# Violated Obligations in a Defeasible Deontic Logic[1]

**Leendert W.N. van der Torre**[2]

**Abstract.** Deontic logic is characterized by the distinction between the actual and the ideal. In this article we discuss the situation where the actual deviates from the ideal, where obligations are violated.

Nonmonotonic logics can be very helpful for the formalization of deontic reasoning, in particular to infer moral cues. It has been argued that the problems related to violated obligations, e.g. the Chisholm 'Paradox', are just instances of problems of defeasible reasoning. We disagree with this claim since we will argue that there is a fundamental difference between a violated and a defeated obligation.

In this article, we analyze violated obligations in Horty's nonmonotonic framework. We extend his definition of deontic consequence in such a way that it covers violated obligations and we give a solution to deal with conflicts between violability and defeasibility.

## 1 Introduction

Deontic logic is characterized by the distinction between what is the case and what should be the case, between the actual and the ideal [5]. In the eighties new interest in deontic logic has arisen among computer scientists, who use deontic logic as a knowledge representation language [11]. Applications of deontic logic can be found in normative systems [5], e.g. in the area of law.

Deontic reasoning is usually formalized by a modal system. A (propositional) base language is extended with a modal operator $O$, which can be read as 'it should be the case that'. For example, when $\alpha_1$ stands for the fact that you eat with your fingers, $O(\neg\alpha_1)$ stands for the obligation not to eat with your fingers. The most familiar modal system is so-called 'standard' deontic logic (SDL), a normal modal system of type *KD* according to the Chellas classification [1]. It satisfies two axioms, $K = O(p \rightarrow q) \rightarrow (O(p) \rightarrow O(q))$ which states that modus ponens holds within the scope of the modal operator, and $D = \neg(O(p) \wedge O(\neg p))$ which states that something cannot be obliged to be the case and obliged to be not the case at the same time.

In a modal approach, it is straightforward to formalize that something is the case although it should not be the case, i.e. that an obligation is violated. For example, $\alpha_1 \wedge O(\neg\alpha_1)$ can represent that you eat with your fingers although you should not eat with your fingers.

SDL is plagued by a large number of 'Paradoxes', sets of sentences that derive sentences with a counterintuitive reading. The most notorious is the Chisholm 'Paradox' [2], caused by the existence of so-called contrary-to-duty (CTD) obligations, obligations conditional to a violation. CTD obligations as a consequence of the fact that you eat with your fingers are that you should wash your hands first, that you should not lick your fingers etc. Despite these CTD obligations, you still have the (primary) obligation not to eat with your fingers at all. In SDL, a conditional obligation is formalized by $\alpha_1 \rightarrow O(\alpha_2)$, where $\alpha_1$ is the condition and $\alpha_2$ the (deontic) conclusion. The conditional obligation $\alpha_1 \rightarrow O(\alpha_2)$ is a CTD obligation of $O(\neg\alpha_1)$, where $\alpha_1$ stands for eating with your fingers and $\alpha_2$ for washing your hands, because $\alpha_1$ and $\neg\alpha_1$ are contradictory. The traditional approach to these problematic CTD obligations is to weaken the notion of implication in such a way that the counterintuitive sentences are no longer derived.

Horty [3] and McCarty [7], among others, have argued that the techniques of nonmonotonic logic may provide a better theoretical framework for deontic reasoning than the usual modal treatment. These nonmonotonic techniques are used to deal with some problematic aspects of CTD obligations in the 'Paradoxes'. Horty focussed on deontic reasoning in the presence of conflicting obligations and reasoning with conditional obligations. McCarty also focussed on conditional obligations, in particular the Reykjavic 'Paradox', a version of the Chisholm 'Paradox' introduced by Loewer and Belzer [6].

A defeasible deontic logic has to formalize – besides violations and CTD obligations – defeasible obligations (traditionally called *prima facie* obligations), obligations which are subject to exceptions. A general approach to defeasibility is to defeat a general obligation by a more specific one. An exception to the general rule not to eat with your fingers is the situation where you are served asparagus; in that specific case, you should eat with your fingers.

