The Temporal Analysis of Chisholm's Paradox

Leendert W.N. van der Torre

IRIT
Paul Sabatier University
118 Route de Narbonne
31062 Toulouse, France
TORRE@IRIT.FR

Yao-Hua Tan

EURIDIS
Erasmus University
P.O. Box 1738
3000 DR Rotterdam, the Netherlands
YTAN@FAC.FBK.EUR.NL

Abstract

Deontic logic, the logic of obligations and permissions, is plagued by several paradoxes that have to be understood before deontic logic can be used as a knowledge representation language. In this paper we extend the temporal analysis of Chisholm's paradox using a deontic logic that combines temporal and preferential notions.[0]

Introduction

Deontic logic is a modal logic in which Op is read as 'p ought to be (done).' Deontic logic has traditionally been used by philosophers to analyze the structure of the normative use of language. In the eighties deontic logic had a revival, when it was discovered by computer scientists that this logic 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 (Wieringa & Meyer 1993)). 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. Another application is the use of deontic logic to represent legal reasoning in legal expert systems in artificial intelligence. Legal expert systems have to be able to reason about legal rules and documents such as for example a trade contract. Deontic notions are essential to represent the meaning of such rules or the content of such contracts. A recent topic is the relation between deontic logic and logics of desires and goals as these are developed in qualitative decision theory. First results seem to indicate that extensions of deontic logic can be used in this type of decision theory (Pearl 1993; van der Torre & Tan 1998a). Another recent development that is currently attracting a lot of attention is the use of deontic logic to specify intelligent agents for the Internet. For example, one of the major challenges in electronic commerce is to develop agents that can automatically draft, negotiate and process trade contracts. Since contracts are legal documents, these agents have to be able to perform deontic reasoning to handle these contracts (for a survey see (Kimbrough & Lee 1997)). Furthermore, deontic logic could be fruitful for the analysis and specification of normative issues about the Internet such as authorization, access regulation, and privacy maintenance (Conte & Falcone 1997).

With the increasing popularity and sophistication of applications of deontic logic the fundamental problems of deontic logic become more pressing. From the early days, when deontic logic was still a purely philosophical enterprise, it is known that it suffers from certain paradoxes. The most notorious one is the so-called Chisholm paradox. The conceptual issue of this paradox 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. For example, if the delivery time is over due the responsible agent might be obliged to pay the extra transport and warehousing costs that result from the delay. If the violation is not too serious, or was not intended by the violating party, 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. Hence, Chisholm's paradox is an important benchmark example of deontic logic, and deontic logics incapable of dealing with it are considered insufficient tools to analyze deontic reasoning. The Chisholm set consists of the following four sentences.

1. 'α ought to be (done),'
2. 'if α is (done), then β ought to be (done),'
3. 'if α is not (done), then β ought not to be (done),'
4. 'α is not (done).'

The formalization of these sentences in Standard Deontic Logic (see below) is either inconsistent or the sentences are logically dependent. The Chisholm set is therefore called a *paradox*. Temporal deontic logic can consistently formalize the set

1'. 'α ought to be (done),'
2'. 'if α has been (done), then β ought to be (done),'
3'. 'if α has not been (done), then β ought not to be (done),'
4'. 'α has not been (done).'

For most α and β the first set can be transformed to the second one without changing the meaning of the sentences. It has been argued that temporal deontic logic therefore *solves* the paradox (van Eck 1982; Loewer & Belzer 1983).[1] However, this 'solution' does not work for the original set given by Chisholm (1963), in which α is read as 'a certain man goes to the assistance of his neighbors' and β as 'the man tells his neighbors that he will come' (Vorobej 1986; Feldmann 1990; Smith 1994; Yu 1995). For example,

2. 'if a certain man goes to the assistance of his neighbors, then the man ought to tell his neighbors that he will come'

means something different than

2'. 'if a certain man has gone to the assistance of his neighbors, then the man ought to tell his neighbors that he will come.'

