It seems attractive to have verifiable truth conditions for deontic statements. Such truth conditions can be given in the form of the facts that constitute the fact expressed in the deontic statement. More concrete, the verifiable truth conditions of deontic statements are given by the reasons why this particular deontic state of affairs obtains. For instance, the truth of the sentence that Jones ought to pay Smith five dollars is shown by pointing out that Jones promised Smith five dollars. Clearly such truth conditions do not give the meaning of the deontic statement, but they do indicate circumstances under which we hold the sentence to be true.

It is also possible to give general truth conditions for deontic statements. Actually these truth conditions hold for all facts which obtain because of their constituting reasons (all reason-based facts), and not only for deontic facts. These truth conditions are that a sentence S which expresses a state of affairs F is true if and only if the reasons which plead for F outweigh the reasons against F. It is these truth conditions which have been elaborated in this paper. Clearly they do not specify the meaning of deontic statements, but in my opinion the enterprise to

specify the meaning of sentences by truth conditions, otherwise than by re-using the same sentence needs reconsideration.

APPENDIX 2: DEONTIC COLLAPSE

The law is on the one hand a social phenomenon, and as such its contents are an empirical matter. On the other hand, the law is a teleological enterprise which aims to guide human behavior [Fuller 1969]. As such it is the result of practical thinking. As a consequence of this dual nature of the law, the questions what the law is and what the law should be cannot be entirely separated [contra: Hart 1958]. In particular in matters of interpretation and where the application of legal rules is concerned, legal reasoning is consequentialist [MacCormick 1978].

Consequentialist reasoning answers the question what we ought to (should) do; it is essentially deontic/practical. Yet the conclusion must be what the law *is*. For instance, we argue from the consequences of a particular interpretation of a legal text to the (in)correctness of that interpretation [Alexy 1978, p. 297]. The 'proper' conclusion of such an argument is which interpretation is to be preferred, which interpretation is to be chosen. The actual conclusion is which interpretation is legally correct, is 'the law'.

It is this dual nature of the law (and other normative systems) which, I think, causes the phenomenon of *deontic collapse*, the phenomenon that sometimes deontic reasoning leads to factual conclusions. It appears to have a counterpart in the phenomenon that might be called *deontic inflation*, where a factual argument about the contents of the law leads to a normative conclusion about what we ought to do. (Cf. Alexy's view that legal reasoning is a species of practical reasoning [Alexy 1978]) .

ACKNOWLEDGEMENTS

REFERENCES

Alexy, R. (1978). *Theorie der juristischen Argumentation*. Suhrkamp Verlag, Frankfurt am Main.

Anscombe, G.E.M. (1958). On Brute Facts, *Analysis*, vol. 18, p. 69-72.

Brewka, G. (1994). Reasoning about preference in default logic. Proceedings of the Twelfth National Conference on Artificial Intelligence, Seattle.

Davidson, D. (1967). Truth and Meaning. in D. Davidson, *Truth and Interpretation*, Clarendon Press, Oxford 1984.

Dworkin, R. (1978). *Taking Rights Seriously*, 2nd ed., Duckworth, London.

Foot, Ph. R. (1958). Moral Arguments. in Ph. R. Foot, *Virtues and Vices and other Essays in Moral Philosophy*, Basil Blackwell, Oxford 1978, p. 96-109.

Fuller, L.L. (1969). *The Morality of Law*, 2nd ed., Yale University Press, New Haven and London.

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

Gordon, Th. F. (1994): The Pleadings Game: An Exercise in Computational Dialectics, *Artificial Intelligence and Law* 2, pp. 239-292.

Hage, J.C. (1993). Monological Reason Based Logic. *Proceedings of the fourth International Conference on Law and Artificial Intelligence*, ACM-press, Amsterdam, p. 30-39.

Hage, J.C. (1994). Two Metaphors of Reasoning. H. Prakken, A.J. Muntjewerff, A. Soeteman and R.G.F. Winkels (eds.), *Legal knowledge based systems; the relation with legal theory*, Koninklijke Vermande, Lelystad, pp. 127-138.

Hage, J.C. (forthcoming). The two layers of the law (in Dutch).

Hage, J.C., R. Leenes, and A. Lodder (1994). Hard cases; a procedural approach. *Artificial Intelligence and Law*, vol. 2, pp. 113-167.