Horty has given a formal framework for defeasible conditional obligations, using a specificity principle to deal with exceptional circumstances. The conditional obligation $\top \rightarrow O(\neg\alpha_1)$ is defeated (overridden) by the conditional obligation $\alpha_2 \rightarrow O(\alpha_1)$, where $\top$ stands for any tautology, $\alpha_2$ for being served asparagus and $\alpha_1$ for eating with your fingers, because the condition of the second one is more specific ($\alpha_2$ logically implies $\top$) and the conclusions are contradictory. His approach is based on a translation of conditional obligations to default rules. His main motivation was to be able to reason with moral dilemma's, deontic inconsistencies like $O(\alpha)$ and $O(\neg\alpha)$. In SDL, $O(\alpha) \wedge O(\neg\alpha)$ is inconsistent because of the D axiom, but in Horty's framework, $\top \rightarrow O(\alpha)$ and $\top \rightarrow O(\neg\alpha)$ do not derive $\top \rightarrow O(\beta)$ for all $\beta$.

One of the disadvantages of Horty's framework is that there is no formal representation of violations. This notion, however, is implicit. The first thing we will do in this article is to extend Horty's notion of deontic consequence in such a way that violations are covered explicitly.

Given our extended notion of deontic consequence, it appears that there is some interference between violability and defeasibility. The same obligation can sometimes be interpreted as a CTD obligation

[1] This research was partially supported by the ESPRIT III Basic Research Working Group No.8319 MODELAGE.

[2] Erasmus University Research Institute for Decision and Information Systems (EURIDIS) and Department of Computer Science, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands.

© 1994 L.W.N. van der Torre
ECAI 94. *11th European Conference on Artificial Intelligence* Edited by A. Cohn
Published in 1994 by John Wiley & Sons, Ltd.

(violability) as well as an exception of another obligation (defeasibility). Some deontic consequences, namely the derived violated obligations, can be different in both interpretations. In Horty's framework these conflicts are not detected since the violated obligations are not represented.

Using Horty's definition of overridden in our extended framework, the notion of defeasibility is always stronger than the notion of violability. By two classic examples, the gentle murderer 'Paradox' and the Reykjavic 'Paradox', we will argue that the notion of violability should sometimes be stronger than the notion of defeasibility. We will weaken Horty's definition of overridden in two steps in such a way that it covers these two examples satisfactorily.

## 2 Horty's framework

In this section we will discuss Horty's nonmonotonic framework. First we will look at his notion of deontic rules, then we will discuss defeasibility and finally we will look at his notion of deontic consequence.

### 2.1 Deontic rules

Deontic rules are conditional obligations, represented by $O(\alpha_1 / \alpha_2)$, which states that if $\alpha_2$ (the *condition*) is the case then $\alpha_1$ (the *conclusion*) should be the case and therefore corresponds to the conditional obligation $\alpha_2 \rightarrow O(\alpha_1)$ in SDL. These conditional obligations are considered as a special kind of inference rules, just like Reiter's default rules [8].

Deontic rules derive obligations from a factual sentence (the conjunction of a set of factual sentences), which represents the factual situation. The factual sentence consists of background knowledge and contingent facts. The background knowledge consists of necessary conditions, for example all penguins are birds.[3] A set of conditional obligations with a factual sentence is called a deontic context. In this article, we assume that the factual sentence and the condition and conclusion of the conditional obligations are sentences of a propositional language.

**Definition 1** *A* conditional obligation $O(\alpha_1 / \alpha_2)$ *is a special inference rule, with condition* $\alpha_2$ *and conclusion* $\alpha_1$*, where* $\alpha_1$ *and* $\alpha_2$ *are sentences of a propositional language. A* deontic context $T = \langle D, F \rangle$ *consists of a set of (deontic) conditional obligations D and a (factual) propositional sentence F. The factual sentence is the conjunction of background knowledge* $F_b$ *and contingent facts* $F_c$*.*

The main difference between these deontic inference rules and modal conditional obligations is that modal operators have a truth value. In a modal language, the factual and deontic sentences are both sentences of the same language and can be combined to form mixed formulae.

### 2.2 Defeasibility

Horty translates conditional obligations to default rules.[4] A default rule is overridden by a second default rule when the conditions of both defaults can be derived, the condition of the second default is more specific and the conclusions of both defaults are contradictory.[5]

**Definition 2** *[3]* $O(\alpha_1 / \alpha_2) \in D$ *is* overridden *in* $\langle D, F \rangle$ *iff there is a* $O(\alpha_3 / \alpha_4) \in D$ *s.t.:*

1. $F \vdash \alpha_4$, $F \vdash \alpha_2$, $\alpha_4 \vdash \alpha_2$, $\alpha_2 \nvdash \alpha_4$, *where* $\vdash$ *stands for classical derivability, and*
2. $F \wedge \alpha_1 \wedge \alpha_3$ *is inconsistent, and*
3. $F \wedge \alpha_3$ *is consistent.*

This notion of overridden determines which conditional obligations are active, i.e. which are relevant in the specific factual situation of the deontic context.

