Dynamic Normative Reasoning Under Uncertainty:

How to Distinguish Between Obligations Under Uncertainty and Prima Facie Obligations

Leendert van der Torre
Department of Artificial Intelligence
Vrije Universiteit Amsterdam
De Boelelaan 1081a
1081 HV Amsterdam
The Netherlands
torre@cs.vu.nl

Yao-Hua Tan
EURIDIS
Erasmus University Rotterdam
P.O. Box 1738
3000 DR Rotterdam
The Netherlands
ytan@fac.fbk.eur.nl

May 23, 2000

Abstract

The deontic update semantics is a dynamic semantics for prescriptive obligations based on Veltman's update semantics, in which the dynamic evaluation of *conflicts* of hierarchic obligations naturally leads to defeasibility. In this paper we use this dynamic semantics to study the *diagnostic problem* of defeasible deontic logic. For example, consider a defeasible obligation 'α ought to be done' together with the fact '$\neg\alpha$ is done' under uncertainty. Is there an exception of the normality claim, or is it a violation of the obligation? We show that to answer this question a distinction has to be made between 'normally α ought to be done' and 'prima facie α ought to be done.'

1 The Logic of Norms

Computer scientists use the logic of obligations, prohibitions and permissions – called deontic logic – since the early eighties to represent and reason with legal knowledge [McC83], and recently it has been used to specify and analyze security issues about electronic networks [CF97], to represent norms in qualitative decision theory [Pea93, Bou94b, Lan96] and to represent rights, duties and commitments in multi-agent systems [vdTT99b]. A further increase may be expected now recently developed prescriptive deontic logics [Mak99, vdTT99d, MvdT99] have delivered some promising approaches for the following long-standing problems in normative reasoning and their notorious deontic paradoxes.

Norms and normative propositions. The conceptual issue of the distinction between norms (that do not have truth values) and normative propositions (that do have them) is how to distinguish between applying and observing norms. This issue is relevant for the formalization of speech acts like declarations and assertions in multi agent communication.

Contrary-to-duty. The conceptual issue of the contrary-to-duty paradoxes is how to proceed once a norm has been violated. Clearly this issue is of great practical relevance, because in most applications norms are violated frequently. In electronic contracting the contracting parties usually do not want to consider a violation as a breach of contract, but simply as a disruption in the execution of the contract that has to be repaired.

Dilemma. The conceptual problem of the dilemma paradoxes is to determine the coherence conditions of a normative system. For example, when drafting regulations a coherence check indicates whether they have this desired property, or whether they should be further modified.

In [vdTT99d] the deontic logic DUS is introduced, a deontic update semantics for prescriptive obligations based on the update semantics of [Vel96]. In DUS meaning becomes a dynamic notion: you know the meaning of a normative sentence, if you know the change it brings about in the deontic state of anyone the news conveyed by the norm applies to. In this paper we show that the dynamic approach not only gives a better analysis of the traditional deontic problems, but it also gives a better analysis of the problems discussed in *defeasible* deontic logic. For example, we can easily formalize conflicts of hierarchic obligations that prima facie exist, but which dynamically can be re-evaluated. In particular we study the diagnostic problem of defeasible deontic logic.

Diagnostic problem. Assume a defeasible obligation 'α ought to be done' together with the fact '$\neg\alpha$ is done' under uncertainty. Is this an exception of the normality claim, or is it a violation of the obligation? Obviously, this is a crucial question for legal knowledge-based systems.

The main complication of defeasible deontic logic is that there are at least two incompatible types of defeasible normative reasoning. To answer the diagnostic question we distinguish in the deontic update semantics between 'prima facie α ought to be done' [Ros30, AB96, Mor96] and obligations under uncertainty like 'normally α ought to be done' [vdTT97]. The system proposed in this paper generalizes the preliminary proposals in [vdTT98b, vdTT99c].

The distinction between the two types of defeasible normative reasoning is illustrated by the following two examples.

1. You have a prima facie obligation to go to a birthday party if you promised to go, but this prima facie obligation does not turn into a proper obligation when you have to save a child from drowning, and analogously

2. Normally you have an obligation not to have a fence around your cottage, but this obligation is defeated in the exceptional circumstances when you own a dog (an example taken from the cottage housing regulations discussed in [PS96, PS97])

The crucial distinction is that the defeasible obligation 'there should not be a fence' is completely cancelled when you own a dog, whereas the obligation to go to the party is still in force as a prima facie obligation when saving the child from drowning. It just did not turn into an actual obligation. Consequently, prima facie obligations have properties obligations under uncertainty considered in this paper do not have, such as reinstatement (if the overriding obligation is violated or overridden, then the obligation it is overriding is again in force) [vdTT97].

This distinction has important consequences for the diagnostic problem. Compare the following variants of the diagnostic problem: "Is it a violation that there is a conflict?"

- Consider the potential disaster in which we must break our promise to go to the party, i.e. the situation in which a child may drown if we do not save it. The crucial question to solve this diagnostic problem is: "Could we have evaded this situation?", i.e. "Do we control whether such a disaster may occur?" In the drowning child example it seems intuitive that this potential disaster was not caused by us, and that we could not have prevented it, and therefor it is not a violation.

- Now assume that we can control that a disaster must be prevented, for example by on purpose leaving the coffee machine on when leaving a house. Is it a valid excuse that you cannot go to the party because you have to go home to turn off the coffee machine? In this case the potential disaster is not an ideal situation, and therefore a violation, because the prima facie obligation to go to the party is violated. If we can evade a conflict between prima facie obligations, we should do so.

- Finally, consider the analogous diagnostic problem under uncertainty whether having a fence and a dog is a violation or not. Clearly we control whether we have a dog or not. However, in contrast to the party example it can be an ideal (though abnormal) situation, because the obligation not to have a fence is completely cancelled. Therefor it is not a violation.

This example illustrates that to solve a diagnostic problem we need more than a set of defeasible obligations and some facts. Both examples consist of the norm 'α_1 should be (done), and if α_2 then α_3 should be (done)' together with the fact α_2 and background knowledge that α_1, α_2 and α_3 cannot all be true true at the same time, but in the prima facie case α_2 is a violation (when it is controllable) and in the uncertainty case it is not. Consequently, to solve this diagnostic problem of defeasible deontic logic we also need to know whether the defeasible obligations are prima facie obligations or obligations under uncertainty.

Makinson [Mak93] observed that 'at the present state of play, it would not seem advisable to try to cover all complicating factors [of deontic logic] at once, but rather to get an initial appreciation of them few at a time, only subsequently putting them together and investigating their interactions.' In [Mak99, vdTT99d, MvdT99] only non-defeasible obligations in a propositional setting have been studied, and in this paper only defeasible obligations in a propositional setting are studied. To keep our analysis as simple as possible we do not extend the language with permissions and prohibitions, a first-order base language, nested conditionals, background knowledge, authorities, agents, actions, and time. In the context of deontic update semantics some of the extensions have been discussed in [vdTT99b].

The layout of this paper is as follows. First, we give a short introduction to deontic logic (Section 2) and deontic update semantics (Section 3). To distinguish controllable and uncontrollable situations we distinguish between two kinds of propositions, which enables us to analyze the diagnostic problem without getting into the difficulties of defining a full-fledged logic of action and ability. Second, we formalize different types of prescriptive defeasible obligations in update semantics (Section 4). Two characteristic properties of the logic of prima facie obligations and the logic of obligations under uncertainty are that obligations are overridden by more specific and conflicting obligations, and that unresolvable strong conflicts like 'p ought to be (done) and $\neg p$ ought to be (done)' are 'inconsistent' in the sense that they derive all sentences of the deontic language. We show that the logic formalizes the specificity principle without introducing an irrelevance problem. For prima facie obligations we formalize the heuristic principle that specific obligations are more important than more general obligations, and for obligations under uncertainty we formalize the heuristic principle that it is a violation, unless there is a more specific overriding obligation. The underlying motivation from legal reasoning is that criminals should have as little opportunities as possible to excuse themselves by claiming that their behavior was exceptional rather than criminal. Finally, we analyze the diagnostic problem above in the deontic update semantics (Section 5).

2 Deontic Logic: a Short Introduction

Deontic logic is a modal logic in which absolute and conditional obligations are represented by the modal formulas $O\alpha$ and $O(\alpha \,|\, \beta)$, where the latter is either read as 'α ought to be (done) if β is (done)' or as 'α ought to be (done) in the context where β is (done).' It can be used for the formal specification and validation of a wide variety of topics in computer science (for an overview and further references see [WM93]). For example, deontic logic can be used to formally specify soft constraints in planning and scheduling problems as norms. The advantage is that norms can be violated without creating an inconsistency in the formal specification, in contrast to violations of hard constraints. With the increasing popularity and sophistication of applications of deontic logic the fundamental problems of deontic logic, observed when deontic logic was still a purely philosophical enterprise, become more pressing. For example, Jones and Sergot argue in [JS93, JS92] that contrary-to-duty reasoning is necessary to represent certain aspects of the legal code in legal expert systems. Unfortunately, this contrary-to-duty reasoning leads to notorious paradoxes of deontic logic. The systems studied in this paper are the result of a study of the four problems discussed below: the distinction between norms and normative propositions, contrary-to-duty reasoning, the inconsistency of dilemmas, and finally the main focus of this paper, the diagnostic problem of defeasible deontic logic.

2.1 Norms and Normative Propositions

One of the first topics discussed in the development of deontic logic was the question whether norms have truth values, see [vW99, Mak99, vdTT99d] for some recent discussions. For example, Von Wright [vW81, vW99] was hesitant to call deontic formulas 'logical truths,' because "it seems to be a matter of extra-logical decision when we shall say that 'there are' or 'are not' such and such norms." Alchourrón and Bulygin [AB81, Alc93] discussed the *possibility* of a logic of norms, which they distinguish from the logic of normative propositions. The distinction between norms and normative propositions is that the former are prescriptive whereas the latter are descriptive. In the second sense, the sentence 'it is obligatory to keep right on the streets' is a description of the fact that a certain normative system contains an obligation to keep right on the streets. In the first sense this statement is the obligation of traffic law itself. Alchourrón explains the distinction with the following box metaphor.

> "We may depict the difference between the descriptive meaning (normative propositions) and the prescriptive meaning (norm) of deontic sentences by means of thinking the obligatory sets as well as the permitted sets as different boxes ready to be filled. When the authority α uses a deontic sentence prescriptively to norm an action, his activity belongs to the same category as *putting something into a box.* When α, or someone else, uses the deontic sentence descriptively his activity belongs to the same category as *making a picture of α putting something into a box.* A proposition is like a picture of reality, so to assert a proposition is like making a picture of reality. On the other hand to issue (enact) a norm is like putting something in a box. It is a way of creating something, of building a part of reality (the normative qualification of an action) with the purpose that the addressees have the option to perform the authorized actions while performing the commanded actions." [Alc93]

In the deontic update semantics prescriptive obligations are formalized as actions in dynamic semantics, as discussed in Section 3, which captures Alchourrón's idea that prescriptive obligations build a part of reality.

2.2 Contrary-to-duty Reasoning

A contrary-to-duty obligation is an obligation that is only in force in a sub-ideal situation. For example, the obligation to apologize for a broken promise is only in force in the sub-ideal situation where the obligation to keep promises is violated. Reasoning structures like 'α_1 should be (done), but if $\neg\alpha_1$ is (done) then α_2 should be (done)' must be formalized without running into the notorious contrary-to-duty paradoxes of deontic logic like Chisholm's and

Forrester's paradoxes [Chi63, For84]. The conceptual issue of these paradoxes is how to proceed once a norm has been violated. Clearly this issue is of great practical relevance, because in most applications norms are violated frequently. Usually it is stipulated in the fine print of a contract what has to be done if a term in the contract is violated. If the violation is not too serious, or was not intended by the violating party, then the contracting parties usually do not want to consider this as a breach of contracts, but simply as a disruption in the execution of the contract that has to be repaired.

