L. Royakkers and F. Dignum. Defeasible reasoning with legal rules. In M.A. Brown and J. Carmo, editors, *Deontic Logic, Agency and Normative Systems (Workshops in Computing)*, pages 174--193. Springer-Verlag, 1996.

Defeasible Reasoning with Legal Rules in a Deontic Logic based on Preferences*

Lambèr Royakkers
Dept. of Legal Informatics
Tilburg University
P.O.box 90153, 5000 LE Tilburg
L.M.M.Royakkers@kub.nl

Frank Dignum
Dept. of Mathematics and Computer Science
Eindhoven University of Technology
P.O.box 513, 5600 MB Eindhoven
dignum@win.tue.nl

Abstract

Over the last few years several defeasible deontic reasoning formalisms have been developed to solve the problem of deontic inconsistency. However, these formalisms are unable to deal with some very common forms of deontic reasoning since, e.g., their expressiveness is restricted. In this paper, which is based on Royakkers and Dignum (1996), we will establish a priority hierarchy of legal rules to solve the problem of deontic conflicts, and we will present a mechanism to reason about nonmonotonicity of legal rules over the priority hierarchy. The theory presented here, based on default logic, and being a modification and extension of Prakken's argumentation framework, adequately deals with some shortcomings of other defeasible deontic reasoning approaches.

1 Introduction

Logical analysis of reasoning with inconsistent rules is a very relevant area for AI-and-Law research, since rules used in the legal domain are often conflicting. 'Prioritised' rules have received attention in research on the formalisation of nonmonotonic reasoning, particularly as a way of modelling the choice crite-

*This research was sponsored by the foundation for Law and Public Administration (Reob), which is part of the Netherlands Organization for Scientific Research (NWO).

rion in dealing with exceptions (cf. Brewka, 1991; Poole, 1988; Prakken, 1993; Shoham, 1988).

To deal with the inconsistencies, various sorts of consistency-based approaches have been developed, such as the nonmonotonic logic of McDermott and Doyle (1980) and the default logic of Reiter (1980). But these approaches fail to reason about conflicting norms, since they are all based on non-modal logics. As a way of solving the problems of deontic conflicts, forms of defeasible reasoning (cf. Pollock, 1987) are adopted, which provide a mechanism to establish preference hierarchies of norms and to select a more applicable norm among conflicting ones in a specific situation (cf. Alchourrón and Makinson, 1981; Royakkers and Dignum, 1994). The existing formalisations of defeasible deontic reasoning approaches (Horty, 1994; Ryu, 1995; Tan and Van der Torre, 1994) are unable to deal with several highly common forms of deontic logic (cf. Prakken, 1993).

A first problem is *the lack of notion of permission* in certain approaches (Horty, 1994; Tan and Van der Torre, 1994). In these approaches, $O(p/q)$: 'p is obligatory in case of q' is treated as a normal default and can be read as a Reiter default $q : p/p$: a non-deontic Reiter default. Inherent in this treatment is the absence of a reasonable Reiter default for the *negated obligation* (permission).

Another problem is *the defeasibility of only a single opposing statement* in some approaches (Horty, 1994; Prakken, 1994; Tan and Van der Torre, 1994). In these approaches, only couples of statements are considered to check inconsistencies. For instance, take the three statements $O(a)$, $O(\neg a \vee b)$ and $\neg O(b)$. No single statement is in conflict with the other single statements. However, the group of statements $O(a)$ and $O(\neg a \vee b)$ implies $O(b)$ in standard deontic logic, which is in conflict with the statement $\neg O(b)$.

The third problem is that most approaches (Horty, 1994; Ryu, 1995; Tan and Van der Torre, 1994) can only deal with defeasible conditionals that are deontic. But deontic defaults are not the only defaults in legal reasoning. Consider the deontic default $a : O(b)/O(b)$. With this default, it is very often the case that a is derived by another default rule, e.g. $c : a/a$, which is called a 'classification rule' or an 'interpretation rule'. In the legal domain, it is accepted that these rules are also defeasible (cf. Hart, 1961). Prakken shows this by extending Hart's standard example on a park regulation that forbids vehicles to enter the park:

> not only this rule itself may turn out to be defeasible, for example, if the vehicle is an ambulance, but also rules on when something counts as a vehicle may be defeasible: imagine that a court says that objects on wheels that are meant for normal transport are vehicles: then roller skates used by people on their way to the office might be recognised as an exception. (Prakken, 1995)

In this paper, we will develop a theory of defeasible deontic reasoning which adequately deals with the above-mentioned problems. The theory is an extension and modification of the argumentation framework in default logic developed by Prakken (1993, 1994). Further, our theory is an extension of the Dung's theory (1993), which only considers argumentation frameworks with one kind of conflict between arguments.

The structure of this paper is as follows. In section 2, we will give the representation of legal rules, which we subdivide in rules and conditional norms. In section 3, we will discuss the argumentation framework for rules. The argumentation framework for norms depending on rules selected from the argumentation framework for rules will be discussed in section 4. We will concentrate here on defeasibility and violation. We will end this paper with some conclusions.

2 Legal Rules: Rules and Conditional Norms

The fundamental logical structure of legal knowledge gives rise to the nonmonotonicity of legal reasoning (Reiter, 1980; Delgrande, 1988): the consequences that may follow from a set of legal and factual premises can be invalidated by further information. This means that rules can be 'defeated' by other or new rules and facts. The principal idea of this paper, which goes back to Rescher (1964), is to allow the rules to be ordered and to use this ordering in such a way that conflicts can be solved in a logical argumentation framework using nonmonotonic logic. Such an ordering can often be discerned when considering the rules in a legal code. The 'Lex Superior' principle, for instance, is based on the general hierarchy of a legal system; the rules are divided along the lines of the hierarchical structure of the normative system. Rules with a lower rank of priority have to respect the consequences that follow from a higher ranked rule (see paper 6). To describe the ordering between the formulas, we use the following notation. Let x and y be legal rules, then '$x \preceq y$' means that y is preferred to x; '$x \sim y$' is an abbreviation for '$x \preceq y$ and $y \preceq x$'; and '$x \prec y$' is an abbreviation for '$x \preceq y$ and $y \not\preceq x$'. The ordering relation $\preceq$ is reflexive and transitive.

Legal rules, or at least most of them, subordinate a legal effect to a legal condition. By legal effect we mean every qualification generated by a legal norm: the ascription of deontic or normative modalities, status, professional titles, other legal qualities of persons and things. By legal condition we mean every condition to which a legal effect is subordinated. The legal rules are represented as conditional statements of the type

$$a_1 \wedge a_2 \wedge \ldots \wedge a_n \Rightarrow \theta,$$

where θ is the legal effect and $a_1, a_2, \ldots, a_n$ are the elements of the antecedent: the conjunction of literals,[1] representing the legal condition. If θ is a norm: an

[1] A literal is any atomic propositional formula and any negation of an atomic propositional formula.

obligation ($O(\phi)$) or a permission ($P(\phi)$), with ϕ a formula of the propositional logic, then the conditional statement is called a *conditional norm.* If θ is a literal, then the statement is called a *rule*.

The statement $A \Rightarrow B$ has to be interpreted as a *normal default* according to Reiter's theory (1980, 1987) $A : B/B$: 'If A, and it can be consistently assumed B, then we can infer B'. This means that $\Rightarrow$ is *not* interpreted as the material implication, but as an inference rule that can be defeated. From A and $A \Rightarrow B$, we can infer B unless $\neg B$ can be proven. This representation corresponds to the formalisations usually proposed by legal theory and legal logic (cf. Sartor, 1993).

In our theory, we distinguish between rules and norms for the following reasons:

- Rules cannot be violated;
- The defeasibility of rules is different from the defeasibility of norms (cf. definition 3.5 and definition 4.8), which is the most important difference.

The most important thing about the difference between rules and norms is not *what* differences there are, but simply *that* there are differences. This is why we discuss our theory on different levels: first, on the level of rules (section 3), and second, on the level of norms based on a given set of rules (section 4).