**Definition 3** $O(\alpha_1/\alpha_2) \in D$ *is* active *in* $\langle D, F \rangle$ *iff* $F \vdash \alpha_2$ *and* $O(\alpha_1/\alpha_2)$ *is not overridden in* $\langle D, F \rangle$*.*

**Example 1** *Consider the following sentences of* $\langle D, F \rangle$*:*

1. $O(\neg \alpha_1 / \top)$*: You should not eat with your fingers.*
2. $O(\alpha_1 / \alpha_2)$*: If you are served asparagus, you should eat with your fingers.*
3. $\alpha_2$*: You are served asparagus.*

$O(\neg \alpha_1 / \top)$ *is overridden by* $O(\alpha_1 / \alpha_2)$ *in* $\langle D, F \rangle$*, so* $O(\neg \alpha_1 / \top)$ *is not active;* $O(\alpha_1 / \alpha_2)$ *is the only active rule.*

Analogous to the notion of extension in Reiter's default logic, Horty introduces a notion of extension, the so-called conditioned extension, which can be considered as a set of conclusions from a deontic context. The definition of conditioned extension consists of two parts. $\mathcal{E}'$ contains the conclusions of some active rules. The corresponding conditioned extension $\mathcal{E}$ is the deductive closure of $\mathcal{E}'$ and the factual sentence $F$.

**Definition 4** *[3] A set of sentences* $\mathcal{E}$ *is a* conditioned extension *of* $\langle D, F \rangle$ *iff there is another (maximal) set* $\mathcal{E}'$ *s.t.:*

1. $\mathcal{E}' = \{\alpha_1 \mid O(\alpha_1 / \alpha_2) \in D$ *is active in* $\langle D, F \rangle$ *and* $\neg \alpha_1 \notin \mathcal{E}\}$
2. $\mathcal{E} = Cn[\{F\} \cup \mathcal{E}']$*, where* $Cn[S]$ *is the consequence set of S.*

**Example 2** *Reconsider the deontic context of the previous example.* $\mathcal{E}'$ *contains only* $\alpha_1$ *and the conditioned extension* $\mathcal{E}$ *contains all logical consequences of* $\alpha_1 \wedge \alpha_2$*.*

Horty shows that every deontic context has at least one extension. He also shows [4] [3] that this defeasibility approach can deal with conflicting obligations, for which he applies a method of Van Fraassen [10]. He benefits from the existence of multiple extensions when there are conflicting obligations, where every extension contains a maximal number of consistent obligations.

### 2.3 Background knowledge

Horty's definition of overridden does not explicitly take the distinction between background knowledge and contingent facts into account, as illustrated by the following example.

**Example 3** *Consider* $O(f / b)$*,* $O(\neg f / p)$ *and* $F = p \wedge (p \rightarrow b)$*, consisting of* $F_c = p$ *and* $F_b = p \rightarrow b$*. All p are b, but* $O(f / b)$ *is not overridden by* $O(\neg f / p)$*.*

To deal with this distinction, we will alter the definition of overridden as follows:

[3] In a modal approach, these background sentences are represented by $\Box\alpha$ where the $\Box$ is interpreted in Kripke models with an *S*5 structure.

[4] Horty acknowledges that his approach can only be seen as a preliminary since it is beset by several problems. These problems have to do with defeasibility and are irrelevant to our use, that is to analyze violations of obligations.

[5] The third condition of the definition ensures that an obligation is not overridden by an unrelated obligation that has a conclusion that is contradictory with the factual situation. For example, the obligation $O(\alpha_1 / \top)$ may not be overridden by $O(\alpha_2 / \alpha_3)$ for $F = \neg\alpha_2 \wedge \alpha_3$, as would be the case if the definition lacked the third condition, because $\alpha_1 \wedge \alpha_2$ is consistent and the two obligations are therefore not related.

**Definition 5** $O(\alpha_1 / \alpha_2) \in D$ *is* overridden *in* $\langle D, F\rangle = \langle D, F_b \wedge F_c\rangle$ *iff there is a* $O(\alpha_3 / \alpha_4) \in D$ *s.t.:*
1. $F \vdash \alpha_4$, $F \vdash \alpha_2$, $F_b \wedge \alpha_4 \vdash \alpha_2$, $F_b \wedge \alpha_2 \nvdash \alpha_4$, *and*
2. $F_b \wedge \alpha_1 \wedge \alpha_3$ *is inconsistent.*

In the following, we will use this definition of overridden. The alteration of the second condition of the definition will also facilitate the definition of violated obligations in Section 4. The definitions of active and conditioned extension remain unchanged.