Recently it was discovered that approaches based on temporal concepts are not able to formalize the benchmark examples satisfactorily [PS96, PS97, vdTT98a]. The approach to formalize contrary-to-duty reasoning in the deontic update semantics [vdTT99d] is an extension of the dyadic representation together with a preference-based semantics [Han71, Lew74], see [vdTT99a] for an extensive survey of the contrary-to-duty problem and a comparison of this dyadic representation with other approaches. The dynamic and iterative approaches [vdTT99d, Mak99, MvdT99, vdT00] are natural successors and generalizations of several approaches based on complex inductive definitions that recently have been proposed, usually applying techniques developed in non-monotonic reasoning, such as Horty's non-monotonic approach [Hor93, Hor94, vdT94] based on Reiter's default logic (Reiter's default rules do not have a truth value either and they iteratively construct extensions), Prakken and Sergot's contextual reasoning [PS96, PS97, vdTT00, vdT99], labeled deontic logic [vdTT97] and phased deontic logic [TvdT96].

2.3 Dilemmas

A dilemma is an unresolved deontic conflict, as for example experienced by Sartre's famous soldier that has the obligation to kill and the (moral) obligation not to kill. In this paper we adopt the perspective that deontic logic formalizes the reasoning of an authority issuing norms, and we assume that such an authority does not intentionally create dilemmas. Consequently, dilemmas only occur in incoherent systems and they should thus be inconsistent. In other words, all deontic conflicts should be resolved, for example by giving higher priority to newer rules, more specific rules, etc. The conceptual problem of dilemma paradoxes is to determine the coherence conditions of a normative system, because a coherent system does not contain dilemmas. In several applications it is relevant to know whether a normative system is coherent. For example, when drafting regulations that should be coherent a consistency check on the formalization of the regulations in deontic logic indicates whether they have this desired property, or whether they should be further modified. This problem and its paradoxes have also been discussed more extensively in [vdTT99a].

2.4 Diagnosis

It has been argued that specific defeasible obligations are more important than more general defeasible obligations, and therefore override them in case of conflict [Hor93, vdT94, AB96, Mor96]. Unfortunately, the analysis of the specificity principle in logics of defeasible reasoning does not apply to defeasible deontic logic, because it may interfere with the violability of norms [vdT94, vdTT97, vdT97]. In other words, the *combination* of reasoning about uncertainty and contrary-to-duty reasoning of Section 2.2 leads to new complications. In this section this interference is illustrated by a diagnostic problem with prima facie obligations and by an analogous problem with obligations under uncertainty.

We first give a short introduction of the notion of so-called prima facie obligations introduced by Ross [Ros30]. In his own words: "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" [Ros30, p.19]. A prima facie duty is a duty proper when it is not overridden by another prima facie duty. When a prima facie duty is overridden, it is not a duty proper, 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" [Ros30, p.28]. Consequently, a prima facie duty is again a duty proper when its overriding duties are violated or themselves overridden [vdTT97]. In such a case we say that the obligation is reinstated.

The diagnostic problem with prima facie obligations contains the following sentences.

1. Prima facie you should not speed.

2. To prevent a possible disaster, prima facie you should speed.

3. If you are speeding, then prima facie you should speed safely.

4. You are speeding safely.

Is the fact 'you are speeding safely' a violation or not? In this paper we formalize the heuristic principle that specific obligations are more important than more general ones. In case of a possible disaster, you should speed according to the second line and the first obligation is cancelled. Moreover, if you are speeding, then you should speed safely according to the third line and the first obligation is overshadowed (terminology of cancelling and overshadowing is taken from [vdTT97]). Only in absence of a possible disaster you have to pay a penalty for speeding, because only in that case the first obligation is a violated proper obligation. Note that the second obligation has to be stronger than the first obligation to cancel it, but the third obligation does not have to be stronger than the first one to overshadow it.

Moreover, the analogous diagnostic problem with obligations under uncertainty contains the following sentences.

1. Normally there should not be a fence around the cottage.

2. If the cottage owner has a dog, then normally there should be a fence.

3. If there is a fence, then normally it should be white.

4. There is a white fence.

Is the fact 'there is a white fence' a violation or an exception? In this paper we formalize the heuristic principle that an obligation under uncertainty 'normally α ought to be (done)' together with the fact '$\neg\alpha$ is (done)' is a violation, unless there is a more specific overriding obligation. The underlying motivation from legal reasoning is that criminals should have as little opportunities as possible to excuse themselves by claiming that their behavior was exceptional rather than criminal. If there is a dog, then there should be a fence according to the second line and the first obligation is cancelled. Moreover, if there is a fence, then there should be a white fence according to the third line, but the first obligation is not cancelled. In absence of a dog you have to pay a penalty for having a fence, because in that case the first obligation is a violated actual obligation. The difference between the antecedent of the second and third obligation is represented in the deontic states of the update semantics by two different orderings: the second gives rise to levels of exceptionality (inspired by preference-based approaches to defeasible reasoning) and the third gives rise to levels of ideality (inspired by preference-based approaches to deontic reasoning).

Summarizing, both the logic of prima facie obligations and the logic of obligations under uncertainty are extended with heuristic principles to distinguish between specificity structures (sentence (1.) and (2.)) and contrary-to-duty reasoning (sentence (1.) and (3.)). With these principles they give the same intuitively desired result for the diagnostic problem in the examples above, as we show in detail later. However, this does not imply that these two defeasible deontic logics are equivalent! As we mentioned in the introduction, if we diagnose whether the condition of sentence (2.) in the examples above is a violation or not, then the result differs. In the first example one should evade the situation in which a disaster must be prevented, and in general conflicts between prima facie obligations should be evaded. In contrast, in the second example there is no reason not to have a dog. The regulations just indicate that this is an unusual situation, for which a special rule has been given, but this unusual situation is not a violation. This crucial distinction between the two types of defeasible normative reasoning is formalized in Section 3 and 4, and in Section 5 it is shown how this formalization analyzes the diagnostic problem.

3 Obligations as Actions

In [vdTT99d] an update semantics for prescriptive obligations is introduced, based on the update semantics of [Vel96]. The system is called DUS for Deontic Update Semantics. In this section the concepts underlying DUS are informally introduced, and in Section 4 different types of prescriptive obligations are defined in DUS.

3.1 Deontic Update Semantics

In the standard definition of logical validity, an argument is valid if its premises cannot all be true without its conclusion being true as well. In update semantics, the slogan 'you know the meaning of a sentence if you know the conditions under which it is true' is replaced by 'you know the meaning of a sentence if you know the change it brings about in the information state of anyone who accepts the news conveyed by it.' Thus meaning becomes a dynamic notion: the meaning of a sentence is an operation on information states. In [vdTT99d] the following definition of meaning for normative statements is proposed. The deontic states of deontic update semantics are instantiations of ideality relations, i.e. preference orderings reflecting different degrees of ideality [Han71, Lew74, Jac85, Gob90, Han90, BMW93, TvdT96, Han97, vdTT99a]. In contrast to Veltman's update semantics, the change in the deontic state is caused by the normative sentence itself, not by the acceptance of the sentence by the agent. In speech act terminology, the prescriptive obligations are declarations instead of assertions.

> You know the meaning of a normative sentence if you know the change it brings about in the ideality relation of anyone the news conveyed by the norm applies to.

This dynamic interpretation of prescriptive obligations is related to Alchourrón's box metaphor in Section 2.1. Alchourrón compares a (monadic) prescriptive obligation with the *action* of putting something in a box. In [vdTT99d] a (dyadic) prescriptive obligation is compared with an *action* of creating an ordering. Updating ideality relations is based on the following two principles, where an ideality relation $\leq_I$ is represented by a possible worlds model, i.e. a set of worlds with a binary relation and a valuation function. We write $w_1 \leq_I w_2$ for w_1 is at least as preferable as (as good as) w_2.

1. In the initial state each world is connected to all other worlds, i.e. for all w_1, w_2 we have $w_1 \leq_I w_2$ as well as $w_2 \leq_I w_1$.
2. Adding obligations is deleting ordered pairs of worlds $w_1 \leq_I w_2$ from the ideality relation.

In [vdTT99d] we propose to formalize the update of an ideality relation by the obligation 'α ought to be (done) if β is (done)' with the action of deleting all pairs $w_1 \leq w_2$ where w_1 satisfies $\neg\alpha\wedge\beta$ and w_2 satisfies $\alpha\wedge\beta$. For example, consider the update of the initial state by the prescriptive obligation 'p ought to be (done)' represented in Figure 1. This figure should

Figure 1: Updating the initial state with 'p ought to be (done)'

be read as follows. A proposition represents a set of worlds that satisfy this proposition and that are connected to at least all other worlds that satisfy this proposition. A box represents an ideality relation between worlds. A single headed arrow from one proposition to another represents that the worlds satisfying the second proposition are better than the worlds satisfying the first proposition. The left box represents the initial state, the ideality relation in which each world is connected to all other worlds. The right box represents the ideality relation that is created by updating the initial state with the obligation 'p ought to be (done).' By deleting all pairs $w_1 \leq w_2$ from the ideality relation where w_1 is a $\neg p$ world and w_2 is a p world, we end up with an ideality relation in which each p world is better than all the $\neg p$ worlds. In [vdTT99d] we also discussed the drawbacks of the dual approach, in which the initial state each world is disconnected from all other worlds, and adding obligations is adding ordered pairs of worlds $w_1 < w_2$ to the ideality relation.

Besides prescriptive obligations also facts are formalized in the deontic logic DUS, because the actual as well as the ideal have to be represented. This is needed to formalize violations of obligations and also to give an adequate analysis of Chisholm's and Forrester's paradoxes. The static interpretation of facts is the set of worlds that satisfy the facts, and the dynamic interpretation is zooming in on the ordering. We call the set of all worlds the context of justification and the subset of worlds that satisfy the facts the context of deliberation. An example of an update with a fact is represented in Figure 2. This figure

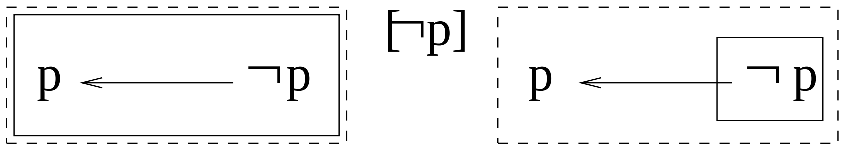

Figure 2: Updating the state 'p ought to be (done)' with 'p is not (done)'

should be read as follows. A dashed box determines the context of justification and the non-dashed box determines the context of deliberation. The left box represents the ideality relation that is created by updating the initial state with the obligation 'p ought to be (done),' see Figure 1. The right box represents the ideality relation that is created by updating this state with the fact 'p is not (done).' After the fact $\neg p$ has been settled, the context of deliberation has shrunk to the right proposition. The only ideal worlds are the leftmost worlds, and the right dashed box therefore indicates a violation.

To distinguish between the context of justification and the context of deliberation we write oblige^* for obligations referring to the latter. The obligations $\mathsf{oblige}^*\alpha$ are called factually detached. For example, from the conditional obligation 'p should be (done) if q is (done)' $\mathsf{oblige}(p|q)$ and the fact q we cannot derive $\mathsf{oblige}\ p$, but we can derive $\mathsf{oblige}^* p$. The context of deliberation is the set of choices when you are looking for practical advice, and only considers the set of future options, given that the past is already fully determined by the settled facts. For example, if Jones already has murdered Smith, it does not make sense for Jones to deliberate whether he should save this person or not. The context of justification is the set of choices for someone who is judging you, and also considers past options which are not realisable anymore.

3.2 Extensions

In the deontic language we distinguish between decision variables and parameters or events [Lan96] (called controllable and uncontrollable propositions in [Bou94b]). The distinction is well-known from decision theory and represents a simple notion of deterministic actions. We use this distinction to formalize a simple notion of ought-to-do obligations, which can be used to analyze the diagnostic problem. In contrast to ought-to-be obligations they refer to actions (see e.g. [vdTT98a] for the indeterministic case). Moreover, we assume complete knowledge of the parameters (see [Bou94b, Lan96] for the general case).