The representation $1' - 4'$ assumes that the antecedent (condition) α occurs before the consequent (conclusion) β, but the contrary is the case for these specific α and β from the Chisholm set! The contribution of this paper is twofold. We introduce a new deontic logic, combining temporal and preferential notions, and we show how the temporal antecedent-before-consequent analysis of the paradox can be extended with preferences on sequences of actions to cover the original Chisholm set. Moreover, we start with a survey of the paradox in so-called Standard Deontic Logic and temporal deontic logic, and, to put our new formalization in context, we end with a formalization in preference-based deontic logic.

Standard Deontic Logic (SDL)

SDL is usually formalized by a normal modal system of type KD[2] according to the Chellas classification, although normal modal systems validate the counterintuitive theorem $O\top$, where $\top$ stands for any tautology like $p \vee \neg p$. We start with some terminology.

- A conditional obligation 'α ought to be (done) if β is (done)' is usually formalized in SDL by $\beta \rightarrow O\alpha$, and sometimes by $O(\beta \rightarrow \alpha)$.
- The conditional obligation $\beta \rightarrow O\alpha$ or $O(\beta \rightarrow \alpha)$ is called a *Contrary-To-Duty* (CTD or *secondary*) obligation of the (*primary*) obligation $O\alpha_1$ when β and α_1 are contradictory. The condition of a CTD obligation is only fulfilled if the primary obligation is violated.
- A conditional obligation $\beta \rightarrow O\alpha$ or $O(\beta \rightarrow \alpha)$ is an *According-To-Duty* (ATD) obligation of $O\alpha_1$ when β logically implies α_1. The condition of an ATD obligation is satisfied only if the primary obligation is fulfilled.

The Chisholm paradox can be represented in SDL by one of the following two sets of SDL formulas, where a is read as 'a certain man goes to the assistance of his neighbors' and t as 'he tells them that he will come.'

$$T_1 = \{Oa, O(a \rightarrow t), \neg a \rightarrow O\neg t, \neg a\}$$

$$T_2 = \{Oa, a \rightarrow Ot, \neg a \rightarrow O\neg t, \neg a\}$$

The second obligation is an ATD obligation and the third obligation is a CTD obligation, see Figure 1.

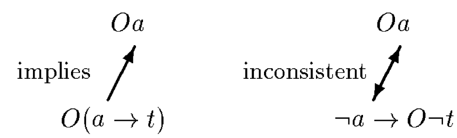

Figure 1: $O(a \rightarrow t)$ is an ATD, $\neg a \rightarrow O\neg t$ is a CTD

Both sets are problematic. T_1 derives two contradictory obligations, although the set of premises is intuitively consistent. Since SDL allows a kind of so-called deontic detachment, i.e. $\models_{\text{SDL}} (O\beta \wedge O(\beta \rightarrow \alpha)) \rightarrow O\alpha$, we have $T \models_{\text{SDL}} Ot$ from the first two sentences. Moreover, since SDL also allows factual detachment, i.e. $\models_{\text{SDL}} (\beta \wedge (\beta \rightarrow O\alpha)) \rightarrow O\alpha$, we have $T \models_{\text{SDL}} O\neg t$ from the last two sentences.

T_2 has logical redundant sentences, because $a \rightarrow Ot$ can be derived from $\neg a$,[3] and, more seriously, it does not derive the obligation 'the man ought to tell his neighbors that he will come and go to their assistance' $O(t \wedge a)$, not even from $\{Oa, a \rightarrow Ot\}$ (i.e. when the truth value of a is not yet fixed). In SDL the obligation $O(t \wedge a)$ is derived if and only if Ot is derived, because SDL has the theorem $O(\alpha \wedge \beta) \leftrightarrow (O\alpha \wedge O\beta)$. Ot cannot be deontically detached from the first two sentences of T_2 in

[1] Traditionally, the deontic paradoxes are understood as phenomena occurring in natural language which cannot be solved, only analyzed. However, several logicians have argued (though not very convincingly) that the deontic paradoxes are not paradoxes in natural language, but only 'paradoxes' in the logical formalization.

[2] System KD is closed under the inference rules modus ponens and necessitation and it satisfies besides the propositional theorems the axioms **K**: $O(\alpha \rightarrow \beta) \rightarrow (O\alpha \rightarrow O\beta)$ that says that the modal operator is closed under modus ponens and **D**: $\neg(O\alpha \wedge O\neg\alpha)$ that says that there are no conflicting obligations.