## 2.4 Deontic consequence

Horty uses the definitions of overridden and conditioned extension to determine which conditional obligations follow from a set of (possibly conflicting) conditional obligations. In [3] he defines the following relation $\vdash_{CF}$ of conditional deontic consequence:

**Definition 6** $D \vdash_{CF} O(\alpha_1 / \alpha_2)$ *iff* $\alpha_1 \in \mathcal{E}$ *for some conditioned extension* $\mathcal{E}$ *of* $\langle D, \alpha_2\rangle$.

Horty [3] shows that this definition of deontic consequence satisfies a restricted form of augmentation (strengthening the antecedent), which defines a notion of irrelevance:

**Example 4** *Consider the following sentences of* $\langle D, F\rangle$*:*
1. $O(\neg\alpha_1 / \top)$*: You should not eat with your fingers.*
2. $O(\alpha_1 / \alpha_2)$*: If you are served asparagus, you should eat with your fingers.*
3. $O(\alpha_3 / \top)$*: You should put your napkin on your lap.*
4. $\alpha_2$*: You are served asparagus.*

*It can be proven that* $D \nvdash_{CF} O(\neg\alpha_1 / \alpha_2)$ *and* $D \vdash_{CF} O(\alpha_3 / \alpha_2)$*, i.e. in the more specific context that you are served asparagus, you do not have the obligation not to eat with your fingers but you do have the obligation to put your napkin on your lap. The fact that you are served asparagus is irrelevant to the obligation to put your napkin on your lap.*

The notion of deontic consequence as defined by Horty has several properties. Firstly, it satisfies a *restricted form of agglomeration*, e.g. the derivation of $O(\alpha_1 \wedge \alpha_2 / \top)$ from $O(\alpha_1 / \top)$ and $O(\alpha_2 / \top)$; restricted because $O(\alpha_1 \wedge \neg\alpha_1 / \top)$ is not derivable from $O(\alpha_1 / \top)$ and $O(\neg\alpha_1 / \top)$. It is therefore weaker than SDL (unrestricted agglomeration), but stronger than non-normal modal logics (no agglomeration).

Secondly, it satisfies what we will call *reflexivity*, i.e. $D \vdash_{CF} O(F/F)$ can be proven for all $D$ and $F$. This follows immediately from Def. 4. and 6., because $F$ is always an element of the conditioned extension.

Finally, it satisfies *consistency of condition and conclusion*, i.e. for all consistent $\alpha_2$ and all derived obligations $D \vdash_{CF} O(\alpha_1 / \alpha_2)$, $\alpha_1 \wedge \alpha_2$ is consistent for all $\alpha_1$ and $D$. This follows from Def. 4. Assume that $\alpha_1 \wedge \alpha_2$ is inconsistent. The extension $\mathcal{E}$ contains $\alpha_1$ and $\alpha_2$ and therefore all sentences, because it is deductively closed. $\mathcal{E}'$ is empty when the extension contains all sentences because of the condition $\neg\alpha_1 \notin \mathcal{E}$, and therefore $\alpha_2$ is itself inconsistent.

# 3 Moral cues

A moral cue is an imperative to act, given a specific factual situation. In SDL, a moral cue is represented by the obligation $O(\alpha)$ and the consistency of the factual counterpart $\alpha$. In Horty's nonmonotonic framework, moral cues correspond to the derived obligations $D \vdash_{CF} O(\alpha/F)$. The consistency of the 'factual counterpart' $\alpha$ with $F$ follows from the consistency of condition and conclusion.

As a result of reflexivity, the set of $\alpha$'s also contains all logical consequences of the factual sentence $F$. We say that $\alpha$ is a moral cue when it does not follow from $F$.

**Definition 7** $\alpha$ *is a* moral cue *of* $\langle D, F\rangle$ *iff* $D \vdash_{CF} O(\alpha / F)$ *and* $F \nvdash \alpha$.

We will not discuss whether this is a satisfactory definition, but focus on the situation where the second condition of the previous definition is not satisfied, i.e. where $\alpha$ can be derived from $F$. The question of the next section is: when is $\alpha$ a violation?