Moreover we extend the deontic states of [vdTT99d] in two respects, to deal with respectively prima facie obligations and obligations under uncertainty. We write these operators as $\mathsf{oblige_{pf}}$ and $\mathsf{oblige^{u}}$ respectively. First, to deal with prima facie obligations we replace the ideality relation ($w_1 \leq_I w_2$) by an integer valued ranking function ($R_I(w_1, w_2)$) on ordered pairs of worlds. The function $R_I(w_1, w_2)$ can be read as 'there is a reason to prefer w_1 to w_2, or, in terms of removing pairs, there is a reason to remove the pair $w_2 \leq_I w_1$. The value zero ($R_I(w_1, w_2) = 0$) corresponds to the existence of a pair of the ideality relation ($w_2 \leq_I w_1$) and the value infinity ($R_I(w_1, w_2) = \infty$) corresponds to its absence ($w_2 \not\leq_I w_1$). Moreover, a value between zero and infinity ($0 < R_I(w_1, w_2) < \infty$) represents there is a (prima facie) reason to prefer one world to another one ($R_I(w_1, w_2)$), but it leaves the possibility open that there is another prima facie reason to prefer the second world to the first ($R_I(w_2, w_1)$). The second prima facie reason may be more important ($R_I(w_2, w_1) > R_I(w_1, w_2)$) such that the first prima facie obligation does not turn into an actual obligation. The updates of the new deontic states by prima facie obligations are increases of the rank of ordered pairs. For example, consider the update of the initial state by the prescriptive obligation 'prima facie p ought to be (done)' represented in Figure 3. In figures in this paper $R_I(w_1, w_2)$ is represented by a value associated with an arrow from the world w_1 to the world w_2. The update of the initial state with the obligation increases the rank of the $(p, \neg p)$ pairs to 1.

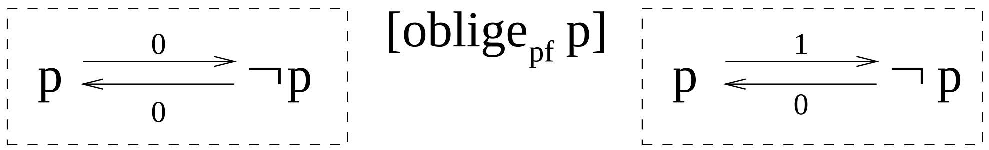

Figure 3: Updating the initial state with 'prima facie p ought to be (done)'

The second way we adapt the deontic states is the extension with a second ranking function for normality (R_N), which is defined on worlds instead of pairs of worlds. This second function corresponds to a preference relation $\leq_N$ (a reflexive, transitive and totally connected relation), given by $w_1 \leq_N w_2$ if and only if $R_N(w_1) \leq R_N(w_2)$. Such a relation is used in many logics to formalize defaults or reasoning about uncertainty, see e.g. [Sho88, Del88, Bou94a]. The minimally ranked elements are the most normal worlds. In the update with $\mathsf{oblige}^{\mathsf{u}} p$ no longer all pairs of p and $\neg p$ worlds are considered, but only pairs of *most normal* p and *most normal* $\neg p$ worlds. In the updates of these extended deontic states new ideality levels may be created when a more specific obligation overrides a more general one. This is explained in detail in Section 4.

4 Prescriptive Obligations in DUS

In this section we define different types of prescriptive obligations in update semantics. The logic of prima facie obligations and the logic of obligations under uncertainty handle conflicts of hierarchic obligations, which might be dynamically re-evaluated. Two characteristic properties of these logics are that obligations are overridden by more specific and conflicting obligations, and that unresolvable strong conflicts like 'p ought to be (done) and $\neg p$ ought to be (done)' are 'inconsistent,' in the sense that they derive all sentences of the deontic language. This is accomplished in two completely different ways. Consequently, the logics have different properties and, as we show in Section 5, for some problems they produce different diagnoses.

4.1 Deontic Update System

We start with the basic definitions of Veltman's update semantics [Vel96]. To define a deontic update semantics for a deontic language L, one has to specify a set Σ of relevant deontic states (called information states in [Vel96]), and a function $[\ ]$ that assigns to each sentence ϕ an operation $[\phi]$ on Σ. If σ is a state and ϕ a sentence, then we write '$\sigma[\phi]$' to denote the result of updating σ with ϕ. We can write '$\sigma[\psi_1]\dots[\psi_n]$' for the result of updating σ with the sequence of sentences $\psi_1, \dots, \psi_n$. Moreover, one of the deontic states has to be labeled as the minimal deontic state, written as $\mathbf{0}$, and another one as the absurd state, written as $\mathbf{1}$.

Definition 1 (Deontic update system) *A deontic update system is a triple $\langle L, \Sigma, [\ ]\rangle$ consisting of a logical language L, a set of relevant deontic states Σ and a function $[\ ]$ that assigns to each sentence ϕ of L an operation. Σ contains the elements* $\mathbf{0}$ *and* $\mathbf{1}$.

Veltman explains what kind of semantic phenomena may successfully be analyzed in update semantics and he gives a detailed analysis of one such phenomenon: default reasoning. To define obligations in update semantics we have to define the deontic language, the deontic states and the deontic updates.

4.2 Deontic Language and Deontic States

The deontic language is based on a propositional language with decision variables and parameters. It contains the dyadic operators $\mathsf{oblige}(\alpha \mid \beta)$, read as '$\alpha$ ought to be (done), if β is (done),' $\mathsf{oblige}_{\mathsf{pf}}(\alpha \mid \beta)$, read as 'prima facie α ought to be (done), if β is (done),' $\mathsf{oblige}^{\mathsf{u}}(\alpha|\beta)$, read as 'normally α ought to be (done), if β is (done)' and $\mathsf{oblige}^{\mathsf{u}}_{\mathsf{pf}}(\alpha|\beta)$, read as 'normally prima facie α ought to be (done), if β is (done).' Moreover, it also contains the operators $\mathsf{oblige}^{*}(\alpha \mid \beta)$, $\mathsf{oblige}^{*}_{\mathsf{pf}}(\alpha \mid \beta)$, $\mathsf{oblige}^{*\mathsf{u}}(\alpha \mid \beta)$ and $\mathsf{oblige}^{*\mathsf{u}}_{\mathsf{pf}}(\alpha \mid \beta)$ that only refer to states the agent holds possible. If the deontic update semantics is extended with sets of authorities (or normative systems, or institutions) and agents, then the operators can be relativised to $\mathsf{oblige}^{s}_{a_1,a_2}(\alpha|\beta)$, read as '$a_1$ ought to do α towards a_2 according to authority s, if β is (done).' This straightforward extension is further developed in [vdTT99b].

Definition 2 (Deontic language) *Let A be a set of atoms and L_0^A a propositional language with A as its non-logical symbols, partitioned in decision variables and parameters. A string of symbols ϕ is a sentence of L_1^A if and only if either ϕ is a sentence of L_0^A or there are two sentences ψ_1 and ψ_2 of L_0^A such that $\phi = X(\psi_1 \mid \psi_2)$ where X is either* oblige, $\mathsf{oblige}_{\mathsf{pf}}$, $\mathsf{oblige}^{\mathsf{u}}$, $\mathsf{oblige}^{\mathsf{u}}_{\mathsf{pf}}$, oblige^{*}, $\mathsf{oblige}^{*}_{\mathsf{pf}}$, $\mathsf{oblige}^{*\mathsf{u}}$ *or* $\mathsf{oblige}^{*\mathsf{u}}_{\mathsf{pf}}$. *We write* $\mathsf{oblige}\ \psi$ *for* $\mathsf{oblige}(\psi|\top)$*, etc., where $\top$ stands for any tautology like $p \vee \neg p$.*

A deontic state contains structures for the context of justification and for the context of deliberation, represented by respectively the possible states of the world and the agent's epistemic state. Moreover, it contains the normative and normality constraints on both structures. It is represented by a possible worlds structure $\sigma = \langle W, W^*, R_I, R_N, V \rangle$.

The context of justification is the set of worlds W with the valuation function V. For propositional ϕ and world $w \in W$ we write $\sigma, w \models \phi$ if the classical interpretation represented by $V(w)$ satisfies ϕ. In this paper we assume that the deontic state contains a world for each interpretation of L_0^A. If we want to represent background knowledge, then this assumption has to be dropped [vdT94, Lan96].

The context of deliberation is a subset W^* of the context of justification W. Whereas in Kripke semantics a unique world is singled out, called the actual world, we single out a set of worlds, called the agent's epistemic state.

The normative constraints are represented by a function on ordered pairs of worlds of W to the non-negative integers including infinity ∞.[1] The rank of a pair of worlds (w_1, w_2) represents the strength of the prima facie obligation that prefers world w_1 to w_2. If there is no such obligation then its rank is 0, and if there are several of such obligations, then its rank is the strength of the strongest of the obligations. Note

[1]The ranking function can also have the value ∞, which is larger than all integers and which has the special property that the addition of any integer number to it results again in ∞. Once a link is ranked ∞, it cannot be updated to another value. In the absurd state, all links are ranked ∞. Formally, the ranking R is a mapping of $W \times W$ to the set of positive integers $I\!N$ plus infinity, $I\!N \cup \{\infty\}$, with infinity larger than any element of $I\!N$, i.e. $\forall x \in I\!N (x \neq \infty \rightarrow x < \infty)$.

that the numbers are only used to codify a qualitative ordering, like numbers on a temperature scale or numbers in possibilistic logic, because we do not calculate with the numbers (i.e. the additive property of numbers is not exploited). We call an ordered pair of worlds a link. In particular, we call an ordered pair (w_1, w_2) of σ an (α_1, α_2) link if $\sigma, w_1 \models \alpha_1$ and $\sigma, w_2 \models \alpha_2$.

The normality constraints are represented by a ranking function R_N on worlds of W to the non-negative integers. It corresponds to the reflexive, transitive and totally connected relation $\leq_N$ defined by $w_1 \leq_N w_2$ if and only if $R_I(w_1) \leq R_I(w_2)$. Note that the relation $\leq_N$ does not contain infinite descending chains.

Finally, in the minimal state nothing is obliged, and everything is normal.

Definition 3 (Deontic state) *Let L_1^A be a deontic language. Assume a set of worlds W and a valuation function V for L_0^A such that for every interpretation of L_0^A there is at least one corresponding $w \in W$. A deontic state is a tuple $\Sigma = \langle W, W^*, R_I, R_N, V\rangle$ consisting of the set of worlds W, a possibly empty subset $W^* \subseteq W$, an integer (or ∞) valued ranking function R_I on $W \times W$, an integer valued ranking function R_N on W, and the valuation function V.*

0, *the* minimal state, *is given by $\langle W, W, W \times W \to 0, W \to 0, V\rangle$, and*

1, *the* absurd state, *is given by $\langle W, \emptyset, W \times W \to \infty, W \to 0, V\rangle$.*

In the following section the updates on these deontic states are defined.

4.3 Deontic Updates

The deontic updates are operations on the deontic states that either zoom in on the deontic state (for facts), increase the ranks of links (for prima facie obligations), or create ideality and normality levels (for obligations under uncertainty). The prescriptive obligations have the dynamic component of creating a new deontic state. For readability we describe each type of deontic operator separately below, in order of increasing complexity. Moreover, we do not yet consider the distinction between decision variables and parameters, which will be used in the analysis of the diagnostic problem.

4.3.1 Facts and Non-Defeasible Obligations

The update by non-defeasible obligations [vdTT99d] consists of two parts: a reduction of the deontic state and a conflict detection test. In Definition 4 the reduction of an ideality relation by an obligation is given.

Definition 4 (Reduction) *Let*

- *$\sigma = \langle W, W^*, R_I, R_N, V\rangle$ be a deontic state,*
- *$W_1 = \{w \in W \mid \sigma, w_1 \models \alpha \wedge \beta\}$ and $W_2 = \{w \in W \mid \sigma, w_2 \models \neg\alpha \wedge \beta\}$, and*

- $W_1^* = \{w \in W^* \mid \sigma, w_1 \models \alpha \wedge \beta\}$, $W_2^* = \{w \in W^* \mid \sigma, w_2 \models \neg\alpha \wedge \beta\}$.

The reduction of σ *by* $\mathsf{oblige}(\alpha \,|\, \beta)$ *is* $\sigma - \mathsf{oblige}(\alpha \,|\, \beta) = \langle W, W^*, R'_I, R_N, V\rangle$*, where* R'_I *is defined as follows.*

$$R'_I(w_1, w_2) = \begin{array}{ll} \infty & \textit{iff } w_1 \in W_1 \textit{ and } w_2 \in W_2 \\ R_I(w_1, w_2) & \textit{otherwise} \end{array}$$

Analogously, the reduction of σ *by* $\mathsf{oblige}^*(\alpha \,|\, \beta)$ *is* $\sigma - \mathsf{oblige}^*(\alpha \,|\, \beta) = \langle W, W^*, R'_I, R_N, V\rangle$ *with*

$$R'_I(w_1, w_2) = \begin{array}{ll} \infty & \textit{iff } w_1 \in W_1^* \textit{ and } w_2 \in W_2^* \\ R_I(w_1, w_2) & \textit{otherwise} \end{array}$$

We now turn to the conflict detection test. In the logic of oblige conflicts cannot be resolved, and each conflict is therefor a dilemma – and inconsistent, see Section 2.3. We use the following two-step conflict detection test.

