DETACHMENT IN NORMATIVE SYSTEMS: EXAMPLES, INFERENCE PATTERNS, PROPERTIES

XAVIER PARENT, LEENDERT VAN DER TORRE
University of Luxembourg
{xavier.parent,leon.vandertorre}@uni.lu

Abstract

There is a variety of ways to reason with normative systems. This partly reflects a variety of semantics developed for deontic logic, such as traditional semantics based on possible worlds, or alternative semantics based on algebraic methods, explicit norms or techniques from non-monotonic logic. This diversity raises the question how these reasoning methods are related, and which reasoning method should be chosen for a particular application. In this paper we discuss the use of examples, inference patterns, and more abstract properties. First, benchmark examples can be used to compare ways to reason with normative systems. We give an overview of several benchmark examples of normative reasoning and deontic logic: van Fraassen's paradox, Forrester's paradox, Prakken and Sergot's cottage regulations, Jeffrey's disarmament example, Chisholm's paradox, Makinson's Möbius strip, and Horty's priority examples. Moreover, we distinguish various interpretations that can be given to these benchmark examples, such as consistent interpretations, dilemma interpretations, and violability interpretations. Second, inference patterns can be used to compare different ways to reason with normative systems. Instead of analysing the benchmark examples semantically, as it is usually done, in this paper we use inference patterns to analyse them at a higher level of abstraction. We discuss inference patterns reflecting typical logical properties such as strengthening of the antecedent or weakening of the consequent. Third, more abstract properties can be defined to compare different ways to reason with normative systems. To define these more abstract properties, we first present a formal framework around the notion of detachment. Some of the ten properties we introduce are derived from the

Thanks to an anonymous reviewer for valuable comments. This work is supported by the European Union's Horizon 2020 research and innovation programme under the Marie Curie grant agreement No: 690974 (Mining and Reasoning with Legal Texts, MIREL).

inference patterns, but others are more abstract: factual detachment, violation detection, substitution, replacements of equivalents, implication, para-consistency, conjunction, factual monotony, norm monotony, and norm induction. We consider these ten properties as desirable for a reasoning method for normative systems, and thus they can be used also as requirements for the further development of formal methods for normative systems and deontic logic.

Keywords: Deontic Logic, Normative Systems, Benchmarks, Inference Patterns, Framework, Properties

1 Introduction

The *Handbook of Deontic Logic and Normative Systems* [5] describes a debate between the traditional or standard semantics for deontic logic and alternative approaches. The traditional semantics is based on possible world models, whereas many alternative approaches refer to foundations in normative systems, algebraic methods, or non-monotonic logic. In particular, whereas Anderson [1] argued to refer explicitly to normative systems and also Åqvist [2] builds on it, various alternative approaches such as input/output logic [13, 14] represent norms explicitly in the semantics.

Proponents of alternative approaches typically refer to limitations in the traditional approach, although the traditional approach has been generalised or extended to handle many of these limitations [10]. The development of formal and conceptual bridges between traditional and alternative approaches is one of the main current challenges in the area of normative systems and deontic logic. The following three limitations are frequently discussed.

Dilemmas. Examples discussed in the literature are those of van Fraassen [30], Makinson [13]'s Möbius strip, Prakken and Sergot [20]'s cottage regulations, and Horty [9]'s priority examples.

Defeasibility. The traditional approach does not distinguish various kinds of defeasibility. Legal norms are often assumed to be defeasible, and there is an increasing interest in philosophy in defeasibility, such as the defeasibility of moral reasons [9, 16].

Identity. Many traditional deontic logics validate the formula $\bigcirc(\alpha|\alpha)$, read as "α is obligatory given α," "whose intuitive standing is open to question" [13]. This has been dismissed as a harmless borderline case by proponents of the traditional semantics,

> but it hinders the representation of fulfilled obligations and violations, playing a central role in normative reasoning. Consider a logic validating identity: the formula $\bigcirc(\alpha|\neg\alpha)$, which represents explicitly that there is a violation, is not satisfiable; the obligation of α disappears, in context $\neg\alpha$. (See Section 2 in this paper.)

Different disciplines and applications have put forward different requirements for the development of formal methods for normative systems and deontic logic. For example, in linguistics compositionality is an important requirement, as deontic statements must be integrated into a larger theory of language. In legal informatics, constitutive and permissive norms play a central role, and legal norms may conflict. It is an open problem whether there can be a unique formal method which can be widely applied across disciplines, or even whether there is a single framework of formal methods which can be used. In this sense, there may be an important distinction between classical and normative reasoning, since there is a unique first order logic for classical logic reasoning about the real world using sets, relations and functions. The situation for normative reasoning may be closer to the situation for non-monotonic reasoning, where also a family of reasoning methods have been proposed, rather than a unique method.

In this paper we do not want to take a stance on these discussions, but we want to provide techniques and ideas to compare traditional and alternative approaches. We focus on inference patterns and proof-theory instead of semantical considerations. In particular, in this paper we are interested in the question:

> Which obligations can be detached from a set of rules or conditional norms in a context?

Our angle is different from the more traditional one in terms of inference rules.

There are many frameworks for reasoning about rules and norms, and there are many examples about detachment from normative systems, many of them problematic in some sense. However, there are few properties to compare and analyse ways to detach obligations from rules and norms, and they are scattered over the literature. We are not aware of a systematic overview of these properties. We address our research question by surveying examples, inference patterns and properties from the deontic logic literature.

Examples: Van Fraassen's paradox, Forrester's paradox, Prakken and Sergot's cottage regulations, Jeffrey's disarmament example, Chisholm's paradox, Makinson's Möbius strip, and Horty's priority example. They illustrate challenges for normative reasoning with deontic dilemmas, contrary-to-duty reasoning, defeasible obligations, reasoning by cases, deontic detachment, prioritised obligations, and combinations of these.

Inference Patterns: Conjunction, weakening of the consequent, forbidden conflict, factual detachment, strengthening of the antecedent, violation detection, compliance detection, reinstatement, deontic detachment, transitivity, and various variants of these patterns.

Framework: We develop a *framework* for deontic logics representing and resolving conflicts. By framework we mean that we do not develop a single logic, but many of them. This reflects that there is not a single logic of obligation and permission, but many of them, and which one is to be used depends on the application.

Properties: Factual detachment, violation detection, substitution, replacements of equivalents, implication, paraconsistency, conjunction, factual monotony, norm monotony, and norm induction.

The term “property" is more general than the term “inference pattern". An inference pattern describes a property of a certain form. The inference patterns listed above appear also in the list properties. For instance, factual monotony echoes strengthening of the antecedent. In some cases, we use the same name for both the inference pattern and the corresponding property.

A formal framework to compare formal methods should make as little assumptions as possible, so it is widely applicable. We only assume that the context is a set of facts $\{a, b, \ldots\}$ and that the conditional norms are of the type “if a is the case, then it ought to be the case that b” where a and b are sentences of a propositional language. This is more general than some rule-based languages based on logic programming, where a is restricted to a conjunction of literals and b is a single literal. However, it is less expressive than many other languages, that contain, for example, modal or first order sentences, constitutive and permissive norms, mixed norms such as “if a is permitted, then b is obligatory," nested operators, time, actions, knowledge, and so on. There are few benchmark examples discussed in the literature for such an extended language (see [6] for a noteworthy exception) and we are not aware of any properties specific for such extended languages. Extending our formal framework and properties to such extended languages is therefore left to further research.

Our framework is built upon the notion of detachment. In traditional approaches “if a, then it ought that b” is typically written as either $a \to \bigcirc b$ or as $\bigcirc(b|a)$, and in alternative approaches it is sometimes written as (a, b). To be able to compare the different reasoning methods, we will not distinguish between these ways to represent normative systems. The challenge for comparing the formal approaches is that traditional methods typically derive conditional obligations, whereas alternative methods typically do not, maybe because they assume norms do not have truth values and thus they cannot be derived from other norms. Instead, they derive only unconditional obligations. To compare these approaches, one may assume that the derivation of a conditional obligation "if a, then it ought that b" is short for "if the context is exactly $\{a\}$, then the obligation $\bigcirc b$ is detached." Alternatively, the detachment of an obligation for b in context a in alternative systems may be written as the derivation of a pair (a, b), as it is done in the proof theory of input/output logics [13, 14]. These issues are discussed in more detail in Section 3 of this paper.

A remark on notation and terminology. We use Greek letters α, β, γ, ...for propositional formulas, and roman letters $a, b, c, \ldots, p, q, \ldots$ for (distinct) propositional atoms. Throughout this paper the terms "rule" and "conditional norm" will be used interchangeably. The term "rule" is most often used in computer science (with reference to so-called rule-systems and expert systems), and the term "conditional norm" in philosophy and linguistics. Readers should feel free to use the term they prefer. The unconditional obligation for α will be written as $\bigcirc\alpha$, while the conditional obligation for α given β will be written as $O(\alpha|\beta)$, or as (β, α). We do not assume a specific semantics for these constructs.

We give two examples below.

Example 1.1 (Deontic explosion)**.** *The deontic explosion requirement says that we should not derive all obligations from a dilemma. Now consider a dilemma with obligations for $\alpha \wedge \beta$ and $\neg\alpha \wedge \gamma$. It may be tempting to think that an obligation for $\beta \wedge \gamma$ should follow:*

$$\dfrac{\dfrac{\bigcirc(\alpha \wedge \beta)}{\bigcirc\beta} \qquad \dfrac{\bigcirc(\neg\alpha \wedge \gamma)}{\bigcirc\gamma}}{\bigcirc(\beta \wedge \gamma)}$$

Assuming that we have replacements by logical equivalents, if we substitute a for α, $a \vee b$ for β, and $\neg a \vee b$ for γ, then we would derive from the obligations for a and $\neg a$ the obligation for c: deontic explosion. We should not derive the obligation for $\beta \wedge \gamma$, because $\alpha \wedge \beta$ and $\neg\alpha \wedge \gamma$ are classically inconsistent. As we show in Section 2.1, the obligation for $\beta \wedge \gamma$ should be derived only under suitable assumptions.

Example 1.2 (Aggregation)**.** *Consider an iterative approach deriving from the two norms "obligatory c given $a \wedge b$" and "obligatory b given a" that in some sense we have in context a that c is obligatory. This derivation of the obligation for c is made by so-called deontic* *detachment, because it is derived from the fact a together with the obligation for b. However, if the input is a together with the negation of b, then (intuitively) c should not be derived. However, we can (still intuitively) make the following two derivations. First, we can derive "obligatory a and b given c," a norm which is accepted by the two norms (Parent and van der Torre [18, 19]).*

$$\frac{\bigcirc(\alpha|\beta\wedge\gamma),\bigcirc(\beta|\gamma)}{\bigcirc(\alpha\wedge\beta|\gamma)} \qquad \frac{(\gamma,\beta),(\gamma\wedge\beta,\alpha)}{(\gamma,\beta\wedge\alpha)}$$

Second, we can also derive the ternary norm "given α, and assuming β, γ is obligatory." However, we would need to extend the language with such expressions as done by van der Torre [27] and Xin & van der Torre [24]. Different motivations for using a ternary operator can be given. For instance, one may want to reason about exceptions to norms. This is the approach taken by van der Torre [27], who works with expressions of the form "given α, γ is obligatory unless β."

This paper is organised as follows. In Section 2 we introduce benchmark examples of deontic logic, and discuss them using inference patterns. In Section 3, we introduce the formal framework and its properties. Our approach is general and conceptual, and we abstract away from any specific system from literature. The reader will find in the *Handbook of Deontic Logic and Normative Systems* sample systems which can serve to exemplify the general considerations offered in this paper.

The present paper does not cover the notion of permission nor does it cover the notion of counts-as conditional. These topics will be a subject for future research. The reader is referred to the chapter by S. O. Hansson and to the chapter by A. Jones and D. Grossi in the aforementioned handbook for an overview of the state-of-the-art and perspectives for future research regarding these notions.