# 4 Violated obligations

A violated obligation is an unfulfilled obligation that is not a moral cue. In SDL, violated obligations are represented by $O(\alpha) \wedge \neg\alpha$. In Horty's nonmonotonic framework, violated obligations are implicitly represented by active rules whose conclusion is not part of a conditioned extension. They are not represented, however, in the notion of deontic consequence $\vdash_{CF}$ because of the consistency of condition and conclusion. To represent them $\vdash_{CF}$ must be liberalized.

We use the fact that a conditional obligation $O(\alpha_1 / \alpha_2) \in D$ (implicitly) represents a violated obligation of a deontic context $\langle D, F\rangle$ iff $O(\alpha_1 / \alpha_2)$ is active, and $\alpha_1 \notin \mathcal{E}$ for all conditioned extensions $\mathcal{E}$ of $\langle D, F\rangle$.

**Definition 8** $D \vdash_{CF*} O(\alpha_1 / \alpha_2)$ *iff* $D \vdash_{CF} O(\alpha_1 / \alpha_2)$ *or some* $O(\alpha_1 / \alpha_3) \in D$ *is active in* $\langle D, \alpha_2\rangle$.

This notion of deontic consequence $\vdash_{CF*}$ does not satisfy the property of consistency of condition and conclusion anymore, but still satisfies reflexivity. It covers moral cues and violated obligations. The derived moral cues as given by Def. 7. are the $\alpha$ s.t. $D \vdash_{CF*} O(\alpha / F)$, $F \nvdash \alpha$ and $F \nvdash \neg\alpha$, because the conclusion of an active conditional that is not contradictory with $F$ is always contained in some conditioned extension.

**Definition 9** $\alpha$ *is a* violated obligation *(and* $\neg\alpha$ *its* violation*) of* $\langle D, F\rangle$ *iff* $D \vdash_{CF*} O(\alpha / F)$ *and* $F \vdash \neg\alpha$.

Notice the difference in representation of moral cues and violated obligations in Def. 8.: moral cues are represented through deductively closed sets (the extensions) but violated obligations are not. There seems to be no straightforward way to define deductively closed sets for violated obligations.

The next example – an instance of the famous Chisholm 'Paradox' [2] – shows the idea of the extended deontic consequence relation $\vdash_{CF*}$:

**Example 5** *Consider the following sentences [9] of* $\langle D, F\rangle$*:*
1. $O(\neg\alpha_1 / \top)$*: X has to see to it that he has no more than six library books in his possession.*
2. $O(\neg\alpha_2 / \neg\alpha_1)$*: If X has no more than six books then disciplinary action should not be taken against X.*
3. $O(\alpha_2 / \alpha_1)$*: If X has more than six books then disciplinary action should be taken against X.*
4. $\alpha_1$*: X has more then six books.*

*There are no overridden obligations and the only conditioned extension* $\mathcal{E}$ *is* $Cn[\{\alpha_1, \alpha_2\}]$*.* $D \vdash_{CF*} O(\neg\alpha_1 / \alpha_1)$ *(a violated obligation) can be proven because* $O(\neg\alpha_1 / \top)$ *is active and* $D \vdash_{CF*} O(\alpha_2 / \alpha_1)$ *(a moral cue) can be proven because* $\mathcal{E}$ *contains* $\alpha_2$*.*[6]

[6] When $F = \top$, the only conditioned extension $\mathcal{E}$ is $Cn[\{\neg\alpha_1\}]$. Horty [3] discusses the possibility to introduce a restricted form of transitivity (so-called deontic detachment) in his extensions. In that case, $\mathcal{E}$ would be $Cn[\{\neg\alpha_1, \neg\alpha_2\}]$ and $D \vdash_{CF*} O(\neg\alpha_2 / \top)$ could also be proven.

**Remark.** Some deontic logics deal with the formalization problem of violated obligations by using temporal notions. These solutions adopt what we call the *deadline principle*, which means that an obligation is only in effect until a certain time point (the deadline). At this time point, it is evaluated whether the obligation is violated or not. After this time point, the obligation is not valid anymore but consequences of the violation (CTD obligations) are. There is never a situation where a violated obligation and a CTD obligation of the violated obligation are true at the same time, so the representation problem does not exist. There is a problem with this principle for nontemporal examples, like Example 5., as Smith [9] explains for so called *standing obligations*. With a deadline principle these standing obligations can not be represented, so it leads to a lack of expressiveness.

## 5 Conflicts between defeasibility and violability

Given the definition of extended deontic consequence, we will now look at the definition of overridden. We will analyze two classic examples of conflicts between defeasibility and violability and stepwise adapt the definition of overridden to solve these conflicts satisfactorily.