1. Construct a (not necessarily connected) preference order $\leq_I$ from R_I by (1a.) $w_1 \leq_I w_2$ if and only if $R_I(w_2, w_1) \neq \infty$, and (1b.) taking the transitive closure;
2. The update $\sigma[\mathsf{oblige}(\alpha \,|\, \beta)]$ is the reduction of σ by $\mathsf{oblige}(\alpha \,|\, \beta)$ if afterwards the best β worlds of the constructed preference order $\leq_I$ are α worlds. Otherwise there is a conflict and the result of the update is the absurd state.

For this conflict detection test we need a test whether the best (i.e. $\leq_I$-minimal) β worlds are α worlds. This test is analogous to the satisfaction test of a dyadic obligation in the Hansson-Lewis semantics [Han71, Lew74], and to the test whether a set of formulas preferentially entails a conclusion in preferential entailment [Sho88]. The following definition deals with infinite descending chains, see [Lew73].

> The best β-worlds of W (W^*) of σ satisfy α if and only if for all worlds w_1 such that $\sigma, w_1 \models \beta$ there is a world $w_2 \leq_I w_1$ such that $\sigma, w_2 \models \beta$ and for all worlds $w_3 \leq_I w_2$ we have $\sigma, w_3 \models \beta \rightarrow \alpha$.

Definition 5 below combines the latter definition with taking the transitive closure. The transitive closure of an ordering can be calculated by adding relations $w_1 \leq_I w_3$ to the ordering for worlds $w_1, w_2, w_3 \in W$ such that $w_1 \leq_I w_2$ and $w_2 \leq_I w_3$. The fix-point of this iterative process is the smallest superset of the ordering such that for all worlds w_1, w_2, w_3 with $w_1 \leq_I w_2$ and $w_2 \leq_I w_3$ we have $w_1 \leq_I w_3$.

Definition 5 (pref) *Let*

- $\sigma = \langle W, W^*, R_I, R_N, V\rangle$ *be a deontic state,*
- $\leq_I$ *be derived from* R_I *by* $w_1 \leq_I w_2$ *if and only if* $R_I(w_2, w_1) \neq \infty$*, and*

- *let $\leq_\beta$ be the transitive closure of $\leq_I$ in $\{w \in W \,|\, \sigma, w \models \beta\}$, i.e. the smallest superset of $\leq_I$ such that for all β-worlds $w_1, w_2, w_3 \in W$ with $w_1 \leq_\beta w_2$ and $w_2 \leq_\beta w_3$ we have $w_1 \leq_\beta w_3$.*

The best β-worlds of W of σ satisfy α, written as $\mathsf{pref}(\sigma \,|\, \beta) \models \alpha$, if and only if for all β-worlds w_1 there is a β-world $w_2 \leq_\beta w_1$ such that for all β-worlds $w_3 \leq_\beta w_2$ we have $\sigma, w_3 \models \alpha$ with $w_1, w_2, w_3 \in W$. $\mathsf{pref}^*(\sigma|\beta) \models \alpha$ is defined analogously for worlds of W^*.

Putting it all together we define the updates. Von Wright's contingency principle, i.e. the obligation 'α ought to be (done) if β is (done)' implies the consistency of $\alpha \wedge \beta$ and $\neg\alpha \wedge \beta$, is formalized by a test on the existence of $\alpha \wedge \beta$ and $\neg\alpha \wedge \beta$ worlds.

Definition 6 (Deontic updates) *Let*

- *$\sigma = \langle W, W^*, R_I, R_N, V \rangle$ be a deontic state, and*
- *$W_1 = \{w \in W \mid \sigma, w_1 \models \alpha \wedge \beta\}$, $W_2 = \{w \in W \mid \sigma, w_2 \models \neg\alpha \wedge \beta\}$, and*
- *$W_1^* = \{w \in W^* \mid \sigma, w_1 \models \alpha \wedge \beta\}$ and $W_2^* = \{w \in W^* \mid \sigma, w_2 \models \neg\alpha \wedge \beta\}$.*

The update function $\sigma[\phi]$ is defined as follows.

- *if ϕ is a sentence of L_0^A, then*
  - *if $W' = \{w \in W^* \mid \sigma, w \models \phi\} \neq \emptyset$, then $\sigma[\phi] = \langle W, W', R_I, R_N, V \rangle$;*
  - *otherwise, $\sigma[\phi] = \mathbf{1}$.*
- *if $\phi = \mathsf{oblige}(\alpha|\beta)$, then*
  - *if $W_1 \neq \emptyset$ and $W_2 \neq \emptyset$ then*
    * *if $\mathsf{pref}(\sigma - \mathsf{oblige}(\alpha|\beta)|\beta) \models \alpha$, then $\sigma[\phi] = \sigma - \mathsf{oblige}(\alpha|\beta)$;*
    * *otherwise, $\sigma[\phi] = \mathbf{1}$.*
  - *otherwise, $\sigma[\phi] = \mathbf{1}$.*
- *if $\phi = \mathsf{oblige}^*(\alpha|\beta)$, then analogously for worlds of W_1^* and W_2^*,*
  - *if $W_1^* \neq \emptyset$ and $W_2^* \neq \emptyset$ then*
    * *if $\mathsf{pref}^*(\sigma - \mathsf{oblige}^*(\alpha|\beta)|\beta) \models \alpha$, then $\sigma[\phi] = \sigma - \mathsf{oblige}^*(\alpha|\beta)$;*
    * *otherwise, $\sigma[\phi] = \mathbf{1}$.*
  - *otherwise, $\sigma[\phi] = \mathbf{1}$.*

The update function is further defined in Definition 7.

We illustrate the definition by the update of the initial state with the four sentences of Chisholm's paradox: 'a certain man should go to the assistance of his neighbors' ($\mathsf{oblige}\ a$), 'if the man goes to their assistance, then he should tell them he will come' ($\mathsf{oblige}(t \mid a)$), 'if the man does not go to the assistance, then he should not tell them he will come ($\mathsf{oblige}(\neg t \mid \neg a)$) and 'the man does not go' ($\neg a$). The updates are visualized in Figure 4, see [vdTT99d] for the full discussion on this paradox in the deontic update semantics. This figure should be read as follows. Each set of propositions represents a set of worlds satisfying the propositions, and falsifying the other propositions. An arrow from one set of propositions to another represents that for all corresponding worlds w_1 and w_2 that are represented by these propositions, we have $w_2 \leq_I w_1$ and thus $R_I(w_1, w_2) \neq \infty$.

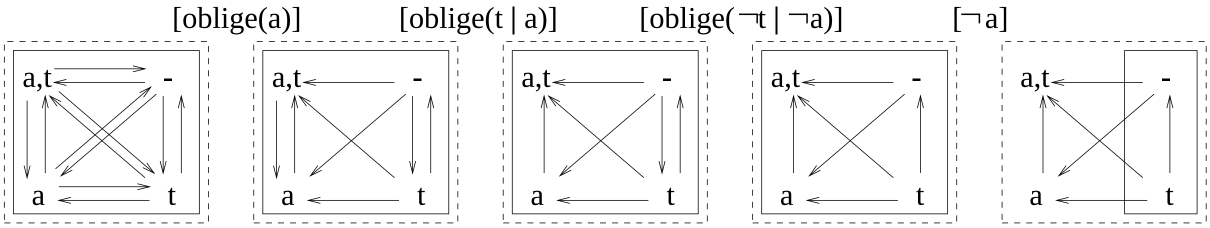

Figure 4: Chisholm's paradox

The right-most deontic state in Figure 4 is transitively closed and totally connected, and represents the order (from best to worst) $a \wedge t$, $a \wedge \neg t$, $\neg a \wedge \neg t$ and $\neg a \wedge t$. This order is the typical semantic representation of Chisholm's set in preference-based semantics, see e.g. [vdTT99a]. Note that the context of deliberation of this deontic state consists of the $\neg a$ worlds, which are only ordered by the last obligation of the Chisholm set.

4.3.2 Prima Facie Obligations

The general principle of our prima facie obligations is that in case of conflict later obligations are stronger than earlier ones. For the update with obligation $\mathsf{oblige}_{\mathsf{pf}}(\alpha \mid \beta)$ there is a conflict if all the reverse links (i.e. $(\neg\alpha \wedge \beta, \alpha \wedge \beta)$ links) are non-zero. If there is no conflict then the rank of the $(\alpha \wedge \beta, \neg\alpha \wedge \beta)$ links is 1. Otherwise, their rank is higher than the minimum of the reverse links.

Definition 7 (Deontic updates, continued from Definition 6) *Let*

- σ, W_1, W_2, W_1^* *and* W_2^* *be defined as in Definition 6, and*
- $\min_{R_I}(\alpha \mid \beta)$ *be the minimum of* $\{R_I(w_2, w_1) \mid w_1 \in W_1$ *and* $w_2 \in W_2\}$ *if this set is non-empty, undefined otherwise.*

The update function $\sigma[\phi]$ *is further defined as follows.*

- *if* $\phi = \mathsf{oblige}_{\mathsf{pf}}(\alpha \mid \beta)$*, then*

– *if* $W_1 \neq \emptyset$ *and* $W_2 \neq \emptyset$ *then* $\sigma[\phi] = \langle W, W^*, R'_I, R_N, V\rangle$ *with*

$$R'_I(w_1, w_2) = \begin{array}{ll} \max(R_I(w_1, w_2), \min_{R_I}(\alpha|\beta) + 1) & \textit{if } w_1 \in W_1 \textit{ and } w_2 \in W_2 \\ R_I(w_1, w_2) & \textit{otherwise;} \end{array}$$

– *otherwise,* $\sigma[\phi] = \mathbf{1}$.

- *if* $\phi = \mathsf{oblige}^*_{\mathsf{pf}}(\alpha|\beta)$, *then* $\sigma[\phi]$ *is defined analogously for worlds of* W_1^* *and* W_2^*.

The update function is further defined in Definition 11.

We illustrate the update with prima facie obligations by the update of the initial state with the two obligations $\mathsf{oblige}_{\mathsf{pf}}(\neg s|\top)$ and $\mathsf{oblige}_{\mathsf{pf}}(s|p)$, where s stands for speeding and p for a potential disaster if you are not speeding. If the more specific obligation comes first then all $(\neg s, s)$ and $(p \wedge s, p \wedge \neg s)$ links have rank 1 and both obligations are equally strong. Otherwise, the $(\neg s, s)$ links have rank 1 and the $(p \wedge s, p \wedge \neg s)$ links have rank 2, i.e. the more specific obligation is stronger and overrides the more general one. The latter case is illustrated in Figure 5 below. For readability only the ranks of non-zero links are represented.

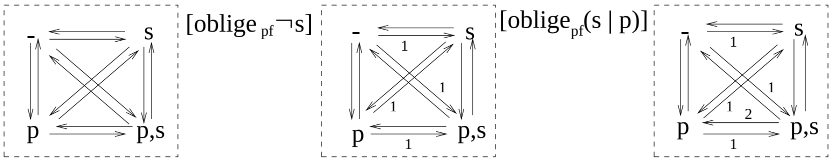

Figure 5: Speeding to prevent a disaster

4.3.3 Obligations Under Uncertainty

In the update by obligations under uncertainty we ignore the updates by prima facie obligations, i.e. we consider all ranks unequal to infinity as equivalent. The following two definitions are extensions of Definition 4 and 5 of the non-defeasible case in Section 4.3.1. In the extended definitions only *the most normal* $\alpha \wedge \beta$ and *the most normal* $\neg\alpha \wedge \beta$ worlds are compared in the ideality relation.

Definition 8 (Reduction) *Let*

- $\sigma = \langle W, W^*, R_I, R_N, V\rangle$ *be a deontic state,*
- $\min_{R_N}(\alpha) = \min\{R_N(w) \mid \sigma, w \models \alpha\}$ *if there is such a* w, *undefined otherwise,*
- $W_1^n = \{w \in W \mid \sigma, w \models \alpha \wedge \beta$ *and* $R_N(w) = \min_{R_N}(\alpha \wedge \beta)\}$, *and*
- $W_2^n = \{w \in W \mid \sigma, w \models \neg\alpha \wedge \beta$ *and* $R_N(w) = \min_{R_N}(\neg\alpha \wedge \beta)\}$.

The reduction of σ *by* $\mathsf{oblige^u}(\alpha|\beta)$ *is* $\sigma - \mathsf{oblige^u}(\alpha|\beta) = \langle W, W^*, R'_I, R_N, V\rangle$*, where* R'_I *is defined as follows.*

$$R'_I(w_1, w_2) = \begin{array}{ll} \infty & \textit{iff } w_1 \in W_1^n \textit{ and } w_2 \in W_2^n \\ R_I(w_1, w_2) & \textit{otherwise} \end{array}$$

$\sigma - \mathsf{oblige^{*u}}(\alpha|\beta)$ *is defined analogously for worlds of* W^**.*

Definition 9 (pref^u) *Let the sets of most normal* $\alpha\wedge\beta$ *worlds* W_1^n *and most normal* $\neg\alpha\wedge\beta$ *worlds* W_2^n *be defined as in Definition 8.* $\mathsf{pref}^u(\sigma\,|\,\beta) \models \alpha$ *is defined analogously to* pref *in Definition 5 for worlds of* $W_1^n \cup W_2^n$*, and* $\mathsf{pref}^{*u}(\sigma\,|\,\beta) \models \alpha$ *is defined analogously for the most normal worlds of* W^**.*

If there is a conflict, then we introduce another exceptionality level by increasing the ranks of (i.e. shifting) all worlds satisfying the condition of the obligation. In this exceptional level all ideality ranks are set to zero, which represents that updates under uncertainty only affect the most normal states.

Definition 10 (Exception) *Let*

- $\sigma = \langle W, W^*, R_I, R_N, V\rangle$ *be a deontic state,*
- $\min_{R_N}(\alpha) = \min\{R_N(w) \mid \sigma, w \models \alpha\}$ *if there is such a* w*, undefined otherwise.*

The introduction of an exceptionality level in σ *by* $\mathsf{oblige^u}(\alpha\,|\,\beta)$ *is* $\sigma -_N \mathsf{oblige^u}(\alpha\,|\,\beta) = \sigma$ *if* $\min_{R_N}(\beta)$ *or* $\min_{R_N}(\neg\beta)$ *is undefined, or* $\min_{R_N}(\neg\beta) < \min_{R_N}(\beta)$*, and otherwise* $\sigma -_N \mathsf{oblige^u}(\alpha|\beta) = \langle W, W^*, R'_I, R'_N, V\rangle$ *with*

$$\begin{array}{lll} R'_I(w_1, w_2) = & 0 & \textit{if } w_1 \models \beta \textit{ or } w_2 \models \beta \\ & R_I(w_1, w_2) & \textit{otherwise} \\ R'_N(w) = & R_N(w) + (1 - \min_{R_N}(\beta) + \min_{R_N}(\neg\beta)) & \textit{if } w \models \beta \\ & R_N(w) & \textit{otherwise} \end{array}$$

Putting it all together we define the updates. The general principle is that in case of conflict when updating with $\mathsf{oblige}(\alpha|\beta)$ a new exceptionality level for β is introduced.

Definition 11 (Deontic updates, continued from Definition 7) *Let* σ*,* W_1*,* W_2*,* W_1^* *and* W_2^* *be as in Definition 7. The update function* $\sigma[\phi]$ *is further defined as follows.*

- *if* $\phi = \mathsf{oblige^u}(\alpha|\beta)$*, then*
  - *if* $\sigma \neq \mathbf{1}$ *and* $W_1 \neq \emptyset$ *and* $W_2 \neq \emptyset$*, then*
    - *if* $\mathsf{pref^u}(\sigma - \mathsf{oblige^u}(\alpha|\beta), \beta) \models \alpha$ *then* $\sigma[\phi] = \sigma - \mathsf{oblige^u}(\alpha|\beta)$
    - *otherwise*

· *if* $\mathsf{pref}^{\mathsf{u}}(\sigma -_N \mathsf{oblige}^{\mathsf{u}}(\alpha|\beta) - \mathsf{oblige}^{\mathsf{u}}(\alpha|\beta), \beta) \models \alpha$
then $\sigma[\phi] = \sigma -_N \mathsf{oblige}^{\mathsf{u}}(\alpha|\beta) - \mathsf{oblige}^{\mathsf{u}}(\alpha|\beta)$

· *otherwise,* $\sigma[\phi] = \mathbf{1}$

– *otherwise,* $\sigma[\phi] = \mathbf{1}$.

- *if* $\phi = \mathsf{oblige}^{*\mathsf{u}}(\alpha|\beta)$, *then* $\sigma[\phi]$ *is defined analogously for worlds of* W^*.

The update function is further defined in Definition 12.

This third part of the update definition is illustrated by the update of the initial state with the two obligations from the cottage housing regulations $\mathsf{oblige}^{\mathsf{u}}(\neg f|\top)$ and $\mathsf{oblige}^{\mathsf{u}}(f|d)$, where f stands for a fence around the cottage and d for a dog in the cottage. We first consider the situation in which the most specific obligation comes first. The ideality ranking is represented in Figure 6. We only represent the ideality ranking, because if the more specific obligation comes first then all worlds remain equivalent in the normality ranking.

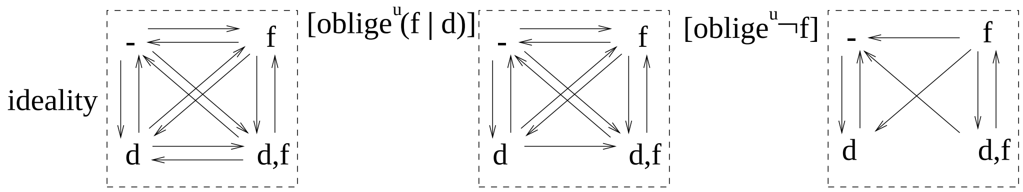

Figure 6: Fence: most specific obligation first

Otherwise, the more general obligation comes before the more specific one, as illustrated in Figure 7 below. Note that in the third ideality relation the arrows from $d \wedge \neg f$ to $d \wedge f$ are restored, because the d worlds have become exceptional.

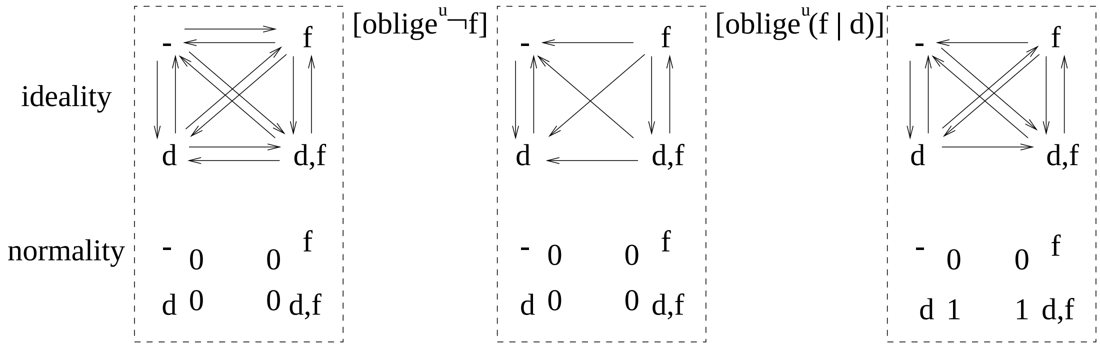

Figure 7: Fence: most general obligation first

4.3.4 Prima Facie Obligations Under Uncertainty

In the logic of prima facie obligations under uncertainty we have two distinct mechanisms to resolve conflicts: by increasing the rank of the ideality links, or by making worlds exceptional. The former is the simplest, and is illustrated in the following definition. The update by $\mathsf{oblige}^{\mathsf{u}}_{\mathsf{pf}}$ is identical to the update by $\mathsf{oblige}_{\mathsf{pf}}$ in Definition 7 restricted to the most normal worlds.

Definition 12 (Deontic updates, continued from Definition 11) *Let σ, W_1, W_2, W_1^*, W_2^* and $\min_{R_I}(\alpha|\beta)$ be defined as in Definition 7. The update function $\sigma[\phi]$ is further defined as follows.*

- *if $\phi = \mathsf{oblige}^{\mathsf{u}}_{\mathsf{pf}}(\alpha|\beta)$, then*
  - *if $W_1^n \neq \emptyset$ and $W_2^n \neq \emptyset$ then $\sigma[\phi] = \langle W, W^*, R_I', R_N, V\rangle$ with*
    $$R_I'(w_1, w_2) = \begin{array}{ll} \max(R_I(w_1, w_2), \min_{R_I}(\alpha|\beta) + 1) & \textit{if } w_1 \in W_1^n \textit{ and } w_2 \in W_2^n \\ R_I(w_1, w_2) & \textit{otherwise;} \end{array}$$
  - *otherwise, $\sigma[\phi] = \mathbf{1}$.*
- *if $\phi = \mathsf{oblige}^{*\mathsf{u}}_{\mathsf{pf}}(\alpha|\beta)$, then $\sigma[\phi]$ is defined analogously for worlds of W^*.*

Since in this paper we are interested in the diagnostic problem we do not further study this update or alternative updates by prima facie obligations under uncertainty, but we turn to the definition of validity relations.

4.4 Acceptance and Validity

A crucial notion of update systems is acceptance. The formula ϕ is accepted in a deontic state σ, written as $\sigma \Vdash \phi$, if the update by ϕ results in the same state. In that case, the information conveyed by ϕ is already subsumed by σ. Acceptance is the counterpart of satisfaction in standard semantics.

Definition 13 (Acceptance) *Let σ be an deontic state and ϕ a formula of the logical language L. $\sigma \Vdash \phi$ if and only if $\sigma[\phi] = \sigma$.*

If an update is accepted, then the deontic state usually has a specific content. For example, it is easily checked that a fact α is accepted if all the worlds of $W^* \neq \emptyset$ satisfy α, or $\sigma = \mathbf{1}$. Moreover, an obligation $\mathsf{oblige}(\alpha|\beta)$ is accepted if the rank of all $(\alpha \wedge \beta, \neg\alpha \wedge \beta)$ links is infinity (and such pairs exist), and an obligation $\mathsf{oblige}_{\mathsf{pf}}(\alpha\,|\,\beta)$ is accepted if the ranking of all $(\alpha \wedge \beta, \neg\alpha \wedge \beta)$ links is higher than the smallest rank of the reversed links. For example, it is easily checked that in Figure 4 we have that the initial state updated with the obligations from Chisholm's set accepts these sentences, because doing the same update twice has the same effect as doing it only once:

$$\mathbf{0}[\mathsf{oblige}\ a][\mathsf{oblige}(t|a)][\mathsf{oblige}(\neg t|\neg a)][\neg a] \Vdash \mathsf{oblige}\ a$$

$\mathbf{0}[\mathsf{oblige}\ a][\mathsf{oblige}(t|a)][\mathsf{oblige}(\neg t|\neg a)][\neg a] \Vdash \mathsf{oblige}(t|a)$
$\mathbf{0}[\mathsf{oblige}\ a][\mathsf{oblige}(t|a)][\mathsf{oblige}(\neg t|\neg a)][\neg a] \Vdash \mathsf{oblige}(\neg t|\neg a)$
$\mathbf{0}[\mathsf{oblige}\ a][\mathsf{oblige}(t|a)][\mathsf{oblige}(\neg t|\neg a)][\neg a] \Vdash [\neg a]$

Moreover, the same figure also illustrates the distinction between the context of justification and the context of deliberation. In the ideal state the man tells that he comes, but in the optimal reachable state the man tells that he will not come. Obviously this distinction is crucial for an adequate analysis of contrary-to-duty reasoning, see [vdTT99d].

$\mathbf{0}[\mathsf{oblige}\ a][\mathsf{oblige}(t|a)][\mathsf{oblige}(\neg t|\neg a)][\neg a] \nVdash \mathsf{oblige}\ \neg t$
$\mathbf{0}[\mathsf{oblige}\ a][\mathsf{oblige}(t|a)][\mathsf{oblige}(\neg t|\neg a)][\neg a] \Vdash \mathsf{oblige}^{*}\neg t$

The notion of acceptance is used to define notions of validity. An argument is $\Vdash_1$ valid if updating the minimal state $\mathbf{0}$ with the premises $\psi_1, \ldots, \psi_n$, in that order, yields a deontic state in which the conclusion is accepted, and an argument is $|\!\!\sim$ valid if all deontic states constructed by updating the minimal state $\mathbf{0}$ with the premises $\psi_1, \ldots, \psi_n$ *in some order such that the premises are accepted*, also accept the conclusion. Note that the order of the premises is only relevant for $\Vdash_1$, not for $|\!\!\sim$.