Hage, J.C. and H.B. Verheij (1994). Towards a logic for reasoning with norms. J. Breuker (ed.) *Proceedings of the ECAI-94 Workshop on Artificial Normative Reasoning*, pp. 160-177.

Hage, J.C. and H.B. Verheij (to appear). Reason-Based Logic: a logic for reasoning with rules and reasons. Paper accepted for *Law, Computers and Artificial Intelligence*.

Hare, R.M. (1952). *The Language of Morals*. Oxford University Press, Oxford.

Hart, H.L.A. (1958). Positivism and the Separation of Law and Morals, in H.L.A. Hart, *Essays in Jurisprudence and Philosophy*, Clarendon Press, Oxford 1983, p.49-87.

Hempel, C.G. (1950). The Empiricist Criterion of Meaning. in A.J. Ayer, *Logical Positivism*, The Free Press, New York 1959.

Koers, A.W. and D. Kracht (1991). A Goal Driven Knowledge Based System For A Domain Of Private International Law, *Proceedings of the Third International Conference on Artificial Intelligence and Law*, ACM Press, p. 81-85.

Lorenz, K. (1977). *Behind the Mirror. A Search for a Natural History of Human Knowledge*. Methuen, London.

Lukaszewicz, W. (1990). *Non-monotonic Reasoning; formalization of commonsense reasoning*. Ellis Horwood, London e.a.

MacCormick, D.N. (1974). Law as institutional fact, *The Law Quarterly Review* 90, p. 102-129.

MacCormick, D.N. (1978). *Legal Reasoning and Legal Theory*, Clarendon Press, Oxford.

MacCormick, D.N. and O. Weinberger (1987), *An Institutional Theory of Law*, Reidel, Dordrecht etc.

McCarty, L.T. (1980). The Taxman Project: Towards a Cognitive Theory of Legal Argument. B. Niblett (ed.), *Computer Science and Law*, Cambridge University Press, Cambridge.

Naess, A. (1966). *Communication and Argument. Elements of Applied Semantics*. Allen & Unwin, London.

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

Prakken, H. (1993). *Logical tools for modelling legal argument*, Ph.-D.-thesis Amsterdam.

Prakken, H. (1995). A semantic view on reasoning about priorities. *Proceedings of the Second Dutch-German workshop on Non-monotonic Reasoning*, Utrecht, pp. 160-167.

Putnam, H. (1976). Realism and Reason. H. Putnam, *Meaning and the Moral Sciences*. Routledge and Kegan Paul, Boston e.a. 1978, p. 123-138.

Raz, J. (1975). *Practical Reason and Norms*, Hutchinson, London.

Ross, A. (1968). *Directives and Norms*, London.

Sanders, K.E. (1991). Representing and reasoning about open-textured predicates. *Proceedings of the third International Conference on Law and Artificial Intelligence*, ACM-press, Oxford, p. 137-144.

Sartor, G. (1994). A Formal Model of Legal Argumentation. *Ratio Juris*, vol. 7, pp. 177-211.

Searle, J.R. (1969). *Speech acts; An essay in the philosophy of language*, Cambridge University Press.

Searle, J. (1975). A taxonomy of Illocutionary Acts, in J. Searle, *Expression and Meaning*, Cambridge University Press, Cambridge e.a., p. 1-29.

Summers, R.S. (1978). Two types of substantive reasons: The core of a theory of common-law justification. *Cornell Law Review* 63 nr. 5, pp. 707-735.

Toulmin, S.E. (1958). *The Uses of Argument*, Cambridge University Press, London, New York.

Verheij, H.B. (1995). Accrual of arguments in defeasible argumentation. Accepted paper for the second Dutch/German Workshop on Nonmonotonic Reasoning.

Valente, A. and J. Breuker, Ontologies: the Missing Link Between Legal Theory and AI & Law. H. Prakken. A.J. Muntjewerff, A. Soeteman, and R.G.F. Winkels (eds.), *Legal knowledge based systems; The Relation with Legal Theory*, Koninklijke Vermande BV, Lelystad 1994, p. 139-149.

Verheij, H.B. and J.C. Hage (1994). Reasoning by analogy; a formal reconstruction. H. Prakken, A.J. Muntjewerff, A. Soeteman and R.F. Winkels (eds.), *Legal knowledge based systems; the relation with legal theory*, Koninklijke Vermande, Lelystad, pp. 65-78.

White, A.R. (1975). *Modal Thinking*. Basil Blackwell, Oxford.