2 Benchmark Examples and Inference Patterns

In this section we discuss benchmark examples of deontic logic. The analysis in this section is based on a number of inference patterns. We do not consider ways in which deontic statements can be given a semantics. These principles must be understood as expressing strict rules. For future reference, we list the inference patterns in Table 1, in the order they are discussed in this section.

pattern	name
$\bigcirc\alpha_1, \bigcirc\alpha_2 \,/\, \bigcirc(\alpha_1 \wedge \alpha_2)$	AND
$\bigcirc\alpha_1, \bigcirc\alpha_2, \Diamond(\alpha_1 \wedge \alpha_2) \,/\, \bigcirc(\alpha_1 \wedge \alpha_2)$	RAND
$\bigcirc\alpha_1 \,/\, \bigcirc(\alpha_1 \vee \alpha_2)$	W
$\bigcirc(\alpha_1\|\beta), \bigcirc(\alpha_2\|\beta), \Diamond(\alpha_1 \wedge \alpha_2) \,/\, \bigcirc(\alpha_1 \wedge \alpha_2\|\beta)$	RANDC
$\bigcirc(\alpha_1\|\beta) \,/\, \bigcirc(\alpha_1 \vee \alpha_2\|\beta)$	WC
$\bigcirc(\alpha_1\|\beta), \bigcirc(\alpha_2\|\beta), \Diamond(\alpha_1 \wedge \alpha_2 \wedge \beta) \,/\, \bigcirc(\alpha_1 \wedge \alpha_2\|\beta)$	RANDC2
$\bigcirc(\alpha_1 \wedge \alpha_2\|\beta_1), \bigcirc(\neg\alpha_1 \wedge \alpha_3\|\beta_1 \wedge \beta_2) \,/\, \bigcirc(\neg\beta_2\|\beta_1)$	FC
$\bigcirc(\alpha\|\beta), \beta \,/\, \bigcirc\alpha$	FD
$\bigcirc(\alpha\|\beta_1) \,/\, \bigcirc(\alpha\|\beta_1 \wedge \beta_2)$	SA
$\bigcirc(\alpha\|\beta_1), \Diamond(\alpha \wedge \beta_1 \wedge \beta_2) \,/\, \bigcirc(\alpha\|\beta_1 \wedge \beta_2)$	RSA
$\bigcirc(\alpha\|\beta) \,/\, \bigcirc(\alpha\|\beta \wedge \neg\alpha)$	VD
$\bigcirc(\alpha\|\beta \wedge \neg\alpha) \,/\, \bigcirc(\alpha\|\beta)$	VD$^-$
$\bigcirc(\alpha\|\beta_1), C \,/\, \bigcirc(\alpha\|\beta_1 \wedge \beta_2)$	RSA$_C$
$\bigcirc(\alpha\|\beta) \,/\, \bigcirc(\alpha\|\beta \wedge \alpha)$	CD
$\bigcirc(\alpha\|\beta \wedge \alpha) \,/\, \bigcirc(\alpha\|\beta)$	CD$^-$
$\bigcirc(\alpha_1\|\beta_1), \bigcirc(\neg\alpha_1 \wedge \alpha_2\|\beta_1 \wedge \beta_2) \,/\, \bigcirc(\alpha_1\|\beta_1 \wedge \beta_2 \wedge \neg\alpha_2)$	RI
$\bigcirc(\alpha_1\|\beta_1), \bigcirc(\neg\alpha_1 \wedge \alpha_2\|\beta_1 \wedge \beta_2),$ $\bigcirc(\neg\alpha_2\|\beta_1 \wedge \beta_2 \wedge \beta_3) \,/\, \bigcirc(\alpha_1\|\beta_1 \wedge \beta_2 \wedge \beta_3)$	RIO
$\bigcirc(\alpha\|\beta_1), \bigcirc(\alpha\|\beta_2) \,/\, \bigcirc(\alpha\|\beta_1 \vee \beta_2)$	ORA
$\bigcirc(\alpha\|\beta), \bigcirc\beta \,/\, \bigcirc\alpha$	DD
$\bigcirc(\alpha\|\beta), \bigcirc(\beta\|\gamma) \,/\, \bigcirc(\alpha\|\gamma)$	T
$\bigcirc(\alpha\|\beta \wedge \gamma), \bigcirc(\beta\|\gamma) \,/\, \bigcirc(\alpha\|\gamma)$	CT
$\bigcirc(\alpha\|\beta \wedge \gamma), \bigcirc(\beta\|\gamma) \,/\, \bigcirc(\alpha \wedge \beta\|\gamma)$	ACT

Table 1: Inference patterns

The letter C in RSA$_C$ stands for the condition: there is no premise $\bigcirc(\alpha' \mid \beta')$ such that $\beta_1 \wedge \beta_2$ logically implies β', β' logically implies β_1 and not vice versa, α and α' are contradictory and $\alpha \wedge \beta'$ is consistent. RSA$_C$ is not a rule in the usual proof-theoretic sense. For it has a statement that quantifies over all other premises as an auxiliary condition. Thus the rule is not on a par with the other rules, like for instance weakening of the output.

2.1 Van Fraassen's Paradox

We first discuss deontic explosion in van Fraassen's paradox, then the trade-off between on the one hand "ought implies can" and on the other hand the representation of violations in the violation detection problem, whether it is forbidden to put oneself into a dilemma, and finally the use of priorities to resolve conflicts.

2.1.1 Deontic Explosion: Conjunction versus Weakening

It is a well-known problem from paraconsistent logic that the removal of all inconsistent formulas from the language is insufficient to reason in the presence of a contradiction, because there may still be explosion in the sense that all formulas of the language are derived from a contradiction. The following derivation illustrates how we can derive q from p and $\neg p$ in propositional logic, where all formulas in the derivation are classically consistent.

$$\dfrac{\dfrac{\dfrac{p}{q \vee p} \qquad \neg p}{q \wedge \neg p}}{q}$$

The rules of replacements of logical equivalents, $\vee$-introduction, $\wedge$-introduction, and $\wedge$-elimination are used in this derivation.

A similar phenomenon occurs in deontic logic, if we reason about deontic dilemmas or conflicts, that is situations where $\bigcirc p$ and $\bigcirc \neg p$ both hold. Van der Torre and Tan [29] call this deontic explosion problem "van Fraassen's paradox," because van Fraassen [30] gave the following (informal) analysis of dilemmas in deontic logic. He rejects the conjunction pattern AND:

$$\text{AND}: \frac{\bigcirc \alpha_1, \bigcirc \alpha_2}{\bigcirc (\alpha_1 \wedge \alpha_2)}$$

This is because AND warrants the move from $\bigcirc p \wedge \bigcirc \neg p$ to $\bigcirc (p \wedge \neg p)$, and such a conclusion is not consistent with the principle 'ought implies can', formalised as $\neg \bigcirc (p \wedge \neg p)$. However, he does not want to reject the conjunction pattern in all cases. In particular, he wants to be able to derive $\bigcirc (p \wedge q)$ from $\bigcirc p \wedge \bigcirc q$ when p and q are distinct propositional atoms. His suggestion is that a restriction should be placed on the conjunction pattern: one derives $\bigcirc (\alpha_1 \wedge \alpha_2)$ from $\bigcirc \alpha_1$ and $\bigcirc \alpha_2$ only if $\alpha_1 \wedge \alpha_2$ is consistent. He calls the latter inference pattern *Consistent Aggregation*, renamed to restricted conjunction (RAND) by van der Torre and Tan in their following variant of van Fraassen's suggestion.

Example 2.1 (Van Fraassen's paradox [29]). *Consider a deontic logic without nested modal operators in which dilemmas like $\bigcirc p \wedge \bigcirc\neg p$ are consistent, but which validates $\neg \bigcirc \bot$, where $\bot$ stands for any contradiction like $p\wedge\neg p$. Moreover, assume that it satisfies replacement of logical equivalents and at least the following two inference patterns* Restricted Conjunction *(*RAND*), also called consistent aggregation, and* Weakening *(*W*), where $\Diamond\phi$ can be read as "ϕ is possible" (possibility is not necessarily the same as consistency).*

$$\text{RAND:}\frac{\bigcirc\alpha_1, \bigcirc\alpha_2, \Diamond(\alpha_1 \wedge \alpha_2)}{\bigcirc(\alpha_1 \wedge \alpha_2)} \qquad \text{W:}\frac{\bigcirc\alpha_1}{\bigcirc(\alpha_1 \vee \alpha_2)}$$

Moreover, assume the two premises 'Honor thy father or thy mother!' $\bigcirc(f \vee m)$ *and* 'Honor not thy mother!' $\bigcirc\neg m$. *The left derivation of Figure 1 illustrates how the desired conclusion* 'thou shalt honor thy father' $\bigcirc f$ *can be derived from the premises. Unfortunately, the right derivation of Figure 1 illustrates that we cannot accept restricted conjunction and weakening in a monadic deontic logic, because we can derive* every $\bigcirc\beta$ *from* $\bigcirc\alpha$ *and* $\bigcirc\neg\alpha$.

$$\dfrac{\dfrac{\bigcirc(f \vee m) \quad \bigcirc\neg m}{\bigcirc(f \wedge \neg m)}\,\text{RAND}}{\bigcirc f}\,\text{W} \qquad\qquad \dfrac{\dfrac{\dfrac{\bigcirc\alpha}{\bigcirc(\alpha \vee \beta)}\,\text{W} \quad \bigcirc\neg\alpha}{\bigcirc(\neg\alpha \wedge \beta)}\,\text{RAND}}{\bigcirc\beta}\,\text{W}$$

Figure 1: Van Fraassen's paradox

Van Fraassen's paradox has a counterpart in dyadic deontic logic. The paradox consists in deriving $\bigcirc(\gamma\,|\,\beta)$ from $\bigcirc(\alpha\,|\,\beta)$ and $\bigcirc(\neg\alpha\,|\,\beta)$ using the following rules of *Restricted Conjunction for the Consequent* (RANDC) and *Weakening of the Consequent* (WC).

$$\text{RANDC}: \frac{\bigcirc(\alpha_1|\beta), \bigcirc(\alpha_2|\beta), \Diamond(\alpha_1 \wedge \alpha_2)}{\bigcirc(\alpha_1 \wedge \alpha_2|\beta)} \qquad \text{WC}: \frac{\bigcirc(\alpha_1|\beta)}{\bigcirc(\alpha_1 \vee \alpha_2|\beta)}$$

2.1.2 Violation Detection Problem: Unrestricted versus Restricted Conjunction

Whereas $p \wedge \neg p$ can not be derived in a paraconsistent logic, we can consistently represent the formula $\bigcirc(p\wedge\neg p)$ in a modal logic, and we can block deontic explosion using a minimal modal logic [3]. This raises the question whether we should accept the conjunction pattern unrestrictedly or in its restricted form.

The choice between the two can be illustrated as follows. Suppose we can derive the obligation $\bigcirc(p \wedge \neg p)$ from $\bigcirc(p)$ and $\bigcirc(\neg p)$ without deriving $\bigcirc f$, or any other counter-intuitive consequence. In that case, is $\bigcirc(p \wedge \neg p)$ by itself a consequence we want to block? This presents us with a choice. On the one hand we would like to block $\bigcirc(p \wedge \neg p)$, because it contradicts the "ought implies can" principle. On the other hand, we would like to allow the derivation of $\bigcirc(p \wedge \neg p)$, because such a formula represents explicitly the fact that there is a dilemma.

This choice is even more subtle in dyadic deontic logic. There is the extra question as to whether the "ought implies can" reading implies that the obligation in the consequent must only be consistent in itself, or consistent with the antecedent too. The latter requirement is represented by the following variant of the *Restricted Conjunction for the Consequent* pattern, which we call RANDC2.

$$\text{RANDC2} : \frac{\bigcirc(\alpha_1|\beta), \bigcirc(\alpha_2|\beta), \Diamond(\alpha_1 \wedge \alpha_2 \wedge \beta)}{\bigcirc(\alpha_1 \wedge \alpha_2|\beta)}$$

On the one hand, given $\bigcirc(p \,|\, \neg p \vee \neg q)$ and $\bigcirc(q \,|\, \neg p \vee \neg q)$, we would like to block the derivation of $\bigcirc(p \wedge q|\neg p \vee \neg q)$ because "ought implies can". On the other hand, we would like to be able to derive it in order to make explicit that $\neg p \vee \neg q$ gives rise to a dilemma, and is not consistent with the fulfillment of the two obligations appearing as premises.

The alternative restricted conjunction pattern RANDC2 highlights the distinction between what we call the violability and the temporal interpretation of dyadic deontic logic. The former interprets the obligation $O(\alpha \,|\, \beta)$ as "given that β has been settled beyond repair, we should do α to make the best out of the sad circumstances" [7] and the latter as "if α is the case now, what should be the case next?" The violability interpretation says that $O(\neg\alpha \,|\, \alpha)$ represents that α is a violation. For example, if you are going to kill, then do it gently. The temporal interpretation says that the present situation must be changed—which may or may not indicate a violation. For example, the temporal interpretation may be used to express a conditional obligation like "if the light is on, turn it off!"

We would like to point out that the violability interpretation is more expressive, in the sense that the temporal interpretation can be represented by introducing distinct propositional letters for what is the case now, and what is the case in the next moment. For example, "if the light is on, turn it off" can be represented by $\bigcirc(\neg on_2 \,|\, on_1)$, where on_1 represents that the light is on now, and on_2 that it is on at the next moment in time. In the temporal interpretation, however, it seems impossible to represent all violations in a natural way. Thus, a temporal interpretation with future directed obligations only seems to be a strong limitation.

We use the name "violation detection problem" to refer to the phenomenon that with the restricted conjunction pattern the representation (and hence the detection) of violations is made impossible. We continue the discussion on the violation detection problem in Section 2.2, where we discuss restricted inference patterns formalising contrary-to-duty reasoning.

2.1.3 Forbidden Conflicts

Here is another question raised by dilemmas: is it forbidden to create a dilemma? The following inference pattern is called *Forbidden Conflict* (FC). If the inference pattern is accepted, then it is not allowed to bring about a conflict, because a conflict is sub-ideal.

$$\text{FC}: \frac{\bigcirc(\alpha_1 \wedge \alpha_2|\beta_1), \bigcirc(\neg\alpha_1 \wedge \alpha_3|\beta_1 \wedge \beta_2)}{\bigcirc(\neg\beta_2|\beta_1)}$$

Here is an example, taken from van der Torre and Tan [28]. Assume the premises $\bigcirc k$ and $\bigcirc(p \wedge \neg k|d)$, where k can be read as 'keeping a promise', p as 'preventing a disaster' and d as 'a disaster will occur if nothing is done to prevent it'. (FC) yields $\bigcirc\neg d$. There are situations where this is the right outcome. Consider a person having the obligation to keep a promise to show up at a birthday party. We have $\bigcirc k$, but also $\bigcirc(p \wedge \neg k \,|\, d)$. She does not want to go, and so before leaving she does something that might result in a disaster later on, like leaving the coffee machine on. During the party, she leaves and goes home, using her second obligation as an excuse. Nobody will contest that leaving the machine on (on purpose) was a violation already, viz. $\bigcirc\neg d$.

An instance of this inference pattern has been discussed in defeasible deontic logic, and we return to it in Section 2.3.

2.1.4 Resolving Dilemmas

To resolve a conflict between an obligation for p and an obligation for $\neg p$, we need additional information. For example, a total preference order on sets of propositions can resolve all dilemmas by picking the preferred set of obligations among the alternatives of the dilemma, and weaker relations on sets of propositions such as a total pre-order or a partial order leaves some dilemmas unresolved.

The most studied source for a preference order over sets of propositions is a preference order over propositions, which is then lifted to an order on sets of propositions. For example, an ordering on obligations can be derived from an ordering on the authorities who created the obligations, or the moment in time they were created. The level of preference of an obligation may reflect its priority.

Consider three obligations with priority 3, 2 and 1, and a dilemma between the first and the latter two. To represent the priority of an obligation, we write it in the $\bigcirc$ notation. A higher number reflects a higher priority.

$$\{③(p \wedge q), ②\neg p, ①\neg q\}$$

In other words, we can either satisfy the most important obligation $③(p \wedge q)$, or two less important obligations $②\neg p$ and $①\neg q$. Can this dilemma be resolved? There are various well known possibilities in the area of non-monotonic logic. Whether they can be used depends on the origin of the priorities and the application.

The issue of lifting priorities from obligations to sets of them gets more challenging when we consider conditional obligations and deontic detachment, as discussed later on in Section 2.7.

2.2 Forrester's Paradox

We first discuss factual detachment in Forrester's paradox, then the problematic derivation of secondary obligations from primary ones, and finally what we call the violation detection problem for Forrester's paradox.

2.2.1 Factual Detachment versus Conjunction

Forrester's paradox consists of the four sentences 'Smith should not kill Jones,' 'if Smith kills Jones, then he should do it gently,' 'Smith kills Jones', and 'killing someone gently logically implies killing him.' The preference based models of dyadic deontic logic give a natural representation of the two obligations: not killing is preferred to gentle killing, and both are preferred to other forms of killing. However, the following example illustrates that it is less clear how to combine dyadic obligation with factual detachment, deriving unconditional obligations from conditional ones.

Example 2.2 (Forrester's paradox)**.** *Assume a dyadic deontic logic without nested modal operators that has at least replacement of logical equivalents, the Conjunction pattern* AND *and the following inference pattern called factual detachment* FD.

$$\text{FD}: \frac{\bigcirc(\alpha|\beta), \beta}{\bigcirc\alpha}$$

Furthermore, assume the following premise set with background knowledge that gentle murder implies murder $\vdash g \rightarrow k$.

$$S = \{\bigcirc(\neg k|\top), \bigcirc(g|k), k\}$$

The set S represents the Forrester paradox when k is read as 'Smith kills Jones' *and g as* 'Smith kills Jones gently.' *We say that the last obligation is a contrary-to-duty obligation with respect to the first obligation, because its antecedent is contradictory with the consequent of the first obligation. Figure 2 visualizes how we can represent the concept of contrary-to-duty as a binary relation among dyadic obligations: the obligation $\bigcirc(\alpha_2|\beta_2)$ is a contrary-to-duty with respect to $\bigcirc(\alpha_1|\beta_1)$ if and only if $\beta_2 \wedge \alpha_1$ is inconsistent.*

$$\begin{array}{c} \bigcirc(\neg k|\top) \\ \text{inconsistent} \ \updownarrow \\ \bigcirc(g|k) \end{array}$$

Figure 2: $\bigcirc(g|k)$ is a contrary-to-duty obligation with respect to $\bigcirc(\neg k|\top)$

The derivation in Figure 3 illustrates how the obligation $\bigcirc(\neg k \wedge g)$, i.e. $\bigcirc(\bot)$, can be derived from S by FD *and* AND.

$$\dfrac{\dfrac{\bigcirc(\neg k|\top) \quad \top}{\bigcirc(\neg k)}\,\text{FD} \qquad \dfrac{\bigcirc(g|k) \quad k}{\bigcirc(g)}\,\text{FD}}{\bigcirc(\neg k \wedge g)}\,\text{AND}$$

Figure 3: Forrester's paradox

Forrester's paradox can be given two interpretations. First, the dilemma interpretation says that the two obligations give rise to a dilemma, just like the obligations $\bigcirc p$ and $\bigcirc\neg p$ in van Fraassen's paradox. Consequently, according to the dilemma interpretation, there is no problem, the derivation of $\bigcirc(\bot)$ just reflects the fact that there is a dilemma.

The coherent interpretation appeals to the independent and seemingly plausible principle 'ought implies can', $\neg \bigcirc (\bot \,|\, \alpha)$. According to this interpretation, the Forrester set is intuitively consistent with the 'ought implies can' principle, and so there is no dilemma, just an obligation to act as good as possible in the sub-ideal situation where the primary obligation has been violated.

There is a consensus in the literature that the example should be given a coherent interpretation, and that the dilemma interpretation is wrong.

2.2.2 Deriving Secondary Obligations from Primary Ones: Strengthening of the Antecedent versus Weakening of the Consequent

The following example shows that Forrester's paradox can be used also to illustrate that combining the desirable inference patterns strengthening of the antecedent and weakening of the consequent is problematic in dyadic deontic logic. For example, strengthening of the antecedent is used to derive 'Smith should not kill Jones in the morning' $\bigcirc(\neg k|m)$ from the obligation 'Smith should not kill Jones' $\bigcirc(\neg k|\top)$ and weakening of the consequent is used to derive 'Smith should not kill Jones' $\bigcirc(\neg k|\top)$ from the obligation 'Smith should drive on the right side of the street and not kill Jones' $\bigcirc(r \wedge \neg k|\top)$.

Example 2.3 (Forrester's paradox, cont'd [29])**.** *Assume a dyadic deontic logic without nested modal operators that has at least replacement of logical equivalents and the following inference patterns* Strengthening of the Antecedent *(*SA*), the* Conjunction pattern for the Consequent *(*ANDC*) and* Weakening of the Consequent *(*WC*)* .

$$\text{SA} : \frac{\bigcirc(\alpha|\beta_1)}{\bigcirc(\alpha|\beta_1 \wedge \beta_2)} \qquad \text{ANDC} : \frac{\bigcirc(\alpha_1|\beta), \bigcirc(\alpha_2|\beta)}{\bigcirc(\alpha_1 \wedge \alpha_2|\beta)} \qquad \text{WC} : \frac{\bigcirc(\alpha_1|\beta)}{\bigcirc(\alpha_1 \vee \alpha_2|\beta)}$$

The derivation in Figure 4 illustrates how the obligation $\bigcirc(\neg k \wedge g|k)$, i.e. $\bigcirc(\bot|k)$, can be derived from S by SA *and* ANDC*. Note that the dyadic obligation $\bigcirc(\neg k|k)$ can be given only a violability interpretation in this example, not a temporal interpretation, because it is impossible to undo a killing. That is, this dyadic obligation can be read only as "if Smith kills Jones, then this is a violation."*

$$\dfrac{\dfrac{\bigcirc(\neg k|\top)}{\bigcirc(\neg k|k)}\,\text{SA} \qquad \bigcirc(g|k)}{\bigcirc(\neg k \wedge g|k)}\,\text{ANDC} \qquad\qquad \dfrac{\dfrac{\dfrac{\bigcirc(\neg k|\top)}{\bigcirc(\neg g|\top)}\,\text{WC}}{\bigcirc(\neg g|k)}\,\text{RSA} \qquad \bigcirc(g|k)}{\bigcirc(\neg g \wedge g|k)}\,\text{ANDC}$$

Figure 4: Forrester's paradox

The derivation is blocked when SA *is replaced by the following inference pattern* Restricted Strengthening of the Antecedent *(*RSA*).*

$$\text{RSA} : \frac{\bigcirc(\alpha|\beta_1), \Diamond(\alpha \wedge \beta_1 \wedge \beta_2)}{\bigcirc(\alpha|\beta_1 \wedge \beta_2)}$$

However, the obligation $\bigcirc(\bot|k)$ can still be derived from S by WC*,* RSA *and* ANDC*. This derivation from the set of obligations is represented on the right hand side of Figure 4. Like in Example 2.2, we can give the set a dilemma or a coherent interpretation.*

The underlying problem of the counterintuitive derivation in Figure 4 is the derivation of $\bigcirc(\neg g|k)$ from the first premise $\bigcirc(\neg k|\top)$ by WC *and* RSA*, because it derives a contrary-to-duty obligation from its own primary obligation.*

Since there is consensus that Forrester's paradox should be given a coherent interpretation, Forrester's paradox in Example 2.3 shows that combining strengthening of the antecedent and weakening of the consequent is problematic for *all* deontic logics.

2.2.3 Violation Detection Problem: Restricted versus Unrestricted Strengthening of the Antecedent

The choice between the unrestricted version and the restricted version of the law of strengthening of the antecedent has some similarity with the choice between the unrestricted version and the restricted version of the law of conjunction. This can be illustrated as follows. Suppose we have the obligation $\bigcirc(\neg k|\top)$. In that case, is $\bigcirc(\neg k|k)$ a consequence we want to block? This presents us with a choice. On the one hand, we would like to block $\bigcirc(\neg k|k)$, because it contradics the "ought implies can" principle. On the other hand, we would like to allow the derivation of $\bigcirc(\neg k|k)$, because this formula represents explicitly that there is a violation. (Cf. our explanatory comments on the violability interpretation, on p. 10.)

The following inference pattern *Violation Detection* (VD) formalizes the intuition that an obligation cannot be defeated by only violating it, and represents a solution to the violation detection problem. The VD pattern models the intuition that after violation the obligation to do α is still in force. Even if you drive too fast, you are still obliged to obey the speed limit.

$$\text{VD}: \frac{\bigcirc(\alpha|\beta)}{\bigcirc(\alpha|\beta\wedge\neg\alpha)} \qquad \text{VD}^-: \frac{\bigcirc(\alpha|\beta\wedge\neg\alpha)}{\bigcirc(\alpha|\beta)}$$

The inverse pattern VD$^-$ says that violations do not come out of the blue. Although this inference pattern may seem intuitive at first sight, it appears too strong on further inspection.

Example 2.4 (Metro)**.** *Consider the following derivation.*

$$\cfrac{\cfrac{\bigcirc(\alpha|\beta)}{\bigcirc(\alpha|\beta\wedge\neg\alpha)}\ \text{VD}}{\bigcirc(\alpha|\alpha\vee\beta)}\ \text{VD}^-$$

For example, assume that if you travel by metro, you must have a ticket. We can derive that traveling by metro without a ticket is a violation. The two inference patterns together would derive that if you travel by metro or you buy a ticket, then you must buy a ticket. This is *counterintuitive, because buying a ticket without traveling by metro does not involve any obligations. The example illustrates how reasoning about violations only can lead to the wrong conclusions.*

Normative systems typically associate sanctions with violations, as an incentive for agents to obey the norms. Such sanctions can sometimes be expressed as contrary-to-duty obligations: the sanction to pay a fine if you do not return the book to the library in time, can be modelled as a contrary-to-duty obligation to pay the fine. By symmetry, though this is less often implemented in normative systems, rewards can be associated with compliance of obligations. In modal logic, an obligation for α is fulfilled if we have $\alpha \wedge \bigcirc\alpha$.

The following inference pattern *Compliance Detection* (CD) formalizes the intuition that an obligation cannot be defeated by only complying with it, analogous to the *Violation Detection* (VD) pattern.

$$\text{CD} : \frac{\bigcirc(\alpha|\beta)}{\bigcirc(\alpha|\beta \wedge \alpha)} \qquad \text{CD}^- : \frac{\bigcirc(\alpha|\beta \wedge \alpha)}{\bigcirc(\alpha|\beta)}$$

The following example illustrates that the inference pattern CD should not be confused with the inverse of CD$^-$, which seems to say that fulfilled obligations do not come out of the blue. Although this inference pattern may seem intuitive at first sight, it is highly counterintuitive on further inspection.

Example 2.5 (Forrester, continued)**.** *Consider the following derivation.*

$$\cfrac{\cfrac{\bigcirc(\alpha \wedge \beta|\alpha)}{\bigcirc(\alpha \wedge \beta|\alpha \wedge \beta)}\ \text{CD}}{\bigcirc(\alpha \wedge \beta|\top)}\ \text{CD}^-$$

You should kill gently, if you kill $\bigcirc(k \wedge g\,|\,k)$. Hence, by CD*, you should kill gently, if you kill gently $\bigcirc(k \wedge g|k \wedge g)$ (a fulfilled obligation). However, this does not mean that there is an unconditional obligation to kill gently $\bigcirc(k \wedge g\,|\,\top)$. Hence, the inference pattern* CD$^-$ *should not be valid.*

Without the CD pattern, we say that the fulfilled obligation "disappears," analogous to violations. A fulfilled obligation also disappears when we have as an axiom of the logic that $\bigcirc(\alpha\,|\,\beta) \leftrightarrow \bigcirc(\alpha \wedge \beta\,|\,\beta)$, because in that case $\bigcirc(\alpha \wedge \beta\,|\,\beta)$ does not hold because β is compliant with a norm.

2.3 Prakken and Sergot's Cottage Regulations

We first discuss the extension of Forrester's paradox with defeasible obligations, then we return to the violation detection problem, and finally we discuss reinstatement.

2.3.1 Violations and Exceptions

The so-called cottage regulations are introduced by Prakken and Sergot [20] to illustrate the distinction between contrary-to-duty reasoning and defeasible reasoning based on exceptional circumstances. It is an extended version of the Forrester or gentle murderer paradox discussed in Section 2.2. The following example is an alphabetic variant of the original example, because we replaced s, to be read as 'the cottage is by the sea,' by d, to be read as 'there is a dog.' Moreover, as is common, instead of representing background knowledge that w implies f, Prakken and Sergot represent a white fence by $w \wedge f$.

Example 2.6 (Cottage regulations [28])**.** *Assume a deontic logic that validates at least replacement of logical equivalents and the inference pattern* $\textsc{rsa}_C$.

$$\textsc{rsa}_C : \frac{\bigcirc(\alpha|\beta_1), C}{\bigcirc(\alpha|\beta_1 \wedge \beta_2)}$$

> *C: there is no premise $\bigcirc(\alpha' \mid \beta')$ such that $\beta_1 \wedge \beta_2$ logically implies β', β' logically implies β_1 and not vice versa, α and α' are contradictory and $\alpha \wedge \beta'$ is consistent. [26]*

$\textsc{rsa}_C$ *formalises a principle of specificity to deal with exceptional circumstances. It is illustrated with Figure 5 (a). Suppose we are given these rules: you ought not to eat with your fingers; if you are served asparagus, you ought to eat with your fingers. One does not want to be able to strengthen the first obligation into: if you are served asparagus, you ought not to eat with your fingers. Such a strengthening is blocked by* $\textsc{rsa}_C$.

Now, assume the obligations

$$S = \{\bigcirc(\neg f|\top), \bigcirc(w \wedge f|f), \bigcirc(w \wedge f|d)\},$$

where f can be read as 'there is a fence around your house,' $w \wedge f$ as 'there is a white fence around your house' and d as 'you have a dog.' Notice that $\bigcirc(w \wedge f \mid f)$ is a contrary-to-duty obligation with respect to $\bigcirc(\neg f \mid \top)$ and $\bigcirc(w \wedge f \mid d)$ is not. If all we know is that there is a fence and a dog ($f \wedge d$), then the first obligation in S is intuitively overridden, and therefore it cannot be violated. Hence, the obligation $\bigcirc(\neg f \mid f \wedge d)$ should not *be derivable.*

However, if all we know is that there is a fence without a dog (f), then the first obligation in S is intuitively not overridden, and therefore it is violated. Hence, the obligation $\bigcirc(\neg f|f)$ should be derivable.

One should be careful not to treat both $\bigcirc(w \wedge f|f)$ and $\bigcirc(w \wedge f|d)$ as more specific obligations that override the obligation $\bigcirc(\neg f \mid \top)$: this does not hold for $\bigcirc(w \wedge f \mid f)$. The latter obligation should be treated as a contrary-to-duty obligation, i.e. as a case of violation. This interference of specificity and contrary-to-duty is represented in Figure 5. This figure should be read as follows. Each arrow is a condition: a two-headed arrow is a consistency check, and a single-headed arrow is a logical implication. For example, the condition C formalizes that an obligation $\bigcirc(\alpha|\beta)$ is overridden by $\bigcirc(\alpha'|\beta')$ if the conclusions are contradictory (a consistency check, the double-headed arrow) and the condition of the overriding obligation is more specific (β' logically implies β). Case (a) represents criteria for overridden defeasibility, and case (b) represents criteria for contrary-to-duty. Case (c) shows that the pair $\bigcirc(\neg f \mid \top)$ and $\bigcirc(w \wedge f \mid f)$ can be viewed as overridden defeasibility as well as contrary-to-duty.

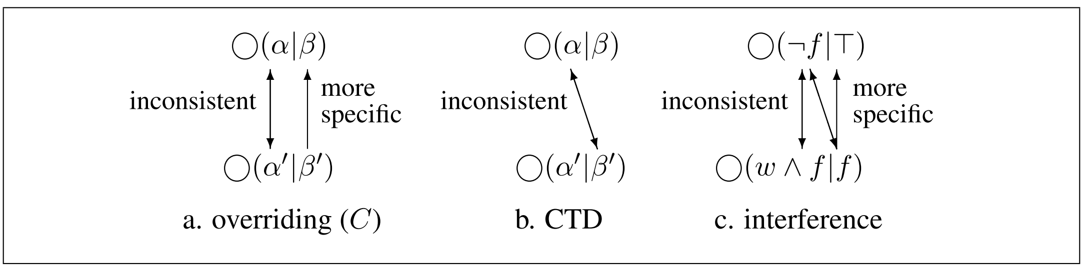

Figure 5: Specificity and CTD

2.3.2 Violation Detection Problem for Defeasible Obligations

What is most striking about the cottage regulations is the observation that when the premise $\bigcirc(\neg f \mid \top)$ is violated by f, then the obligation for $\neg f$ should be derivable, but not when $\bigcirc(\neg f \mid \top)$ is overridden by $f \wedge d$. In other words, we have to distinguish violations from exceptions.

In approaches where $\bigcirc(\alpha|\beta)$ implies that $\alpha \wedge \beta$ is consistent, we cannot represent this difference by deriving $\bigcirc(\neg f|f)$ and not deriving $\bigcirc(\neg f|d \wedge f)$. In this sense, this is again an example of the violation detection problem.

We can use priorities to represent the specificity example, by giving the more specific obligation a higher priority. Many conditional logics have specificity built in, but this must be combined with other conflict resolution methods, for example based on time or authority. This is an issue of reasoning about uncertainty, default reasoning, and nonmonotonic logic.

2.3.3 Reinstatement

The question raised by the inference pattern *Reinstatement* (RI) is whether an obligation can be overridden by an overriding obligation that itself is violated. The obligation $\bigcirc(\alpha_1|\beta_1)$ is overridden by $\bigcirc(\neg\alpha_1 \wedge \alpha_2|\beta_1 \wedge \beta_2)$ for $\beta_1 \wedge \beta_2$, but is it also overridden for $\beta_1 \wedge \beta_2 \wedge \neg\alpha_2$? If the last conclusion is not accepted, then the first obligation α_1 should be in force again. Hence, the original obligation is reinstated.

$$\text{RI} : \frac{\bigcirc(\alpha_1|\beta_1), \bigcirc(\neg\alpha_1 \wedge \alpha_2|\beta_1 \wedge \beta_2)}{\bigcirc(\alpha_1|\beta_1 \wedge \beta_2 \wedge \neg\alpha_2)}$$

Suppose you are in the street, and see a child's bike unattended. As a general rule, you should not take the bike, viz. $\bigcirc\neg t$ where t is for taking the bike. Now, suppose you also observe an elderly neighbor collapse with what might be a heart attack. You are a block away from the nearest phone from which you could call for help. In that more specific situation, you should take the bike and go call for help, $\bigcirc(t \wedge h \mid e)$, where e and h are for an elderly neighbor collapses and go call for help, respectively. The obligation $\bigcirc\neg t$ is overriden by $\bigcirc(t \wedge h|e)$ for e. But it is not overriden for $e \wedge \neg g$. Of course, if you do not go for help, then the prohibition of t remains.

The following inference pattern RIO is a variant of the previous inference pattern RI, in which the overriding obligation is not factually defeated but overridden. The obligation $\bigcirc(\alpha_1|\beta_1)$ is overridden by $\bigcirc(\neg\alpha_1 \wedge \alpha_2|\beta_1 \wedge \beta_2)$ for $\beta_1 \wedge \beta_2$, and the latter is overridden by $\bigcirc(\neg\alpha_2|\beta_1 \wedge \beta_2 \wedge \beta_3)$ for $\beta_1 \wedge \beta_2 \wedge \beta_3$. The inference pattern RIO says that an obligation cannot be overridden by an obligation that is itself overridden. Hence, an overridden obligation becomes reinstated when its overriding obligation is itself overridden.

$$\text{RIO} : \frac{\bigcirc(\alpha_1|\beta_1), \bigcirc(\neg\alpha_1 \wedge \alpha_2|\beta_1 \wedge \beta_2), \bigcirc(\neg\alpha_2|\beta_1 \wedge \beta_2 \wedge \beta_3)}{\bigcirc(\alpha_1|\beta_1 \wedge \beta_2 \wedge \beta_3)}$$

Example: you should not kill; if you find yourselves in a situation of self-defence, you should kill; if you find yourselves in a situation of self-defence, but your opponent is weak, you should not kill.

Van der Torre and Tan [28] argue that Reinstatement does not hold in general, for example it does not hold for obligations under uncertainty. However, they argue also that these patterns hold for so-called prima facie obligations. The notion of prima facie obligation was introduced by Ross [21]. He writes: 'I suggest '*prima facie* duty' or 'conditional duty' as a brief way of referring to the characteristic (quite distinct from that of being a duty proper) which an act has, in virtue of being of a certain kind (e.g. the keeping of a promise), of being an act which would be a duty proper if it were not at the same time of another kind which is morally significant' [21, p.19]. A prima facie duty is a duty proper when it is not overridden by another prima facie duty. When a prima facie obligation is overridden, it is not a proper duty but it is still in force: 'When we think ourselves justified in breaking, and indeed morally obliged to break, a promise [...] we do not for the moment cease to recognize a prima facie duty to keep our promise' [21, p.28].

Van der Torre and Tan argue also that the inference pattern Forbidden Conflict, discussed in Section 2.1.3, does not hold in general, but it holds for prima facie obligations. If the inference pattern is accepted, then it is not allowed to bring about a conflict, because a conflict is sub-ideal, even when it can be resolved.

2.4 Jeffrey's Disarmament Paradox

In general, reasoning by cases is a desirable property of reasoning with conditionals. In this reasoning scheme, a certain fact is proven by proving it for a set of mutually exclusive and exhaustive circumstances. For example, assume that you want to know whether you want to go to the beach. If you desire to go to the beach when it rains, and you desire to go to the beach when it does not rain, then you may conclude by this scheme 'reasoning by cases' that you desire to go to the beach under all circumstances. The two cases considered here are rain and no rain. This kind of reasoning schemes can be formalized by the following derivation: *If* 'α if β' *and* 'α if not β,' *then* 'α regardless of β.' Formally, if we write the conditional 'α if β' by $\beta > \alpha$, then it is represented by the following disjunction pattern for the antecedent.

$$\text{ORA:}\frac{\beta > \alpha, \neg\beta > \alpha}{\top > \alpha}$$

The following example illustrates that the disjunction pattern for the antecedent combined with strengthening of the antecedent derives counterintuitive consequences in dyadic deontic logic. Example 2.7 is based on the following classic illustration of Jeffrey [11], see also the discussion by Thomason and Horty [25].

Example 2.7 (Disarmament paradox [29])**.** *Assume a deontic logic that validates at least replacement of logical equivalents and the two inference patterns* RSA *and the* Disjunction pattern for the Antecedent *(*ORA*),*

$$\text{ORA}: \frac{\bigcirc(\alpha|\beta_1), \bigcirc(\alpha|\beta_2)}{\bigcirc(\alpha|\beta_1 \vee \beta_2)}$$

and assume as premises the obligations ‘we ought to be disarmed if there will be a nuclear war’*,* ‘we ought to be disarmed if there will be no war’ *and* ‘we ought to be armed if we have peace if and only if we are armed’. *They may be formalized as $\bigcirc(d \mid w)$, $\bigcirc(d \mid \neg w)$ and $\bigcirc(\neg d \mid d \leftrightarrow w)$, respectively. The derivation in Figure 6 shows how we can derive the counterintuitive $\bigcirc(d \wedge \neg d|d \leftrightarrow w)$. The derived obligation is inconsistent in most deontic logics, whereas intuitively the set of premises is consistent. The derivation of $\bigcirc(d|d \leftrightarrow w)$ is counterintuitive, because it is not possible to fulfill this obligation together with the obligation $\bigcirc(d \mid \neg w)$ it is derived from. The contradictory fulfillments are respectively $d \wedge w$ and $d \wedge \neg w$.*

$$\dfrac{\dfrac{\dfrac{\bigcirc(d|w) \quad \bigcirc(d|\neg w)}{\bigcirc(d|\top)}\ \text{ORA}}{\bigcirc(d|d \leftrightarrow w)}\ \text{RSA} \qquad \bigcirc(\neg d|d \leftrightarrow w)}{\bigcirc(d \wedge \neg d|d \leftrightarrow w)}\ \text{AND}$$

Figure 6: The disarmament paradox

In other words, in this derivation the obligation $\bigcirc(d \mid d \leftrightarrow w)$ is considered to be counterintuitive, because it is not grounded in the premises. If $d \leftrightarrow w$ and w (the antecedent of the first premise) are true then d is trivially true, and if $d \leftrightarrow w$ and $\neg w$ (the antecedent of the second premise) are true then d is trivially false. In other words, if $d \leftrightarrow w$ then the first premise cannot be violated and the second premise cannot be fulfilled. Hence, the two premises do not ground the conclusion that for arbitrary $d \leftrightarrow w$ we have that $\neg d$ is a violation.

The example is difficult to interpret, because it makes use of a bi-implication. An alternative set of premises, also based on bi-implications, with analogous counterintuitive conclusions is $\{\bigcirc(d|d \leftrightarrow w), \bigcirc(d|\neg d \leftrightarrow w), \bigcirc(\neg d|w)\}$.

ORA also plays a role in the so-called miners’ scenario introduced recently by Kolodny and MacFarlane [12].

2.5 Chisholm's Paradox

The second contrary-to-duty paradox we consider is Chisholm [4]'s paradox. We first discuss the choice between deontic versus factual detachment, and then the representation of deontic detachment. We discuss the violation detection problem for deontic detachment only in Section 2.6 after we have introduced Makinson's Möbius strip example.

2.5.1 Deontic versus Factual Detachment

Chisholm's paradox consists of the three obligations of a certain man 'to go to his neighbours assistance,' 'to tell them that he comes if he goes,' and 'not to tell them that he comes if he does not go,' together with the fact 'he does not go.' The preference-based models of dyadic deontic logic again give a natural representation of the three sentences, just like for Forrester's paradox. For example, going to the assistance and telling is preferred to all the other possibilities, and not going to the assistance and not telling is preferred to not going and telling. It seems that the going and not telling and not going and telling may be ordered in various ways. However, the following example illustrates that it is difficult to combine factual with deontic detachment, and to derive unconditional obligations from conditional and unconditional ones.

Example 2.8 (Chisholm's paradox). *Assume a dyadic deontic logic without nested modal operators that has at least replacement of logical equivalents, the Conjunction pattern* AND *factual detachment* FD *and the following inference pattern deontic detachment* DD.

$$\text{DD}: \frac{\bigcirc(\alpha|\beta), \bigcirc\beta}{\bigcirc\alpha}$$

Furthermore, consider the following premise set S.

$$S = \{\bigcirc(a|\top), \bigcirc(t|a), \bigcirc(\neg t|\neg a), \neg a\}$$

The set S formalizes Chisholm's paradox when a is read as 'a certain man goes to the assistance of his neighbors' *and t as* 'the man tells his neighbors that he will come.' *Chisholm's paradox is more complicated than Forrester's paradox, because it also contains an* According-To-Duty *(ATD) obligation. We can represent the notion of according-to-duty as a binary relation among conditional obligations, just like the notion of contrary-to-duty. A conditional obligation $\bigcirc(\alpha \mid \beta)$ is an ATD obligation of $\bigcirc(\alpha_1 \mid \beta_1)$ if and only if β logically implies α_1. The condition of an ATD obligation is satisfied only if the primary obligation is fulfilled. The definition of ATD is analogous to the definition of CTD* *in the sense that an ATD obligation is an obligation conditional upon the fulfilment of an obligation and a CTD obligation is an obligation conditional upon a violation. The second obligation is an ATD obligation and the third obligation is a CTD obligation with respect to the first obligation, see Figure 7.*

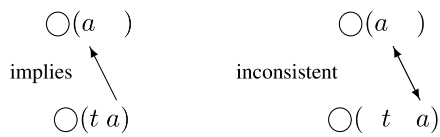

Figure 7: $\bigcirc(t|a)$ is an ATD of $\bigcirc(a|\top)$ and $\bigcirc(\neg t|\neg a)$ is a CTD of $\bigcirc(a|\top)$

The derivation in Figure 8 shows how the counterintuitive obligation $\bigcirc(t \wedge \neg t)$, or $\bigcirc\bot$, can be derived from S by FD, DD *and* AND. *Just like in Forrester's paradox, we can give a dilemma and a coherent interpretation to the scenario, and there is consensus that the latter one is preferred. This is not surprising, as Forrester's paradox shows that factual detachment and conjunction are problematic in themselves.*

$$\dfrac{\dfrac{\bigcirc(t|a) \qquad \dfrac{\bigcirc(a|\top) \quad \top}{\bigcirc(a)}\,\text{FD}}{\bigcirc t}\,\text{DD} \qquad \dfrac{\bigcirc(\neg t|\neg a) \quad \neg a}{\bigcirc(\neg t)}\,\text{FD}}{\bigcirc(t \wedge \neg t)}\,\text{AND}$$

Figure 8: Chisholm's paradox

2.5.2 Deriving Secondary Obligations from Primary Ones: Three Kinds of Transitivity

Deontic detachment is related to the following three variants of transitivity: plain transitivity T, cumulative transitivity CT, and what Parent and van der Torre [18, 19] call aggregative cumulative transitivity ACT.

$$\text{T}: \frac{\bigcirc(\alpha|\beta), \bigcirc(\beta|\gamma)}{\bigcirc(\alpha|\gamma)} \qquad \text{CT}: \frac{\bigcirc(\alpha|\beta \wedge \gamma), \bigcirc(\beta|\gamma)}{\bigcirc(\alpha|\gamma)} \qquad \text{ACT}: \frac{\bigcirc(\alpha|\beta \wedge \gamma), \bigcirc(\beta|\gamma)}{\bigcirc(\alpha \wedge \beta|\gamma)}$$

The left derivation illustrates that T can be derived from ACT together with SA and WC, and likewise CT can be derived from T and SA, and T can be derived from CT and SA. The right derivation illustrates how ANDC can be derived from SA and ACT. RANDC can be derived analogously from RSA and ACT.

$$\dfrac{\dfrac{\dfrac{\bigcirc(\alpha|\beta)}{\bigcirc(\alpha|\beta\wedge\gamma)}\,\text{SA} \qquad \bigcirc(\beta|\gamma)}{\bigcirc(\alpha\wedge\beta|\gamma)}\,\text{ACT}}{\bigcirc(\alpha|\gamma)}\,\text{WC} \qquad \dfrac{\dfrac{\bigcirc(\alpha_1|\beta)}{\bigcirc(\alpha_1|\beta\wedge\alpha_2)}\,\text{SA} \qquad \bigcirc(\alpha_2|\beta)}{\bigcirc(\alpha_1\wedge\alpha_2|\beta)}\,\text{ACT}$$

The following variant of Chisholm's paradox illustrates that only ACT can be combined with restricted strengthening of the antecedent.

Example 2.9 (Chisholm's paradox, continued)**.** *Assume a dyadic deontic logic that validates at least replacement of logical equivalents and the (intuitively valid) inference patterns* RSA *(or* SA*),* T *(or* CT*), and* ANDC*.*

The left derivation in Figure 9 illustrates how the counterintuitive $\bigcirc(\bot \mid \neg a)$ can be derived from S. Again we can give a dilemma and a coherent interpretation, and there is consensus in the literature that it should get a coherent interpretation. The underlying problem is the derivation of $\bigcirc(t \mid \neg a)$, which seems counterintuitive since it derives a contrary-to-duty obligation from the primary $\bigcirc(a \mid \top)$. If we accept RSA*, then we cannot accept* T *or* CT*.*

$$\dfrac{\dfrac{\dfrac{\bigcirc(t|a) \quad \bigcirc(a|\top)}{\bigcirc(t|\top)}\,\text{T/CT}}{\bigcirc(t|\neg a)}\,\text{RSA} \qquad \bigcirc(\neg t|\neg a)}{\bigcirc(t\wedge\neg t|\neg a)}\,\text{AND} \qquad \dfrac{\dfrac{\dfrac{\dfrac{\bigcirc(t|a) \quad \bigcirc(a|\top)}{\bigcirc(a\wedge t|\top)}\,\text{ACT}}{\bigcirc(t|\top)}\,\text{WC}}{\bigcirc(t|\neg a)}\,\text{RSA} \qquad \bigcirc(\neg t|\neg a)}{\bigcirc(t\wedge\neg t|\neg a)}\,\text{AND}$$

Figure 9: Chisholm's paradox

Assume a dyadic deontic logic that validates at least replacement of logical equivalents and the (intuitively valid) inference patterns RSA*,* ANDC*,* WC *and* ACT*. The right derivation of Figure 9 illustrates how the counterintuitive $\bigcirc(\bot|\neg a)$ can be derived from S. However, without* WC *the counterintuitive obligation cannot be derived.*

When we compare the two derivations of the contrary-to-duty paradoxes in dyadic deontic logic, we find the following similarity. The underlying problem of the counterintuitive derivations is the derivation of the obligation $\bigcirc(\alpha_1|\neg\alpha_2)$ from $\bigcirc(\alpha_1\wedge\alpha_2|\top)$ by WC and

RSA. It is respectively the derivation of $\bigcirc(\neg g|k)$ from $\bigcirc(\neg k|\top)$ in Figure 3 and $\bigcirc(t|\neg a)$ from $\bigcirc(a \wedge t|\top)$ in Figure 9. The underlying problem of the contrary-to-duty paradoxes is that a contrary-to-duty obligation can be derived from its primary obligation. It is no surprise that this derivation causes paradoxes. The derivation of a secondary obligation from a primary obligation confuses the different contexts found in contrary-to-duty reasoning. The context of primary obligation is the ideal state, whereas the context of a contrary-to-duty obligation is a violation state. Preference-based deontic logics were developed to semantically distinguish the different violation contexts in a preference ordering, but it appears more challenging to represent these contexts in derivations.

2.6 Makinson's Möbius Strip

Makinson [13]'s Möbius strip illustrates that dilemmas and deontic detachment can also be combined, leading to new challenges and distinctions. We discuss also the violation detection problem for deontic detachment.

2.6.1 Iterated deontic detachment

The so-called Möbius strip (whose name comes from the shape of the example in Figure 10) arises when we allow for deontic detachment to be iterated. We give the version of the example presented by Makinson and van der Torre in their input/output logic, though we use the dyadic representation.

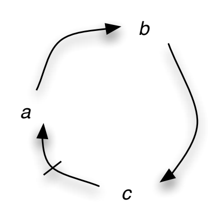

Figure 10: Möbius strip

Example 2.10 (Möbius strip)**.** *Consider three conditional obligations stating $\neg a$ is obligatory given c, that c is obligatory given b, and that b is obligatory given a, together with the* *fact that a is true.*

$$\bigcirc(\neg a|c), \bigcirc(c|b), \bigcirc(b|a), a$$

For instance, a, b, c could represent "Alice (respectively Bob,Carol) is invited to dinner." The obligation $\bigcirc(b|a)$ says that if Alice is invited then Bob should be, and so on.

Makinson [13] gives what we call here the coherent interpretation. He mentions that "intuitively, we would like to have" that under condition a, each of b and c is obligatory, even though we may not want to conclude for $\neg a$ under the same condition. He also indicates that "an approach inspired by maxi choice in AGM theory change" (like the one described in the paper in question) leads to three possible outcomes: both b and c are obligatory; only b is obligatory; neither of b and c is obligatory. The three sets of obligations corresponding to these outcomes are linearily ordered under set-theoretical inclusion.

In their input/output logic framework, Makinson and van der Torre [15] present what we call here the dilemma interpretation of the example. They change the definitions such that precisely the dilemma among these three alternatives is the desired outcome of the example.

There does not seem to be consensus in the literature on which interpretation is the intuitive answer for this example. Deontic detachment has been severely criticised in the literature, so it may be questioned whether full transitivity is natural. However, the choice between coherent and dilemma interpretation is general and can be found in other examples, such as the following variant of Chisholm's paradox.

Example 2.11 (Chisholm's paradox, continued)**.** *Consider this variant of the Möbius strip:*

$$\{\bigcirc(d|c), \bigcirc(c|b), \bigcirc(b|a), a, \neg d\}$$

By symmetry with the dilemma interpretation of Möbius strip, the dilemma interpretation gives three alternatives, $\{\bigcirc b, \bigcirc c\}$, $\{\bigcirc b\}$ and $\emptyset$. Now consider deontic detachment in Chisholm's paradox, together with the fact that we do not tell.

$$\bigcirc(t|a), \bigcirc(a|\top), \neg t$$

Again by symmetry, the dilemma interpretation gives two alternatives, $\{\bigcirc a\}$ and $\emptyset$.

The following example has been introduced by Horty [9] in a prioritised setting, and we will consider it again in the section that comes next. Again the question is raised whether one solution can be a subset of another solution.

Example 2.12 (Order). *Consider the following set of obligations. a is for putting the heating on, and b is for opening the window.*

$$\bigcirc(a|\top), \bigcirc(b|\top), \bigcirc(\neg b|a)$$

The example is a dilemma, but the question is whether there are two or three alternatives. According to the first interpretation, the only two alternatives are the obligations for a and b, and the obligations for a and $\neg b$. According to the second interpretation, there is also the alternative of an obligation for b, without an obligation for a. The latter alternative is a subset of another alternative, analogous to the dilemma interpretation of the Möbius strip example.

2.6.2 Violation detection problem and transitivity

In the previous subsections, like most authors we have assumed that in the Möbius strip the derivation of the obligation for $\neg a$ is intuitively not desirable. However, one can also view it as being intuitively desirable, for the following reason.

Example 2.13 (Möbius strip, continued). *Consider first the coherent interpretation of the Möbius strip, deriving obligations for b and c, but not for $\neg a$. With the transitivity* T *pattern, one may consider the derivation of the obligation for $\neg a$. This represents that a was actually a violation. With* ACT*, the violation can be represented by an obligation for $b \wedge c \wedge \neg a$.*

Consider now the dilemma interpretation, presenting three possible outcomes, either $\{\bigcirc b, \bigcirc c\}$, or $\{\bigcirc b\}$, or $\emptyset$. In that case, a leads to a choice, and we may thus have an instance of the forbidden conflict pattern FC *that derives that a is forbidden.*

2.7 Priority

We are given a set S of conditional obligations along with a priority relation defined on them.

Example 2.14 (Order [9], continued from Example 2.12). *Numbers represent the priority of the obligation, as in Section 2.1.4. Consider*

$$\{③(\neg b|a), ②(b|\top), ①(a|\top)\}$$

①, ②, and ③ can be thought of as expressing commands uttered by a priest, a bishop, and a cardinal, respectively. There are three interpretations. The greedy interpretation derives *obligations for a and b. It looks strange, because complying with ①$(a|\top)$ triggers the most important norm ③$(\neg b|a)$, which in turn cancels ②$(b|\top)$. To put it another way, complying with ①$(a|\top)$ and ②$(b|\top)$ results in violating ③$(\neg b|a)$.*

The last link interpretation derives $\bigcirc a$ and $\bigcirc\neg b$. This looks strange too, because ②$(b|\top)$ takes precedence over ①$(a|\top)$, and ③$(\neg b|a)$ will not be triggered (and ②$(b|\top)$ cancelled) unless ①$(a|\top)$ is fulfilled.

The weakest link interpretation derives $\bigcirc b$ only. In order not to trigger ③$(\neg b|a)$, and avoid being in a violation state with respect to it, the agent goes for ②$(b|\top)$ only.

The idea underpinning Parent [16]'s next example is similar. Parent argues that different outcomes are expected depending on whether the example is instantiated in the deontic or epistemic domain.

Example 2.15 (Cancer [16])**.** *Assume we have*

$$\{ ③(c|b), ②(b|a), ①(\neg b|a)\}$$

a is for the set of data used to set up a treatment against cancer, b is for receiving chemo as per the protocol, and c is for keeping WBCs (White Blood Cells) count to a safe level using a drug. In a diagram:

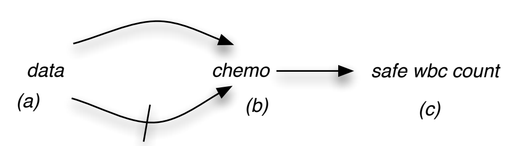

Figure 11: Cancer

Assume the input is a. In that case, we get ②$(b|a)$ and ③$(c|b)$, which derives $\bigcirc b$ and $\bigcirc c$. Given a, both ①$(\neg b\,|\,a)$ and ②$(b\,|\,a)$ are triggered. These two conflict. The stronger obligation takes precedence over the weaker one.

Assume the input is $\{a, \neg c\}$. In that case, we get ①$(\neg b\,|\,a)$ which derives $\bigcirc\neg b$. The reason why may be explained as follows. Following one of Hansson [7]'s suggestions, one might think of the input as someting settled as true. The question is: shall the agent do b or not? The ordering ② $>$ ① says that b has priority over $\neg b$. So it would seem to follow that he should do b. But, in reply, it can be said that the ordering ③ $>$ ② tells us that compliance with the stronger of the two conflicting norms triggers an obligation of even higher rank, namely the obligation to do c. Furthermore, c is already (settled as) false. Hence if the agent goes for b he will put himself in a violation state with respect to a norm with an even higher rank. The only way to avoid the violation of the most important norm is to go for $\neg b$. This is fully in line with what practitioners do: if the WBCs count cannot be maintained at a safe level, chemo is postponed.

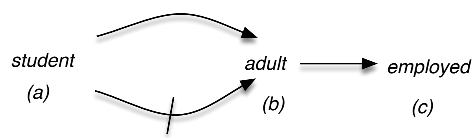

Figure 12: Student example

In the epistemic domain, a different outcome is expected. This can be seen using the reliability interpretation discussed by Horty [9, p. 391] among others. Under the latter interpretation, an epistemic conditional indicates something like a high conditional probability that its conclusion is satisfied, and the priority ordering measures relative strength of these conditional probabilities. For illustration purposes, assume that these conditional probabilities encode statistical assertions about some population groups, and instantiate a, b and c into (this is the example often used to illustrate the non-transitivity of default patterns) *being a student*, *being an adult*, and *being employed*. This is shown in Figure 12. Given input $\{a, \neg c\}$, the expected output remains b.

3 Formal Framework

We extract ten basic properties from the examples, falling in three groups. We believe that the properties of factual detachment and violation detection, the logical properties of substitution, replacement by logical equivalents, implication and paraconsistency are desirable for methods to reason with normative systems, and that the properties of aggregation, factual and norm monotony, and norm induction are optional.

In this section we use the detachment terminology instead of the inference rules terminology.

3.1 Norms, Obligations and Factual Detachment

The distinction between norms and obligations is fundamental in the modern approach to deontic logic. They are related via factual detachment, the detachment of an obligation from a norm.

3.1.1 Representing Norms and Imperatives Explicitly

There are two traditions in normative reasoning, as witnessed by the two historical chapters in the *Handbook on Deontic Logic and Normative Systems* [5]. The first tradition of deontic logic is concerned with logical relations between obligations and permissions, or between the actual and the ideal. The second tradition of normative systems is concerned with normative reasoning, including reasoning about imperatives. Many people suggested a more comprehensive approach, by bringing the two traditions closer to each other, or proposing a uniform approach. For example, when van Fraassen [30] is asking himself whether restricted conjunction can be formalized to reason about dilemmas, he suggests to represent imperatives explicitly.

> "But can this [...] be reflected in the logic of the ought-statements alone? Or can it be expressed only in a language in which we can talk directly about the imperatives as well? This is an important question, because it is the question whether the inferential structure of the 'ought' language game can be stated in so simple a manner that it can be grasped in and by itself. Intuitively, we want to say: there are simple cases, and in the simple cases the axiologist's logic is substantially correct even if it is not in general—but can we state precisely when we find ourselves in such a simple case? These are essentially technical questions for deontic logic, and I shall not pursue them here." [30]

The distinction between norms and obligations was most clearly put forward by Makinson [13], and we follow his notational conventions. To detach an obligation from a norm, there must be a context, and the norms must be conditional. Consequently, norms are a particular kind of rules.

3.1.2 Formal Representation

In this section, a set of norms is represented by a set of pairs of formulae from a base logic, (a_1, x_1), ..., (a_n, x_n). A norm (a, x) can be read as "if a is the case, then x ought to be the case." A normative system contains at least one set of norms, the regulative norms from which obligations and prohibitions can be detached. It often contains also permissive norms, from which explicit permissions can be detached, and constitutive norms, from which institutional facts can be detached.

The context is represented by a set of formulae of the same logic. A deontic operator $\bigcirc$ factually detaches obligations, represented by a set of formulae of the base logic, from a set of norms N in a context A, written as $\bigcirc(N, A)$. Unless there is a need for it, we adopt the convention that we do not prefix the detached formula with a modal operator. For example, from a norm that if you travel by metro, you must have a valid ticket $(\textit{metro}, \textit{ticket})$ in the context where you travel by metro, we derive $\textit{ticket} \in \bigcirc(\{(\textit{metro}, \textit{ticket})\}, \{metro\})$, but *ticket* itself is not prefixed with a deontic modality. Note that there is no risk of confusing facts and obligations. We know that *ticket* represents an obligation for *ticket*, because it is factually detached by the $\bigcirc$ operator.

To facilitate presentation and proofs, in this paper we assume propositional logic as the base logic. We write $\beta \in \bigcirc(N, \alpha)$ for $\beta \in \bigcirc(N, \{\alpha\})$, and $\gamma \in \bigcirc((\alpha, \beta), A)$ for $\gamma \in \bigcirc(\{(\alpha, \beta)\}, A)$.

3.1.3 Arguments

Maybe the most important technical innovation of the modern approach is the following convention of writing an argument for α supported by A, traditionally written as $A \therefore \alpha$, as a pair (A, α):

$$(A, \alpha) \in \bigcirc(N) = \alpha \in \bigcirc(N, A)$$

We can move between $\bigcirc(N)$ and $\bigcirc(N, A)$ as we move between $\vdash$ and Cn in classical logic.

It is crucial to understand that the representation of arguments by a pair (A, α) is just a technical method to develop logical machinery: we use it to give more compact representations, to provide proof systems, and to make relations with other branches of logic. However, if you want to know what the argument $(A, \alpha) \in \bigcirc(N)$ *means*, then you always have to translate it back to $\alpha \in \bigcirc(N, A)$.

We reserve the term “norms" to explicit norms, in N. Obviously, one does not derive norms from norms.

In this section we give both the long and the short version of the properties we discuss, to prevent misreading.

3.1.4 Factual Detachment

Factual detachment says that if there is a norm with precisely the context as antecedent, then the output contains the consequent. On the one hand this is relatively weak, as we require the context to be *precisely* the antecedent. A much stronger detachment principle imposes detachment when the antecedent is *implied* by the context. Between these two extremes, we can have that most obligations are detached, or in the most normal cases the obligation is detached. On the other hand the factual detachment principle is also quite strong, as in context a from the norm $(a, \bot)$ the contradiction $\bot$ is detached, and in case of a dilemma of (a, x) and $(a, \neg x)$, in context a both x and $\neg x$ are detached.

Definition 3.1 (Factual detachment)**.** *A deontic operator $\bigcirc$ satisfies the factual detachment property if and only if for all sets of norms N and all sentences α and β we have:*

$$\frac{(\alpha, \beta) \in N}{\beta \in \bigcirc(N, \alpha)}\text{FD} \qquad \frac{(\alpha, \beta) \in N}{(\alpha, \beta) \in \bigcirc(N)}\text{FD} \qquad \frac{(\alpha, \beta) \in N}{(\alpha, \beta)}\text{FD}$$

3.2 Violation Detection

The distinctive feature of norms and obligations with respect to other types of rules and modalities is that they can be violated. Obligations which cannot be violated are not real obligations, but obligations of a degenerated kind. It is not only that ought implies can, but more importantly, ought implies can-be-violated. Issues concerning violations can be found in most deontic examples. For example, dilemma examples arise because some obligation has to be violated, and contrary-to-duty examples arise because some obligation has been violated.

Modal logic offers a simple representation for violations. An obligation for α has been violated if and only we have $\neg\alpha \wedge \bigcirc\alpha$. In our notation with explicit norms, this is $\alpha \in \bigcirc(N, A)$ with $\neg\alpha \in Cn(A)$.

To make sure that violated obligations do not drown, we use the violation detection inference pattern, which we already discussed in Section 2.2.3.

Definition 3.2 (Violation Detection)**.** *A deontic operator $\bigcirc$ satisfies the violation detection property if and only for all sets of norms N, all sets of sentences A and all sentences α we have:*

$$\frac{\alpha \in \bigcirc(N, A)}{\alpha \in \bigcirc(N, A \cup \{\neg\alpha\})}\text{VD} \qquad \frac{(A, \alpha)}{(A \cup \{\neg\alpha\}, \alpha)}\text{VD}$$

Consequently, the restricted strengthening of the antecedent pattern is too weak.

3.3 Substitution

Whereas the first two properties define what is special about *deontic* logic, namely factual detachment and violation detection, the next four properties of substitution, replacements of logical equivalence, implication and paraconsistency say something about *logic*.

The first logical requirement is substitution, well known from classical propositional logic. It says that we can uniformly replace propositional letters by propositional formulae.

Definition 3.3 (Substitution). *Let a uniform substitution map each proposition letter to a propositional formula. A deontic operator $\bigcirc$ satisfies substitution if and only for all sets of norms N, all sets of formulae A, all sentences α and all uniform substitutions σ we have:*

$$\frac{\alpha \in \bigcirc(N, A)}{\alpha[\sigma] \in \bigcirc(N[\sigma], A[\sigma])}\,\mathrm{SUB}$$

For example, it allows to replace propositional letters by distinct new letters, thus renaming them. This is an example of irrelevance of syntax, a core property of logic.

3.4 Replacement of Logical Equivalents

The following definition introduces two stronger types of irrelevance of syntax.

Definition 3.4 (Irrelevance of Syntax). *Let Cn be closure under logical consequence, and Eq closure under logical equivalence: $\alpha \in Eq(S)$ if and only if there is a β in S such that $Cn(\alpha) = Cn(\beta)$. We write $Eq(a_1, \ldots, a_n)$ for $Eq(\{a_1, \ldots, a_n\})$, and $Cn(a_1, \ldots, a_n)$ for $Cn(\{a_1, \ldots, a_n\})$. Here Cn is the consequence operation of the base logic on top of which the deontic operator $\bigcirc$ operates.*

A deontic operator $\bigcirc$ satisfies formula input (output) irrelevance of syntax if and only for all sets of norms N and all sets of formulae A we have:

$$\bigcirc(N, A) = \bigcirc(N, Eq(A)) \qquad (\bigcirc(N, A) = Eq(\bigcirc(N, A)))$$

and it satisfies set input (output) irrelevance of syntax if and only if for all sets of norms N and all sets of formulae A we have:

$$\bigcirc(N, A) = \bigcirc(N, Cn(A))) \qquad (\bigcirc(N, A) = Cn(\bigcirc(N, A)))$$

The following example illustrates the various types of irrelevance of syntax.

Example 3.5 (Irrelevance of syntax)**.** *Let* $N = \{(a,x),(a,y)\}$ *and* $A = \{a\}$*. The following table lists some possibilities for* $\bigcirc(N,A)$*:*

$\emptyset$	$\{x,y\}$	$\{x,y,x\wedge y\}$
$\{x\wedge y, y\wedge x\}$	$\{x\wedge y, y\wedge x, x, y\}$	$\{x\wedge y, y\wedge x, x, y, x\vee y, y\vee x\}$
$Eq(x\wedge y)$	$Eq(x\wedge y, x, y)$	$Eq(x\wedge y, x, y, x\vee y)$
$Cn(x)\cup Cn(y)$	$Cn(x\wedge y)$	

The first row gives some deontic operators which do not satisfy basic properties. For example, $\emptyset$ *does not satisfy factual detachment,* $\{x,y\}$ *does not satisfy conjunction, and* $\{x,y,x\wedge y\}$ *does not satisfy variable renaming. That is, if we replace* x *and* y *in* N*, then we end up with the same set, but if we replace* x *and* y *in the output, we obtain* $y\wedge x$*. This violates the most basic property of irrelevance of syntax.*

The second row gives some examples satisfying variable renaming for x *and* y*. The set of obligations* $\{x\wedge y, y\wedge x\}$ *does not satisfy factual detachment again, and the set* $\{x\wedge y, y\wedge x, x, y, x\vee y, y\vee x\}$ *satisfies besides closure under conjunction also closure under disjunction. Whether this is desired depends on the application. However, all three examples do not satisfy formula output irrelevance of syntax. For example, they all three derive* $x\wedge y$*, but they do not derive the logically equivalent* $x\wedge x\wedge y$*.*

The third and fourth row close the output under logical equivalence and logical consequence, respectively. $Cn(x\wedge y)$ *in the last row satisfies set output irrelevance of syntax.*

Input irrelevance is analogous to output irrelevance. For example, when the input is $a\wedge a$ *rather than* a*, it may or may not derive again the same output. If it does not, then the operator violates formula input irrelevance of syntax. Moreover, if it does not treat* $\{a,b\}$ *and* $\{a\wedge b\}$ *the same, then it violates input set irrelevance of syntax.*

The following example illustrates that output set irrelevance of syntax is too strong in the context of dilemmas, because it may lead to deontic explosion.

Example 3.6 (Irrelevance of syntax, continued)**.** *Let*

$$N = \{(a, x\wedge y),(a,\neg x\wedge y)\}$$

and $A = \{a\}$*. The following table lists some possibilities for* $\bigcirc(N,A)$*. We only list options closed under logical equivalence, i.e. which satisfy output formula irrelevance of syntax.*

$Eq(x\wedge y, \neg x\wedge y)$	$Eq(x\wedge y, \neg x\wedge y, x\wedge\neg x\wedge y)$
$Eq(x\wedge y, \neg x\wedge y, y)$	$Eq(x\wedge y, \neg x\wedge y, y, x\wedge\neg x\wedge y)$
$Cn(x\wedge y)\cup Cn(\neg x\wedge y)$	$Cn(x\wedge y)\cup Cn(\neg x\wedge y)\cup Eq(x\wedge\neg x\wedge y)$
$Cn(x\wedge y, \neg x\wedge y)$	

The last set $Cn(x \wedge y, \neg x \wedge y)$ derives the whole language, and thus gives rise to explosion. Hence we cannot accept it. The example illustrates that we cannot accept set output irrelevance of syntax.

The difference between the left and right column is that the right column is closed under conjunction, and represents with inconsistent formulae that there is a dilemma.

The difference between the first and the second row is that the second row is closed under disjunction. The difference between the second and the third row is that consistent formulae are closed under logical consequence.

$Cn(x \wedge y) \cup Cn(\neg x \wedge y) \cup Eq(x \wedge \neg x \wedge y))$ has the feature that violations and other obligations are treated in a distinct way.

In this paper we require set input irrelevance of syntax, and formula output irrelevance of syntax. In addition, along the same lines we require that we can replace formulae within the norms by logically equivalent ones. All together, it corresponds to the following property of replacement of logical equivalents.

Definition 3.7 (Replacement of logically equivalent expressions)**.** *We say that two norms ar similar, written as $(\alpha_1, \beta_1) \approx (\alpha_2, \beta_2)$, if and only if $Cn(\alpha_1) = Cn(\alpha_2)$, and $N \approx M$ if and only if for all $(\alpha_1, \beta_1) \in N$ there is a $(\alpha_2, \beta_2) \in M$ such that $(\alpha_1, \beta_1) \approx (\alpha_2, \beta_2)$, and vice versa. A deontic operator $\bigcirc$ satisfies the replacement of Logical Equivalents property if and only if for all sets of norms N and M, all sets of formulae A and B, and all sentences α and β we have:*

$$\frac{N \approx M, Cn(A) = Cn(B), Cn(\alpha) = Cn(\beta), \alpha \in \bigcirc(N, A)}{\beta \in \bigcirc(M, B)}\text{RLE}$$

The examples illustrate that there are other options in between formula and set output irrelevance of syntax, such as requiring that the output is closed under conjunction, or under disjunction, or both. We consider them in Section 3.7.

The principle of irrelevance of syntax has been criticized in belief revision theory. It is discussed by [23] in the context of a study of the notion of revision of a normative system. This notion falls outside the scope of the present paper, and must be left as a topic for future research.

3.5 Implication

The four properties FD, VD, SUB and RLE defined thus far may be called positive properties, in the sense that they require something to be obligatory. That is why we could represent them as Horn rules: given a set of conditions, we require some obligation to be derivable. This contrasts with the examples in Section 2, where typically too much is derived.

The implication requirement in this section and the paraconsistency requirement in the following section may be called negative properties, in the sense that they forbid something to be obligatory. The first requirement makes use of the so-called materialisation of a normative system, which means that each norm (a, x) is interpreted as a material conditional $a \rightarrow x$, i.e. as the propositional sentence $\neg a \vee x$. The implication requirement says that if the materializations of N, written as $m(N)$, do not imply $a \rightarrow x$, then $(a, x) \notin \bigcirc(N)$. This represents the idea that we cannot derive more than we can derive in propositional logic. In general, implication in the base logic is the upper bound.

Definition 3.8 (Implication). *Let $m(N) = \{a \rightarrow x \mid (a, x) \in N\}$ be the set of materializations of N. A deontic operator $\bigcirc$ satisfies the implication property if and only if for all sets of norms N and all sets of sentences A we have $\bigcirc(N, A) \subseteq Cn(m(N) \cup A)$.*

The elements $(\{\alpha\}, \beta)$ of $\bigcirc(N)$ are a subset of $\{(\alpha, \beta) \mid \alpha \rightarrow \beta \in Cn(m(N))\}$. In most systems, the base logic is classical propositional logic, but it need not be so. For instance, Cn may be the consequence relation of intuitionistic propositional logic, as in [17]. Cn may also be what Makinson calls a pivotal consequence relation Cn_K, defined by $Cn_K(A) = C(A \cup K)$, where K is a set of formulas, and C is the consequence relation of classical propositional logic. [22] defines and studies two such input/output operations. They are aimed to model the interplay between norms and so-called material dependencies. We have $\bigcirc(N, A) \subseteq Cn_K(m(N) \cup A)$.

3.6 Paraconsistency

To prevent explosion we do not want to derive the whole language, unless maybe in pathological cases in which the normative system contains a norm for each propositional formula. A consequence relation may be said to be paraconsistent if it is not explosive, though there are various ways to make this formal.

To define our paraconsistency requirement, we distinguish obligations representing violations from other obligations. That is, we decompose an operator $\bigcirc(N, A)$ into two operators $V(N, A)$ and $\overline{V}(N, A)$, such that we have $V(N, A) = \{x \in \bigcirc(N, A) \mid \neg x \in Cn(A)\}$ and $\overline{V}(N, A) = \bigcirc(N, A) \setminus V(N, A)$. Trivially, we have

$$\bigcirc(N, A) = V(N, A) \cup \overline{V}(N, A)$$

The basic idea of our paraconsistency requirement is that obligations in $\overline{V}$ can be derived from a set of norms M in N, such that this set of norms M does not explode.

Definition 3.9 (Paraconsistency)**.** *A deontic operator $\bigcirc$ satisfies the paraconsistency property if and only if for all sets of norms N, all sets of formulae A and all sentences α, if $\alpha \in \overline{V}(N, A)$, then there is a $M \subseteq N$ such that $\alpha \in \bigcirc(M, A)$ and $\bigcirc(M, A) \cup A$ is classically consistent.*

Implication and paraconsistency together imply that if $\alpha \in \overline{V}(N, A)$, then there is a $M \subseteq N$ such that $\alpha \in Cn(m(N) \cup A)$ and $\bigcirc(M, A) \cup A$ is classically consistent. This suggest an additional condition: if $\alpha \in \overline{V}(N, A)$, then there is a $M \subseteq N$ such that $\alpha \in Cn(m(N) \cup A)$ and $m(N) \cup A$ is classically consistent.

The underlying intuition to restrict to a set of norms was already raised in Example 1.1 in the introduction. There we observe that if we can derive $\bigcirc(\beta \wedge \gamma)$ from $\bigcirc(\alpha \wedge \beta)$ and $\bigcirc(\neg\alpha \wedge \gamma)$, and we have substitution and replacements of logical equivalents, then we also derive $\bigcirc(\beta)$ from $\bigcirc(\alpha)$ and $\bigcirc(\neg\alpha)$, in other words, we have deontic explosion. This can be verified by replacing β by $\alpha \vee \beta$ and γ by $\neg\alpha \vee \beta$. Therefore, we restrict the set of norms we use to a set of norms which is in some sense "consistent" with the input A.

3.7 Aggregation

The last four properties of aggregation, factual and norm monotony, and norm induction determine the kind of deontic logics we are going to study in our framework. We believe that other choices at this point may be of interest too, but we do not pursue them in this paper.

Aggregation is a core issue in van Fraassen's paradox.

Definition 3.10 (Aggregation)**.** *A deontic operator $\bigcirc$ satisfies the aggregation property if and only if for all sets of norms N, sets of sentences A and sentences α and β we have*

$$\frac{\alpha, \beta \in \bigcirc(N, A)}{\alpha \wedge \beta \in \bigcirc(N, A)}\text{AND} \qquad\qquad \frac{(A, \alpha), (A, \beta)}{(A, \alpha \wedge \beta)}\text{AND}$$

Van Fraassen's paradox shows that therefore we cannot accept weakening of the consequent. In the context of our present framework, we prefer to call it weakening of the output.

Definition 3.11. *A deontic operator $\bigcirc$ satisfies the weakening of the output property if and only if for all sets of norms N, sets of sentences A and sentences α and β we have*

$$\frac{\alpha \wedge \beta \in \bigcirc(N, A)}{\alpha, \beta \in \bigcirc(N, A)}\text{WO} \qquad \frac{(A, \alpha \wedge \beta)}{(A, \alpha), (A, \beta)}\text{WO}$$

Proposition 3.12. *There is no operator $\bigcirc$ satisfying simultaneously paraconsistency, aggregation, and weakening of the output.*

Proof. Assume the statement does not hold, so there is a deontic $\bigcirc$ satisfying paraconsistency, aggregation and weakening of the output. Consider van Fraassen's paradox $N = \{(\top, p), (\top, \neg p)\}$. According to aggregation and weakening of the output, we have $(\top, q) \in \bigcirc(N)$. According to paraconsistency, $(\top, q) \notin \bigcirc(N)$. Contradiction. □

3.8 Factual Monotony

In this paper we are interested in monotonic logics. Though non-monotonic logics may have their applications too, we believe they should be build on top of the monotonic ones.

Definition 3.13 (Factual monotony)**.** *The factual monotony property holds for $\bigcirc$ if and only if for all sets of norms N, and all sets of sentences A and B, we have $\bigcirc(N, A) \subseteq \bigcirc(N, A \cup B)$.*

As this implies strengthening of the antecedent, Forrester's paradox illustrates that we cannot accept weakening of the consequent.

Proposition 3.14. *There is no operator $\bigcirc$ satisfying simultaneously paraconsistency, factual monotony, and weakening of the output.*

Proof. Assume the statement does not hold, so there is a deontic $\bigcirc$ satisfying paraconsistency, factual monotony and weakening of the output. Consider the first norm of Forrester's paradox $N = \{(\top, \neg k)\}$. According to factual monotony and weakening of the output, we have $(k, \neg k \vee g) \in \bigcirc(N)$. According to paraconsistency, $(k, \neg k \vee g) \notin \bigcirc(N)$. Contradiction. □

3.9 Norm Monotony

Definition 3.15 (Norm monotony)**.** *A deontic operator $\bigcirc$ satisfies the property of norm monotony if and only if for all sets of norms N and M we have $\bigcirc(N) \subseteq \bigcirc(N \cup M)$.*

A deontic operator $\bigcirc$ satisfies the property of monotony if and only if it satisfies those of factual and norm monotony, i.e. for all N, M, A, B we have $\bigcirc(N, A) \subseteq \bigcirc(N \cup M, A \cup B)$.

3.10 Norm Induction

Norm induction says that if there is an output β for an input α, and we add the norm (α, β) to the normative system, then for all inputs, the output of the normative system stays the same. We call it norm induction, because the norm is induced from the relation between facts and obligations. The norm induction requirement considers a set M of such pairs (α, β).

Definition 3.16 (Norm induction)**.** *A deontic operator $\bigcirc$ verifies the property of norm induction if and only if for all sets of norms N and M and all sets of sentences A we have $M \subseteq \bigcirc(N) \Rightarrow \bigcirc(N) = \bigcirc(N \cup M)$*

The strong norm induction principle strengthens the norm induction principle to expansion of the normative system with new norms.

Definition 3.17 (Strong norm induction)**.** *A deontic operator $\bigcirc$ satisfies the property of strong norm induction if and only if for all sets of norms N, N', M, and all sets of sentences A we have $M \subseteq \bigcirc(N) \Rightarrow \bigcirc(N \cup N') = \bigcirc(N \cup N' \cup M)$*

Clearly we have that the strong norm induction property implies the norm induction property.

Together, factual detachment, monotony and norm induction are equivalent to requiring that $\bigcirc$ is a closure operator.

Definition 3.18 (Closure operator)**.** *$\bigcirc$ is a closure operator if and only if it satisfies the following three properties:*

INCLUSION $N \subseteq \bigcirc(N)$

MONOTONY $N \subseteq M$ *implies* $\bigcirc(N) \subseteq \bigcirc(M)$

IDEMPOTENCE $\bigcirc(N) = \bigcirc(\bigcirc(N))$

Their counterparts in terms of Cn are knowns as the "Tarskian" conditions, after A. Tarski. They can each be rephrased in terms of $\vdash$ ('proves') as follows.

REFLEXIVITY $A \vdash x$ for all $x \in A$

MONOTONY $A \vdash x$ implies $A \cup B \vdash x$

TRANSITIVITY $A \vdash x$ for all $x \in B$ and $B \vdash y$ imply $A \vdash y$

Inclusion for Cn translates into reflexivity of $\vdash$. Monotony for Cn translates into monotony of $\vdash$. Idempotence of Cn corresponds to the transitivity of $\vdash$.

4 Summary

Table 2 lists the examples we discussed in this paper. Given that the world is full of conflicts, we have that normative systems are developed by humans and full of inconsistencies. We need to represent dilemmas consistently, if only to consider their resolution. Van Fraassen's paradox illustrates that doing so presents a basic dilemma: do we accept aggregation or closure under consequence? Forrester's paradox seems to indicate a dilemma too, as it presents two alternatives. In the cottage regulations, such a dilemma interpretation makes sense: either remove the fence, or paint it white. However, in Forrester's gentle murderer example, you cannot undo killing someone. So only the coherent interpretation makes sense. Dilemmas can be resolved by explicit priorities, for example reflecting the authority creating the obligation, or it can be derived from the specificity of the obligations. In the latter case, as illustrated by the cottage regulations, we have to be careful to distinguish violations from exceptions. Jeffrey's disarmament illustrates the problem of reasoning by cases in deontic reasoning. When conditions have an epistemic reading, reasoning by cases may not be valid. Deontic detachment and transitivity originate from Chisholm's paradox, though it is known in the literature as a contrary-to-duty paradox rather than a deontic detachment paradox. Chisholm's paradox illustrates that an alternative representation of the transitivity pattern makes it analogous to Forrester's paradox. Makinson's Möbius strip illustrates many of the problems of reasoning with transitivity. In particular, the dilemma interpretation highlights that we can have solutions being a strict subset of other solutions. More priority examples are introduced in the area of epistemic reasoning, and reasoning with defaults.

Ex.	obligations	patterns	
2.1	Fraassen	$\bigcirc p, \bigcirc \neg p$	AND, WC
2.2	Forrester	$\bigcirc(\neg k\|\top), \bigcirc(g\|k), \vdash g \to k$	FD, (R)AND
2.3	Forrester	$\bigcirc(\neg k\|\top), \bigcirc(g\|k), \vdash g \to k$	(R)SA, ANDC, WC
2.6	Cottage	$\bigcirc(\neg f\|\top), \bigcirc(w \wedge f\|f), \bigcirc(f\|d)$	RSA$_o$
2.7	Jeffrey	$\bigcirc(d\|w), \bigcirc(d\|\neg w), \bigcirc(\neg d\|d \leftrightarrow w)$	RSA, ORA
2.8	Chisholm	$\bigcirc(a\|\top), \bigcirc(t\|a), \bigcirc(\neg t\|\neg a), \neg a$	AND, FD, DD
2.9	Chisholm	$\bigcirc(a\|\top), \bigcirc(t\|a), \bigcirc(\neg t\|\neg a), \neg a$	T/ CT / ACT, ANDC
2.10	Möbius	$\bigcirc(\neg a\|c), \bigcirc(c\|b), \bigcirc(b\|a), a$	T/ CT
2.14	Priority	③$(\neg b\|a)$, ②$(b\|\top)$, ①$(a\|\top)$	T/ CT

Table 2: Summary of the examples

Maybe the most important technical innovation of our formal framework is the convention of writing an argument for α supported by A as a pair (A, α) with $(A, \alpha) \in \bigcirc(N)$, which means the same as $\alpha \in \bigcirc(N, A)$. We can move between $\bigcirc(N)$ and $\bigcirc(N, A)$ as we move between $\vdash$ and Cn in classical logic.

The ten properties of our formal framework listed in Table 3. We believe that all deontic logics have to satisfy the deontic properties of factual detachment and violation detection, and the logical properties of substitution, replacement by logical equivalents, implication and paraconsistency. Moreover, we discussed the optional properties of aggregation, factual and norm monotony, and norm induction.

FD	$(\alpha, \beta) \in N \Rightarrow \beta \in \bigcirc(N, \alpha)$	Factual detachment
VD	$(A, \beta) \Rightarrow (A \cup \{\neg\beta\}, \beta)$	Violation detection
SUB	$\alpha \in \bigcirc(N, A) \Rightarrow \alpha[\sigma] \in \bigcirc(N[\sigma], A[\sigma])$	Substitution
RLE	$N \approx M, Cn(A) = Cn(B), Cn(\alpha) = Cn(\beta),$ $(A, \alpha) \in \bigcirc(N) \Rightarrow (B, \beta) \in \bigcirc(M)$	Replacement of equivalents
IMP	$\bigcirc(N, A) \subseteq Cn(m(N) \cup A)$	Implication
PC	$\alpha \in \overline{V}(N, A) \Rightarrow \exists M \subseteq N : \alpha \in \bigcirc(M, A)$ and $\bigcirc(M, A) \cup A$ consistent	Paraconsistency
AND	$(A, \alpha)(A, \beta) \Rightarrow (A, \alpha \wedge \beta)$	Conjunction
FM	$(A, \alpha) \Rightarrow (A \cup B, \alpha)$	Factual monotony
NM	$\bigcirc(N) \subseteq \bigcirc(N \cup M)$	Norm monotony
NI	$M \subseteq O(N) \Rightarrow O(N) = O(N \cup M)$	Norm induction

Table 3: Properties

There are two ways to look at the operator $\bigcirc$. First, given a set of norms, it derives sentences from sentences: $\alpha \in \bigcirc_N(A)$. This is the classical way deontic logics considered normative systems: facts go in, obligations go out. Secondly, it derives arguments from norms: $(A, \alpha) \in \bigcirc(N)$. These two views can be used to summarise our properties as follows.

First, the operator in $(A, \alpha) \in \bigcirc(N)$ must be a closure operator, which means that it satisfies factual detachment, norm monotony and norm induction. In addition, it must satisfy substitution and replacement of logical equivalents. Secondly, the operator in $\alpha \in \bigcirc_N(A)$ must satisfy violation detection, implication, paraconsistency, factual monotony, and aggregation.

The properties of norm monotony and norm induction have the effect that our logics will behave classically as Tarskian consequence operators. However, it is important to realise that the closure properties on $\bigcirc(N)$ are not as innocent as they are in other branches of philosophical logic. In particular norm induction is very strong, because it says that every argument (A, α) can itself be used as a norm. This may be true of some branches of case law, but it is probably too strong to be accepted as a universal law for norms. We therefore expect that future studies will first relax this requirement, before relaxing the others.

Finally, we may consider our ten properties as requirements for the further development of reasoning methods for normative systems and deontic logic. We have recently presented two logics satisfying all ten properties [19], which shows that the ten properties are consistent in the sense that they can be satisfied simultaneously.

References

[1] A. R. Anderson. The formal analysis of normative systems. In N. Rescher, editor, *The Logic of Decision and Action*, pages 147–213. Univ. Pittsburgh, 1967, 1956.

[2] L. Åqvist. Deontic logic. In D. Gabbay and F. Guenthner, editors, *Handbook of philosophical logic*, volume 8, pages 147–264. Kluwer Academic publisher, 2002.

[3] B.F. Chellas. *Modal Logic: An Introduction*. Cambridge University Press, 1980.

[4] R.M. Chisholm. Contrary-to-duty imperatives and deontic logic. *Analysis*, 24:33–36, 1963.

[5] D. Gabbay, J. Horty, R. van der Meyden, X. Parent, and L. van der Torre, editors. *Handbook of Deontic Logic and Normative Systems*, volume 1. College Publications, London, UK, 2013.

[6] G. Governatori and M. Hashmi. Permissions in deontic event-calculus. In *Legal Knowledge and Information Systems - JURIX 2015: The Twenty-Eighth Annual Conference, Braga, Portual, December 10-11, 2015*, volume 279, pages 181–182. IOS Press, 2015.

[7] B. Hansson. An analysis of some deontic logics. *Noûs*, 3:373–398, 1969. Reprinted in [8, pp 121-147].

[8] R. Hilpinen, editor. *Deontic Logic: Introductory and Systematic Readings*. Reidel, Dordrecht, 1971.

[9] J. Horty. Defaults with priorities. *Journal of Philosophical Logic*, 36:367–413, 2007.

[10] J. Horty. Deontic modals: why abandon the classical semantics? *Pacific Philosophical Quarterly*, 95:424–460, 2014.

[11] R. Jeffrey. *The Logic of Decision*. University of Chicago Press, 2nd edition, 1983.

[12] N. Kolodny and J. MacFarlane. Iffs and oughts. *Journal of Philosophy*, 107(3):115–143, 2010.

[13] D. Makinson. On a fundamental problem in deontic logic. In P. Mc Namara and H. Prakken, editors, *Norms, Logics and Information Systems*, Frontiers in Artificial Intelligence and Applications, pages 29–54. IOS Press, Amsterdam, 1999.

[14] D. Makinson and L. van der Torre. Input/output logics. *Journal of Philosophical Logic*, 29(4):383–408, 2000.

[15] D. Makinson and L. van der Torre. Constraints for input-output logics. *Journal of Philosophical Logic*, 30(2):155–185, 2001.

[16] X. Parent. Moral particularism in the light of deontic logic. *Artificial Intelligence and Law*, 19(2-3):75–98, 2011.

[17] X. Parent, D. Gabbay, and L. van der Torre. Intuitionistic basis for input/output logic. In S. O. Hansson, editor, *David Makinson on Classical Methods for Non-Classical Problems*, pages 263–286. Springer Netherlands, Dordrecht, 2014.

[18] X. Parent and L. van der Torre. Aggregative deontic detachment for normative reasoning (short paper). In T. Eiter, C. Baral, and G. De Giacomo, editors, *Principles of Knowledge Representation and Reasoning. Proceedings of the 14th International Conference - KR 14*. AAAI Press, 2014.

[19] X. Parent and L. van der Torre. “Sing and dance!": Input/output logics without weakening. In F. Cariani, D. Grossi, J. Meheus, and X. Parent, editors, *Deontic Logic and Normative Systems - 12th International Conference, DEON 2014, Ghent, Belgium, July 12-15, 2014. Proceedings*, volume 8554 of *Lecture Notes in Computer Science*, pages 149–165. Springer, 2014.

[20] H. Prakken and M.J. Sergot. Contrary-to-duty obligations. *Studia Logica*, 57:91–115, 1996.

[21] D. Ross. *The Right and the Good.* Oxford University Press, 1930.

[22] A. Stolpe. Normative consequence: The problem of keeping it whilst giving it up. In G. Governatori and G. Sartor, editors, *Deontic Logic in Computer Science, 10th International Conference, DEON 2010. Proceedings*, volume 6181 of *Lecture Notes in Computer Science*, pages 174–188. Springer, 2008.

[23] A. Stolpe. Norm-system revision: Theory and application. *Artif. Intell. Law*, 18(3):247–283, 2010.

[24] X. Sun and L. van der Torre. Combining constitutive and regulative norms in input/output logic. In F. Cariani, D. Grossi, J. Meheus, and X. Parent, editors, *Deontic Logic and Normative Systems - 12th International Conference, DEON 2014. Proceedings*, volume 8554 of *Lecture Notes in Computer Science*, pages 241–257. Springer, 2014.

[25] R. Thomason and R. Horty. Nondeterministic action and dominance: foundations for planning and qualitative decision. In *Proceedings of the Sixth Conference on Theoretical Aspects of Rationality and Knowledge (TARK’96)*, pages 229–250. Morgan Kaufmann, 1996.

[26] L. van der Torre. Violated obligations in a defeasible deontic logic. In *Proceedings of the Eleventh European Conference on Artificial Intelligence (ECAI’94)*, pages 371–375. John Wiley & Sons, 1994.

[27] L. van der Torre. Contextual deontic logic: Normative agents, violations and independence. *Ann. Math. Artif. Intell.*, 37(1-2):33–63, 2003.

[28] L. van der Torre and Y.-H. Tan. The many faces of defeasibility in defeasible deontic logic. In

D. Nute, editor, *Defeasible Deontic Logic*, pages 79–121. Kluwer, 1997.

[29] L. van der Torre and Y.-H. Tan. Two-phase deontic logic. *Logique et analyse*, 43(171-172):411–456, 2000.

[30] B.C. van Fraassen. Values and the heart command. *Journal of Philosophy*, 70:5–19, 1973.

Received 23 June 2016