The conflicts occur when an obligation is overridden by a conditional obligation that is also a CTD obligation, i.e. an obligation conditional to a violation.

**Definition 10** *$O(\alpha_3 / \alpha_4)$ is a* Contrary-To-Duty obligation *(CTD) of the (primary) obligation $O(\alpha_1 / \alpha_2)$ in $\langle D, F\rangle$ iff $F \vdash \alpha_2$, $F \vdash \alpha_4$, and $F_b \wedge \alpha_1 \wedge \alpha_4$ is inconsistent.*

### 5.1 Gentle murderer 'Paradox'

The following example is an instance of the gentle murderer 'Paradox' (you should not kill, but if you kill you should do it gently):

**Example 6** *Consider the following sentences of $\langle D, F\rangle$:*
1. *$O(\neg\alpha_1 / \top)$: You should not eat with your fingers.*
2. *$O(\alpha_2 / \alpha_1)$: If you eat with your fingers, you should eat with clean fingers.*
3. *$F_b = (\alpha_2 \rightarrow \alpha_1)$: If you eat with clean fingers, you eat with your fingers.*
4. *$F_c = \alpha_1$: You eat with your fingers.*

*Given the definition of overridden, these sentences are given a defeasibility interpretation.*

***Defeasibility interpretation.*** *$O(\neg\alpha_1 / \top)$ is overridden by $O(\alpha_2 / \alpha_1)$, and therefore $D \not\vdash_{CF*} O(\neg\alpha_1 / (\alpha_2 \rightarrow \alpha_1) \wedge \alpha_1)$. $\alpha_1$ in $O(\alpha_2 / \alpha_1)$ is an exception to $O(\neg\alpha_1 / \top)$, i.e. the fact that you eat with your fingers is an exception to the obligation not to eat with your fingers.*

*Given the definition of CTD obligations, they can also be given a violability interpretation:*

***Violability interpretation.*** *$O(\alpha_2 / \alpha_1)$ is considered as a CTD obligation of $O(\neg\alpha_1 / \top)$; $\alpha_1$ in $O(\alpha_2 / \alpha_1)$ is a violation and therefore not an exception to $O(\neg\alpha_1 / \top)$. In this interpretation, $O(\neg\alpha_1 / \top)$ should not be overridden and $D \vdash_{CF*} O(\neg\alpha_1 / (\alpha_2 \rightarrow \alpha_1) \wedge \alpha_1)$ should be provable.*

In the defeasibility interpretation, $\alpha_1$ is an exception to the rule $O(\neg\alpha_1/\top)$. In general, exceptions of an obligation that logically imply the violation of the obligation are highly counterintuitive. In these cases, we will therefore follow the violability interpretation.

We can easily weaken the definition of overridden to moderately overridden to solve this conflict in favor of the violability interpretation by introducing one more condition (related to the last condition of Def. 10.):

**Definition 11** *$O(\alpha_1 / \alpha_2)$ is* moderately overridden *in $\langle D, F\rangle$ iff there is a $O(\alpha_3 / \alpha_4) \in D$ s.t.:*
1. *$F \vdash \alpha_4$, $F \vdash \alpha_2$, $F_b \wedge \alpha_4 \vdash \alpha_2$, $F_b \wedge \alpha_2 \not\vdash \alpha_4$, and*
2. *$F_b \wedge \alpha_1 \wedge \alpha_3$ is inconsistent, and*
3. *$F_b \wedge \alpha_1 \wedge \alpha_4$ is consistent.*

The definitions of active, conditioned extension and the notion of deontic consequence can be adapted by replacing overridden by moderately overridden. This only affects (increases the number of) the derivable violated obligations (Def. 9.), not the moral cues.

### 5.2 The Reykjavic 'Paradox'

We will now look at a more complicated example of conflicts between violability and defeasibility. Loewer and Belzer [6] introduced the following example:

**Example 7** *Consider the following sentences of $\langle D, F\rangle$:*
1. *$O(\neg\alpha_1 / \top)$: X should not tell the secret to Reagan.*
2. *$O(\neg\alpha_2 / \top)$: X should not tell the secret to Gorbatsjov.*
3. *$O(\alpha_2 / \alpha_1)$: If X tells Reagan, then X should tell Gorbatsjov.*
4. *$O(\alpha_1 / \alpha_2)$: If X tells Gorbatsjov, then X should tell Reagan.*
5. *$\alpha_1 \wedge \alpha_2$: X tells Reagan and Gorbatsjov.*

*There are several interpretations of this 'Paradox'.*[7] *The definitions of overridden and moderately overridden give these conditional obligations a defeasibility interpretation.*

***Defeasibility interpretation.*** *The agent's primary obligation is not to tell Reagan or Gorbatsjov. When he tells Reagan, he should not tell Reagan but he should tell Gorbatsjov $D \not\vdash_{CF*} O(\neg\alpha_2 / \alpha_1)$, a case of defeasibility. When he tells both, he does not violate any obligations because $\alpha_1$ and $\alpha_2$ are considered as exceptions: the third sentence is an exception to the second sentence, and the fourth sentence is an exception to the first sentence.*[8]

*McCarty [7] argues for a mixed defeasibility-violability interpretation, which we will adopt here too.*

***Mixed defeasibility-violability interpretation.*** *When X tells Reagan, it is identical to the defeasibility interpretation, i.e. a case of defeasibility. When X tells both, he should tell neither of them, $D \vdash_{CF*} O(\neg\alpha_1 / \alpha_1 \wedge \alpha_2)$ and $D \vdash_{CF*} O(\neg\alpha_2 / \alpha_1 \wedge \alpha_2)$, a case of violability: the third sentence is a CTD obligation of the first sentence and the fourth sentence is a CTD obligation of the second sentence.*

The solution of this conflict in favor of the mixed interpretation can be found when sets of obligations are considered simultaneously. Given $F = \alpha_1 \wedge \alpha_2$, the last two sentences can be considered as exceptions of the first two sentences but they are also CTD obligations of the first two sentences. This is a general case of the gentle murderer 'Paradox',

[7] Besides the two interpretations mentioned here, there is also a violability interpretation which will be discussed in Section 5.3.

[8] According to the defeasibility interpretation, it might be argued that the 'Paradox' is not well modeled by the deontic context. When the last two conditional obligations should be interpreted as CTD obligations when X tells both, the first two obligations should be represented by one conditional obligation $O(\neg\alpha_1 \wedge \neg\alpha_2 / \top)$. In that case, the last two sentences are considered as CTD obligations when the definition of moderately overridden is used.

where one obligation could be considered as both an exception and a CTD obligation of another obligation. Based on a similar argument as in the previous section, we will give priority to violability instead of defeasibility.

We can use the definition of overridden (Def. 5.) to identify cases like this, where a set of conditional obligations is overridden by a set of other conditionals, but each conditional of the second set can also be seen as a CTD obligation of an obligation of the first set. We call these obligations CTD overridden.

**Definition 12** *Given $\langle D, F\rangle$. Let $S \subset D$ and $T \subset D$ be two nonempty sets s.t. $S \cap T = \emptyset$. The elements of $S$ are* CTD overridden *by the elements of $T$ if:*

1. *each $O(\alpha_3 / \alpha_4) \in T$ is overriding a $O(\alpha_1 / \alpha_2) \in S$, and each $O(\alpha_1 / \alpha_2) \in S$ is overridden by some $O(\alpha_3 / \alpha_4) \in T$ in $\langle D, F\rangle$, and*
2. *each $O(\alpha_3 / \alpha_4) \in T$ is a CTD obligation of a $O(\alpha_1 / \alpha_2) \in S$, and each $O(\alpha_1 / \alpha_2) \in S$ has some CTD obligation $O(\alpha_3 / \alpha_4) \in T$ in $\langle D, F\rangle$.*

Given this notion of CTD overridden, we can give a definition of weakly overridden which gives the Reykjavic 'Paradox' a mixed interpretation.

**Definition 13** *$O(\alpha_1 / \alpha_2) \in D$ is* weakly overridden *in $\langle D, F\rangle$ iff there is a $O(\alpha_3 / \alpha_4) \in D$ s.t. $O(\alpha_1 / \alpha_2)$ is overridden by $O(\alpha_3 / \alpha_4)$ and $O(\alpha_1 / \alpha_2)$ is not CTD overridden.*

The definitions of active, conditioned extension and the notion of deontic consequence can be adapted by replacing overridden by weakly overridden. Again, this only affects the derivable violated obligations and not the moral cues.

### 5.3 Open problem: an example

There is one more potential conflict between violability and defeasibility we have to consider:

**Example 8** *Consider the following sentences of $\langle D, F\rangle$:*

1. *$O(\neg\alpha_1 / \top)$: You should not eat with your fingers.*
2. *$O(\alpha_1 / \alpha_2)$: If you are served asparagus, you should eat with your fingers.*
3. *$O(\neg\alpha_2 / \top)$: You should not be served asparagus.*
4. *$\alpha_2$: You are served asparagus.*

*$O(\alpha_1 / \alpha_2)$ is a CTD obligation of $O(\neg\alpha_2 / \top)$ and it is also an exception to $O(\neg\alpha_1 / \top)$ given one of the definitions of overridden. Given the fact that it is a CTD obligation, we might wonder whether it is still meant to be an exception.*

We have not found a convincing natural language example in which violability should be stronger than defeasibility in this deontic theory. A candidate is the Reykjavic 'Paradox' with the violability interpretation.

**Example 9** *Consider the sentences of Example 7 with the following interpretation:*

***Violability interpretation.*** *The violability interpretation works according to the mixed interpretation when X tells neither or both. When X tells only Reagan, he has the CTD obligation to tell Gorbatsjov $O(\alpha_2 / \alpha_1)$ (a moral cue) but also the primary obligation not to tell Gorbatsjov $O(\neg\alpha_2 / \alpha_1)$ (a violated obligation?).*

A drawback of the formal representation just given, $O(\alpha_2 / \alpha_1)$ and $O(\neg\alpha_2 / \alpha_1)$, is that it has become impossible to tell which of the obligations is the moral cue given the fact that $X$ tells Reagan: according to Def. 7., both are moral cues.

There is another approach to represent violated obligations: by obligations with a more general condition. A possible solution is to extend Horty's notion of deontic consequence $\vdash_{CF}$ in such a way that it infers conditional obligations from a deontic context instead of only conditional obligations. In the case of the Reykjavic 'Paradox' this extended notion of deontic consequence would infer for $F = \alpha_1$ the conditional obligations $\langle D, F\rangle \vdash_{CF+} O(\neg\alpha_2 / \top)$ and $\langle D, F\rangle \vdash_{CF+} O(\alpha_2 / \alpha_1)$. The difference between this notion of deontic consequence $\vdash_{CF+}$ and Horty's notion $\vdash_{CF}$ would be that conditional obligations that are overridden in $\langle D, F\rangle$ are not derived by $\vdash_{CF+}$.

## 6 Summary

In this article, we have extended Horty's nonmonotonic framework for deontic reasoning in two ways. Firstly we have extended his notion of deontic consequence in such a way that violated obligations are represented explicitly and secondly we have weakened his definition of overridden to deal with conflicts between contrary-to-duty obligations and defeasible obligations.

We have indicated that there might be more complicated conflicts between violability and defeasibility, which is subject of further research.


## ACKNOWLEDGEMENTS

Thanks to Patrick van der Laag, Yao-Hua Tan and members of Euridis' FANS club for discussions on the issues raised in this paper and to Henry Prakken for showing me the difference between a defeated and a violated norm and introducing the work of Horty and McCarty to me.


## REFERENCES


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

[2] R.M. Chisholm, 'Contrary-to-duty imperatives and deontic logic', *Analysis*, **24**, 33–36, (1963).

[3] J.F. Horty, 'Nonmonotonic techniques in the formalization of commonsense normative reasoning', in *Proceedings Workshop on Nonmonotonic Reasoning*, pp. 74–84, Austin, Texas, (1993).

[4] J.F. Horty, 'Moral dilemmas and nonmonotonic logic', *Journal of Philosophical Logic*, **23**, 35–65, (1994).

[5] A.J.I. Jones and M. Sergot, 'On the characterisation of law and computer systems: The normative systems perspective', in *Deontic logic in computer science*, John Wiley & Sons, (1993).

[6] B. Loewer and M. Belzer, 'Dyadic deontic detachment', *Synthese*, **54**, 295–318, (1983).

[7] L.T. McCarty, 'Defeasible deontic reasoning', in *Fourth International Workshop on Nonmonotonic Reasoning*, Plymouth, (1992).

[8] R. Reiter, 'A logic for default reasoning', *Artificial Intelligence*, **13**, 81–132, (1980).

[9] T. Smith, 'Violation of norms', in *Proceedings of the Fourth ICAIL*, pp. 60–65, New York, (1993). ACM.

[10] B.C. van Fraassen, 'Values and the heart command', *Journal of Philosophy*, **70**, 5–19, (1973).

[11] R.J. Wieringa and J.-J.Ch. Meyer, 'Applications of deontic logic in computer science: A concise overview', in *Deontic logic in computer science*, 17–40, John Wiley & Sons, Chichester, England, (1993).