Definition 14 (Validity) *Let $\psi_1, \ldots, \psi_n$ and ϕ be sentences of the deontic language L_1^A. The argument of ϕ from the premises $\psi_1, \ldots, \psi_n$ is valid, written as $\psi_1, \ldots, \psi_n \Vdash_1 \phi$, if and only if $\mathbf{0}[\psi_1] \ldots [\psi_n] \Vdash \phi$. The argument of ϕ from the premises $\psi_1, \ldots, \psi_n$ is non-monotonically valid, written as $\psi_1, \ldots, \psi_n \,|\!\!\sim \phi$, if and only if for all permutations π of $1 \ldots n$ such that $\psi_{\pi(1)}, \ldots, \psi_{\pi(n)} \Vdash_1 \psi_i$ for $1 \leq i \leq n$ we have $\psi_{\pi(1)}, \ldots, \psi_{\pi(n)} \Vdash_1 \phi$.*

The notions of acceptance and validity are illustrated by our two running examples. First, reconsider the update of the initial state with the two obligations $\mathsf{oblige}_{\mathsf{pf}}(\neg s|\top)$ and $\mathsf{oblige}_{\mathsf{pf}}(s|p)$ in Figure 5. The only accepted order of the premises in Definition 14 is that the more general obligation comes before the more specific one, because only in that case the premises are accepted.

$\mathsf{oblige}_{\mathsf{pf}}(\neg s|\top), \mathsf{oblige}_{\mathsf{pf}}(s|p) \Vdash_1 \mathsf{oblige}_{\mathsf{pf}}(\neg s|\top)$
$\mathsf{oblige}_{\mathsf{pf}}(\neg s|\top), \mathsf{oblige}_{\mathsf{pf}}(s|p) \Vdash_1 \mathsf{oblige}_{\mathsf{pf}}(s|p)$
$\mathsf{oblige}_{\mathsf{pf}}(s|p), \mathsf{oblige}_{\mathsf{pf}}(\neg s|\top) \Vdash_1 \mathsf{oblige}_{\mathsf{pf}}(\neg s|\top)$
$\mathsf{oblige}_{\mathsf{pf}}(s|p), \mathsf{oblige}_{\mathsf{pf}}(\neg s|\top) \nVdash_1 \mathsf{oblige}_{\mathsf{pf}}(s|p)$

The second deontic state in Figure 5 accepts $\mathsf{oblige}_{\mathsf{pf}}(\neg s|p)$, but the third does not. Consequently we have the following.

$\mathsf{oblige}_{\mathsf{pf}}(\neg s|\top) \,|\!\!\sim \mathsf{oblige}_{\mathsf{pf}}(\neg s|p)$
$\mathsf{oblige}_{\mathsf{pf}}(\neg s|\top), \mathsf{oblige}_{\mathsf{pf}}(s|p) \,|\!\!\not\sim \mathsf{oblige}_{\mathsf{pf}}(\neg s|p)$

Hence, the logic formalizes the specificity principle. Moreover, the logic does not have an irrelevance problem, because we have for example that you should not speed in the weekend (w).

$$\mathsf{oblige}_{\mathsf{pf}}(\neg s\,|\,\top), \mathsf{oblige}_{\mathsf{pf}}(s\,|\,p) \mid\!\!\sim \mathsf{oblige}_{\mathsf{pf}}(\neg s\,|\,w)$$

We now continue with the updates represented in Figure 6 and 7, and we show that the logic of obligations under uncertainty exhibits exactly the same behavior as the logic of prima facie obligations *on this example*, as desired. The only accepted order of the premises in Definition 14 is again that the more general obligation comes before the more specific one, because only in that case the premises are accepted.

$$\begin{array}{l}
\mathsf{oblige}^{\mathsf{u}}(\neg f\,|\,\top), \mathsf{oblige}^{\mathsf{u}}(f\,|\,d) \Vdash_1 \mathsf{oblige}^{\mathsf{u}}(\neg f\,|\,\top) \\
\mathsf{oblige}^{\mathsf{u}}(\neg f\,|\,\top), \mathsf{oblige}^{\mathsf{u}}(f\,|\,d) \Vdash_1 \mathsf{oblige}^{\mathsf{u}}(f\,|\,d) \\
\mathsf{oblige}^{\mathsf{u}}(f\,|\,d), \mathsf{oblige}^{\mathsf{u}}(\neg f\,|\,\top) \Vdash_1 \mathsf{oblige}^{\mathsf{u}}(\neg f\,|\,\top) \\
\mathsf{oblige}^{\mathsf{u}}(f\,|\,d), \mathsf{oblige}^{\mathsf{u}}(\neg f\,|\,\top) \nVdash_1 \mathsf{oblige}^{\mathsf{u}}(f\,|\,d)
\end{array}$$

The second deontic state in Figure 7 accepts the obligation $\mathsf{oblige}^{\mathsf{u}}(\neg f\,|\,d)$, but the third does not. Consequently we again have the following.

$$\begin{array}{l}
\mathsf{oblige}^{\mathsf{u}}(\neg f\,|\,\top) \mid\!\!\sim \mathsf{oblige}^{\mathsf{u}}(\neg f\,|\,d) \\
\mathsf{oblige}^{\mathsf{u}}(\neg f\,|\,\top), \mathsf{oblige}^{\mathsf{u}}(f\,|\,d) \mid\!\!\not\sim \mathsf{oblige}^{\mathsf{u}}(\neg f\,|\,d)
\end{array}$$

Hence, the logic formalizes the specificity principle. Moreover, the logic does not have an irrelevance problem, because we have for example that there ought to be no fence in the weekend (w).

$$\mathsf{oblige}^{\mathsf{u}}(\neg f\,|\,\top), \mathsf{oblige}^{\mathsf{u}}(f\,|\,d) \mid\!\!\sim \mathsf{oblige}^{\mathsf{u}}(\neg f\,|\,w)$$

4.5 Properties

The logic of prima facie obligations and the logic of obligations under uncertainty have many defeasible variants of the properties of the logic of non-defeasible obligations discussed in [vdTT99d]. Some examples are given below, such as strengthening of the antecedent (**SA**), a kind of deontic detachment (**DD**) and a conjunction and disjunction rule (**AND** and **ORC**). Obviously, some of these derivations are only valid non-monotonically. In particular, the examples in the previous section illustrate that strengthening of the antecedent is blocked by a more specific and conflicting obligation. Moreover, a strongly conflicting set derives a contradiction (**D**). We assume that p, q, r and s are atoms of A in $\mathcal{L}_0^A$, such that sets like W_1 and W_2 in the update definitions are non-empty.

$\mathbf{SA}_{\mathsf{pf}}$ $\quad \mathsf{oblige}_{\mathsf{pf}}\, p \mid\!\sim \mathsf{oblige}_{\mathsf{pf}}(p|q)$
$\mathbf{DD}_{\mathsf{pf}}$ $\quad \mathsf{oblige}_{\mathsf{pf}}(p|q \wedge r) \wedge \mathsf{oblige}_{\mathsf{pf}}(q|r) \mid\!\sim \mathsf{oblige}_{\mathsf{pf}}(p \wedge q|r)$
$\mathbf{AND}_{\mathsf{pf}}$ $\quad \mathsf{oblige}_{\mathsf{pf}}\, p, \mathsf{oblige}_{\mathsf{pf}}\, q \mid\!\sim \mathsf{oblige}_{\mathsf{pf}}(p \wedge q)$
$\mathbf{ORC}_{\mathsf{pf}}$ $\quad \mathsf{oblige}_{\mathsf{pf}}\, p, \mathsf{oblige}_{\mathsf{pf}}\, q \mid\!\sim \mathsf{oblige}_{\mathsf{pf}}(p \vee q)$
$\mathbf{D}_{\mathsf{pf}}$ $\quad \mathsf{oblige}_{\mathsf{pf}}\, p, \mathsf{oblige}_{\mathsf{pf}}\, \neg p \mid\!\sim \bot$

$\mathbf{SA}^{\mathsf{u}}$ $\quad \mathsf{oblige}^{\mathsf{u}} p \mid\!\sim \mathsf{oblige}^{\mathsf{u}}(p|q)$
$\mathbf{DD}^{\mathsf{u}}$ $\quad \mathsf{oblige}^{\mathsf{u}}(p|q \wedge r) \wedge \mathsf{oblige}^{\mathsf{u}}(q|r) \mid\!\sim \mathsf{oblige}^{\mathsf{u}}(p \wedge q|r)$
$\mathbf{AND}^{\mathsf{u}}$ $\quad \mathsf{oblige}^{\mathsf{u}} p, \mathsf{oblige}_{\mathsf{pf}}\, q \mid\!\sim \mathsf{oblige}^{\mathsf{u}}(p \wedge q)$
$\mathbf{ORC}^{\mathsf{u}}$ $\quad \mathsf{oblige}^{\mathsf{u}} p, \mathsf{oblige}^{\mathsf{u}} q \mid\!\sim \mathsf{oblige}^{\mathsf{u}}(p \vee q)$
$\mathbf{D}^{\mathsf{u}}$ $\quad \mathsf{oblige}^{\mathsf{u}} p, \mathsf{oblige}^{\mathsf{u}} \neg p \mid\!\sim \bot$

Moreover, the logics of defeasible obligations have several non-derivations of the logic of non-defeasible obligations, such as the following weakening of the consequent (**WC**), the disjunction rule for the antecedent (**OR**), or the obligation to evade conflicts (**FC**). Moreover, in contrast to the logic of non-defeasible obligations, a specificity set does not imply a contradiction (**DS**).

$\mathbf{WC}_{\mathsf{pf}}$ $\quad \mathsf{oblige}_{\mathsf{pf}}\, p \mid\!\not\sim \mathsf{oblige}_{\mathsf{pf}}(p \vee q)$
$\mathbf{OR}_{\mathsf{pf}}$ $\quad \mathsf{oblige}_{\mathsf{pf}}(p|q), \mathsf{oblige}_{\mathsf{pf}}(p|\neg q) \mid\!\not\sim \mathsf{oblige}_{\mathsf{pf}}\, p$
$\mathbf{FC}_{\mathsf{pf}}$ $\quad \mathsf{oblige}_{\mathsf{pf}}\, p, \mathsf{oblige}_{\mathsf{pf}}(\neg p|q) \mid\!\not\sim \mathsf{oblige}_{\mathsf{pf}}\, \neg q$
$\mathbf{DS}_{\mathsf{pf}}$ $\quad \mathsf{oblige}_{\mathsf{pf}}\, p, \mathsf{oblige}_{\mathsf{pf}}(\neg p|q) \mid\!\not\sim \bot$

$\mathbf{WC}^{\mathsf{u}}$ $\quad \mathsf{oblige}^{\mathsf{u}} p \mid\!\not\sim \mathsf{oblige}^{\mathsf{u}}(p \vee q)$
$\mathbf{OR}^{\mathsf{u}}$ $\quad \mathsf{oblige}^{\mathsf{u}}(p|q), \mathsf{oblige}^{\mathsf{u}}(p|\neg q) \mid\!\not\sim \mathsf{oblige}^{\mathsf{u}} p$
$\mathbf{FC}^{\mathsf{u}}$ $\quad \mathsf{oblige}^{\mathsf{u}} p, \mathsf{oblige}^{\mathsf{u}}(\neg p|q) \mid\!\not\sim \mathsf{oblige}^{\mathsf{u}} \neg q$
$\mathbf{DS}^{\mathsf{u}}$ $\quad \mathsf{oblige}^{\mathsf{u}} p, \mathsf{oblige}^{\mathsf{u}}(\neg p|q) \mid\!\not\sim \bot$

The main distinctions between the logic of prima facie obligations and the logic of obligations under uncertainty are related to the different ways they deal with conflicts, such as for example the following two types of reinstatement (**RI** and **RIO**).