[3] Chisholm argued that this logical dependence is counterintuitive, and several logicians have demanded that a solution of the Chisholm paradox should represent the sentences such that they are logically independent. However, Tomberlin (1981) observes that the criterion is a 'rather glaring theoretical commitment' which 'would be a case of flagrant methodological question-begging.' Moreover, this logical dependence is easily solved by introducing a weaker notion of implication. For example, the two conditional obligations can be represented by $a > Ot$ and $\neg a > O\neg t$ where '$>$' is a so-called strict implication. For example, we can represent the obligations by $\Box(a \rightarrow Ot)$ and $\Box(\neg a \rightarrow O\neg t)$ where $\Box$ is a so-called alethic modal operator that satisfies at least axiom **T**: $\Box\alpha \rightarrow \alpha$ (reflexivity). This solves the logical dependence, because the formula $\neg\alpha \rightarrow (\alpha > \beta)$ is in contrast to the formula $\neg\alpha \rightarrow (\alpha \rightarrow \beta)$ not a theorem.

SDL, because we have $\not\models_{\text{SDL}} (O\beta \wedge (\beta \rightarrow O\alpha)) \rightarrow O\alpha$. The problem is that the following intuitive deontic reasoning pattern is not supported.

> Assume that although the man is able to go to the assistance of his neighbors, he has no intention of doing so. He argues: 'I ought to change my mind, tell them, and go to their assistance. So I ought to tell them. My present fulfillment of this obligation will help to make up for my sinfully not going to the assistance!'[4]

Summarizing, the SDL analysis of the Chisholm paradox (the inconsistency of T_1) is based on rejection of one of the detachment principles (e.g. rejection of deontic detachment in T_2). However, both principles seem intuitive in most cases. Rejection of one of the principles because they *seem* to be problematic in a *very few* cases is a solution that seems like overkill. These 'solutions' miss the point of the paradox. Since the problems are caused by the second obligation of the set, we prefer to call it an ATD paradox instead of, as it is usually called, a CTD paradox.

Many alphabetic variants of the Chisholm paradox have been proposed, see e.g. (van Eck 1982; Loewer & Belzer 1983). However, a crucial distinction with the original Chisholm set is that the consequents of the CTD and ATD obligation occur later than the primary obligation! For example, Section 79 subsection 4 of the United Nations Convention on Contracts for the International Sale of Goods reads as follows (see (Smith 1994, p.127)): "The party who fails to perform must give notice to the other party of the impediment and its effect on his ability to perform. If the notice is not received by the other party within a reasonable time after the party who fails to perform knew or ought to have known of the impediment, he is liable for damages resulting from such non-receipt." Here we have a double contrary-to-duty construction: first a contrary-to-duty obligation (to give notice), and then a prevision of what the consequences will be if that contrary-to-duty obligation remains unfulfilled (liability for damages). Obviously, there are many practical issues involved in representing such 'real' systems of norms, for example the formalization of domain knowledge (e.g. exact conditions when a party fails to conform), and the formalization of the protocols involved.[5] However, as far as we are concerned, a typical case between the parties A and B can be formalized by the SDL theory $T_3 = \{O_A p, p \rightarrow O_A \neg n, \neg p \rightarrow O_A n, \neg p\}$ where p stands for 'party A performs' and p for 'it gives notice to party B.' These SDL sentences have the same logical structure as the SDL sentences of T_2.

At first sight it may seem that a solution for T_3 also solves T_2. However, the two examples are not the same, because they have different temporal references. Their logical representations are only the same, because we left the temporal representation implicit. In the following section we show that this makes a fundamental difference.

Temporal Deontic Logic (TDL)

Since the late seventies, several temporal deontic logics and deontic action logics were introduced, which formalize satisfactorily a special type of CTD obligations, see for example (Thomason 1981; van Eck 1982; Loewer & Belzer 1983). In this section we illustrate the TDL analysis of the paradox using a logic recently proposed by Horty (Horty & Belnap 1995; Horty 1996), based on a seeing-to-it-that (stit) operator. A stit-frame $\langle Tree, <, Agent, Choice, Ought\rangle$ is based on a picture of moments ordered into a tree-like structure $(Tree, <)$, with forward branching representing the openness or indeterminacy of the future, and the absence of backward branching representing the determinacy of the past. $Choice^m_\alpha$ is the partition of the set of histories through moment m for agent α, which represents that at any moment in time, an agent can choose between several sets of histories. $Ought(m)$ is a subset of the set of histories through moment m, representing the good histories.

There are several ways to define actions and obligations in these stit-frames.[6] We say that two histories are m-indistinguishable if at moment m they are in the same equivalence class of the partition. A history is preferred to another one if all histories m-indistinguishable from the first are good, and there is a bad history m-indistinguishable from the second (Horty & Belnap 1995, p.592). Finally, we conditionalize on β by only considering histories in which β is true. We denote the

[4] In other words, the man ought to tell his neighbors that he will come, because *otherwise he will violate an obligation.* If he does not tell them and later he goes to the assistance, then he violates the second obligation. If he does not tell them and later he does not go, then he violates the first obligation. In the ideal state the man tells his neighbors that he will come and he goes. According to the semantics of SDL, the obligations $O(t \wedge a)$ and Ot should be derived.

[5] The most serious practical problem is caused by the fact that some norms are defined vaguely (called open texture) such that they are applicable in unforeseen circumstances. As a consequence, criminal law is more difficult to formalize than for example contract law.

[6] A simple definition in the style of SDL is the following. It is obligatory to see to it that α if and only if on all good histories $\mathsf{stit}_H : \alpha$ is true (Horty & Belnap 1995, p.616). This is a very strong definition. For example, consider the choice between the two sets $\{h_1, h_2\}$ and $\{h_3, h_4\}$ where h_4 is the only bad history. It seems that we have to choose the first set, because then we always end up in a good history. However, this is not derived from the definition above, because we have $M, m, h_3 \not\models \mathsf{stit}_H : \alpha$, whereas h_3 is a good history. The example hints at a definition based on comparing different choices by a so-called dominance function. It is obligatory to see to it that α if and only if all histories in which $\mathsf{stit}_H : \alpha$ is true dominate histories in which $\mathsf{stit}_H : \alpha$ is false. See (Horty 1996; Horty & Belnap 1995) for different notions of dominance.

obligations by O_{ABC}, because they presuppose that the Antecedent occurs Before the Consequent, as is illustrated by the example below.

Definition 1 **(stit_H and O_{ABC})** *A stit_H-frame $\langle Tree, <, Agent, Choice, Ought \rangle$ consists of a picture of moments ordered into a tree-like structure with forward branching and the absence of backward branching $(Tree, <)$, a set of agents Agent, a partition of the set of histories through moment m for agent α $Choice^m_\alpha(h)$, and a subset of the set of histories through moment m called $Ought(m)$. Two histories are m-indistinguishable if at moment m they are in the same equivalence class of the partition $Choice^m_\alpha$. The agent sees to it that α at moment m on history h, denoted by $M, m, h \models \mathsf{stit}_H : \alpha$, if and only if all histories m-indistinguishable from h make α true, and there is a history through m that does not make α true. We have $M \models O_{ABC}(\mathsf{stit}_H : \alpha \mid \beta)$ if and only if for all moments m and histories h_1, h_2 such that $M, m, h_1 \models \beta \wedge \mathsf{stit}_H : \alpha$ and $M, m, h_2 \models \beta \wedge \neg \mathsf{stit}_H : \alpha$ we have that all β-histories m-indistinguishable from h_1 are good and there is a bad β-history m-indistinguishable from h_2, and such m, h_1 and h_2 exist.*

Horty's logic satisfactorily formalizes the Convention on Contracts example, but not the assistance of neighbors example. Let M be the stit_H-model of $T = \{O_{ABC}(\mathsf{stit}_H : p \mid \top), O_{ABC}(\mathsf{stit}_H : \neg n \mid p), O_{ABC}(\mathsf{stit}_H : n \mid \neg p)\}$ represented in Figure 2. We leave it implicit that

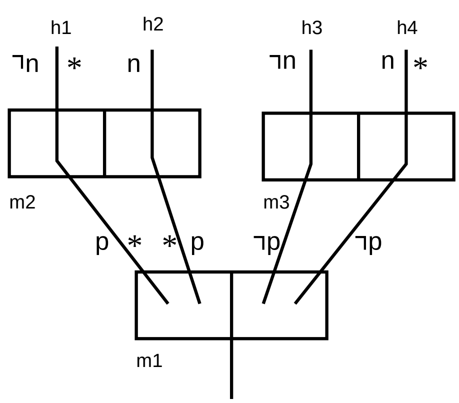

Figure 2: Convention on Contracts

this model only represents the obligations of agent A. This figure should be read as follows. The upward direction represents the forward direction of time, and a box represents a partition of histories at a moment. Good histories are represented by an asterisk '$*$' (for the moment just below it).

First consider the action model. The tree structure represents that party A first has to choose between '$\mathsf{performing}$' p and '$\mathsf{not\ performing}$' $\neg p$, and secondly between '$\mathsf{giving\ notice}$' n and '$\mathsf{not\ giving\ notice}$' $\neg n$. At moment m_1 party A sees to it that p or it sees to it that $\neg p$. For example, we have $M, m_1, h_1 \models \mathsf{stit}_H : p$, because p is true on histories m_1-indistinguishable from h_1 (i.e. h_1 and h_2) and there is a history at m that does not make p true (histories h_3 and h_4). Analogously, at moment m_2 or m_3 it sees to it that n or it sees to it that $\neg n$. Given this action model, consider the deontic model. We have $M \models O_{ABC}(\mathsf{stit}_H : p \mid \top)$ because at moment m_1 the histories h_3 and h_4 are bad histories, $M \models O_{ABC}(\mathsf{stit}_H : \neg n \mid p)$ because at moment m_2 history h_2 is bad, and $M \models O_{ABC}(\mathsf{stit}_H : n \mid \neg p)$ because at moment m_3 history h_3 is bad. Summarizing, history h_1 is the only good history, and history h_3 is a double-violation history. At each moment it is clear what should be done, because for each moment there is an obligation that prescribes the choice of party A.

Now consider the action model of the assistance of neighbors example represented in Figure 3. Notice that

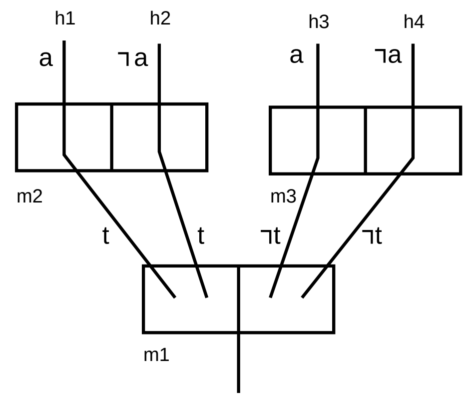

Figure 3: The Chisholm paradox

the deontic part of the model (i.e. $Ought(m)$) has not been specified in the figure. First, the agent has to choose between '$\mathsf{telling}$' t and '$\mathsf{not\text{-}telling}$' $\neg t$, and secondly the agent has to choose between '$\mathsf{going\ to\ the\ assistance}$' a and '$\mathsf{not\text{-}going\ to\ the\ assistance}$' $\neg a$. Given this action model, the problem is how to define the deontic part of the model such that the three obligations $O_{ABC}(\mathsf{stit}_H : a \mid \top)$, $O_{ABC}(\mathsf{stit}_H : t \mid a)$ and $O_{ABC}(\mathsf{stit}_H : \neg t \mid \neg a)$ are true. The fundamental problem is that the model cannot validate the premises, regardless of the choice of $Ought(m)$. For example, we have $M \not\models O_{ABC}(\mathsf{stit}_H : t \mid a)$, because once a is settled (moment m_2 or m_3) the man can no longer see to it that t. The truth value of t is already fixed.

There is an underlying problem in the action model. The man does not see to it that $(t \wedge a)$ at moment m_1 on history h_1, i.e. we have $M, m_1, h_1 \not\models \mathsf{stit}_H : (t \wedge a)$, because not all histories m_1-indistinguishable from h_1 make $(t \wedge a)$ true. However, intuitively the man can see to it that $(t \wedge a)$ by first choosing h_1 and h_2 at moment m_1, and thereafter choosing h_1 at moment m_2. Moreover, the man is able to see to it that the ideal state is reached in this way. This intuitive deontic reasoning in the Chisholm paradox can be formalized, if the man can reason about sequences of actions. In the following section we show how this can be achieved.

Pref.-based Temporal Deontic Logic (PTDL)

In this section we analyze the Chisholm paradox in Preference-based Temporal Deontic Logic (PTDL). We show how the temporal deontic logic discussed in the previous section can be extended with preferential notions by formalizing a suggestion from Horty (Horty 1996, Section 7.1), see also (Horty 1997). To reason about preferred strategies we adapt the definitions introduced in the previous section in three ways.

stit_H First, we adapt the definition of stit_H such that strategies are taken into account. We define an action which not only considers the choices at a moment, but also the choices the agent can make in the future. We replace the definition of m-indistinguishable by a global definition of indistinguishable. We consider two histories indistinguishable if there is not any moment in which we can distinguish them. We call the operator stit_S, where 's' stands for strategies.

O_{ABC} Second, we adapt the definition of O_{ABC} with a notion of dominance for strategies. We therefore need to adapt *Ought* of the stit_H-frame such that it represents a preference ordering on histories instead of a binary distinction between good and bad (Horty & Belnap 1995, p.617). We change the dominance function to the following one: a set of histories is preferred to a second one if each history in the first set is at least as good as each history in the latter set, and there is a history in the second set which is worse than all histories in the first set. This is an arbitrary choice, and other (more complicated) definitions may be preferred, see (Horty 1996, Section 7.2).

O_{ABC} Third, we also adapt the definition of O_{ABC} for a conditionalization on β such that the antecedent can be later than the consequent. We call a history a β-history if β is true at some moment of it. It is implicitly assumed that propositions formalize facts that cannot change over time. For example, we cannot write s for 'Ron is smoking,' but we have to use 'Ron is smoking at moment t.'

Definition 2 (stit_S and O) *A stit_S-frame $\langle Tree, <, Agent, Choice, Ought\rangle$ is a stit_H-frame, where $Ought(m)$ is a preference ordering on the set of histories through moment m. Two histories are indistinguishable if at any moment m they are m-indistinguishable. The agent sees to it that (stit_S) α at moment m on history h, denoted by $M, m, h \models \mathsf{stit}_S : \alpha$, if and only if all histories indistinguishable from h make α true, and there is a history through m that does not make α true. We have $M \models O(\mathsf{stit}_S : \alpha|\beta)$ if and only if for all moments m and all β-histories h_1, h_2 such that $M, m, h_1 \models \mathsf{stit}_H : \alpha$ and $M, m, h_2 \models \neg\mathsf{stit}_S : \alpha$ we have that all β-histories indistinguishable from h_1 are at least as good as each β-history indistinguishable from h_2, and there is a β-history indistinguishable from h_2 which is worse than each β-history indistinguishable from h_1, and such m, h_1 and h_2 exist.*

Consider the set of conditional obligations $T = \{O(\mathsf{stit}_S : a \mid \top), O(\mathsf{stit}_S : t \mid a), O(\mathsf{stit}_S : \neg t \mid \neg a), \neg a\}$ and let M be a model of T represented in Figure 3 with the *Ought* ordering $h_1 > h_3 > h_4 > h_2$. We have $M, m_1, h_1 \models \mathsf{stit}_S : (t \wedge a)$ and $M, m_2, h_1 \models \mathsf{stit}_S : (t \wedge a)$. Moreover, we have $M \models O(\mathsf{stit}_S : a \mid \top)$, because we have the obligation 'go to the assistance' deontically prefers history h_1 and h_3, $M \models O(\mathsf{stit}_S : t \mid a)$, because the obligation 'tell that you go if you go' prefers history h_1 to h_3, and $M \models O(\mathsf{stit}_S : \neg t \mid \neg a)$, because the obligation 'do not tell that you go if you do not go' prefers history h_4 to h_2. We have that the agent first ought to see to it that t and thereafter ought to see to it that a, $M \models O(\mathsf{stit}_S : t \wedge a \mid \top)$, because history h_1 is preferred to all other histories. We do not have that the agent ought to see to it that t, $M \not\models O(\mathsf{stit}_S : t|\top)$, because at m_1 history h_2 is not as good as history h_3 and h_4.

Summarizing, taking two moments together in consideration we can derive the obligation to tell as part of a more complex action, but there is not an obligation to tell simpliciter.

Preference-based Deontic Logic (PDL)

Hansson (1971) argues that the fundamental problem underlying the CTD paradoxes is that the type of possible world semantics of SDL is not flexible enough. In these semantics only two types of worlds are distinguished in a model; *actual* and *ideal* ones. The ideal worlds have to satisfy all obligations in a deontic theory T. Clearly, if these obligations contradict each other, then a problem arises. We must use more complicated value structures that somehow bear information about comparisons or gradations of value.

Definition 3 *Let $M = \langle W, \leq, V\rangle$ be a standard possible worlds model with W a nonempty set of worlds, $\leq$ a binary reflexive, transitive and connected accessibility relation, and V a valuation function of the propositions at the worlds. We have $M \models O_{HL}(\alpha|\beta)$ if and only if there is a world w_1 such that $M, w_1 \models \alpha \wedge \beta$ and for all worlds $w_2 \leq w_1$ we have $M, w_2 \models \beta \rightarrow \alpha$.*

A typical preference-based model M of the Chisholm set $T = \{O_{HL}(a \mid \top), O_{HL}(t \mid a), O_{HL}(\neg t \mid \neg a), \neg a\}$ is represented in Figure 4. This figure should be

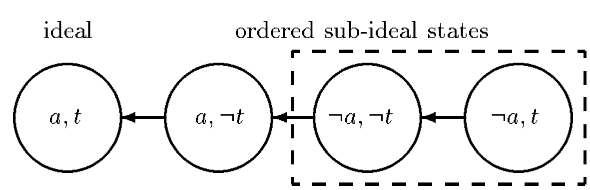

Figure 4: Preference relation of Chisholm's paradox

read as follows. A circle represents a nonempty set of worlds that satisfies the propositions written within them. An arrow represents strict preference for all the worlds represented by the circle. The transitive closure is left implicit. The dashed box represents the set of worlds which might be the actual world, see e.g. (Hansson 1971) for a discussion on the interpretation of circumstances in preference-based logics. We have $M \models O_{HL}(a \wedge t | \top)$ and $M \models O_{HL}(t | \top)$, because the logic has the theorem $\models O_{HL}(\alpha|\beta) \wedge O_{HL}(\beta|\top) \rightarrow O_{HL}(\alpha|\top)$.

In previous papers (Tan & van der Torre 1996; van der Torre & Tan 1997) we argued that PDL is the *minimal* logic to analyze CTDs. The analyses given in this paper are in accord with this arguments. In particular, it is clear that the PTDL analysis of the Chisholm paradox in the previous section is analogous to the PDL analysis, in the sense that a history is a world in which we added temporal structure explicitly. Moreover, the analysis of Chisholm's paradox in Hansson's PDL and the analysis in our new PTDL are both based on rejection of factual detachment. There are several ways in which our PTDL can be extended with alternative notions of factual detachment developed in TDL and PDL (for the latter see e.g. (van der Torre 1997)).

Conclusions

In previous work we introduced preference-based frameworks for deontic reasoning and in this paper we propose a deontic logic that combines preferential and temporal notions. We used the logic to analyze the backward version of Chisholm's paradox. We showed how the paradox can be analyzed if temporal and preferential notions are represented explicitly. Moreover, the analysis of the interaction between preferential and temporal notions is a first step towards the analysis of the dynamics of obligations. In (van der Torre & Tan 1998b) we discuss an alternative way to combine preferences and time by formalizing prescriptive obligations in update semantics. However, the logics are not expressive enough yet to formalize all aspects of preferences that change in time. This is subject of present research.

References

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

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

Feldmann, F. 1990. A simpler solution to the paradoxes of deontic logic. In Tomberlin, J., ed., *Philosophical perspectives 4: Action theory and Philosophy of Mind.* