An Update Semantics for Deontic Reasoning (extended paper)*

L.W.N. van der Torre
Max-Planck-Institute for Computer Science
Im StadtWald,
D-66123 Saarbrücken, Germany
torre@mpi-sb.mpg.de

Y.-H. Tan
EURIDIS, Erasmus University Rotterdam
P.O. Box 1738,
3000 DR Rotterdam, The Netherlands
ytan@euridis.fbk.eur.nl

Abstract

In this paper we propose the deontic logic DUS, that formalizes reasoning about prescriptive obligations in update semantics. In DUS the definition of logical validity of obligations is not based on truth values but on action dynamics. You know the meaning of a normative sentence if you know the change it brings about in the betterness relation of anyone who is subjected to the news conveyed by it.

1 The logic of norms

One of the first topics discussed in the development of deontic logic was the question whether norms have *truth values*. For example, von Wright 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" [vW81]. Alchourrón and Bulygin discussed the *possibility* of a logic of norms, which they distinguish from the logic of normative propositions.

> "One such issue is the problem of the possibility of a logic of norms. Some authors think that there are logical relations between the norms, and so favor the development of a specific logic of norms (sometimes called 'deontic logic', though 'normative logic' would perhaps be a more appropriate name). Other authors deny the very possibility of such a logic because in their view there are no logical relations between norms. According to them deontic logic can only assume the form of a logic of normative propositions, i.e. (true or false) propositions about (the existence of) norms." [AB81]

*Section 1 to 4 appeared as [vdTT98d].

The distinction between norms and normative propositions is that the former are prescriptive whereas the latter are descriptive. 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]

Alchourrón and Bulygin distinguish between statements that describe a normative system, and statements that prescribe a certain behavior or state of affairs. In the first sense, 'it is obligatory to keep right on the streets' is a description of the fact that a certain normative system (e.g. Dutch traffic law) contains an obligation to keep right on the streets. In the second sense this statement is the obligation of Dutch traffic law itself. Alchourrón and Bulygin continue to develop an inference relation conditional to a normative system NS, written as $\vdash_{\mathrm{NS}}$. The normative system NS consists of a set of formulas, representing the set of norms, and the deontic formulae are re-interpreted as propositions about norms. In Alchourrón and Bulygin's proposal, the normative system is represented by a set of propositional sentences, representing the set of (explicitly) promulgated obligations, and something similar for permissions. In this paper we only consider obligations, and leave permissions for further research. Unlike Alchourrón and Bulygin, we do not introduce a new logic $\vdash_{\mathrm{NS}}$ but we introduce a logic that formalizes reasoning about the normative system NS. Two questions immediately arise: what does the normative system look like and what kind of derivations are possible when reasoning about normative systems.

What does the normative system look like? There are several options. We can represent the normative system by a set of propositional sentences (like Alchourrón and Bulygin), a set of prescriptive obligations, or more complex structures. For example, in [vdTT97a] we formalized the normative system by a tuple $\langle \mathrm{NORMS}, \mathrm{ND} \rangle$ where NORMS is a set of objects representing the norms, and ND a set of first-order sentences describing the norms. A typical norm description of the norm n: 'α ought to be (done) if β is (done)' is the formula $(V(n) \leftrightarrow (\beta \wedge \neg\alpha)) \wedge (F(n) \leftrightarrow (\beta \wedge \alpha))$, that expresses that the norm n is violated when $\beta \wedge \neg\alpha$ is true, and fulfilled when $\beta \wedge \alpha$ is true. Another example of a norm description is $A(n) \leftrightarrow \beta$, that expresses that the norm n is applicable when β is true. This formalization of the normative system clearly makes a distinction between norms, which do *not* have truth values, and descriptions about norms, which do. In this paper we do not consider such complex structures and consider normative systems that consist of a set of prescriptive obligations.

What kind of derivations are made when we reason about NS**?** We first give a simple example of such a derivation. Assume that a prescriptive obligation that 'α ought to be (done)' is represented by $\mathsf{oblige}\ \alpha$, and that an inference relation $\vdash_{\mathrm{NS}}$ of normative system NS$= \{\mathsf{oblige}\ p, \mathsf{oblige}\ q\}$ is equivalent to $\vdash_{\mathrm{NS}'}$ of NS$' = \{\mathsf{oblige}\ p, \mathsf{oblige}\ q, \mathsf{oblige}(p \wedge q)\}$. In other words, we have $\phi \vdash_{\mathrm{NS}} \psi$ if and only if $\phi \vdash_{\mathrm{NS}'} \psi$, where ϕ and ψ are sentences of a (here unspecified) descriptive deontic logic. In that case, we say that the prescriptive obligation $\mathsf{oblige}(p \wedge q)$ is *implied* by the normative system NS. Adding this new prescriptive obligation to the two prescriptive obligation already in NS does not influence the description of the normative system. In general, a set of prescriptive obligations – a normative system NS – implies an obligation if adding this latter obligation to NS does not change the inference relation $\vdash_{\mathrm{NS}}$.

The distinction between the logic of norms and the logic of normative propositions can be studied from a semantic and a syntactic-axiomatic point of view. The first distinction is the most obvious, because norms do not have truth values whereas normative propositions do.

Semantic presentation. Alchourrón and Bulygin argue that the logic of prescriptive obligations and the logic of descriptive obligations are different from a semantic point of view. Alchourrón [Alc93] explains that when it is the case that an authority has commanded (authorized) an action, he usually performs his normative action (of enacting a norm) by means of an utterance of a deontic sentence (or of a context-equisignificant form, e.g. an imperative). But it should also be clear that in such situations deontic sentences do not stand for the assertions concerning acts of command (authorization) of an authority. The authority has used the deontic sentence not to assert anything at all, but to command (authorize) the action. So when, for example, the Dutch traffic authority commands drivers to drive on the right by using the sentence 'Drivers should drive on the right' it will be a complete misunderstanding if someone replied: 'no, you are wrong, no authority has commanded so' which would be a quite reasonable reply (it does not matter whether true or false) had the deontic sentence be used to assert (the normative proposition) that some authority or other has delivered such a command. As a consequence, the logic of prescriptive obligations does not have a truth functional semantics – we have to look for another semantics. A clue for the semantics of norms is that prescriptive obligations are interpreted *dynamically* whereas descriptive obligations are interpreted *statically*. This follows directly from Alchourrón's box metaphor, in which norms are characterized as a kind of actions. Prescriptive obligations must be formalized by a dynamic semantics. The semantic distinction between normative propositions (truth values) and norms (performance values of actions?) has consequences for the deontic language, in particular for the kind of operators it can contain and whether it can contain nested operators. The operators $\wedge$, $\vee$ and $\neg$ can be applied to normative propositions and the operators of dynamic logic can be applied to norms as actions. Moreover, normative propositions can be nested if the factual sentences are propositions, and norms as actions can be nested if the factual sentences are action descriptions. For example, the formula $\mathsf{oblige}\ \mathsf{oblige}\ p$ might formalize the action of obliging someone to oblige someone else to do p. In this paper we do not consider complex combinations of norms.

Syntactic-axiomatic presentation. Alchourrón and Bulygin asked themselves the question whether the logic of prescriptive obligations and the logic of descriptive obligations are different from a syntactic-axiomatic point of view. They argue that there is an important difference related to the consistency of dilemmas. Assume that the inference relations of the two logics are represented by $\Vdash$ and $\vdash$ respectively, the two types of obligations are represented by $\mathsf{oblige}\ \alpha$ and $O\alpha$, and that we have the symbol $\perp$ in the languages to express inconsistency. They argue that we have $\mathsf{oblige}\ p, \mathsf{oblige}\ \neg p \Vdash \perp$ and $Op, O\neg p \not\vdash \perp$. The former is valid because the logic of norms formulates rationality constraints for a normgiver's intentions, and part of his intention is to leave open the possibility of the joint fulfillments of obligations that stem from this command. The latter is invalid, because it is possible to give a consistent description of an inconsistent system. In other words, it is possible (consistent) to norm inconsistently, as it is possible to postulate, assert, believe, etc, inconsistently. Besides this distinction, Alchourrón and Bulygin mention that the former is not relative to an authority whereas the latter is, and that the former only has one kind of permission whereas the latter distinguishes between two non-equivalent kinds of permission. In this paper we introduce a logic of prescriptive obligations that has the property $\mathsf{oblige}\ p, \mathsf{oblige}\ \neg p \Vdash \perp$ as desired by Alchourrón and Bulygin, but it should be observed that the system can easily be adapted such that this property is blocked.

2 Obligations as actions

In this paper we introduce an update semantics to formalize prescriptive obligations. This update semantics is based on the update semantics of Veltman [Vel96]. We call the resulting system DUS for 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 this paper, we propose the following definition of meaning for normative statements. The information states of update semantics are instantiations of betterness relations, i.e. preference orderings reflecting different degrees of ideality [Han71, Lew74, Jac85, Gob90b, Han90, BMW93, TvdT96, vdTT97c, Han97].

> You know the meaning of a normative sentence if you know the change it brings about in the betterness relation of anyone who is subjected to the news conveyed by it.

This dynamic interpretation of prescriptive obligations is related to Alchourrón's box metaphor. Alchourrón compares a (monadic) prescriptive obligation with the *action* of putting something in a box. In our proposal a (dyadic) prescriptive obligation is compared with an *action* of creating an ordering. Updating betterness relations is based on the following two principles.

1. In the initial state each world is connected to all other worlds.

2. Adding obligations is deleting preference relations between worlds.

A betterness relation $\leq$ can be represented by a possible worlds model, i.e. a set of worlds with a binary relation and a valuation function. We propose to formalize the update of a betterness relation by the obligation 'α ought to be (done) if β is (done)' with the action of deleting the betterness relations between all worlds w_1 and w_2 (i.e. $w_1 \leq w_2$) 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 be

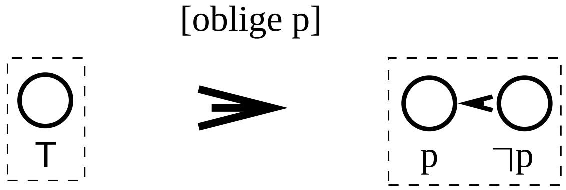

Figure 1: Updating the initial state with 'p ought to be (done)'

read as follows. A circle is a set of worlds that satisfy the propositions written within the circle and that are connected to at least all other worlds of the circle. A box represents a betterness relation between worlds. A single headed arrow from one circle to another represents that the worlds of the first circle are better than the worlds of the second circle, but not vice versa. The symbol $\top$ stands for any tautology, for example $p \vee \neg p$. The left box represents the initial state, the betterness relation in which each world is connected to all other worlds. The right box represents the betterness 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 betterness relation where w_1 is a $\neg p$ world and w_2 is a p world, we end up with a betterness relation in which each p world is preferred to all the $\neg p$ worlds.