The set of rules will be denoted by W and the set of conditional norms by Δ. Furthermore, we have a factual sentence F representing the factual situation, which consists of background knowledge and contingent facts. The background knowledge consists of necessary conditions, for example, a human being is mortal. A set of conditional norms, a set of rules and a factual sentence will be called a deontic context.

Definition 2.1 *A deontic context* $T = (\Delta, W, F, \prec)$ *consists of a set* Δ *of conditional norms, a set* W *of rules, a factual propositional sentence* F*: the conjunction of background knowledge* F_b *and contingent facts* F_c*, and an ordering* $\prec$ *over rules and conditional norms.*

3 Rules

Facts (formalised by the sentence F) can contain material implications. Rules, however, are represented by normal defaults written as a conditional statement of the type $a_1 \wedge a_2 \wedge \ldots \wedge a_n \Rightarrow \theta$, with θ the legal effect formalised by a literal. The theory of defeasible reasoning for rules here is based on four notions:

- the notion of an *argument* (definition 3.4);
- the notion of *defeating* (definition 3.5);
- the notion of a *defeasibility chain* (definition 3.7);

- the notion of *justified, defensible and overruled arguments* (definition 3.9).

At the end of this section, we will define *maximal coherent argument sets* of rules that we will use for the notion of the applicability of norms and the violation of obligations in section 4.

Before we discuss the notion of argument, we will give three definitions which we will use in the sequel.

Definition 3.1 *Let F be the factual propositional sentence, V be a set of rules and r a literal, then V explains r ($V \cup \{F\} \models r$) iff*

$$\{F\} \vdash r \text{ or } \exists_{a_1 \wedge a_2 \wedge \ldots \wedge a_n \Rightarrow r \in V} \{V \cup \{F\} \models a_i \mid i \in \{1, 2, \ldots, n\}\}.$$

Intuitively, *explaining* is the same as logical consequence, except that we now deal with defaults and not with implications.

Definition 3.2 *Let V be a set of rules, then the consequences of $V \cup \{F\}$ ($Cons(V \cup \{F\})$) is defined as*

$$Cons(V \cup \{F\}) := \{r \mid r \text{ is a literal and } V \cup \{F\} \models r\}.$$

Thus, the *Cons* relation is a transitive closure of the explain. It gives the set of all literals, that can be consistently derived from V and $\{F\}$.

Definition 3.3 *Let V be a set of rules. Then $V \cup \{F\}$ is coherent iff*

$$\neg\exists_{r \text{ is a literal}}(r \in Cons(V \cup \{F\}) \wedge \neg r \in Cons(V \cup \{F\})).$$

The notion of argument can now be defined as follows:

Definition 3.4 *Let $M \subseteq W$, ϕ a literal and $M \cup \{F\}$ coherent. Then M explains ϕ minimally iff*

- *$\{F\} \cup M \models \phi$ and*
- *$\neg\exists_{\phi_1 \in M}(\{F\} \cup M \setminus \{\phi_1\}) \models \phi$.*

We call M a minimally explaining set or an argument. The set of all arguments will be denoted as $\mathcal{M}$. The ϕ-relevant set of W, denoted by $[\phi]\mathcal{M}$, is the set of all arguments in $\mathcal{M}$ that explain ϕ minimally.
M_1 is a subargument of M iff $M_1 \subset M$ and M_1 is an argument. If there is an argument for ϕ, thus $[\phi]\mathcal{M} \neq \emptyset$, then ϕ is called an outcome.

Definition 3.5 *Let $M_1 \in [\phi]\mathcal{M}$ and $M_2 \in [\phi']\mathcal{M}$. Then M_1 is defeated by M_2 ($M_1 \prec^* M_2$) iff*

$$\exists_{\phi_1 \Rightarrow \phi_2 \in M_1} \exists_{\phi_3 \Rightarrow \phi_4 \in M_2} \{\phi_3 \Rightarrow \phi_4 \succeq \phi_1 \Rightarrow \phi_2\} \wedge \{\phi_2\} \cup \{\phi_4\} \cup \{F\} \text{ is inconsistent.}$$

Thus, an argument M_2 defeats an argument M_1 iff M_1 and M_2 have contradictory conclusions ϕ_2 and ϕ_4 with respect to the factual sentence F, and the rule $\phi_3 \Rightarrow \phi_4 \in M_2$ (responsible for the conflict) does not have a lower priority than the rule $\phi_1 \Rightarrow \phi_2 \in M_1$. Note that $\{\phi_2\} \cup \{F\}$ and $\{\phi_4\} \cup \{F\}$ are consistent, which directly follows from definition 3.4.

Relation $\prec^*$ is not transitive and not asymmetric. It is possible that $M_1 \prec^* M_2$ and $M_2 \prec^* M_1$ both hold. The following example illustrates this point:

Example 3.6

(1) $a \Rightarrow b$
(2) $\qquad\qquad c \Rightarrow \neg a$
(3) $d \Rightarrow a$
(4) $b \Rightarrow \neg c$
(5) $\qquad\qquad e \Rightarrow c$

F: $f \wedge (f \rightarrow d) \wedge e$
with $(5) \prec (4) \prec (3) \prec (2) \prec (1)$.

Let $M_1 = \{e \Rightarrow c, c \Rightarrow \neg a\}$ and $M_2 = \{d \Rightarrow a, a \Rightarrow b, b \Rightarrow \neg c\}$. Then $M_1 \prec^* M_2$, since $e \Rightarrow c \prec b \Rightarrow \neg c$, and $M_2 \prec^* M_1$, since $d \Rightarrow a \prec c \Rightarrow \neg a$.

Definition 3.7 *A defeasibility chain is a sequence of arguments in* $\mathcal{M}$*:*

$$M_1 \prec^* M_2 \prec^* \ldots \prec^* M_n$$

with the following conditions:

- $\forall_{k,l=1,2,\ldots,n \wedge k<l} M_k \not\subseteq M_l$;
- $\neg\exists_{M_{n+1} \in \mathcal{M}} \{(M_1, \ldots, M_n \not\subseteq M_{n+1}) \wedge (M_n \prec^* M_{n+1})\}$.

We define $Ch(\mathcal{M})$ *as the set of all defeasibility chains of arguments in* $\mathcal{M}$*.*

The first condition ensures that cycles in defeasibility chains are avoided. Suppose that $M_1 \prec^* M_2$ and $M_2 \prec^* M_3$, with $M_1 \subset M_3$. We would thus end up with the endless chain $M_1 \prec^* M_2 \prec^* M_3 \prec^* M_2 \prec^* M_3 \ldots$. This would also be accomplished by '$\neq$' instead of '$\not\subseteq$'. The reason why we do need '$\not\subseteq$' becomes clear in proposition 3.14.

The second condition provides that a chain stops if there is no 'stronger' argument than the last argument in the chain.

Take the example above, then

$$\begin{gathered} Ch(\mathcal{M}) = \{\{d \Rightarrow a, a \Rightarrow b\} \prec^* \{e \Rightarrow c, c \Rightarrow \neg a\}, \\ \{d \Rightarrow a, a \Rightarrow b, b \Rightarrow \neg c\} \prec^* \{e \Rightarrow c, c \Rightarrow \neg a\}, \\ \{e \Rightarrow c\} \prec^* \{d \Rightarrow a, a \Rightarrow b, b \Rightarrow \neg c\}\}, \\ \{d \Rightarrow a\} \prec^* \{e \Rightarrow c, c \Rightarrow \neg a\}, \\ \{e \Rightarrow c, c \Rightarrow \neg a\} \prec^* \{d \Rightarrow a, a \Rightarrow b, b \Rightarrow \neg c\}. \end{gathered}$$

Definition 3.8 *$Ch(M)$ is the set of all defeasibility chains in $Ch(\mathcal{M})$ starting with M.*

The defeasibility chains in $Ch(\mathcal{M})$ take the set of all possible arguments and their mutual relations of defeat as input. They produce a distinction between arguments in three classes as output:[2]

1. *justified* arguments;
2. *overruled* arguments;
3. *defensible* arguments.