$\mathbf{RI}_{\mathsf{pf}}$ $\quad \mathsf{oblige}_{\mathsf{pf}}\, p, \mathsf{oblige}_{\mathsf{pf}}(\neg p \wedge \neg q|r) \mid\!\sim \mathsf{oblige}_{\mathsf{pf}}(p|r \wedge q)$
$\mathbf{RIO}_{\mathsf{pf}}$ $\quad \mathsf{oblige}_{\mathsf{pf}}\, p, \mathsf{oblige}_{\mathsf{pf}}(\neg p \wedge \neg q|r), \mathsf{oblige}_{\mathsf{pf}}(q|r \wedge s) \mid\!\sim \mathsf{oblige}_{\mathsf{pf}}(p|r \wedge s)$

$\mathbf{RI}^{\mathsf{u}}$ $\quad \mathsf{oblige}^{\mathsf{u}} p, \mathsf{oblige}^{\mathsf{u}}(\neg p \wedge \neg q|r) \mid\!\not\sim \mathsf{oblige}^{\mathsf{u}}(p|r \wedge q)$
$\mathbf{RIO}^{\mathsf{u}}$ $\quad \mathsf{oblige}^{\mathsf{u}} p, \mathsf{oblige}^{\mathsf{u}}(\neg p \wedge \neg q|r), \mathsf{oblige}^{\mathsf{u}}(q|r \wedge s) \mid\!\not\sim \mathsf{oblige}^{\mathsf{u}}(p|r \wedge s)$

The verification of these properties is analogous to the analysis of the examples given in the previous section, but there are some additional complications. We give the main details below and leave the instructive construction of the deontic states and verification of the accepted formulas to the reader.

RI$_{\mathsf{pf}}$ Consider the $p \wedge r \wedge q$ worlds W_1 and the $\neg p \wedge r \wedge q$ worlds W_2 of the update of $\mathbf{0}$ by the obligations $\mathsf{oblige}_{\mathsf{pf}}\, p$ and $\mathsf{oblige}_{\mathsf{pf}}(\neg p \wedge \neg q | r)$, in any order. In these deontic states we have for $w_1 \in W_1$ and $w_2 \in W_2$ that $R_I(w_1, w_2) \geq 1$ and $R_I(w_2, w_1) = 0$, because the former is increased by the first obligation and the latter is not increased by any obligation. Consequently the deontic state accepts $\mathsf{oblige}_{\mathsf{pf}}(p|r \wedge q)$.

RIO$_{\mathsf{pf}}$ Consider the $p \wedge r \wedge s$ worlds W_1 and the $\neg p \wedge r \wedge s$ worlds W_2 (each set divided in q and $\neg q$ worlds) of the update of $\mathbf{0}$ by the obligations $\mathsf{oblige}_{\mathsf{pf}}\, p$, $\mathsf{oblige}_{\mathsf{pf}}(\neg p \wedge \neg q | r)$ and $\mathsf{oblige}_{\mathsf{pf}}(q \,|\, r \wedge s)$, in any order. In this deontic state we have for $w_1 \in W_1$ and $w_2 \in W_2$ that $R_I(w_1, w_2) \geq 1$, because it is at least increased by the update with the first obligation. Moreover, there are $p \wedge q \wedge r \wedge s$ worlds $w_1 \in W_1$ and $\neg p \wedge q \wedge r \wedge s$ worlds $w_2 \in W_2$ such that $R_I(w_2, w_1) = 0$, because it is not increased by the update with any obligation. Consequently the deontic state accepts $\mathsf{oblige}_{\mathsf{pf}}(p|r \wedge s)$.

RI$^{\mathsf{u}}$ Consider the deontic state $\mathbf{0}[\mathsf{oblige}^{\mathsf{u}} p][\mathsf{oblige}^{\mathsf{u}}(\neg p \wedge \neg q \mid r)]$. The second obligation introduces an exceptionality level, and the ideality rank of $r \wedge q$ worlds are therefore only increased by this second obligation. Consequently this state does not accept $\mathsf{oblige}^{\mathsf{u}}(p|r \wedge q)$.

RIO$^{\mathsf{u}}$ Consider the deontic state $\mathbf{0}[\mathsf{oblige}^{\mathsf{u}} p][\mathsf{oblige}^{\mathsf{u}}(\neg p \wedge \neg q | r)][\mathsf{oblige}^{\mathsf{u}}(q | r \wedge s)]$. The third obligation introduces an exceptionality level, and the ideality rank of $r \wedge s$ worlds are therefore only increased by this third obligation. Consequently this state does not accept $\mathsf{oblige}^{\mathsf{u}}(p|r \wedge s)$.

The relations between the four types of prescriptive obligations are as follows. The derivations follow directly from the definitions, and the non-derivibility of $\mathsf{oblige}^{\mathsf{u}}_{\mathsf{pf}}(p \mid q)$ from $\mathsf{oblige}_{\mathsf{pf}}(p|q)$ follows from the fact that the minimal rank of the most normal 'reverse' $(\neg p \wedge q, p \wedge q)$ links may be higher than the minimal rank of all 'reverse' $(\neg p \wedge q, p \wedge q)$ links.

if $\sigma \Vdash \mathsf{oblige}(p|q)$ then $\sigma \Vdash \mathsf{oblige}_{\mathsf{pf}}(p|q)$
if $\sigma \Vdash \mathsf{oblige}(p|q)$ then $\sigma \Vdash \mathsf{oblige}^{\mathsf{u}}(p|q)$
if $\sigma \Vdash \mathsf{oblige}_{\mathsf{pf}}(p|q)$ then $\sigma \nVdash \mathsf{oblige}^{\mathsf{u}}_{\mathsf{pf}}(p|q)$
if $\sigma \Vdash \mathsf{oblige}^{\mathsf{u}}(p|q)$ then $\sigma \Vdash \mathsf{oblige}^{\mathsf{u}}_{\mathsf{pf}}(p|q)$

These relations between the operators are represented in Figure 8 below.

5 The Diagnostic Problem

In descriptive deontic logics, p is a violation if and only if $p \wedge O\neg p$ is true. Likewise, at first sight it may seem that a deontic state σ accepts a violation for p if σ accepts p and σ accepts $\mathsf{oblige}\neg p$. This, however, does not work for the following reasons. First, such a definition does not take the distinction between decision variables and parameters (controllable and uncontrollable propositions) of Section 3.2 into account. Secondly, and more

$$\begin{tikzcd}
& \mathsf{oblige}_{\mathsf{pf}}(p|q) & \\
\mathsf{oblige}(p|q) \arrow[ur] \arrow[dr] & & \\
& \mathsf{oblige}^{\mathsf{u}}(p|q) \arrow[r] & \mathsf{oblige}^{\mathsf{u}}_{\mathsf{pf}}(p|q)
\end{tikzcd}$$

Figure 8: Relations between the four types of prescriptive obligations

problematically, this solution does not work with prescriptive obligations. For example, assume this definition of violations and the single obligation $\mathsf{oblige}\ \neg p$. The fact $p \wedge q$ is not a violation, because $\mathsf{oblige}\ \neg p, p \wedge q \not\Vdash \mathsf{oblige}\ \neg(p \wedge q)$ (remember that the logic does not satisfy weakening of the consequent **WC**), whereas the fact p is. One obvious but technically complicated solution to this problem is to reduce a deontic state to a set of preference models of descriptive obligations, as done for example in the definition of pref in Definition 5, or in [vdTT99b]. In this paper we formalize a violation detection test directly on the deontic states.

To analyze the diagnostic problem we use the distinction between decision variables and parameters. We restrict ourselves to the distinction between worlds which still can be realized and worlds which can no longer be realized due to some action of the agent. There is a violation of a duty proper if and only if for every world that still can be realized there is a better world which can no longer be realized, and that has the same truth values for the parameters.

Definition 15 (Violation) *Let W_{p+f} be the set of worlds in which the parameters and the facts are both true, and let W_{p-f} be the worlds in which the parameters are true but some fact is false. There is a violation of a duty proper if and only if for every world $w \in W_{p+f}$ there is a world $w' \in W_{p-f}$ such that $R_I(w, w') < R_I(w', w)$.*

The following example illustrates the diagnostic problem discussed in Section 2.4. Consider the update of the initial state by the obligations $\mathsf{oblige}_{\mathsf{pf}}(\neg s \mid \top)$, $\mathsf{oblige}_{\mathsf{pf}}(s \mid p)$ and $\mathsf{oblige}_{\mathsf{pf}}(s \wedge a \mid s)$, in any order such that the premises are accepted. The $(s, \neg s)$ links of the constructed deontic state have rank 1, the $(p \wedge \neg s, p \wedge s)$ links have rank 2, and the $(s \wedge \neg a, s \wedge a)$ links have rank 1, because the links of the third obligation do not conflict with the links of the first two obligations and therefore the only constraint on the ordering of the premises is that the second obligation comes later than the first one. This is exactly the desired behavior discussed in Section 2.4. Now consider the following cases.

p is a parameter: 'you are speeding to prevent a disaster' ($s \wedge p$) is not a violation. W_{p+f} is the set of $s \wedge p$ worlds and W_{p-f} is the set of $\neg s \wedge p$ worlds. It is not the case that for every $w \in W_{p+f}$ there is a $w' \in W_{p-f}$ such that $R(w, w') < R(w', w)$. On the contrary, $\neg s \wedge p$ is a violation.

p is a decision variable: 'You are speeding safely' ($s \wedge a$) is a violation. W_{p+f} is the set of $s \wedge a$ worlds and W_{p-f} is the set of $\neg(s \wedge a)$ worlds. For every world $w \in W_{p+f}$ there is a $w' \in W_{p-f}$ such that $R(w, w') < R(w', w)$.

p is a decision variable: 'there is a disaster to be prevented' (p) is a violation. W_{p+f} is the set of p worlds and W_{p-f} is the set of $\neg p$ worlds. For every world $w \in W_{p+f}$ there is a $w' \in W_{p-f}$ such that $R(w, w') < R(w', w)$. Note that 'you are speeding to prevent a disaster' ($s \wedge p$) and 'you are speeding safely to prevent a disaster' ($s \wedge a \wedge p$) are violations for similar reasons.

This is exactly the desired result, because if you can evade a conflict then you should. The underlying idea is that agents cannot escape their responsibilities by creating exceptional circumstances, because exceptional circumstances are sub-ideal. Although there is no violation of a duty proper, there is still a violation of a prima facie duty.

We now illustrates how the deontic update semantics for obligations under uncertainty deals with the analogous diagnostic problem. The definition of violation again restricts to the most normal worlds. There is a violation of a duty proper if and only if for every world that still can be realized and that is most normal there is a better world which can no longer be realized and which is as least as normal, and that has the same truth values for the parameters.

Definition 16 (Violation under uncertainty) *Let W^n_{p+f} be the set of most normal worlds in which the parameters and the facts are both true, and let W^n_{p-f} be the most normal worlds in which the parameters are true but some fact is false. There is a violation of a duty proper if and only if for every world $w \in W^n_{p+f}$ there is a world $w' \in W^n_{p-f}$ such that $R_I(w, w') < R_I(w', w)$.*

Consider the update of the initial state by the obligations $\mathsf{oblige}^{\mathsf{u}}(\neg f \,|\, \top)$, $\mathsf{oblige}^{\mathsf{u}}(f \,|\, d)$ and $\mathsf{oblige}^{\mathsf{u}}(w \wedge f \,|\, f)$, in any order such that the premises are accepted. The only orders that accept all premises are again orders in which the second obligation is later than the first obligation. The constructed states are identical to the third state in Figure 7, with the exception that the $\neg d \wedge f$ worlds are now divided into $\neg d \wedge f$ and $\neg d \wedge f$ worlds. Consider the following cases.

d is a parameter: $d \wedge f$ is not a violation. W^n_{p+f} is the set of most normal $d \wedge f$ worlds, i.e. the $d \wedge f$ worlds, and W^n_{p-f} is the set of most normal $d \wedge \neg f$ worlds, i.e. the $d \wedge \neg f$ worlds. It is not the case that for every world $w \in W^n_{p+f}$ there is a $w' \in W^n_{p-f}$ such that $R(w, w') < R(w', w)$. On the contrary, $d \wedge \neg f$ is a violation.

d is a decision variable: 'a white fence' $w \wedge f$ is a violation. W^n_{p+f} is the set of most normal $\wedge w \wedge f$ worlds, i.e. the $\neg d \wedge w \wedge f$ worlds, and W_{p-f} is the set of most normal $\neg(w \wedge f)$ worlds, i.e. the $\neg d \wedge \neg(w \wedge f)$ worlds. For every world $w \in W^n_{p+f}$ there is a

$w' \in W^n_{p-f}$ such that $R(w, w') < R(w', w)$, the worlds are as normal, and they have the same truth values for the parameter d.

d is a decision variable: 'there is a dog' d is not a violation. W^n_{p+f} is the set of most normal d worlds, i.e. the d worlds, and W_{p-f} is the set of most normal $\neg d$ worlds, i.e. the $\neg d$ worlds. It is not the case that for every world $w \in W^n_{p+f}$ there is a $w' \in W^n_{p-f}$ such that $R(w, w') < R(w', w)$, because the ranks of all links between $\neg d$ and d worlds are 0. Note that 'there is a dog and a fence' $(d \wedge f)$ and 'there is a dog and a white fence' $(d \wedge w \wedge f)$ is are not violations either for similar reasons.

Summarizing, if the antecedent of the most specific overriding obligation is a parameter then both logics of defeasible obligations have the same desirable behavior, but if this antecedent is a decision variable then their behavior differs.

6 Conclusions

In this paper we extended our dynamic approach to formalizing norms [vdTT99d]. Moreover, in [Mak99] an iterative approach is proposed. Besides taking into account the philosophical issue that norms do not have a truth value, the dynamic and iterative approaches also give better analyses of the benchmark examples of deontic logic, in particular the deontic paradoxes. In this paper we introduced a logic of different types of defeasible prescriptive obligations, in which more specific obligations override more general ones. The specificity problem of prima facie obligations is solved by giving more specific and conflicting obligations a higher strength, and the specificity problem of obligations under uncertainty is solved by making specific circumstances exceptional. The expressive framework can be used to analyze the relation between the two types of defeasible obligations, as well as a semantics to define new types of operators such as the prima facie obligation under uncertainty.

If we use the deontic update semantics for applications, then we have to add some additional operators. For example, we may choose to add:

Test operators $\mathsf{ideal}(\alpha \mid \beta)$ for the test 'ideally, α is (done), if β is (done)' [vdTT99c, vdTT99b]. (It is shown in [TvdT96] that we have to introduce a separate operator for weakening of the consequent to combine it with strengthening of the antecedent.) The interaction between oblige and ideal is analogous to the interaction between $\mathsf{normally}$ and $\mathsf{presumably}$ operators in Veltman's update semantics [Vel96].

Descriptive obligations $O\alpha$ can be used to discuss the relation between prescriptive and descriptive obligations [vdTT99b]. In dynamic logic we can write formulas like $[\mathsf{oblige}\ \alpha]O\alpha$ representing that after α has been promulgated, the obligation that 'α should be (done)' is true.

Permission operators permit for strong permissions, see [vdTT99b]. In contrast to weak permissions like the descriptive $Pp =_{def} \neg O \neg p$, strong permissions cannot be defined in terms of obligations.

Default operators PRESUMABLE, NORMALLY, as formalized in Veltman's update semantics [Vel96]. Alternatively one may choose to adapt the default operators under the 'epistemic constructibility' paradigm of [Wey98].

References

[AB81] C.E. Alchourrón and Bulygin. The expressive conception of norms. In R. Hilpinen, editor, *New Studies in Deontic Logic: Norms, Actions and the Foundations of Ethics*, pages 95–124. D. Reidel, 1981.

[AB96] N. Asher and D. Bonevac. Prima facie obligation. *Studia Logica*, 57:19–45, 1996.

[Alc93] C.E. Alchourrón. Philosophical foundations of deontic logic and the logic of defeasible conditionals. In J.-J. Meyer and R. Wieringa, editors, *Deontic Logic in Computer Science: Normative System Specification*, pages 43–84. John Wiley & Sons, 1993.

[BMW93] A.L. Brown, S. Mantha, and T. Wakayama. Exploiting the normative aspect of preference: a deontic logic without actions. *Annals of Mathematics and Artificial Intelligence*, 9:167–203, 1993.

[Bou94a] C. Boutilier. Conditional logics of normality: a modal approach. *Artificial Intelligence*, 68:87–154, 1994.

[Bou94b] C. Boutilier. Toward a logic for qualitative decision theory. In *Proceedings of the Fourth International Conference on Principles of Knowledge Representation and Reasoning (KR'94)*, pages 75–86, 1994.

[CF97] R. Conte and R. Falcone. ICMAS'96: Norms, obligations, and conventions. *AI Magazine*, 18,4:145–147, 1997.

[Chi63] R.M. Chisholm. Contrary-to-duty imperatives and deontic logic. *Analysis*, 24:33–36, 1963.

[Del88] J.P. Delgrande. An approach to default reasoning based on a first-order conditional logic: revised report. *Artificial Intelligence*, 36:63–90, 1988.

[For84] J.W. Forrester. Gentle murder, or the adverbial Samaritan. *Journal of Philosophy*, 81:193–197, 1984.

[Gob90] L. Goble. A logic of good, would and should, part 2. *Journal of Philosophical Logic*, 19:253–276, 1990.

[Han71] B. Hansson. An analysis of some deontic logics. In R. Hilpinen, editor, *Deontic Logic: Introductory and Systematic Readings*, pages 121–147. D. Reidel Publishing Company, Dordrecht, Holland, 1971.

[Han90] S.O. Hansson. Preference-based deontic logic (PDL). *Journal of Philosophical Logic*, 19:75–93, 1990.

[Han97] S.O. Hansson. Situationist deontic logic. *Journal of Philosophical Logic*, 26:423–448, 1997.

[Hor93] J.F. Horty. Deontic logic as founded in nonmonotonic logic. *Annals of Mathematics and Artificial Intelligence*, 9:69–91, 1993.

[Hor94] J.F. Horty. Moral dilemmas and nonmonotonic logic. *Journal of Philosophical Logic*, 23:35–65, 1994.

[Jac85] F. Jackson. On the semantics and logic of obligation. *Mind*, 94:177–196, 1985.

[JS92] A.J.I. Jones and M. Sergot. Deontic logic in the representation of law: Towards a methodology. *Artificial Intelligence and Law*, 1:45–64, 1992.

[JS93] A.J.I. Jones and M. Sergot. On the characterisation of law and computer systems: The normative systems perspective. In J.-J. Meyer and R. Wieringa, editors, *Deontic Logic in Computer Science*. John Wiley & Sons, 1993.

[Lan96] J. Lang. Conditional desires and utilities - an alternative approach to qualitative decision theory. In *Proceedings of the Twelfth European Conference on Artificial Intelligence (ECAI'96)*, pages 318–322, 1996.

[Lew73] D. Lewis. *Counterfactuals*. Blackwell, Oxford, 1973.

[Lew74] D. Lewis. Semantic analysis for dyadic deontic logic. In S. Stunland, editor, *Logical Theory and Semantical Analysis*, pages 1–14. D. Reidel Publishing Company, Dordrecht, Holland, 1974.

[Mak93] D. Makinson. Five faces of minimality. *Studia Logica*, 52:339–379, 1993.

[Mak99] D. Makinson. On a fundamental problem of deontic logic. In P. McNamara and H. Prakken, editors, *Norms, Logics and Information Systems. New Studies on Deontic Logic and Computer Science*, pages 29–54. IOS Press, 1999.

[McC83] L.T. McCarty. Permissions and obligations. In *Proceedings of the Eighth International Joint Conference on Artificial Intelligence (IJCAI'83)*, pages 287–294, 1983.

[Mor96] M. Morreau. *Prima Facie* and seeming duties. *Studia Logica*, 57:47–71, 1996.

[MvdT99] D. Makinson and L. van der Torre. Input/output logics. *Submitted*, 1999.

[Pea93] J. Pearl. From conditional oughts to qualitative decision theory. In *Proceedings of the Ninth Conference on Uncertainty in Artificial Intelligence (UAI'93)*, pages 12–20, 1993.

[PS96] H. Prakken and M.J. Sergot. Contrary-to-duty obligations. *Studia Logica*, 57:91–115, 1996.

[PS97] H. Prakken and M.J. Sergot. Dyadic deontic logic and contrary-to-duty obligations. In D. Nute, editor, *Defeasible Deontic Logic*, volume 263 of *Synthese Library*, pages 223–262. Kluwer, 1997.

[Ros30] D. Ross. *The Right and the Good.* Oxford University Press, 1930.

[Sho88] Y. Shoham. *Reasoning About Change.* MIT Press, 1988.

[TvdT96] Y. Tan and L. van der Torre. How to combine ordering and minimizing in a deontic logic based on preferences. In *Deontic Logic, Agency and Normative Systems (ΔEON'96)*, Workshops in Computing, pages 216–232. Springer, 1996.

[vdT94] L. van der Torre. Violated obligations in a defeasible deontic logic. In A.Cohn, editor, *Proceedings of the Eleventh European Conference on Artificial Intelligence (ECAI'94)*, pages 371–375. Wiley, 1994.

[vdT97] L. van der Torre. *Reasoning about Obligations: Defeasibility in Preference-based Deontic Logic.* PhD thesis, Erasmus University Rotterdam, 1997.

[vdT99] L. van der Torre. Violation contexts and deontic independence. In *Modeling and Using Context. Proceedings of the CONTEXT'99*, Lecture Notes in Artificial Intelligence 1688, pages 361–374. Springer, 1999.

[vdT00] L. van der Torre. The logic of reusable propositional output with the fulfilment constraint. In D. Basin *et. al.*, editor, *Labelled Deduction*, pages 247–268. Kluwer, 2000.

[vdTT97] L. van der Torre and Y. Tan. The many faces of defeasibility in defeasible deontic logic. In D. Nute, editor, *Defeasible Deontic Logic*, volume 263 of *Synthese Library*, pages 79–121. Kluwer, 1997. Revised and extended version of L. van der Torre and Y. Tan. Cancelling and overshadowing: two types of defeasibility in defeasible deontic logic. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence (IJCAI'95)*, 1525–1532, 1995.

[vdTT98a] L. van der Torre and Y. Tan. The temporal analysis of Chisholm's paradox. In *Proceedings of 15th National Conference on Artificial Intelligence (AAAI'98)*, pages 650–655. AAAI Press / The MIT Press, 1998.

[vdTT98b] L. van der Torre and Y. Tan. An update semantics for prima facie obligations. In H. Prade, editor, *Proceedings of the Thirteenth European Conference on Artificial Intelligence (ECAI'98)*, pages 38–42. Wiley, 1998.

[vdTT99a] L. van der Torre and Y. Tan. Contrary-to-duty reasoning with preference-based dyadic obligations. *Annals of Mathematics and Artificial Intelligence*, 27:49–78, 1999.

[vdTT99b] L. van der Torre and Y. Tan. Rights, duties and commitments between agents. In T. Dean, editor, *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence (IJCAI'99)*, pages 1239–1244. Morgan Kaufmann, 1999.

[vdTT99c] L. van der Torre and Y. Tan. An update semantics for defeasible obligations. In K. Laskey and H. Prade, editors, *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence (UAI'99)*, pages 631–638. Morgan Kaufmann, 1999.

[vdTT99d] L. van der Torre and Y. Tan. An update semantics for deontic reasoning. In P. McNamara and H. Prakken, editors, *Norms, Logics and Information Systems. New Studies on Deontic Logic and Computer Science*, pages 73–90. IOS Press, 1999.

[vdTT00] L. van der Torre and Y. Tan. Contextual deontic logic: violation contexts and factual defeasibility. In P. Bonzon, M. Cavalcanti, and R. Nossum, editors, *Formal Aspects in Context*, pages 173–192. Kluwer, 2000.

[Vel96] F. Veltman. Defaults in update semantics. *Journal of Philosophical Logic*, 25:221–261, 1996.

[vW81] G.H. von Wright. On the logic of norms and actions. In R. Hilpinen, editor, *New Studies in Deontic Logic: Norms, Actions and the Foundations of Ethics*, pages 3–35. D.Reidel, 1981.

[vW99] G.H. von Wright. Deontic logic - as I see it. In P. McNamara and H. Prakken, editors, *Norms, Logics and Information Systems. New Studies on Deontic Logic and Computer Science*, pages 15–25. IOS Press, 1999.

[Wey98] E. Weydert. System JZ: How to build a canonical ranking model of a default knowledge base. In *Proceedings of the Seventh International conference on Knowledge Representation and Reasoning (KR'98)*, pages 190–201, 1998.

[WM93] R.J. Wieringa and J.-J.Ch. Meyer. Applications of deontic logic in computer science: A concise overview. In J.-J. Meyer and R. Wieringa, editors, *Deontic Logic in Computer Science*, pages 17–40. John Wiley & Sons, Chichester, 1993.