Besides prescriptive obligations we also formalize facts in our deontic logic DUS, because we want to represent the actual as well as the ideal. This is needed to formalize violations of obligations and also to give an adequate analysis of the Chisholm and Forrester paradoxes. The static interpretation of facts is that they determine a subrelation of the betterness relation, namely the subrelation that consists of all the worlds that satisfy the facts. The dynamic interpretation of facts is that the update with a fact is zooming in on the ordering. For these interpretations we make the distinction between the context of justification and the context of deliberation explicit in DUS. The essential contrast between justification and deliberation is a difference in what we take as being settled [Tho81]. In this paper we assume that propositional formulas are facts that are settled, and these facts determine the context of deliberation. 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. To distinguish between the context of justification and the context of deliberation we write oblige* for the latter. An example of an update with a fact is represented in Figure 2. This figure should be read as follows. A dashed box

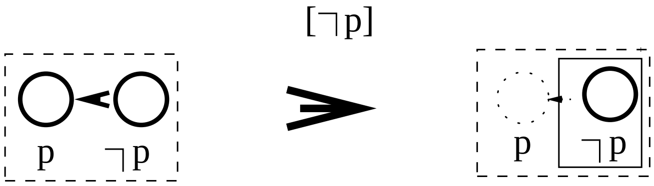

Figure 2: Updating the state 'p ought to be (done)' with 'p is not (done)'

determines the context of justification and the non-dashed box determines the context of deliberation. The left box represents the betterness 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 betterness 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 circle. The only ideal worlds are the leftmost worlds, and the right dashed box therefore indicates a violation.

3 Obligations in update semantics

In this section we define prescriptive obligations in update semantics, and in the following section we give several examples. The obligations do not allow for exceptions, i.e. they are absolute and not prima facie obligations, and they make dilemmas like $Op \wedge O\neg p$ inconsistent.

3.1 Update semantics

We start with the basic definitions of Veltman's update semantics [Vel96]. To define an update semantics for a language L, one has to specify a set Σ of relevant information states, 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 '$\sigma[\psi_1]\ldots[\psi_n]$' for the result of updating σ with the sequence of sentences $\psi_1, \ldots, \psi_n$. Moreover, one of the information states has to be labeled as the minimal information state, written as $\mathbf{0}$, and another one as the absurd state, written as $\mathbf{1}$.

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

We discriminate between successful and unsuccessful updates. An update is successful if it does not result in the absurd state, and unsuccessful otherwise.

Definition 2 (Success) *Let σ be an information state and ϕ a formula of the logical language L. The update $\sigma[\phi]$ is successful if and only if $\sigma[\phi] \neq \mathbf{1}$.*

A crucial notion of update systems is acceptance. The formula ϕ is accepted in an information 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. If an update is accepted, then the information state usually has a specific content. We give some examples of that behavior in Section 3.3 for our obligations in update semantics. Acceptance is the counterpart of satisfaction in standard semantics.

Definition 3 (Acceptance) *Let σ be an information state and ϕ a formula of the logical language L. $\sigma \Vdash \phi$ if and only if $\sigma[\phi] = \sigma$.*

Different notions of validity can be based on the notion of acceptance (see Veltman's paper [Vel96] for an overview). We are interested in the following one. An argument is valid if updating the minimal state $\mathbf{0}$ with the premises $\psi_1, \ldots, \psi_n$, in that order, yields an information state in which the conclusion is accepted.

Definition 4 (Validity) *Let $\psi_1, \ldots, \psi_n$ and ϕ be formulas of the logical language L. $\psi_1, \ldots, \psi_n \Vdash \phi$ if and only if $\mathbf{0}[\psi_1]\ldots[\psi_n] \Vdash \phi$.*

3.2 Prescriptive obligations in update semantics

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. In this section we define prescriptive obligations in update semantics using Veltman's basic definitions. To define obligations in update semantics we have to define the deontic language, the deontic information states and the deontic updates. The deontic language is a propositional language with the dyadic operators $\mathsf{oblige}(\alpha|\beta)$ and $\mathsf{oblige}^*(\alpha|\beta)$. They refer to the obligations as reducing betterness relations, respectively in the context of justification and the context of deliberation.

Definition 5 (Deontic language) *Let A be a set of atoms and L_0^A a propositional language with A as its non-logical symbols. 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 $\phi = \mathsf{oblige}(\psi_1|\psi_2)$ or $\phi = \mathsf{oblige}^*(\psi_1|\psi_2)$. We write $\mathsf{oblige}\ \psi$ for $\mathsf{oblige}(\psi|\top)$ and $\mathsf{oblige}^*\ \psi$ for $\mathsf{oblige}^*(\psi|\top)$.*

A deontic information state is a betterness relation which we represent by a possible worlds model $\langle W, \leq, V\rangle$, i.e. a binary relation $\leq$ on the worlds W (a subset of $W \times W$) with a valuation function V for propositions at the worlds. We use possible worlds models, but this notion of truth does not play a role in the update semantics. That is, we write $\sigma, w \models \phi$ to denote that world w in σ satisfies the propositional formula ϕ, but we leave $\sigma \models \phi$ undefined! We add the following three unusual features to these deontic states.

Explicit sub-relation. We extend the possible world model with a second betterness relation, which is a subrelation of the first one. The complete relation is used for the context of justification and the subrelation is used for the context of deliberation. Whereas in Kripke semantics a unique world is singled out, called the actual world, we single out a set of worlds, called the context of deliberation.

Background knowledge. We define an update system for a specific A, W and V. We call the formulas which are true in all worlds the *background* *knowledge* of the deontic system, and they represent the context of reasoning. Hence, our formal model distinguishes two contexts. First, the context of factual reasoning – the background knowledge – and second the context of deontic reasoning – the context of justification versus the context of deliberation. Unless explicitly mentioned, there is no background knowledge and the deontic state contains a world for each interpretation of L_0^A. The deontic state may be identified with the set of interpretations and a binary relation on it if the language is infinite. Otherwise, this assumption that the logical language is expressive enough to distinguish all worlds would imply that the set of worlds is finite, which is in general not a desirable property. In the following Definition 6 we say that a formula of L_0^A is consistent when it is consistent with respect to the background knowledge. An example of a deontic state that does not contain at least all interpretations of L_0^A is found in the gentle murderer's paradox in Section 6.1, where we have that gentle killing implies killing $g \rightarrow k$. In that case, we have that the deontic state contains all worlds that satisfy $g \rightarrow k$. We do not mention this any further in this paper but leave it implicitly understood.

Non-transitive relation. We assume that the binary relation is reflexive, but we do *not* assume that it is transitive or total. There is a technical problem related to the formalization of conditional obligations, which we discuss in Section 4.4. A consequence of this problem is that we cannot have transitivity for the relations in the deontic states. We take the transitive closure of this ordering only when we determine the preferred worlds. In Section 4.4 we show that a deontic state can be interpreted as a set of betterness relations, one for each factual sentence, instead of a unique betterness relation. However, this technical problem is not relevant for the intuitions of our deontic update system and the interpretation of most examples in this paper, and the deontic state can usually be identified with a single transitive betterness relation.

Definition 6 (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 consistent ϕ of L_0^A there is a $w \in W$ such that $\mathbf{0}, w \models \phi$. A deontic state is a tuple $\Sigma = \langle W, W^*, \leq, V \rangle$ consisting of the set of worlds W, a possibly empty subset $W^* \subseteq W$, a reflexive – but not necessarily transitive – binary relation $\leq$ on W and the valuation function V.*

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

$\mathbf{1}$, *the* absurd state, *is the state given by $\langle W, \emptyset, \emptyset, V \rangle$.*

The deontic updates are operations on the deontic states that either zoom in on the betterness relation (for facts), or reduce the betterness relation (for obligations). The prescriptive obligations have the dynamic component of creating a new deontic state with a new betterness relation. We first define the reduction of a betterness relation by an obligation.

Definition 7 (Reduction) *Let* $\sigma = \langle W, W^*, \leq, V\rangle$ *be a deontic state. The reduction of* σ *by* $\mathsf{oblige}(\alpha|\beta)$ *(*$\mathsf{oblige}^*(\alpha|\beta)$*), denoted by the symbol* $-$*, is defined as follows.*

$$\sigma - \mathsf{oblige}(\alpha|\beta) =$$
$$\langle W, W^*, \leq -\{w_1 \leq w_2 \mid \sigma, w_1 \models \neg\alpha \wedge \beta \textit{ and } \sigma, w_2 \models \alpha \wedge \beta \textit{ and } w_1, w_2 \in W\}, V\rangle$$
$$\sigma - \mathsf{oblige}^*(\alpha|\beta) =$$
$$\langle W, W^*, \leq -\{w_1 \leq w_2 \mid \sigma, w_1 \models \neg\alpha \wedge \beta \textit{ and } \sigma, w_2 \models \alpha \wedge \beta \textit{ and } w_1, w_2 \in W^*\}, V\rangle$$

Finally we deal with dilemmas.[1] The update $\sigma[\mathsf{oblige}(\alpha|\beta)]$ is the reduction of σ by $\mathsf{oblige}(\alpha|\beta)$ if afterwards the preferred β worlds are α worlds. Otherwise there is a dilemma and the result of the update is the absurd state. An example of a dilemma in our update semantics is given in Section 4.3. For this definition of deontic updates we need a test whether the preferred (or 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 example illustrates that we have to take the transitive closure of the ordering before we can minimize. Otherwise it is unclear what 'the minimal worlds' are.

Example 1 *Consider a Kripke model that consists of three worlds* w_1, w_2, w_3 *with* $w_1 \leq w_2$ *and* $w_2 \leq w_3$ *but not* $w_1 \leq w_3$*. Notice that the accessibility relation is not transitive. At first sight world* w_1 *seems to be the minimal world. However, it is not smaller than world* w_3*, unless we take the transitive closure.*

If the ordering is transitively closed and there are no infinite descending chains, then the minimality test can be defined as follows.

> The preferred β-worlds of W (W^*) of σ satisfy α if and only if for all worlds w_1 such that
>
> 1. $\sigma, w_1 \models \beta$, and
> 2. there is no w_2 such that $w_2 \leq w_1$, $w_1 \not\leq w_2$ and $\sigma, w_2 \models \beta$,

[1] We can distinguish between deontic logics with and without what we call the no-dilemma assumption. Deontic logics with the no-dilemma assumption make dilemmas like $O\alpha \wedge O\neg\alpha$, $O\alpha_1 \wedge O(\neg\alpha_1 \wedge \alpha_2)$ and $O(\alpha|\beta_1) \wedge O(\neg\alpha|\beta_1 \wedge \beta_2)$ inconsistent. They are the most standard. For example, in the Chellas classification the axiom $\neg(O\alpha \wedge O\neg\alpha)$ of Standard Deontic Logic is called the Deontic or D axiom. In this paper, we discuss the more interesting (and more complicated!) logics with the no-dilemma assumption, which is in accordance with Alchourrón's requirement for prescriptive obligations. The formalization of weaker deontic logics without the assumption is straightforward: simply delete the test after the reduction.

we have that $\sigma, w_1 \models \alpha$.

If there is no preferred β-world then the definition above is trivially true, even if there is an infinite descending chain of $\beta \wedge \neg\alpha$ worlds. The definition above can be generalized to include infinite descending chains as follows, see [Lew73].

> The preferred 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 $w_2 \leq w_1$ such that $\sigma, w_2 \models \beta$ and for all worlds $w_3 \leq w_2$ we have $\sigma, w_3 \models \beta \rightarrow \alpha$.

Definition 8 below combines the latter definition with taking the transitive closure. The transitive closure of an ordering can be calculated by adding relations $w_1 \leq w_3$ to the ordering for worlds $w_1, w_2, w_3 \in W$ such that $w_1 \leq w_2$ and $w_2 \leq w_3$. The fixed point of this iterative process is the smallest superset of the ordering such that for all worlds w_1, w_2, w_3 with $w_1 \leq w_2$ and $w_2 \leq w_3$ we have $w_1 \leq w_3$.

Definition 8 (pref) *Let $\sigma = \langle W, W^*, \leq, V\rangle$ be a deontic state and let $\leq_\beta$ be the transitive closure of $\leq$ in $\{w \in W \mid \sigma, w \models \beta\}$, i.e. the smallest superset of $\leq$ 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 preferred β-worlds of W of σ satisfy α, written as* $\mathsf{pref}(\sigma, \beta) = \alpha$, *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 $w_1, w_2, w_3 \in W$.* $\mathsf{pref}^*(\sigma, \beta) = \alpha$ *is defined analogously for worlds of W^*.*

Putting it all together we define the updates.

Definition 9 (Deontic updates) *Let $\sigma = \langle W, W^*, \leq, V\rangle$ be a deontic state. The update function $\sigma[\phi]$ is defined as follows.*

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

Having defined the deontic language, the deontic information states and the deontic updates we have defined the deontic update system. In the following section we consider some properties of the obligations in update semantics.

3.3 Properties of obligations in update semantics

The following proposition states the relation between the absurd state and inconsistency for our update system. If we update by a contradiction then we get the absurd state. Hence, the update with a contradiction is an example of an unsuccessful update. Moreover, the absurd state is the only state that accepts contradictions.

Proposition 1 (Contradiction) *Let σ be a deontic state and ϕ a formula of L_1^A. We have (1) $\mathbf{1} \Vdash \phi$ for all $\phi \in L_1^A$, (2) $\sigma[\bot] = \mathbf{1}$, and (3) $\sigma \Vdash \bot$ if and only if $\sigma = \mathbf{1}$.*

Proof *We prove the three items separately.*

1. *For $\phi \in L_0^A$, the only subset of $W^* = \emptyset$ is the empty set. Analogously, for $\phi = \mathsf{oblige}(\alpha|\beta)$, the only subset of $\leq = \emptyset$ is again the empty set.*
2. *There is no $w \in W^*$ such that $\sigma, w \models \bot$.*
3. *$\Rightarrow \sigma = \sigma[\bot] = \mathbf{1}$, see 2. $\Leftarrow \mathbf{1} \Vdash \bot$, see 1.*

The following theorem states that in our update system acceptance corresponds to some simple tests on the deontic states.

Theorem 1 (Acceptance) *Let σ be a deontic state. We have $\sigma \Vdash \phi$ if and only if*

- *if ϕ is a sentence of L_0^A, then $\forall w \in W^*$ we have $\sigma, w \models \phi$;*
- *if $\phi = \mathsf{oblige}(\alpha|\beta)$, then for all worlds $w_1, w_2 \in W$ such that $\sigma, w_1 \models \neg\alpha \wedge \beta$ and $\sigma, w_2 \models \alpha \wedge \beta$ we have $w_1 \not\leq w_2$, and $\mathsf{pref}(\sigma, \beta) = \alpha$;*
- *if $\phi = \mathsf{oblige}^*(\alpha|\beta)$, then for all worlds $w_1, w_2 \in W^*$ s. t. $\sigma, w_1 \models \neg\alpha \wedge \beta$ and $\sigma, w_2 \models \alpha \wedge \beta$ we have $w_1 \not\leq w_2$, and $\mathsf{pref}^*(\sigma, \beta) = \alpha$.*

Proof *From Proposition 1 follows the proof for the absurd state, because in the absurd state everything is accepted. Moreover,*

- *if ϕ is a sentence of L_0^A, then $\sigma[\phi] = \langle W, \{w \in W^* \mid \sigma, w \models \phi\}, \leq, V\rangle = \sigma = \langle W, W^*, \leq, V\rangle$, i.e. $\{w \in W^* \mid \sigma, w \models \phi\} = W^*$.*
- *if ϕ is a sentence $\mathsf{oblige}(\alpha \mid \beta)$, then $\sigma[\phi] = \sigma - \mathsf{oblige}(\alpha \mid \beta) = \sigma$, and $\mathsf{pref}(\sigma - \mathsf{oblige}(\alpha \mid \beta), \beta) = \mathsf{pref}(\sigma, \beta) = \alpha$. The first part is equivalent to: $\leq -\{w_1 \leq w_2 \mid \sigma, w_1 \models \neg\alpha \wedge \beta$ and $\sigma, w_2 \models \alpha \wedge \beta$ and $w_1, w_2 \in W\} = \leq$.*

The proof for $\phi = \mathsf{oblige}^*(\alpha|\beta)$ is analogous to the proof for $\phi = \mathsf{oblige}(\alpha|\beta)$.

The following corollary follows from Theorem 1.

Corollary 1 *For propositional ϕ_1, ϕ_2 we have that $\sigma \Vdash \phi_1$ and $\sigma \Vdash \phi_2$ imply $\sigma \Vdash \phi_1 \wedge \phi_2$, but $\sigma \nVdash \phi_1$ does not imply $\sigma \Vdash \neg\phi_1$.*

In the following section we explain the basic properties of the validity relation by several examples. In this paper we do not give a complete axiomatization of the logic DUS. However, we can prove[2] the following relation between Prohairetic Deontic Logic PDL [vdTT97c] and DUS. Assume a *consistent* set of obligations $S = \{O(\alpha_1|\beta_1), \ldots, O(\alpha_n|\beta_n)\}$ of PDL. We have $S \vdash_{PDL} O(\alpha|\beta)$ if and only if for any list T that contains exactly $\mathsf{oblige}(\alpha_1|\beta_1), \ldots, \mathsf{oblige}(\alpha_n|\beta_n)$ we have $T \Vdash \mathsf{oblige}(\alpha|\beta)$. Hence, the properties of the prescriptive obligations of DUS correspond to the properties of descriptive obligations of PDL, and PDL can be used as a basis to axiomatize DUS. PDL does not take Alchourrón's distinction between the logic of descriptive and descriptive obligations into account, because dilemmas of descriptive obligations are not consistent. However, it is an easy exercise to adapt the logic PDL such that dilemmas become consistent.

4 Examples

The deontic update semantics is illustrated by four examples.

4.1 Violations

The following example illustrates that the order of premises is important.

Example 2 *Consider the obligation* oblige p *and the fact* $\neg p$*, as represented in Figure 1 and 2. The fact* $\neg p$ *represents that the obligation* oblige p *has been violated. It can easily be verified that we have the following.*

$\mathsf{oblige}\ p, \neg p \nVdash \bot$
$\mathsf{oblige}^*\ p, \neg p \nVdash \bot$
$\neg p, \mathsf{oblige}\ p \nVdash \bot$
$\neg p, \mathsf{oblige}^*\ p \Vdash \bot$

A contradiction $\bot$ is derivable from $\neg p$ and $\mathsf{oblige}^*\ p$, $\mathbf{0}[\neg p][\mathsf{oblige}^*\ p] = \mathbf{1}$, *because* $\mathsf{pref}^*(\mathbf{0}[\neg p] - \mathsf{oblige}^*\ p, \top) \neq p$. *If the context of deliberation only consists of* $\neg p$ *worlds, then it makes no longer sense to oblige someone to do* p.

[2] For this proof, we have to distinguish between the fact that the initial state is the state in which each world is connected to each other world, and the fact that updating by an obligation is formalized by reducing the betterness relation. The latter part is related to PDL, because the acceptance condition of DUS is equivalent to the truth condition of PDL. The former condition is formalized by the notion of preferential entailment of PDL, based on maximally connected models. Due to lack of space we cannot give the proof in this paper.

In Example 6 we show that the relation is sensitive to the order of the list of obligations: the premises oblige ϕ_1, oblige ϕ_2 may derive different consequences than oblige ϕ_2, oblige ϕ_1. This is quite surprising given the result mentioned at the end of the previous section, that a list of DUS obligations has the same properties as a *set* of PDL obligations. The difference is that a set of obligations that is inconsistent in PDL may be consistent in DUS.

4.2 Three obligations

The following example illustrates unconditional obligations in the deontic update system. In particular, it illustrates that obligations are not deductively closed, but they satisfy the conjunction and disjunction rule. Moreover, it illustrates the distinction between the context of justification and the context of deliberation.

Example 3 (Three unrelated obligations) *Consider the deontic states in Figure 3, which we write as* $\sigma_1, \ldots, \sigma_6$ *respectively. This figure should be read*

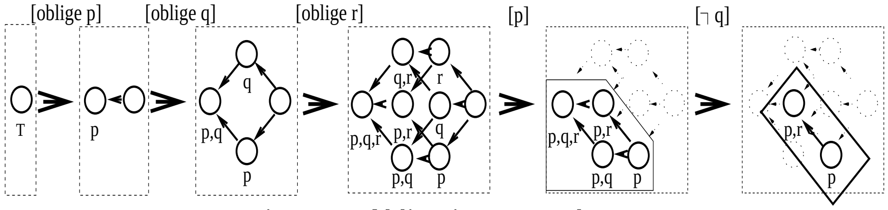

Figure 3: Obligations as updates

as follows. The six dashed boxes are the deontic information states, i.e. the betterness relations. They contain all worlds, and they determine the context of justification. The circles within the boxes are sets of worlds. For each set of worlds, we only write the positive atoms below. The non-dashed boxes contain only the worlds of W^**, i.e. the worlds of the context of deliberation. The arrows represent strict preferences between the worlds represented by the circles. We leave the transitive closure implicit. The figure also represents five updates: three updates by obligations* oblige p, oblige q *and* oblige r*, and two updates by facts* p *and* $\neg q$*. We have* $\sigma_1 \not\Vdash$ oblige p, $\sigma_2 \Vdash$ oblige p, *and* $\sigma_2 \not\Vdash$ oblige$(p \vee q)$. *The latter shows that* oblige *is not closed deductively, i.e. we have (see Definition 4)*

oblige $p \not\Vdash$ oblige$(p \vee q)$

Moreover, we have $\sigma_3 \Vdash$ oblige$(p \wedge q)$ *and* $\sigma_3 \Vdash$ oblige$(p \vee q)$, *which shows that the obligations have the conjunction and disjunction rule.*

oblige p, oblige $q \Vdash$ oblige$(p \wedge q)$

$\mathsf{oblige}\ p, \mathsf{oblige}\ q \Vdash \mathsf{oblige}(p \vee q)$

The figure also shows the distinction between the context of deliberation and the context of justification. The latter two updates illustrate that the context of deliberation shrinks when we add facts, whereas the context of justification remains the same. In the latter deontic state σ_6 we have that the ideal worlds in the context of justification are $p \wedge q \wedge r$ worlds, whereas the ideal worlds in the context of justification are exactly the $p \wedge \neg q \wedge r$ worlds.

$\mathsf{oblige}\ p, \mathsf{oblige}\ q, \mathsf{oblige}\ r, p, \neg q \Vdash \mathsf{oblige}(p \wedge q \wedge r)$
$\mathsf{oblige}\ p, \mathsf{oblige}\ q, \mathsf{oblige}\ r, p, \neg q \Vdash \mathsf{oblige}^*(p \wedge \neg q \wedge r)$

A disadvantage of the representation of the three obligations in Example 3 is that we did not represent any temporal information in the three obligations. For example, we did not specify which obligation has to be fulfilled first. The following example illustrates how temporal information can be represented by conditional obligations in update semantics.

Example 4 (Three unrelated obligations, continued) *Consider the three conditional obligations* $\mathsf{oblige}\ p$*,* $\mathsf{oblige}(q|p)$ *and* $\mathsf{oblige}(r|p \wedge q)$*. Figure 4 represents the updates of the three obligations and of the facts that* p *is done and that* q *is not done. The crucial distinction between this example and the previous one is that there are no obligations left once an obligation has been violated.*

$\mathsf{oblige}\ p, \mathsf{oblige}\ q, \mathsf{oblige}\ r, p, \neg q \Vdash \mathsf{oblige}^*\ r$
$\mathsf{oblige}\ p, \mathsf{oblige}(q|p), \mathsf{oblige}(r|p \wedge q), p, \neg q \not\Vdash \mathsf{oblige}^*\ r$

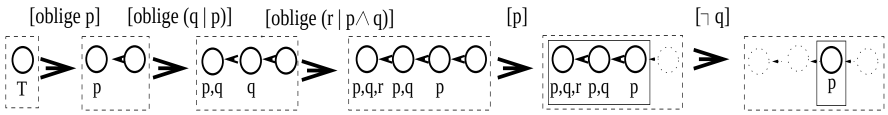

Figure 4: Conditional obligations as updates

Example 3 and 4 illustrate that the number of equivalence classes is a measure of the number of obligations, because the number of equivalence classes increases when we update by obligations. When we compare Figure 3 and 4 we observe that with conditional obligations in Example 4 the set of equivalence classes only increases linearly, whereas with unconditional obligations in Example 3 it increases exponentionally. Obviously this attractive feature of conditional obligations is not true in general (if only because absolute obligations are a special type of conditional obligations).

4.3 Dilemmas

To deal with dilemmas the distinction between successful and unsuccessful updates is used. The following example illustrates that if we update by a dilemma, then we end up in the absurd state. This is the update semantics variant of making a dilemma inconsistent (see Proposition 1 for the formal relation between the absurd state and inconsistency).

Example 5 (Dilemma) *Consider the dilemma 'p ought to be (done)' and 'p ought not to be (done).' The update of the initial state with* $\mathsf{oblige}\ p$ *and thereafter* $\mathsf{oblige}\ \neg p$ *results in the state* $\mathbf{0}[\mathsf{oblige}\ p][\mathsf{oblige}\ \neg p]$ *in which all connections between p and ¬p worlds are deleted. The p and ¬p worlds have become incomparable, see Figure 5 below. The test* $\mathsf{pref}(\mathbf{0}[\mathsf{oblige}\ p] - \mathsf{oblige}\ \neg p, \top) = \neg p$ *checks if all the ideal worlds are ¬p worlds. This is not the case, thus we end up in the absurd state,* $\mathbf{0}[\mathsf{oblige}\ p][\mathsf{oblige}\ \neg p] = \mathbf{1}$*, and we have the following.*

$$\mathsf{oblige}\ p, \mathsf{oblige}\ \neg p \Vdash \bot$$

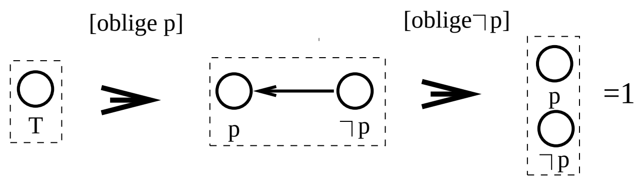

Figure 5: Dilemma

A more complex dilemma is given by 'c ought not to be (done)' and 'c ought to be done if k is (done),' which is Prakken and Sergot's considerate assassin example [PS96] when we read c as 'offering someone (the victim) a cigarette' and k as 'killing someone' (and we assume that it is possible to kill without offering cigarettes). They argue that this is a true dilemma which should be inconsistent.

> "Is this acceptable? In our opinion it is: what is crucial is that $O(c|k)$ is not a contrary-to-duty rule of $O(\neg c|\top)$ but of $O(\neg k|\top)$, for which reason $O(c \mid k)$ and $O(\neg c|\top)$ are unrelated obligations. Now one may ask how this conflict should be resolved and, of course, one plausible option is to regard $O(c \mid k)$ as an exception to $O(\neg c \mid \top)$ and to formalize this with a suitable non-monotonic defeat mechanism. However, it is important to note that this is a separate issue, which has nothing to do with the contrary-to-duty aspects of the example." [PS96]

The following example illustrates that DUS sometimes makes a dilemma inconsistent, dependent on the order of the premises.

Example 6 (Considerate assassin) *If we update the initial state with* $\mathsf{oblige}\ \neg c$ *and thereafter with* $\mathsf{oblige}(c|k)$*, then we end up with the absurd state. However, if we first update the initial state with* $\mathsf{oblige}(c|k)$ *and thereafter with* $\mathsf{oblige}\ \neg c$*,' then we do not end up with the absurd state.*

$\mathbf{0}[\mathsf{oblige}\ \neg c][\mathsf{oblige}(c|k)] \Vdash \bot$
$\mathbf{0}[\mathsf{oblige}(c|k)][\mathsf{oblige}\ \neg c] \not\Vdash \bot$

The latter result is quite interesting. If we first give a general obligation and then a more specific one, then it is considered to be irrational. However, if we first give the more specific one and then the more general one, then there is no problem. At first sight, this seems counterintuitive. However, the semantics give a natural temporal explanation of this phenomenon. To block this (a bit exotic) behavior we have to change the test after the update. In that case it is not sufficient to test $\mathsf{pref}(\sigma, \beta) = \alpha$ for the last update with $\mathsf{oblige}(\alpha|\beta)$, but we have to test all previous updates too. The technical problem is how to know the previous updates. For example, we might want to have

$\mathbf{0}[\mathsf{oblige}\ p][\mathsf{oblige}\ q] \neq \mathbf{1}$
$\mathbf{0}[\mathsf{oblige}\ p][\mathsf{oblige}(\neg q|(p \wedge \neg q) \vee (\neg p \wedge q))][\mathsf{oblige}\ q] = \mathbf{1}$

But this behavior cannot be modeled in the present set-up, because we have $\sigma[\mathsf{oblige}\ p] = \sigma[\mathsf{oblige}\ p][\mathsf{oblige}(\neg q \,|\, (p \wedge \neg q) \vee (\neg p \wedge q))]$. A simple solution is to record all previous updates in the deontic information state, but that is not very elegant. Another solution is to change the definition of our validity relation to the following one (not given in [Vel96]). An argument is valid if updating the minimal state $\mathbf{0}$ with the premises $\psi_1, \ldots, \psi_n$ *in some order* yields an information state in which the conclusion is accepted.

Definition 10 (Validity, continued) *Let* $\psi_1, \ldots, \psi_n$ *and* ϕ *be sentences of the deontic language* L_1^A*.* $\psi_1, \ldots, \psi_n \Vdash^\pi \phi$ *if and only if there is a permutation* π *of* $1 \ldots n$ *such that* $\mathbf{0}[\psi_{\pi(1)}] \ldots [\psi_{\pi(n)}] \Vdash \phi$*.*

However, whereas it may be philosophically satisfactory this solution is computationally not very attractive. We leave it an open issue to find a satisfactory solution.

4.4 Transitivity

In the examples thus far, all betterness relations were transitive, and we left the transitive closure implicit in the figures. The following example illustrates that this does not have to be the case.

Example 7 (Transitivity) *Consider the update of the initial state with the obligation* $[\mathsf{oblige}(p \,|\, q)]$ *illustrated in Figure 6. This figure should be read as*

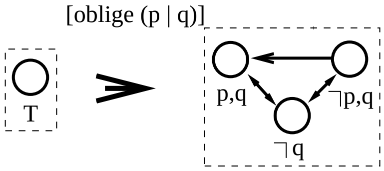

Figure 6: Lack of transitivity

follows. A single-headed arrow from a circle to another one represents that each world of the former is connected to all worlds of the latter but not vice versa (as before), and a double-headed arrow between two circles represents that each world is connected to all worlds of both circles. The reflexive closure is left implicit. In contrast to the previous figures, in this figure the transitive closure is not implicitly assumed! The upper two circles of the right deontic state represent the $p \wedge q$ *and* $\neg p \wedge q$ *worlds respectively. The former are strictly preferred to the latter. The lower circle represents the* $\neg q$ *worlds. The* $\neg q$ *worlds are connected to the* $p \wedge q$ *and the* $\neg p \wedge q$ *worlds, and vice versa. We have for a* $\neg p \wedge q$ *world* w_1*, a* $\neg q$ *world* w_2 *and a* $p \wedge q$ *world* w_3 *that* $w_1 \leq w_2$*,* $w_2 \leq w_3$ *but not* $w_1 \leq w_3$*. Hence, the relation is not transitive.*

The previous example illustrates that the update with a conditional obligation may lead to a relation that is not transitive. As we already remarked in Section 3.2, there is a technical problem with conditional obligations. We cannot use transitive closures, because if we take the transitive closure of the relation in Figure 6, then we get the initial state. Moreover, we cannot use a subset of $\leq$ which is transitively closed, because such subsets are not unique. A consequence of this problem is that we cannot have transitivity for the relations in the deontic states. We take the transitive closure of this relation only when we determine the preferred worlds. The following example illustrates that the deontic state is not a single unique betterness relation, but can be best interpreted as a *set* of betterness relations.

Example 8 (Transitivity, continued) *Reconsider Example 7, where we update the initial state with* $\mathsf{oblige}(p|q)$*. Consider the test* $\mathsf{pref}(\mathbf{0}[\mathsf{oblige}(p|q)], q) = p$ *represented in the left part of Figure 7. The vertical representation in this figure represents the steps that are to be taken for the semantic evaluation of* pref*. We zoom in on* q *worlds, and take the transitive closure of the* q *worlds. The resulting betterness relation prefers the* q *worlds to the* $\neg q$ *worlds. Now consider the test of* $\mathsf{pref}(\mathbf{0}[\mathsf{oblige}(p|q)], q \vee r) = p$ *also represented in Figure 7. We zoom in on the* $q \vee r$ *worlds, and take the transitive closure. By taking the transitive closure, all* $q \vee r$ *worlds become equivalent. The* q *worlds are no longer preferred to the* $\neg q$ *worlds, because by taking the transitive closure they became equivalent. The deontic state implicitly represents a betterness relation in which the* q *worlds are strictly preferred to the* $\neg q$ *worlds, and a betterness relation in which all* q

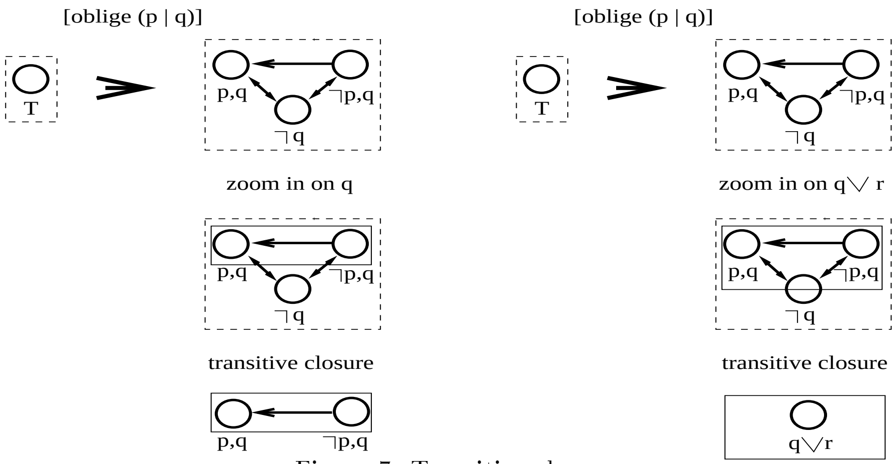

Figure 7: Transitive closures

worlds are equivalent.

The previous example illustrated how a preference ordering is constructed from a deontic state. First we select the worlds in which the antecedent is true (we zoom in on the relation) and then we take the transitive closure of the relation. We construct the preference ordering – and implicitly we work with a set of orderings – because we want to refer to the most preferred (the ideal or optimal) worlds, which were used to check if there is a dilemma.[3] In the following section, we define new operators that also refer to the preferred worlds, and that are used for two different purposes. There are two alternative ways to refer to the preferred worlds. First, we can interpret the links between worlds as one step of a path, and consider whether there is a path from one world to another. In that case, we have to give an interpretation to loops (see Section 6.3) and we have to deal with infinite chains. Second, we can explicitly represent the deontic information state as a set of orderings, instead of as a single relation. This is considered in Section 9 when we compare our deontic update semantics with Veltman's update semantics for default reasoning.

5 Tests

We distinguish between prescriptive obligations and tests. The tests evaluate what the norms are in a particular deontic state depending on its betterness

[3]Another reason to work with a set of orderings instead of a single non-transitive relation is that orderings are well studied objects, as well as from a technical point (what are their properties) as a philosophical point (what they mean). In fact, for this reason it may even be preferred to work with a set of set of connected orderings instead of a set of partial orderings.

relation. In this paper we write $\mathsf{oblige}(\alpha \,|\, \beta)$ for the prescriptive obligation 'α ought to be (done), if β is (done)' and we write $\mathsf{ideal}(\alpha|\beta)$ for the test 'ideally, α is (done), if β is (done).' Notice that the test operator is not a descriptive obligation, because it does not have a truth value (see below). However, the prescriptive obligations have the dynamic component of creating a new deontic state with a new betterness relation, whereas the tests evaluate what the norms are in a particular deontic state depending on its betterness relation. The deontic language L_1^A is extended with the two dyadic operators $\mathsf{ideal}(\alpha|\beta)$ and $\mathsf{ideal}^*(\alpha|\beta)$.

Definition 11 (Deontic language, continued) *Let A, L_0^A and L_1^A be as defined in Definition 5. A string of symbols ϕ is a sentence of L_2^A if and only if either ϕ is a sentence of L_1^A or there are two sentences ψ_1, ψ_2 of L_0^A such that $\phi = \mathsf{ideal}(\psi_1 \,|\, \psi_2)$ or $\phi = \mathsf{ideal}^*(\psi_1 \,|\, \psi_2)$. We write $\mathsf{ideal}\ \psi$ for $\mathsf{ideal}(\psi \,|\, \top)$ and $\mathsf{ideal}^*\ \psi$ for $\mathsf{ideal}^*(\psi|\top)$.*

The distinction between prescriptive obligations and tests should not be confused with the distinction between prescriptive obligations and descriptive obligations, where the latter have the usual truth-functional semantics and are typically represented by $O\alpha$ for α 'ought to be (done).' We introduced descriptive obligations related to the prescriptive obligations of DUS in the two-phase framework for deontic reasoning 2DL [TvdT96, vdTT98c]. The 2DL motto is that we have to order before we can minimize. There are two phases in the static framework 2DL: the construction of the preference ordering and minimizing in the ordering. Two phases are necessary to combine strengthening of the antecedent and weakening of the consequent in the contrary-to-duty examples. Phase-1 obligations are ordering obligations and phase-2 obligations are minimizing obligations. Phase-1 of 2DL is formalized by a variant of Prohairetic Deontic Logic PDL [vdTT97c, vdTT98b] and phase-2 of 2DL is formalized by Hansson-Lewis dyadic deontic logic [Han71, Lew74]. These descriptive obligations also formalize reasoning about betterness relations, but they neglect the dynamic construction of these relations. The distinction between the prescriptive obligations $\mathsf{oblige}(\alpha|\beta)$ and tests $\mathsf{ideal}(\alpha|\beta)$ corresponds to the distinction between phase-1 ordering and phase-2 minimizing made in [TvdT96, vdTT98c].

In this paper, our formal language does not contain descriptive obligations $O\alpha$ and we therefore do not discuss the relation between prescriptive and descriptive obligations. To extend the logic, we need a mixed language in a dynamic logic like approach, in which we could write formula like $[\mathsf{oblige}\ \alpha]O\alpha$. The latter formula represents that after α has been promulgated, the obligation that 'α should be (done)' is true, i.e. the deontic state that arises after updating the initial deontic state with $\mathsf{oblige}\ \alpha$ is a betterness relation that satisfies $O\alpha$. Obviously, this relation between oblige and O is quite different from the relation between oblige and ideal. Hence, the distinction between oblige and ideal is different from the distinction between prescriptive and descriptive obligations.

For the dynamic interpretation of a test we define it analogously to the test operator in dynamic logic, and to the might and presumably operators in Veltman's update semantics [Vel96]. If the test is successful then the information conveyed by the test is already subsumed by the information state and the test update simply returns the state. If the test is unsuccessful, then the test update returns the absurd state.

Definition 12 (Deontic updates, continued) *Let $\sigma = \langle W, W^*, \leq, V\rangle$ be a deontic state. The update function $\sigma[\phi]$ defined in Definition 9 is extended as follows.*

- *if $\phi = \mathsf{ideal}(\alpha|\beta)$, then*
  - *if $\mathsf{pref}(\sigma,\beta) = \alpha$, then $\sigma[\phi] = \sigma$;*
  - *otherwise, $\sigma[\phi] = \mathbf{1}$.*
- *if $\phi = \mathsf{ideal}^*(\alpha|\beta)$, then*
  - *if $\mathsf{pref}^*(\sigma,\beta) = \alpha$, then $\sigma[\phi] = \sigma$;*
  - *otherwise, $\sigma[\phi] = \mathbf{1}$.*

The acceptance theorem for ideal operators follows directly, because the test explicitly determines when a formula is accepted.

Theorem 2 (Acceptance, continued) *Let σ be a deontic state. We have $\sigma \Vdash \phi$ if and only if given by Theorem 1 or*

- *if $\phi = \mathsf{ideal}(\alpha|\beta)$, then $\mathsf{pref}(\sigma,\beta) = \alpha$;*
- *if $\phi = \mathsf{ideal}^*(\alpha|\beta)$, then $\mathsf{pref}^*(\sigma,\beta) = \alpha$.*

Proof *We have $\mathbf{1} \Vdash \mathsf{ideal}(\alpha \mid \beta)$, because we have $\sigma[\phi] = \sigma = \mathbf{1}$. From this extension of Proposition 1 follows the proof for the absurd state, because in the absurd state everything is accepted. Otherwise, it follows directly that if $\phi = \mathsf{ideal}(\alpha|\beta)$, then $\mathsf{pref}(\sigma,\beta) = \alpha$. The proof for $\phi = \mathsf{ideal}^*(\alpha|\beta)$ is analogous to the proof of $\phi = \mathsf{ideal}(\alpha|\beta)$.*

The following corollary of Theorem 1 and 2 gives a relation between the two operators oblige and ideal.