A justified argument is a 'winning' argument. Such an argument can be defeated by another argument, but that argument will be overruled. An overruled argument is a 'losing' argument. A defensible argument is an argument that is neither justified nor overruled. In other words, an 'undeciding' argument.

Definition 3.9 *Let $M \in \mathcal{M}$. Then*

- *M is a justified argument iff for all chains $M \prec^* M_1 \prec^* \ldots \prec^* M_n \in Ch(\mathcal{M})$ it holds that*

  $$n \text{ is even } \wedge \neg\exists_{M' \in \mathcal{M}} M_n \prec^* M' \wedge \forall_{k \text{ is even}} M_k \text{ is a justified argument.}$$

- *M is an overruled argument iff there is a chain $M \prec^* M_1 \prec^* \ldots \prec^* M_n \in Ch(\mathcal{M})$*

  $$n \text{ is odd } \wedge \neg\exists_{M' \in \mathcal{M}} M_n \prec^* M'.$$

- *M is a defensible argument iff M is neither a justified argument nor an overruled argument.*

Note that M_n in the chain of definition 3.9 is a justified argument, since $Ch(M_n) = \{M_n\}$, which is equivalent to $\neg\exists_{M' \in \mathcal{M}} M_n \prec^* M'$.

Let $M \prec^* M_1 \prec^* \ldots \prec^* M_n$ be a chain in $Ch(\mathcal{M})$, then we call the arguments M_i with i is odd *odd arguments*, and the arguments M and M_i with i is even *even arguments*.

In a defeasibility chain $M \prec^* M_1 \prec^* \ldots \prec^* M_n$ with n is even (odd), we stipulate that the odd (even) arguments are the *attacked* arguments and the even (odd) arguments the *non-attacked* arguments. If the chain ends with M_n, then M_{n-1} is an attacked argument, because it is defeated by a non-attacked argument. M_{n-2} is not attacked, because it is defeated by an attacked argument (M_{n-1}), and so on. For example, M is defeated and overruled if M_1 is a non-attacked argument and this follows if n is odd.

[2] The terms justified, overruled and defensible argument were introduced by Prakken and Sartor (1995).

Proposition 3.10

1. *Let M be a justified argument. Then all odd arguments in the chains of $Ch(M)$ are overruled arguments.*
2. *Let M be an overruled argument. Then there is a chain in $Ch(M)$ with all even arguments overruled or defensible.*

Proof

1. Let M be a justified argument. Then for all chains $M \prec^* M_1 \prec^* \ldots \prec^* M_n$ in $Ch(M)$ it holds that n is even and $\neg\exists_{M' \in \mathcal{M}} M_n \prec^* M'$. Let M_k with k is odd be an argument of a chain $M \prec^* M_1 \prec^* \ldots \prec^* M_n$ in $Ch(M)$. Then this chain without the first k arguments, thus $M_k \prec^* M_{k+1} \prec^* \ldots \prec^* M_n$, is a chain in $Ch(M_k)$ which satisfies the conditions of an overruled argument M_k, since the chain contains an even number of arguments and $Ch(M_n) = M_n$. Thus, all odd arguments are overruled arguments.
2. Let M be an overruled argument. Then there is a chain $M \prec^* M_1 \prec^* \ldots \prec^* M_n$ with n is odd and $\neg\exists_{M' \in \mathcal{M}} M_n \prec^* M'$. Suppose an even argument M_k is justified, then the chain $M \prec^* M_1 \prec^* \ldots \prec^* M_n$ without the first k arguments, i.e., $M_k \prec^* M_{k+1} \prec^* \ldots M_n$ is a chain in $Ch(M_k)$ and does not satisfy the conditions for a justified argument M_k, since the chain contains an even number of arguments. Thus, the even arguments in such chains in $Ch(M)$ are overruled or defensible.

The condition that all even arguments in the chains of $Ch(M)$ with M being a justified argument have to be justified arguments, is necessary, since otherwise we obtain some undesirable results:

Example 3.11

(1)	$a \Rightarrow \neg b$				
(2)	$c \Rightarrow a$				
(3)		$b \Rightarrow \neg d$			
(4)		$e \Rightarrow b$			
(5)			$\neg g \Rightarrow \neg f$		
(6)				$d \Rightarrow \neg h$	
(7)				$f \Rightarrow d$	
(8)				$i \Rightarrow f$	
(9)					$h \Rightarrow \neg j$
(10)					$k \Rightarrow h$
(11)			$j \Rightarrow \neg g$		
(12)			$l \Rightarrow j$		

F: $c \wedge e \wedge i \wedge k \wedge l$
with $(12) \prec (11) \prec \ldots \prec (1)$.

Let $M_1 = \{a \Rightarrow \neg b, c \Rightarrow a\}$;
$M_2 = \{b \Rightarrow \neg d, e \Rightarrow b\}$;
$M_3 = \{\neg g \Rightarrow \neg f, j \Rightarrow \neg g, l \Rightarrow j\}$;
$M_4 = \{d \Rightarrow \neg h, f \Rightarrow d, i \Rightarrow f\}$;
$M_5 = \{h \Rightarrow \neg j, k \Rightarrow h\}$.
$Ch(M_3) = \{M_3 \prec^* M_5 \prec^* M_4 \prec^* M_2 \prec^* M_1\}$ and $Ch(M_1) = \{M_1\}$, thus without the condition that all even arguments have to be justified arguments, then M_3 would be a justified argument. However, M_3 is a defensible argument, since M_4 is not a justified argument. The set of defeasibility chains starting with M_4 is $Ch(M_4) = \{M_4 \prec^* M_2 \prec^* M_1, M_4 \prec^* M_3 \prec M_5\}$. Since M_5 is an overruled argument (the chain $M_5 \prec^* M_4 \prec^* M_2 \prec^* M_1$ satisfies the conditions for the overruled argument M_5), the chain $M_4 \prec^* M_3 \prec^* M_5$ does not satisfy the conditions for a justified argument M_4. Thus, M_4 is not a justified argument. Note that M_4 is a defensible argument, since neither of the two chains in $Ch(M_4)$ satisfies the conditions for an overruled argument M_4.

Corollary 3.12

1. *A justified argument can only be defeated by an overruled argument.*
2. *If there is no justified argument, then there is no overruled argument.*
3. *There is no justified argument iff all arguments are defensible.*
4. *There is a justified argument iff there is a defeasibility chain with one argument.*

Proof

1. Let M be a justified argument, and defeated by M_1. We have to prove that M_1 is an overruled argument. For all chains $M \prec^* M_1 \ldots \prec^* M_n$ in $Ch(M)$ it holds that n is even and $Ch(M_n) = \{M_n\}$. For all these chains without the first argument M, i.e., $M_1 \prec^* \ldots \prec^* M_n$, it holds that they are elements of $Ch(\mathcal{M})$ and satisfy the conditions of an overruled argument. Thus, M_1 is an overruled argument.

2. Suppose that there is no justified argument. Then there is no chain $M \prec^* M_1 \prec^* \ldots \prec^* M_n$ in $Ch(\mathcal{M})$ with $Ch(M_n) = \{M_n\}$. Hence, there is no overruled argument.
   The converse does not hold. For example, let $W = \{a \Rightarrow b\}$ and a a fact. Then the only argument is $\{a \Rightarrow b\}$, and this argument is justified.

3. Suppose that there is no justified argument. Then there is no overruled argument, thus all arguments are defensible.
   Evidently, if all arguments are defensible, then there are no justified arguments.

4. Suppose that M is a justified argument. Then for all chains $M \prec^* M_1 \prec^* \ldots \prec^* M_n$ in $Ch(M)$ it holds that $Ch(M_n) = \{M_n\}$. Thus, there is a defeasibility chain with one argument.
   If there is a chain with one argument, say M, then M is a justified argument. Hence, there is a justified argument.

The converse of 3.12.1 does not hold: an overruled argument need not necessarily be defeated by a justified argument.

Example 3.13

(1)	$a \Rightarrow b$				
(2)		$c \Rightarrow d$			
(3)			$\neg b \Rightarrow \neg e$		
(4)			$f \Rightarrow \neg b$		
(5)				$e \Rightarrow \neg d$	
(6)				$g \Rightarrow e$	
(7)					$h \Rightarrow d$

F: $a \wedge c \wedge f \wedge g \wedge h$
with $(7) \prec (6) \prec \ldots \prec (1)$.

Let $M_1 = \{a \Rightarrow b\}$;
$M_2 = \{c \Rightarrow d\}$;
$M_3 = \{\neg b \Rightarrow \neg e, f \Rightarrow \neg b\}$;
$M_4 = \{g \Rightarrow e, e \Rightarrow \neg d\}$;
$M_5 = \{h \Rightarrow d\}$.
Then $Ch(M_5) = \{M_5 \prec^* M_4 \prec^* M_3 \prec^* M_1, M_5 \prec^* M_4 \prec^* M_2\}$. M_5 is an overruled argument, since there is a chain $M_5 \prec^* M_4 \prec^* M_3 \prec^* M_1$ with $Ch(M_1) = \{M_1\}$. Further, M_5 is only defeated by argument M_4, which is an overruled argument, since $M_4 \prec^* M_2 \in Ch(M_4)$ and $Ch(M_2) = \{M_2\}$. Thus, an overruled argument can be defeated by an overruled argument.

Proposition 3.14 *All subarguments of a justified argument are justified arguments.*

Proof. Suppose that M is a justified argument and M' is a subargument of M. We have to prove that M' is a justified argument, and without loss of generality we assume that M' is a 'largest' subargument of M, i.e., $\neg\exists_{M'' \in \mathcal{M}} M' \subset M'' \subset M$. For if this argument is justified, we can repeat this process for M' to prove that all subarguments of M' are justified. Suppose that M' is not a justified argument, then M' is an overruled or a defensible argument.

- Suppose that M' is an overruled argument. Then there is a chain $M' \prec^* M_1 \prec^* \ldots \prec^* M_n$ with $Ch(M_n) = \{M_n\}$, and n is odd. However, then the chain $M \prec^* M_1 \prec^* \ldots \prec^* M_n$ is a chain of $Ch(M)$, which is in contradiction with the assumption that M is a justified argument.

- Suppose now that M' is a defensible argument. Then there is a chain $M' \prec^* M_1 \prec^* \ldots \prec^* M_n$ and $Ch(M_n) \neq \{M_n\}$. Now, the chain $M \prec^* M_1 \prec^* \ldots \prec^* M_n$ is not an element of $Ch(M)$, but part of a chain in $Ch(M)$. M_n can only be followed by M', thus $M \prec^* M_1 \prec^* \ldots \prec^* M_n \prec^* M' \prec^* M'_1 \prec^* \ldots \prec^* M'_n$, with $Ch(M'_n) = \{M'_n\}$. However, now it follows that M' is an overruled argument (if it is an odd argument in the chain (proposition 3.10.2)) or a justified argument (if it is an even argument (definition 3.9)), and this is in contradiction with the assumption that M' is defensible.

Thus, M' is a justified argument.

Example 3.15

(1)	$a \Rightarrow b$				
(2)		$c \Rightarrow \neg b$			
(3)			$d \Rightarrow \neg e$		
(4)				$e \Rightarrow \neg f$	
(5)				$b \Rightarrow e$	
(6)				$h \Rightarrow b$	
(7)					$f \Rightarrow \neg d$
(8)					$i \Rightarrow f$
(9)			$j \Rightarrow d$		

F: $a \wedge c \wedge h \wedge i \wedge j$
with $(9) \prec (8) \prec \ldots \prec (1)$.

Let
$M_1 = \{a \Rightarrow b\}$
$M_2 = \{c \Rightarrow \neg b\}$
$M_3 = \{d \Rightarrow \neg e, j \Rightarrow d\}$
$M_4 = \{e \Rightarrow \neg f, b \Rightarrow e, h \Rightarrow b\}$
$M_5 = \{f \Rightarrow \neg d, i \Rightarrow f\}$

1. $Ch(M_1) = \{M_1\}$, thus M_1 is a justified argument.

2. $Ch(M_2) = \{M_2 \prec^* M_1\}$, thus M_2 is an overruled argument.

3. $Ch(M_3) = \{M_3 \prec^* M_5 \prec^* M_4 \prec^* M_2 \prec^* M_1\}$. M_3 is not overruled, since otherwise the chain contains an even number of arguments. It depends on arguments M_5 and M_4 whether M_3 is justified or defensible.

4. $Ch(M_4) = \{M_4 \prec^* M_2 \prec^* M_1, M_4 \prec^* M_3 \prec^* M_5\}$. M_3 is not an overruled argument, thus M_4 cannot be a justified argument. M_4 cannot be an overruled argument, since there is no chain in $Ch(M_4)$ with an even number of arguments. Thus, M_4 is a defensible argument. Hence, from the chain in $Ch(M_3)$ it also follows that M_3 is a defensible argument.

5. $Ch(M_5) = \{M_5 \prec^* M_4 \prec^* M_2 \prec^* M_1, M_5 \prec^* M_4 \prec^* M_3\}$. M_5 is an overruled argument, since the chain $M_5 \prec^* M_4 \prec M_2 \prec^* M_1$ contains an even number of arguments and $Ch(M_1) = \{M_1\}$.

The following three problems form the main problems in the literature on defeasible arguments. We will show that these problems can be adequately dealt with in the theory as it follows from definition 3.6.

Example 3.16 *The intermediate conclusion*

(1) $a \Rightarrow b$
(2) $c \Rightarrow \neg b$
(3) $d \Rightarrow a$

F: $c \wedge d$
with $(3) \prec (2) \prec (1)$.

Here, a conflict arises between rules (1) and (2). By definition 3.5, the choice is made between the norms which are certainly in conflict with each other. Rule (3) with *the intermediate conclusion* a, necessary to derive the outcome b, is irrelevant to the conflict.

The minimally explaining sets (arguments) are $M_1 = \{d \Rightarrow a, a \Rightarrow b\}$, $M_2 = \{c \Rightarrow \neg b\}$ and $M_3 = \{d \Rightarrow a\}$. The sets of defeasibility chains are $Ch(M_1) = \{M_1\}$, $Ch(M_2) = \{M_2 \prec^* M_1\}$ and $Ch(M_3) = \{M_3\}$. M_1 and M_3 are justified, since they are not defeated by an argument. M_2 is overruled, because it is defeated by justified argument M_1. Thus, a and b are the outcomes.

Example 3.17 *Iterated conflicts*

(1) $c \Rightarrow a$
(2) $d \Rightarrow \neg a$
(3) $\neg a \Rightarrow \neg b$
(4) $a \Rightarrow b$

F: $c \wedge d$
with $(4) \prec (3) \prec (2) \prec (1)$.

Here, a conflict arises between rules (1) and (2) and between the rules (3) and (4). Rules (1) and (2) have intermediate conclusions a and $\neg a$ for the final conclusions $\neg b$ and b. This type of problem is called *iterated conflicts*: conflicts on both intermediate and final conclusions.

The minimally explaining sets are $M_1 = \{c \Rightarrow a\}$, $M_2 = \{d \Rightarrow \neg a\}$, $M_3 = \{d \Rightarrow \neg a, \neg a \Rightarrow \neg b\}$ and $M_4 = \{c \Rightarrow a, a \Rightarrow b\}$. The sets of defeasibility chains are $Ch(M_1) = \{M_1\}$, $Ch(M_2) = \{M_2 \prec^* M_1\}$, $Ch(M_3) = \{M_3 \prec^* M_1\}$ and $Ch(M_4) = \{M_4 \prec^* M_3 \prec^* M_1\}$. M_1 and M_4 are justified arguments. M_2 and M_3 are overruled arguments, because they are defeated by justified argument M_1: a and b are the outcomes. Note that M_4 is defeated by M_3, though is not overruled, since M_3 is an overruled argument (defeated by M_1).

Example 3.18 *Circular conflicts*

(1) $a \Rightarrow b$		
(2)	$c \Rightarrow \neg b$	
(3)		$b \Rightarrow \neg a$
(4) $d \Rightarrow a$		
(5)		$d \Rightarrow b$

F: $c \wedge d$
with $(5) \prec (4) \prec (3) \prec (2) \prec (1)$.

Here, we have conflicts between rules (1) and (2), rules (3) and (4) and rules (2) and (5). The applicability of rule (1) depends on rule (4), with the intermediate conclusion a. Rule (4) is in conflict with a higher rule, (3), (iterated conflict), and the applicability of (3) depends on (5), which is in conflict with (2). Rule (2) is in conflict with rule (1), and the applicability of (1) depends on rule (4), and so on. This will never stop, therefore we call this *the problem of circular conflicts.*

The minimally explaining sets are $M_1 = \{d \Rightarrow a, a \Rightarrow b\}$, $M_2 = \{c \Rightarrow \neg b\}$, $M_3 = \{d \Rightarrow b, b \Rightarrow \neg a\}$, $M_4 = \{d \Rightarrow a\}$ and $M_5 = \{d \Rightarrow b\}$. The sets of defeasibility chains are $Ch(M_1) = \{M_1 \prec^* M_3 \prec^* M_2\}$, $Ch(M_2) = \{M_2 \prec^* M_1 \prec^* M_3\}$, $Ch(M_3) = \{M_3 \prec^* M_2 \prec^* M_1\}$, $Ch(M_4) = \{M_4 \prec^* M_3 \prec^* M_2\}$ and $Ch(M_5) = \{M_5 \prec^* M_2 \prec^* M_1\}$. All the arguments are defensible, since there is no justified argument. Thus, b, $\neg b$, $\neg a$ and a are the outcomes.

Definition 3.19 *Let $\mathcal{M}^*$ be the set of all defensible and justified arguments in $\mathcal{M}$. Then w is a maximal coherent argument set iff w is the set of all rules of the arguments in $\mathcal{M}^*$ with*

- *$Cons(w \cup \{F\})$ is consistent and*
- *$\neg\exists_{M \in \mathcal{M}^* \setminus w}(Cons(w \cup \{F\} \cup M)$ is consistent).*

$\mathcal{W}$ is defined as the set of all maximal coherent argument sets $w \subseteq \mathcal{M}$. $Out(w)$ is defined as the set of all outcomes of the arguments $M \subseteq w$.

Consider example 3.16. Then $\mathcal{W} = \{w_1, w_2, w_3\}$ with $w_1 = \{d \Rightarrow a, a \Rightarrow b, d \Rightarrow b\}$, $w_2 = \{c \Rightarrow \neg b, d \Rightarrow a\}$ and $w_3 = \{b \Rightarrow \neg a, d \Rightarrow b\}\}$. It follows that $Out(w_1) = \{a, b\}$, $Out(w_2) = \{a, \neg b\}$ and $Out(w_3) = \{\neg a, b\}$. There is no $w \in \mathcal{W}$ with $Out(w) = \{\neg a, \neg b\}$, since M_3 is the argument for $\neg a$ and $M_3 \cup \{F\} \models b$. Thus, if $\neg a \in Out(w)$, then b must also be an element in $Out(w)$, since otherwise we cannot derive $\neg a$: the literal b is necessary to derive $\neg a$. In other words, if an argument M is a subset of a maximal coherent argument set w, then all outcomes of the subarguments of M are elements of the set $Out(w)$.

Corollary 3.20 *Let $w \in \mathcal{W}$. Then for all justified arguments M in $\mathcal{M}$ it holds that $M \subseteq w$.*

The following definition is needed for the definition of arguments of norms in the next section.

Definition 3.21 *Let W be the set of rules and $w \in \mathcal{W}$, then $W(w)$ is defined as the maximal coherent set of rules in W with respect to w and F. Let $\phi \Rightarrow \phi_1 \in W$, then*

$$\phi \Rightarrow \phi_1 \in W(w) \text{ iff } \{\phi \Rightarrow \phi_1\} \cup w \cup \{F\} \text{ is coherent.}$$

Note that all rules in w are in $W(w)$.

Example 3.22
(1) $c \Rightarrow a$
(2) $d \Rightarrow \neg a$
(3) $d \Rightarrow b$
(4) $\neg c \Rightarrow \neg b$
F: $c \wedge d$
with $(4) \prec (3) \prec (2) \prec (1)$.

Then $w = \{c \Rightarrow a, d \Rightarrow b\}$ and $W(w) = \{c \Rightarrow a, \neg c \Rightarrow \neg b, d \Rightarrow b\}$. Rule $d \Rightarrow \neg a$ is not an element of $W(w)$, since the rule is incoherent with $w \cup \{F\}$. Furthermore, note that $\neg c \Rightarrow \neg b$ is an element of $W(w)$, though not of w. We will see in the following section that such a rule can be applicable for the derivation of a certain norm.

4 Norms

Defeasible deontic reasoning is based on five notions: the notion of *applicable norms* based on a set $w \in \mathcal{W}$ (definition 4.1) and the same four notions of defeasible reasoning for rules (definitions 4.2, 4.7, 4.9 and 4.11).

We will also define *maximal coherent argument sets* for norms that we will use for the definition of violated obligations in subsection 4.2. At the end of this section, we will give some examples of defeasible deontic reasoning with violated norms.

Definition 4.1 *Let $w \in \mathcal{W}$ and $a_1 \wedge a_2 \wedge \ldots \wedge a_n \Rightarrow \psi$ be a conditional norm in Δ. Then*

$$\psi \in \Delta(w, F) \text{ iff } \forall_{i \in \{1,2,\ldots,n\}} w \cup \{F\} \models a_i.$$

ψ is called an applicable norm in Δ with respect to w and F. Thus, $\Delta(w, F)$ is the set of all applicable norms in Δ with respect to w and F.

Example 4.2

W: *(1)* $a \Rightarrow b$
(2) $b \Rightarrow c$
(3) $e \Rightarrow d$
(4) $f \Rightarrow g$
(5) $h \Rightarrow \neg c$

Δ: *(1')* $b \Rightarrow O(\neg a)$
(2') $d \wedge h \Rightarrow O(i \vee l)$
(3') $g \Rightarrow O(b)$

F: $a \wedge e \wedge h$

with $(5) \prec (4) \prec (3) \prec (2) \prec (1)$ *and* $(3') \prec (2') \prec (1')$.

The arguments are:
$M_1 = \{a \Rightarrow b\}$;
$M_2 = \{a \Rightarrow b, b \Rightarrow c\}$;
$M_3 = \{e \Rightarrow d\}$;
$M_4 = \{h \Rightarrow \neg c\}$.
It is easy to see that M_1, M_2 and M_3 are justified arguments, since they are not defeated by an argument. M_4 is defeated by M_2, thus M_4 is an overruled argument. There is one maximal coherent argument set w: $M_1 \cup M_2 \cup M_3$. Thus $w = \{a \Rightarrow b, b \Rightarrow c, e \Rightarrow d\}$. Now we can give the applicable norms:

1. $O(\neg a) \in \Delta(w, F)$, since $w \cup \{F\} \models b$;
2. $O(i \vee l) \in \Delta(w, F)$, since $w \cup \{F\} \models d$ and $w \cup \{F\} \models h$;
3. $O(b) \notin \Delta(w, F)$, since $w \cup \{F\} \not\models g$.

Thus $\Delta(w, F) = \{O(\neg a), O(i \vee l)\}$.

For $\Delta(w, F)$, we will use the standard deontic logic (SDL) satisfying the two axioms:

(K) $O(p \rightarrow q) \rightarrow (O(p) \rightarrow O(q))$;

(D) $\neg(O(p) \wedge O(\neg p))$.

We use the following two inference rules:[3]

(RK1)

$$\frac{O(p_1), O(p_2), \ldots, O(p_n), p_1 \wedge p_2 \wedge \ldots \wedge p_n \rightarrow q}{O(q)}$$

(RK2)

$$\frac{O(p_1), O(p_2), \ldots, O(p_n), p_1 \wedge p_2 \wedge \ldots \wedge p_n \Rightarrow q}{O(q)}$$

Analogous with the definition of arguments for rules, we define arguments for norms.

[3] For a brief discussion about these rules, see section 5.

Definition 4.3 *Let $N \subseteq \Delta(w, F)$, ψ a norm and $N \cup \{F_b\} \cup W(w)$ consistent. Then N explains ψ minimally iff*

- $\{F_b\} \cup W(w) \cup N \vdash \psi$ *and*
- $\neg\exists_{\psi_1 \in N}\{F_b\} \cup W(w) \cup N \setminus \{\psi_1\} \vdash \psi$.

We call N a minimally explaining set or an argument. The set of all arguments will be denoted as $\mathcal{N}(w, F)$. The ψ-relevant set of $\Delta(w, F)$, denoted by $[\psi]\mathcal{N}(w, F)$ is the set of all arguments, that explain ψ minimally.
Let $\mathcal{N} \subseteq \mathcal{N}(w, F)$. If there is an argument for ψ, thus $[\psi]\mathcal{N} \neq \emptyset$, then ψ is called an outcome.

We do not use F, but F_b in this definition, because otherwise we would derive ridiculous conclusions. Consider the following example.

Example 4.4
(1) $O(a)$: It is obligatory to go to school.
(2) $O(b)$: It is obligatory to behave.
$F = F_c$: $a \wedge \neg b$: You go to school and you do not behave,
with $(2) \prec (1)$.

Now we can derive $O(\neg b)$ from $O(a)$ and F_c, because $\{F_c\} \vdash a \rightarrow \neg b$ and by (RK1) $O(a) \rightarrow O(\neg b)$. Thus, if we use F instead of F_b, then $\{O(a)\}$ would be an argument for the outcome $O(\neg b)$, and this is not a desirable result.

Definition 4.5 *Let $\mathcal{N} \subseteq \mathcal{N}(w, F)$ and $N_1, N_2 \in \mathcal{N}$. Then N_1 and N_2 are in conflict iff*

$$\{F_b\} \cup W(w) \cup N_1 \vdash \psi \text{ and } \{F_b\} \cup W(w) \cup N_2 \vdash \neg\psi.$$

Example 4.6
(1) $O(a \vee b)$
(2) $O(\neg a)$
(3) $O(c \vee \neg b)$
(4) $O(\neg c)$
with $(4) \prec (3) \prec (2) \prec (1)$.

Let $N_1 = \{O(a \vee b), O(c \vee \neg b), O(\neg c)\}$ and $N_2 = \{O(\neg a)\}$ be arguments in $\mathcal{N}$. Then N_1 is an argument for $O(a)$ and N_2 an argument for $\neg O(a)$, since $O(\neg a) \rightarrow \neg O(a)$ is deduced by axiom (OD). Thus, N_1 and N_2 are in conflict.

Example 4.7
(1) $O(a)$
(2) $O(\neg b)$
$(2) \prec (1)$ *and* $W(w) = \{a \Rightarrow b\}$.

Let $N_3 = \{O(a)\}$ and $N_4 = \{O(\neg b)\}$ be arguments in $\mathcal{N}$. Then N_3 is an argument for $O(b)$, since $\{a \Rightarrow b\} \cup N_3 \vdash O(b)$ and N_4 is an argument for $\neg O(b)$. Thus, N_3 and N_4 are in conflict.

4.1 Defeasibility

Because the arguments in $\mathcal{N}(w, F)$ can be in conflict, we resolve these deontic conflicts by adopting defeasible reasoning. The defeasibility of norms in $\Delta(w, F)$ determines the validity of these norms. The validity depends on the rules in w. For instance, let $\Delta(w, F) = \{O(a), O(b)\}$, $O(b) \prec O(a)$ and $w = \{a \Rightarrow \neg b\}$. Then $\{O(a)\}$ is an argument for $\neg O(b)$, since $w \cup \{O(a)\} \vdash \neg O(b)$. Now we say that $\{O(a)\}$ defeats $\{O(b)\}$, and that $O(b)$ is not valid. Suppose that $w = \emptyset$, then $O(a)$ and $O(b)$ are both valid, because $\{O(a)\}$ and $\{O(b)\}$ are not in conflict.

Definition 4.8 *Let N_1 and N_2 be arguments. Then N_1 is defeated by N_2 ($N_1 \prec^* N_2$) iff N_1 and N_2 are in conflict and*

$$\exists_{\psi_1 \in N_1} \forall_{\psi_2 \in N_2} \psi_1 \preceq \psi_2.$$

The main difference between this definition and definition 3.5 is that we look here at the statements of the two arguments with the lowest priority, and in definition 3.5 we looked at the statements of the two arguments with conflicting conclusions. Note that it is possible that $N_1 \prec^* N_2$ and $N_2 \prec^* N_1$ both hold, but only if the statements with the lowest priority of the two arguments have the same priority. $\prec^*$ is not transitive and not asymmetric.

Example 4.9 *Consider example 4.6. N_1 is defeated by N_2, since $O(\neg c)$ has a lower priority than all the norms in N_2: $O(\neg c) \preceq O(\neg a)$.*
Now consider example 4.7. N_4 is defeated by N_3, since $O(\neg b) \preceq O(a)$. If $(1) \sim (2)$, then $N_3 \prec^* N_4$ and $N_4 \prec^* N_3$.

The following three definitions for arguments in $\mathcal{N}$ concerning defeasibility chains, $Ch(N)$ and justified, defensible and overruled arguments are exactly the same as the definitions for arguments in $\mathcal{N}$ (cf. definitions 3.7, 3.8 and 3.9).

Definition 4.10 *Let $\mathcal{N} \subseteq \mathcal{N}(w, F)$. A defeasibility chain is a sequence of arguments in $\mathcal{N}$: $N_1 \prec^* N_2 \prec^* \ldots \prec^* N_n$ with the following conditions:*

- $\forall_{k,l=1,2,\ldots,n \wedge k<l} N_k \not\subseteq N_l$;
- $\neg \exists_{N_{n+1} \in \mathcal{N}} \{(N_1, \ldots, N_n \not\subseteq N_{n+1}) \rightarrow (N_n \prec^* N_{n+1})\}$.

We define $Ch(\mathcal{N})$ as the set of all defeasibility chains in $\mathcal{N}$.

Definition 4.11 *Let $\mathcal{N} \subseteq \mathcal{N}(w, F)$. Then $Ch(N)$ is the set of all defeasibility chains in $Ch(\mathcal{N})$ starting with N.*

Definition 4.12 *Let $\mathcal{N} \subseteq \mathcal{N}(w, F)$ and $N, N_1, \ldots N_n \in \mathcal{N}$. Then*

- *N is a justified argument iff for all chains $N \prec^* N_1 \prec^* \ldots \prec^* N_n$ it holds that*

$$n \text{ is even} \wedge \neg\exists_{N' \in \mathcal{N}} N_n \prec^* N' \wedge \forall_{k \text{ is even}} N_k \text{ is a justified argument.}$$

- *N is an overruled argument iff there is a chain $N \prec^* N_1 \prec^* \ldots \prec^* N_n$ with*

$$n \text{ is odd } \wedge \neg\exists_{N' \in \mathcal{N}} N_n \prec^* N'.$$

- *N is a defensible argument iff N is neither a justified argument nor an overruled argument.*

Note that $\neg\exists_{N' \in \mathcal{N}} N_n \prec^* N'$ is equivalent to $Ch(N_n) = \{N_n\}$.

Example 4.13 *Consider example 4.6. Let $\mathcal{N} = \{N_1, N_2, \ldots, N_6\}$, with*
$N_1 = \{O(a \vee b), O(c \vee \neg b), O(\neg c)\} \in [O(a)]\mathcal{N}$
$N_2 = \{O(\neg a)\} \in [\neg O(a)]\mathcal{N}$
$N_3 = \{O(a \vee b), O(\neg a)\} \in [O(b)]\mathcal{N}$
$N_4 = \{O(c \vee \neg b), O(\neg c)\} \in [\neg O(b)]\mathcal{N}$
$N_5 = \{O(a \vee b), O(c \vee \neg b), O(\neg a)\} \in [O(c)]\mathcal{N}$
$N_6 = \{O(\neg c)\} \in [\neg O(c)]\mathcal{N}$

Argument N_1 is defeated by justified argument N_2, since $O(\neg c) \prec O(\neg a)$ and $Ch(N_2) = \{N_2\}$, thus N_1 is an overruled argument. N_3 and N_5 are justified arguments, since $Ch(N_3) = \{N_3\}$ and $Ch(N_5) = \{N_5\}$. N_4 and N_6 are overruled arguments, since $N_4 \prec^* N_3 \in Ch(N_4)$ and $N_6 \prec^* N_5 \in Ch(N_6)$.
If (4) $\sim$ (3) $\prec$ (2) $\prec$ (1), then N_5 and N_6 are defensible arguments, since $Ch(N_5) = \{N_5 \prec^* N_6\}$ and $Ch(N_6) = \{N_6 \prec^* N_5\}$.

Example 4.14 *Consider example 4.7. N_4 is defeated by justified argument N_3 and $Ch(N_3) = \{N_3\}$, thus N_4 is an overruled argument.*

Definition 4.15 *Let $\mathcal{N} \subseteq \mathcal{N}(w, F)$ and $\mathcal{N}^*$ be the set of all defensible and justified arguments in $\mathcal{N}$. Then o is a maximal consistent argument set iff o is a union of some of the arguments in $\mathcal{N}^*$ with*

- *$o \cup \{F_b\} \cup W(w)$ is consistent and*
- *$\neg\exists_{N \in \mathcal{N}^* \setminus o}(o \cup \{F_b\} \cup W(w) \cup N$ is consistent).*

$\mathcal{O}(\mathcal{N})$ is defined as the set of all maximal consistent argument sets o. $Out(o)$ is defined as the set of all outcomes of the arguments $N \subseteq o$.

Corollary 4.16 *Let $o \in \mathcal{O}(\mathcal{N})$. Then for all justified arguments N in $\mathcal{N}$ it holds that $N \subseteq o$.*

Example 4.17 *Consider example 4.13 with the justified arguments N_2, N_3 and N_5. There is one maximal consistent argument set o: $o = N_2 \cup N_3 \cup N_5 = N_5 = \{O(a \vee b), O(c \vee \neg b), O(\neg a)\}$.*

If $(4) \sim (3) \prec (2) \prec (1)$, then N_5 and N_6 are defensible (see example 4.13). Then there are two maximal consistent argument sets: $o_1 = N_2 \cup N_3 \cup N_5 = N_5$ and $o_2 = N_2 \cup N_3 \cup N_6 = \{O(a \vee b), O(\neg a), O(\neg c)\}$, thus $\mathcal{O}(\mathcal{N}) = \{o_1, o_2\}$.

4.2 Violation

An obligation is violated iff the obligation is not fulfilled. In SDL, we can represent the violated obligation by $O(p) \wedge \neg p$. Analogously, we define the violated obligation in our theory.

Definition 4.18 *Let $\mathcal{N} \subseteq \mathcal{N}(w, F)$, $o \in \mathcal{O}(\mathcal{N})$ and $o \cup W(w) \cup \{F_b\} \vdash O(\phi)$. The norm $O(\phi)$ is violated in o iff $w \cup \{F\} \models \neg\phi$. The set of violated norms in o will be denoted as $V(o)$.*

Example 4.19 *Let $o = \{O(a), O(b)\}$, $W(w) = \{a \Rightarrow c\}$, $\{F_b\} = w = \emptyset$ and F_c: $\neg b \wedge \neg c$. Then $O(c) \in V(o)$, since $o \cup W(w) \vdash O(c)$ and $\{F\} \vdash \neg c$ and $O(b) \in V(o)$, since $o \cup W(w) \vdash O(b)$ and $\{F\} \vdash \neg b$.*

Definition 4.20 *Let $\mathcal{N} \subseteq \mathcal{N}(w, F)$, $o \in \mathcal{O}(\mathcal{N})$ and $V(o)$ be the set of violated norms. Then $N \subseteq o$ is a violated argument iff*

$$\exists_{\psi \in V(o)} N \cup W(w) \cup \{F_b\} \vdash \psi.$$

$\mathcal{N}(V(o))$ is the set of all the violated arguments with respect to $V(o)$.

Example 4.21 *Consider example 4.6:*
(1) $O(a \vee b)$
(2) $O(\neg a)$
(3) $O(c \vee \neg b)$
(4) $O(\neg c)$
with $(4) \prec (3) \prec (2) \prec (1)$, $W(w) = \{F_b\} = \emptyset$ and F_c: $\neg c$. Suppose further (see example 4.13) that $\mathcal{N} = \{N_1, N_2, \ldots, N_6\}$, with
$N_1 = \{O(a \vee b), O(c \vee \neg b), O(\neg c)\} \in [O(a)]\mathcal{N}$
$N_2 = \{O(\neg a)\} \in [\neg O(a)]\mathcal{N}$
$N_3 = \{O(a \vee b), O(\neg a)\} \in [O(b)]\mathcal{N}$
$N_4 = \{O(c \vee \neg b), O(\neg c)\} \in [\neg O(b)]\mathcal{N}$
$N_5 = \{O(a \vee b), O(c \vee \neg b), O(\neg a)\} \in [O(c)]\mathcal{N}$
$N_6 = \{O(\neg c)\} \in [\neg O(c)]\mathcal{N}$.

From examples 4.13 and 4.17 it follows that $o = N_2 \cup N_3 \cup N_5$, thus $\mathcal{O}(\mathcal{N}) = \{N_5\}$. $O(c) \in V(o)$, since $N_5 \vdash O(c)$ and $\{F\} \vdash \neg c$. Therefore, N_5 is a violated argument. N_2 and N_3 are not violated arguments, since $N_2 \nvdash O(c)$ and $N_3 \nvdash O(c)$.

Definition 4.22 *Let* $\mathcal{N} \subseteq \mathcal{N}(w, F)$ *and* $\psi_1, \ldots, \psi_n \in \Delta(w, F)$. *Then*

$$\mathcal{N} \setminus^* \{\psi_1, \ldots, \psi_n\} = \cup\{N | N \in \mathcal{N} \wedge \forall_{i=1,\ldots,n} \psi_i \notin N\}.$$

Thus $\mathcal{N} \setminus^* \{\psi_1, \ldots, \psi_n\}$ is the set of arguments in $\mathcal{N}$ without the arguments containing a norm of the set $\{\psi_1, \ldots \psi_n\}$.

Example 4.23 *A programme committee requires that conference submissions (papers) are sent in by mail. However, if your paper is not sent by mail, then your paper should be sent by fax. And if no faxmachine is available, one should try to send it by e-mail. The facts are that the paper is sent by fax and if you send your paper by fax, then you do not send the paper by e-mail. The formalisation is as follows:*

W: (1) $e - mail \Rightarrow \neg mail$	Δ*: (1')* $O(mail)$	F_c*:* fax
(2) $fax \Rightarrow \neg mail$	*(2')* $\neg mail \Rightarrow O(fax)$	F_b *:* $fax \rightarrow \neg e - mail$
	(3') $\neg fax \Rightarrow O(email)$	

with $(2) \prec (1)$ *and* $(3') \prec (2') \prec (1')$.

Let us consider this example. If the paper is sent by mail, then there is no violation; this is the best situation. However, the norm $O(mail)$ is violated, because the paper is not sent by mail. Then the second-best situation is that your paper is sent by fax. We can formalise this process as follows:
$Out(w) = \{\neg mail\}$, $W(w) = \{fax \Rightarrow \neg mail, e - mail \Rightarrow \neg mail\}$
$\Delta(w, F) = \{O(mail), O(fax)\}$
$\mathcal{N}(w, F) = \{N_1, N_2\}$, with $N_1 = \{O(mail)\}$ and $N_2 = \{O(fax)\}$.
Note that $O(e - mail)$ is not an element of $\Delta(w, F)$, since it is not an applicable norm: $w \cup \{F\} \not\models \neg fax$.

N_1 is a justified argument, since $Ch(N_1) = \{N_1\}$, and N_2 is an overruled argument, since it is defeated by N_1. N_1 and N_2 are in conflict, since $\{F_b\} \cup W(w) \cup N_1 \vdash O(mail)$ and $\{F_b\} \cup W(w) \cup N_2 \vdash \neg O(mail)$ and $O(fax) \prec O(mail)$. Thus, $o = \{N_1\}$.

The norm $O(mail)$ has a higher priority than the norm $O(fax)$. However, the norm $O(mail)$ is violated:

$$V(o) = \{O(mail)\} \text{ and } (\mathcal{N}(w, F))(V(o)) = \{N_1\}.$$

Now we consider the norms again without the violated norm, since they can become valid because of the violation of the norm $O(mail)$. Let $\mathcal{N} = \mathcal{N}(w, F) \setminus^* \{O(mail)\}$. Then $\mathcal{N} = \{N_2\}$. Argument N_2 is a justified argument. Let $o' \in \mathcal{O}(\mathcal{N})$, then it follows that $o' = \{N_2\}$. N_2 (with as an outcome $O(fax)$) is a justified and non-violated argument with respect to the violation of the norm $O(mail)$. Thus, sending in your paper by fax is the second-best situation.

5 Conclusions

In this paper, we presented a theory of defeasible deontic reasoning dealing with some very common problems of other approaches:

- defeasibility between *groups of conditional norms*;
- the combination of defeasibility of rules and norms;
- the lack of notion of permission.

However, we do not claim that our theory is completely free from the above-mentioned problems. Adopting defeasible reasoning for non-deontic constraints (rules) is not trivial, such as the study of the validity of deducing of $O(b)$ from $O(a)$ and $a \Rightarrow b$. This corresponds with the choice of rule (RK2):

$$\frac{O(p_1), O(p_2), \ldots, O(p_n), p_1 \wedge p_2 \wedge \ldots \wedge p_n \Rightarrow q}{O(q)}.$$

However, it is also possible to add the default rule schema (RK)

$$(p_1 \wedge p_2 \wedge \ldots \wedge p_n \Rightarrow q) \rightarrow (O(p_1) \wedge O(p_2) \wedge \ldots O(p_n) \Rightarrow O(q)),$$

which also allows defaults between norms.

The consequence of the use of the (RK2)-rule is that norms are dependent on rules and that rules are not dependent on norms. This means that rules have a higher priority than norms. Thus, an argument of rules cannot be defeated by an argument of norms. If we had not made this distinction between rules and conditional norms (i.e., if we had opted for (RK) instead of (RK2)), then we would get, for example, the following situation:

$W \cup \Delta$: (1) $O(a)$ F: $\emptyset$
(2) $O(b)$
(3) $a \Rightarrow \neg b$
with $(3) \prec (2) \prec (1)$.

From $O(a)$ and $a \Rightarrow \neg b$ we can deduce $O(\neg b)$, thus $\{O(a), a \Rightarrow \neg b\}$ would be an argument for $O(\neg b)$. Thus, the set $\{O(a), (O(b), a \Rightarrow \neg b\}$ is incoherent. We cannot deduce $\neg(a \Rightarrow \neg b)$ from $O(a)$ and $O(b)$ with the (RK2)-rule. However, with (RK) we could derive $\neg(a \Rightarrow \neg b)$ from $O(a)$ and $O(b)$:
$O(a) \wedge O(b) \rightarrow O(a) \wedge \neg O(\neg b)$;
$O(a) \wedge \neg O(\neg b) \rightarrow \neg(\neg O(a) \vee O(\neg b))$;
$\neg(\neg O(a) \vee O(\neg b)) \rightarrow \neg(O(a) \rightarrow O(b))$;
$\neg(O(a) \rightarrow O(b)) \rightarrow \neg(a \Rightarrow \neg b)$.
Thus, $\{O(a), O(b)\}$ is an argument for $\neg(a \Rightarrow \neg b)$.[4]

[4]As a consequence, we allow negations of default rules. It is not trivial to decide the meaning of these formulas.

The advantage of replacing the (RK2)-rule by a stronger rule is that rules and conditional norms do not have to be separated in the arguments, and that rules do not have a higher priority than conditional norms by definition, which is a consequence from the theory presented in this paper. At first glance, this seems a good concept for solving the problem of the separation of rules and norms. However, this concept gives rise to some serious problems, since the definition of defeating arguments for rules is different from the definition of norms. In the definition of defeating arguments for rules, we only look at single statements with conflicting final conclusions, whereas in the definition of norms we not only look at final conclusions but at groups of statements deriving conflicting conclusions. Furthermore, the applicability of a conditional norm depends on the facts and the rules, thus this means that we have to separate rules and conditional norms. We leave this issue of the separation of arguments for rules and conditional norms in deontic reasoning as a future research topic.

Most deontic defeasible reasoning approaches are based on specificity considering the amount of relevant information or supporting evidence. Our approach is based on the more general idea of priority (authority). Other defeasibility criteria can easily be converted to some form of defeasibility on the basis of priority.

References

[1] Alchourrón C and Makinson D. *Hierarchies of regulations and their logic.* In: Hilpinen R (ed) New studies in deontic logic. D. Reidel Publishing Company, 1981, pp 125-148.

[2] Brewka G. *Preferred subtheories: An extended logical framework for default reasoning.* In: Proceedings IJCAI-1991, pp 1043-1048.

[3] Delgrande J. *An approach to default reasoning based on a first-order conditional logic.* Artificial intelligence 1988; 36:63-90.

[4] Dung P. *On the acceptability of arguments and its fundamental role in nonmonotonic reasoning and logic programming.* In: Proceedings IJCAI-93, pp. 852-857.

[5] Hart H. *The concept of law.* Clarendon Press, Oxford, 1961.

[6] Horty J. *Moral dilemmas and nonmonotonic logic.* Journal of philosophical logic 1994; 23:35-65.

[7] Kelsen H. *Reine rechtslehre.* Zweite auflage, Wien, 1975.

[8] McDermott D and Doyle J. *Non-monotonic logic I.* Artificial intelligence 1980; 13:41-72.

[9] Pollock J. *Defeasible reasoning.* Cognitive Science 1987; 11(4):481-518.

[10] Poole D. *A logical framework for default reasoning.* Artificial intelligence 1988; 36:27-47.

[11] Prakken H. *Logical tools for modelling legal argument.* PhD thesis, Amsterdam, 1993.

[12] Prakken H and Sartor G. *On the relation between legal language and legal argument: assumptions, applicability and dynamic priorities.* In: Proceedings of the fifth international conference on artificial intelligence and law. University of Maryland, MD USA, 1995, pp 1-11.

[13] Prakken H. *Two approaches to the formalisation of defeasible deontic reasoning*, to be published, 1995.

[14] Reiter R. *A Logic for default reasoning.* Artificial Intelligence 1980; 13:81-132.

[15] Reiter R. *Nonmonotonic reasoning.* Annual review of computer science 1987; 2:147-186.

[16] Rescher N. *Hypothetical reasoning.* North-Holland, Amsterdam, 1964.

[17] Royakkers L and Dignum F. *Deontic inconsistencies and authorities.* In: Breuker J (ed) Normative reasoning, 1994.

[18] Ryu Y. *Conditional deontic logic augmented with defeasible reasoning*, submitted to Data & knowledge engineering.

[19] Sartor G. *Defeasibility in legal reasoning.* Rechtstheorie 1993; 24:281-316.

[20] Shoham Y. *Reasoning about change*, MIT Press, Cambridge, 1988.

[21] Tan Y.-H and Van der Torre L. *Multi preference semantics for a defeasible deontic logic.* In: Prakken H (ed) Legal knowledge based systems. The relation with legal theory. Koninklijke Vermande, Lelystad, 1994, pp 115-126.