Corollary 2 *Let σ be a deontic state and α and β two sentences of L_0^A. We have $\sigma[\mathsf{oblige}(\alpha|\beta)] \Vdash \mathsf{ideal}(\alpha|\beta)$ and $\sigma[\mathsf{oblige}^*(\alpha|\beta)] \Vdash \mathsf{ideal}^*(\alpha|\beta)$.*

Proof *The test $\mathsf{pref}(\sigma',\beta) = \alpha$ after the reduction of the deontic information state σ by the obligation $\mathsf{oblige}(\alpha|\beta)$ is exactly the same test as used for the test $\mathsf{ideal}(\alpha|\beta)$.*

With oblige and ideal operators, in many cases the relation is sensitive to the order of premises. For example, oblige ϕ_1, ideal ϕ_2 often has different consequences than ideal ϕ_2, oblige ϕ_1. In the following two sections we give several examples of the extended update system. In Section 6 we use the new operators to test the betterness relations, and in Section 7 the operators are used as conditions in derivations.

6 Examples

In many problems we can restrict ourselves to derivations of the following canonical form.

$$\mathsf{oblige}(\alpha_1|\beta_1), \ldots, \mathsf{oblige}(\alpha_n|\beta_n), \phi_1, \ldots, \phi_m \Vdash \mathsf{ideal}^* \psi$$

For example, this structure is used in many examples in this paper. If we use this structure, we can discriminate between the following three phases.

Ordering The descriptive obligations $\mathsf{oblige}(\alpha_i \mid \beta_i)$ create a unique ordering reflecting different degrees of ideality.

Zooming in The facts ϕ_i select the part of the ordering which is used for the context of deliberation, because facts are assumed to be settled.

Minimizing The test $\mathsf{ideal}^* \psi$ considers the ideal state of the context of deliberation. Distinctions between states worse than the ideal state are ignored during this process of minimizing.

Obviously, this canonical form has limited expressive power. In particular, it does not represent the violations and it does not represent contrary-to-duty structures. For the first we have to use ideal operator at the right hand side of $\Vdash$, and for contrary-to-duty structures a dyadic operator (see below).

6.1 Contrary-to-duty paradoxes in DUS

Preference semantics were introduced in deontic logic to model contrary-to-duty reasoning [Han71, Lew74], see also the discussions in [TvdT96, vdTT97c, vdTT97a]. The deontic contrary-to-duty paradoxes can be modeled without any problems in DUS. The following two examples illustrate that the dynamic representation is in some respects more insightful than the static one. We start with the Forrester (or gentle murderer) paradox [For84].

Example 9 (Forrester) *Consider the sentences 'Smith should not kill Jones' (*oblige $\neg k$*), 'if Smith kills Jones, then he should do it gently' (*$\mathsf{oblige}(g|k)$*) and 'Smith kills Jones' (k), given that gentle killing implies killing ($g \to k$). The* *fact that gentle killing implies killing is formalized by the fact that we consider the deontic states that only contain worlds that satisfy $g \rightarrow k$. The three updates of the initial state are represented in Figure 8. In the ideal state of the context of justification we have that Smith does not kill Jones ($\neg k$),*

$$\mathsf{oblige}\ \neg k, \mathsf{oblige}(g|k), k \Vdash \mathsf{ideal}\ \neg k$$

and in the ideal state of the context of deliberation we have that Smith kills Jones gently (g).

$$\mathsf{oblige}\ \neg k, \mathsf{oblige}(g|k), k \Vdash \mathsf{ideal}^* g$$

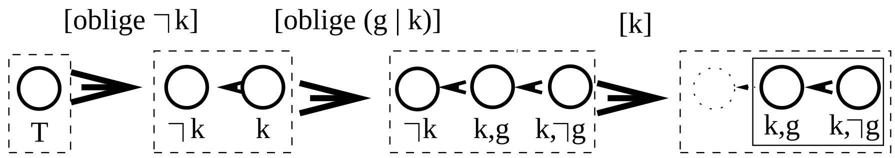

Figure 8: Forrester paradox

Our second example of contrary-to-duty paradoxes in DUS is the Chisholm paradox [Chi63].

Example 10 (Chisholm) *Consider the sentences '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' ($\mathsf{oblige}(t|a)$), 'if the man does not go to the assistance, then he should not tell them he will come ($\mathsf{oblige}(\neg t|\neg a)$) and 'the man does not go' ($\neg a$). The four updates of the initial state are represented in Figure 9. In the context of justification we have that in the ideal state the man tells his neighbors (t),*

$$\mathsf{oblige}\ a, \mathsf{oblige}(t|a), \mathsf{oblige}(\neg t|\neg a), \neg a \Vdash \mathsf{ideal}\ t$$

and in the context of deliberation we have that in the ideal state the man does not tell his neighbors ($\neg t$).

$$\mathsf{oblige}\ a, \mathsf{oblige}(t|a), \mathsf{oblige}(\neg t|\neg a), \neg a \Vdash \mathsf{ideal}^* \neg t$$

The two contrary-to-duty examples above illustrate that we cannot use our general form

$$\mathsf{oblige}(\alpha_1|\beta_1), \ldots, \mathsf{oblige}(\alpha_n|\beta_n), \phi_1, \ldots, \phi_m \Vdash \mathsf{ideal}^* \psi$$

in all cases, because for violations we have to use ideal instead of ideal^*. To derive all aspects of contrary-to-duty structures it is necessary to use the dyadic

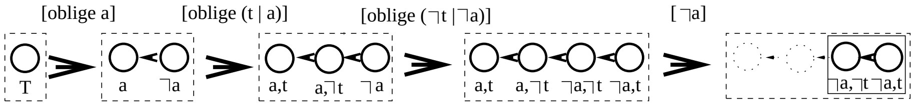

Figure 9: Chisholm paradox

operators at the right hand side of the inference relation, either a dyadic oblige or a dyadic ideal.

6.2 Non-monotonicity

In this section we show that conclusions can be lost by updating. However, first we show that the obligations have strengthening of the antecedent (if there are no conflicting obligations).

Example 11 (Strengthening of the antecedent) *Assume two propositional sentences α and β of L_0^A such that $\alpha \wedge \beta$ is consistent. We have*

$\mathsf{oblige}\ \alpha \Vdash \mathsf{oblige}(\alpha|\beta)$

Example 12 (Polite and honest) *Consider 'be polite' ($\mathsf{oblige}\ p$) and 'be honest' ($\mathsf{oblige}\ h$). The most ideal states in which we are impolite or dishonest are $\neg p \wedge h$ and $p \wedge \neg h$, in which we are at least not impolite as well as dishonest. Hence, we do not have ideally p. However, if we only have the first premise, then we have ideally p.*

$\mathsf{oblige}\ p \Vdash \mathsf{ideal}(p|\neg(p \wedge h))$
$\mathsf{oblige}\ p, \mathsf{oblige}\ h \not\Vdash \mathsf{ideal}(p|\neg(p \wedge h))$

Hence, by addition of a premise we sometimes loose conclusions.

The following example is one of the most notorious examples from the deontic logic literature given by von Wright in [vW71], see also [Alc93] for a discussion. The problem is that this premise set (shaped as a 'diamond' when we add the fact that it rains and the sun shines) is inconsistent in several deontic logics, whereas intuitively it is consistent.

Example 13 (Window) *Consider the obligations 'the window ought to be closed if it rains' ($\mathsf{oblige}(c|r)$) and 'it ought to be open if the sun shines' ($\mathsf{oblige}(\neg c|s)$). If it rains and the sun shines then we have a dilemma. On the one hand the window ought to be closed because it rains, and on the other hand it ought to be* *open because the sun shines. If we only have the first premise, no such dilemmas exist.*

$$\mathsf{oblige}(c|r) \Vdash \mathsf{ideal}(c|r \wedge s)$$
$$\mathsf{oblige}(c|r), \mathsf{oblige}(\neg c|s) \nVdash \mathsf{ideal}(c|r \wedge s).$$

Again, by addition of a premise we sometimes loose conclusions.

Example 12 and 13 illustrate that strengthening of the antecedent is restricted. This is a kind of defeasibility, because conditionals with restricted strengthening of the antecedent are defeasible conditionals [Alc93]. This face of defeasibility is caused by violations and unresolved conflicts. It should therefore not be confused with overridden defeasibility and prima facie obligations, see [vdTT95, vdTT97b] for a discussion on this distinction.

6.3 Loops

Example 14 (Strengthening of the antecedent) *Assume three propositional sentences p, q and r of L_0^A that are mutually exclusive, i.e. we have $\neg(p \wedge q)$ etc. Consider the deontic state $\mathbf{0}[\mathsf{oblige}(p|p \vee q)][\mathsf{oblige}(r|r \vee p)][\mathsf{oblige}(q|q \vee r)]$, as represented in Figure 10. When we take the transitive closure, all worlds are equivalent.*

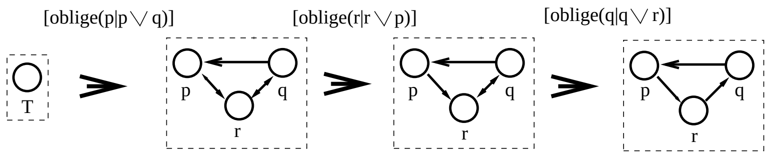

Figure 10: Loop

7 Monotonic fragment

Consider the following two inference relations defined in [Vel96].

Definition 13 (Validity, continued) *Let $\psi_1, \ldots, \psi_n$ and ϕ be sentences of the deontic language L_2^A. $\psi_1, \ldots, \psi_n \Vdash^\sigma \phi$ if and only if for all deontic states σ we have $\sigma[\psi_1]\ldots[\psi_n] \Vdash \phi$. $\psi_1, \ldots, \psi_n \Vdash^m \phi$ if and only if $\sigma \Vdash \phi$ for every σ such that $\sigma \Vdash \psi_1, \ldots, \sigma \Vdash \psi_n$.*

Veltman shows that σ^m is monotonic and $\Vdash^\sigma$ is left monotonic. In fact, for L_1^A the inference relation σ^m is exactly Prohairetic Deontic Logic [vdTRFT97, vdTT98b] with classical entailment. Preferential entailment based on maximally connected models models moreover, that *only* the premises are known. The definitions are increasingly general, i.e we have:

$\psi_1, \ldots, \psi_n \Vdash^\sigma \phi$ implies $\psi_1, \ldots, \psi_n \Vdash \phi$ but not vice versa.
$\psi_1, \ldots, \psi_n \Vdash^m \phi$ implies $\psi_1, \ldots, \psi_n \Vdash^\sigma \phi$ but not vice versa.

The following example illustrates the second use of the operators $\mathsf{ideal}(\alpha|\beta)$. The operators are used as conditions in derivations.

Example 15 *It can easily be shown that we have the following.*
$\mathsf{oblige}\ p, \mathsf{oblige}\ q, \mathsf{ideal}(p \wedge q) \Vdash^m \mathsf{oblige}(p \wedge q)$
$\mathsf{oblige}(p|q), \mathsf{oblige}q|r), \mathsf{ideal}(p \wedge q|r) \Vdash^m$

8 Further research

There are several issues for further research.

Other updates The logic has to be further extended by introducing different kinds of updates.

Permission Most importantly, other deontic operators like permissions and prohibitions have to be formalized. A prohibition for α can be formalized as an obligation for $\neg\alpha$. Permissions are formalized analogous to permissions in 2DL. Strong permissions are defined in 2DL by $M \models P(\alpha \mid \beta)$ if and only if for all worlds $w_1, w_2 \in W$ such that $M, w_1 \models \alpha \wedge \beta$ and $M, w_2 \models \neg\alpha \wedge \beta$ we have $w_2 \not< w_1$, see [vdT97, vdTT98a]. We define a permission as the deletion of $w_2 < w_1$ when $\sigma, w_1 \models \alpha \wedge \beta$ and $\sigma, w_2 \models \neg\alpha \wedge \beta$.

Speech acts Since an example of the dynamic [$\mathsf{oblige}\ \alpha$] is an imperative[4] (a speech act), we can formalize other speech acts in the update semantics framework. Consider Dignum and Weinand's speech acts [DW95].

[4] In [TvdT96] we interpreted the ordering and minimizing obligations *loosely* as respectively imperatives and obligations, because weakening is not an intuitive property of imperatives but it might be accepted for obligations. Ordering obligations do not have weakening of the consequent, whereas minimizing obligations do, expressed by the following theorem $\mathbf{WC}_{HL}$ of the logic.

$$\mathbf{WC}_{HL} \quad O_{\mathrm{HL}}(\alpha_1|\beta) \rightarrow O_{\mathrm{HL}}(\alpha_1 \vee \alpha_2|\beta)$$

Unlike some authors we do not consider obligations and imperatives to be the same. However, in our present framework we can say that imperatives are examples of prescriptive obligations.

Laying guilt on someone Another kind of update can be used to formalize the phenomenon of 'laying guilt' on someone. For example, assume your wife says to you 'you did not bring me flowers.' You are guilty of violating an obligation you did not know it existed. The distinction is that the speech act does not create a deontic cue, but a violation. The set of equivalence classes increases and at the same time we zoom in on the ordering.

Simultaneous updates The language can be extended with the operators of dynamic logic like oblige $p \cap$ oblige q for simulatneous updates and oblige $p \cup$ oblige q for choice.

Defeasible (prima facie) obligations One possibility to formalize defeasible obligations in update semantics is to give Veltman's normally-presumably logic a deontic reading. However, there is no semantic enhancement of the system, it has al. been dealt with by the proof theory, There are at least three different extensions of our preference-based semantics for defeasible obligations. In Figure 11 we give the semantics for 'you should not eat with your fingers oblige$(\neg f \mid \top)$, but you should eat with your fingers if you are served asperagus oblige$(f \mid a)$. The ordering are represented by integers: something is preferred if it has a higher number.

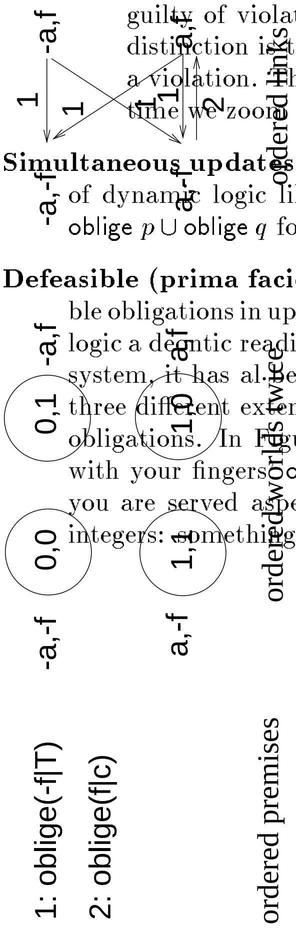

Figure 11: Three types of defeasibility

Ordering premises The simplest way to introduce defeasibility in preference-based semantics is to introduce a priority order on the premises. Andreka et al [ARS95] show given some postulates there are only prioritized strategies to combine non-transitive orderings. This extends Arrow's famous theorem [Arr50], that shows there are only dominant strategies to combine transitive preference orderings. Arrow starts with a set of postulates for combining relations, for example if one relation has $w_1 < w_2$ and all other relations have $w_1 \sim w_2$, then we have in the combined ordering $w_1 < w_2$. Veltman's normally-presumably logic violates the latter postulate for combining preference relations.[5]

Ordering worlds twice Besides the ideality ordering, we can introduce a normality ordering, see [TvdT95, vdTT97b]. In Figure 11, the first number of a circle represents the ideality ordering and the second number represents the normality ordering. In the ideality ordering, the $\neg a$ worlds are ordered such that $\neg f$ is preferred to f, and the a worlds such that f is preferred to $\neg f$. In the normality ordering, the $\neg a$ worlds are preferred to the a worlds, which represents that normally you are not served asparagus.

Ordering links The ordered links can be used to give a semantics to the ordered premises. In this case, the links from f and $\neg f$ are higher than the other links as a result of the fact that normally you sould not eat with your fingers. The exception is the link from $a \wedge \neg f$ to $a \wedge f$, which has the highest number as the result of the obligation that you should eat with your fingers if you eat asparagus.

There are different interpretation of these semantics. In [vdTT97b] we argued that ordered premises can be used for prima facie obligations and the ideality normality orderings for specificity.[6] The ordered links seems to be more general and useful for both approaches.

First-order obligations See Veltman's discussion in [Vel96]. A first-order logic can be used to represent act-types in the von Wright tradition. For example, consider the similar extension to first-order logic of Veltman's normally-presumably logic [Vel96].

> "So far, we have been thinking of the language as a propositional language, but we can also give a predicate logical interpretation

[5] The following example illustrates the distinction between priorities and Veltman's logic.

Example 16 *Consider the set of obligations $S = \{O(p|\top), O(\neg p \wedge q|a)\}$. In defeasible deontic logic we have $S \not\models O(p|a)$ because of specificity principles, i.e. the obligation $O(p|\top)$ is not strengthened to condition a. The crucial question is whether we want to have $S \models O(p|a \wedge \neg q)$, i.e. do we want to reinstate the most general obligation in case the overriding obligation is itself violated? According to Arrow's postulates the obligation is derivable. The argument runs as follows. The $a \wedge \neg q$ worlds are not ordered by $O(\neg p \wedge q|a)$, because according to the constraint $w_2 \not\leq w_1$ for $M, w_1 \models \neg p \wedge q \wedge a$ and $M, w_2 \models \neg(\neg p \wedge q) \wedge a$ all $a \wedge \neg q$ worlds can be equivalent. However, these worlds are obviously ordered by $O(p|\top)$. According to Veltman's logic (and our intuitions) the obligation is not derivable.*

[6] We say that priorities are based on an argument-based interpretation, and Veltman's normally-presumably logic is based on a probability-based interpretation. The argumentation-based interpretation may be useful for prima facie obligations or legal rules, but the probability-based interpretation is the most useful for general AI applications, because it is based on first principles. For further details on the distinction between the argument-based interpretation and the probability-based interpretation, see [vdTT97b].

to it. Think of p, q etc. as monadic predicates rather than atomic sentences. Each such predicate specifies a property and each well-formed expression specifies a boolean combination of properties. Think of W as the set of possible objects rather than the set of possible worlds. A possible object $i \in W$ has the property expressed by the atom p if and only if $p \in i$. Note that different possible objects have different properties. Therefore it would be more precise to call the elements of W possible *types* of objects: in reality there can be more than one or no object fitting the description of a given possible object in W."

The update semantics seem to be able to solve the problem has been discussed by Edelberg [Ede91], because they are sed for similar problems in the natural language analysis. The same counts for definite descriptions, as discussed by Lou Goble **??**. We prefer a first-order framework, because some problems disappear and new problems appear.

Nested obligations A descriptive logic of nested obligations is the deontic interpretation of Weydert's hyperrational conditional logic [Wey92]. For example, the axiom $\mathsf{oblige}(\alpha \mid \beta \wedge \mathsf{oblige}(\alpha \mid \beta))$. Nested obligations are important in a multi-agent environment. For example, $O_M O_m e$ can be used to express that the minister M is obliged to see to it that the mayor m is obliged to see to it that there is a state of emergency. It is an open question how nested obligations can be incorporated in the dynamic framework. In other preference-based logics nested operators are already important in a single agent case, for example for desires: $\mathsf{desiredesire}\neg s$: 'I desire that I desire not to smoke.'

One idea is to define an preference order on action sequences by using a predicate 'done.' The idea is illustrated in Figure 12.

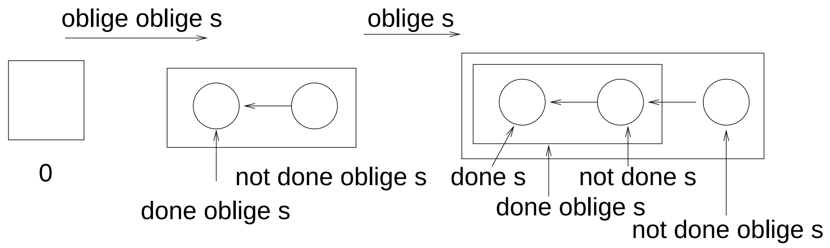

Figure 12: Different levels of zooming in on the ordering

Agents With explicit agents we can formalize who (which authority) obliges us to do something. Likewise, other agents can be introduced, like the claimants in case of contractual obligations (the agents towards who the obligation is directed).

Action model An action model with causal relations, dependencies between actions etc has to be introduced, see e.g. [Pea93].

Algorithms The dynamic approach is probably also a good basis for the development of an efficient proof method for deontic reasoning based on model checking. The idea is first to explicitly construct the unique deontic state for a list of premises, and then check whether a formula is accepted in this state.

Consider for example $\mathsf{oblige}(\alpha_1|\beta_1), \ldots, \mathsf{oblige}(\alpha_n|\beta_n) \Vdash^m \mathsf{oblige}(\alpha|\beta)$. We have to prove that there is no world w_1 and w_2 such that $M, w_1 \models \alpha \wedge \beta$ and $M, w_2 \models \neg\alpha \wedge \beta$, and for all i, $M, w_1 \models \alpha_i \wedge \beta_i$ and $M, w_2 \models \neg\alpha_i \wedge \beta_i$. We look for γ_1 and γ_2 such that:

$$\forall \gamma_1, \gamma_2 \vee_{i \in N} (\alpha \wedge \beta \wedge \gamma_1 \rightarrow \alpha_i \wedge \beta_i) \wedge (\neg\alpha \wedge \beta \wedge \gamma_2 \rightarrow \neg\alpha_i \wedge \beta_i))$$

or

$$\nexists \gamma_1, \gamma_2 \wedge_{i \in N} (\alpha \wedge \beta \wedge \gamma_1 \wedge \neg(\alpha_i \wedge \beta_i)) \vee (\neg\alpha \wedge \beta \wedge \gamma_2 \wedge \neg(\neg\alpha_i \wedge \beta_i)))$$

9 Related research

The work of Horty [Hor94, Hor93] can be related to our work. $\vdash_{\mathrm{NS}}$ is like $\vdash_{\langle D,W \rangle}$ defaults do not have a truth value deriving defaults from a set of defaults Horty: NS is set $O(\alpha|\beta)$ Inference relation $\{O(\alpha_i|\beta_i) \mid i = 1 \ldots n\} \vdash O(\alpha|\beta)$.

Veltman's normally-presumably operators are closely related to the two deontic operators oblige and ideal, respectively. The most important distinction between $\mathsf{normally}$ and oblige is that the former can deal with specificity, whereas the latter cannot. Hence, $\mathsf{normally}$ should be compared with a defeasible or prima facie obligation instead of with a non-defeasible obligation. Moreover, Veltman's operators are formalized by an update system, of which the information state is a set of orderings, one ordering for each possible antecedent (i.e. an equivalence class of the formulas of the propositional language under logical implication). He therefore can find the preferred worlds directly, i.e. he does not have to zoom in and take the transitive closure first. A drawback is that his system does not support any reasoning by cases, which is a direct and obvious result of the fact that he does represent the information state by a set.

Example 17 *We have:*
$\mathsf{oblige}(p|p \vee q), \mathsf{oblige}(p|p \vee \neg q) \Vdash \mathsf{oblige}\ p$
$\mathsf{oblige}(p|q), \mathsf{oblige}(p|\neg q) \nVdash \mathsf{oblige}\ p$

Give formal definitions too!

10 Conclusions

In this paper we introduced a logic of prescriptive obligations. All deontic logics known to us are static in the sense that they only formalize reasoning about *descriptive* obligations. They are logics of normative propositions. From Alchourrón's box metaphor follows that prescriptive obligations are interpreted *dynamically* whereas descriptive obligations are interpreted *statically*. We believe that a logic of prescriptive obligations – a logic of norms, a logic without truth values – like DUS can therefore be used to analyze the dynamics of obligations. Whereas descriptive deontic logics only formalize reasoning about obligations – which obligations follow from a set of obligations – prescriptive deontic logics also reason with obligations [vdTT97a]. Such a logic can therefore be used for applications of deontic logic in computer science.

In this paper we show how obligations in update semantics DUS combine temporal and preferential notions. Two important classes of deontic logics are either based on temporal [vE82, Tho81, HB95] or preferential notions [Jac85, Gob90a, Han90, BMW93, TvdT96], but the combination of both is rare (see [Hor96] for an exception). The formalization of obligations in update semantics proposed in this paper is a natural extension with temporal notions of the preference-based deontic logics proposed in the two-phase framework for deontic reasoning 2DL [TvdT96]. The dynamic approach DUS has something to offer over the static approach 2DL, because it explains nonmonotonicity related to violability and unresolved conflicts, see Section 6.2.[7] From our present analysis it follows that non-monotonicity is caused by the fact that we attempt to simulate dynamic behavior in a static framework.

Acknowledgments. We thank Henry Prakken and Emil Weydert for discussing the issues raised in this paper, and an anonymous referee for several

[7] In [TvdT96, vdTT97c] we proposed the following three desiderata for deontic logic that formalizes reasoning about obligations *which are not subject to exceptions*:

- it has to be able to formalize all types of contrary-to-duty reasoning, in particular the Forrester paradox,
- it has the no-dilemma assumption, in particular $O(\alpha|\top) \wedge O(\neg\alpha|\beta)$ has to be inconsistent (the considerate assassin example [PS96]),
- it does not have an irrelevance problem, i.e. it has (possibly restricted) strengthening of the antecedent to derive $O(\neg k|m)$ from $O(\neg k|\top)$.

For example, the Hansson-Lewis dyadic deontic logics do not satisfy the second and third of our desiderata, because $O_{\mathrm{HL}}(\alpha|\top) \wedge O_{\mathrm{HL}}(\neg\alpha \mid \beta)$ is consistent and we do not have $O_{\mathrm{HL}}(\neg k| \top) \models O_{\mathrm{HL}}(\neg k|m)$. In [TvdT96, vdTT97c] we introduced preferential entailment based on maximally connected models to solve the irrelevance problem (the third desiderata) by adding strengthening of the antecedent to the conditional obligations.

helpful comments.

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.

[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.

[Arr50] K.J. Arrow. A difficulty in the concept of social welfare. *Journal of Political Economy*, 58:328–346, 1950.

[ARS95] H. Andreka, M. Ryan, and P.-Y. Schobbens. Operators and laws for combining preference relations. In *Information Systems: Correctness and Reusability (Selected Papers)*. World Publishing Co, 1995.

[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.

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

[DW95] F. Dignum and H. Weigand. Modelling communication between cooperative systems. In K. Lyytinen, J.Iivari, and M. Rossi, editors, *Advanced Information Systems Engineering (LNCS-932)*, pages 140–153, Berlin, 1995. Springer Verlag.

[Ede91] W. Edelberg. A case for a heretical deontic semantics. *Journal of Philosophical Logic*, 20:1–35, 1991.

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

[Gob90a] L. Goble. A logic of good, would and should, part 1. *Journal of Philosophical Logic*, 19:169–199, 1990.

[Gob90b] 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.

[HB95] J.F. Horty and N. Belnap. The deliberative stit: a study of action, omission, ability, and obligation. *Journal of Philosophical Logic*, pages 583–644, 1995.

[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.

[Hor96] J.F. Horty. Agency and obligation. *Synthese*, 108:269–307, 1996.

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

[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.

[Pea93] J. Pearl. A calculus of pragmatic obligation. In *Proceedings of 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.

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

[Tho81] R. Thomason. Deontic logic as founded on tense logic. In R. Hilpinen, editor, *New Studies in Deontic Logic: Norms, Actions and the Foundations of Ethics*, pages 165–176. D. Reidel, 1981.

[TvdT95] Y.-H. Tan and L.W.N. van der Torre. Why defeasible deontic logic needs a multi preference semantics. In *Symbolic and Quantitative Approaches to Reasoning and Uncertainty. Proceedings of* *the ECSQARU'95. LNAI 946*, pages 412–419. Springer Verlag, 1995.

[TvdT96] Y.-H. Tan and L.W.N. van der Torre. How to combine ordering and minimizing in a deontic logic based on preferences. In *Deontic Logic, Agency and Normative Systems. Proceedings of the Δeon'96. Workshops in Computing*, pages 216–232. Springer Verlag, 1996.

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

[vdTRFT97] L.W.N. van der Torre, P. Ramos, J.L. Fiadeiro, and Y.-H. Tan. The role of diagnosis and decision theory in normative reasoning. In *Proceedings of the Second Modelage Workshop on Formal Models of Agents (Modelage'97)*, 1997. To appear.

[vdTT95] L.W.N. van der Torre and Y.H. 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)*, pages 1525–1532. Morgan Kaufman, 1995.

[vdTT97a] L.W.N. van der Torre and Y.-H. Tan. Distinguishing different roles in normative reasoning. In *Proceedings of the Sixth International Conference on AI and Law (ICAIL'97)*, pages 225–232. ACM Press, 1997.

[vdTT97b] L.W.N. van der Torre and Y.H. Tan. The many faces of defeasibility in defeasible deontic logic. In D. Nute, editor, *Defeasible Deontic Logic*, pages 79–121. Kluwer, 1997.

[vdTT97c] L.W.N. van der Torre and Y.H. Tan. Prohairetic deontic logic and qualitative decision theory. In *Proceedings of the AAAI Spring Symposium on Qualitative Approaches to Deliberation and Reasoning*, 1997. To appear.

[vdTT98a] L.W.N. van der Torre and Y.-H. Tan. Extensions of two-phase deontic logic. Technical report, 1998. In preparation.

[vdTT98b] L.W.N. van der Torre and Y.-H. Tan. Prohairetic deontic logic. Technical report, 1998. In preparation.

[vdTT98c] L.W.N. van der Torre and Y.-H. Tan. Two-phase deontic logic. Technical report, 1998. In preparation.

[vdTT98d] L.W.N. van der Torre and Y.-H. Tan. An update semantics for deontic reasoning. In *Proceedings of the Δeon'98*, 1998.

[vE82] J. van Eck. A system of temporally relative modal and deontic predicate logic and its philosophical application. *Logique et Analyse*, 100:249–381, 1982.

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

[vW71] G.H. von Wright. A new system of deontic logic. In R. Hilpinen, editor, *Deontic Logic: Introductory and Systematic Readings*, pages 105–120. D. Reidel Publishing Company, Dordrecht, Holland, 1971. A reprint of 'A New System of Deontic Logic,' *Danish Yearbook of Philosophy* 1:173–182, 1964, and 'A Correction to a New System of Deontic Logic', *Danish Yearbook of Philosophy* 2:103–107, 1965.

[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 Publishing company, 1981.

[Wey92] E. Weydert. Hyperrational conditional logic. In *Proceedings of the ECAI'92 Workshop on Theoretical Foundations of Knowledge Representation and Reasoning*. Springer Verlag, 1992.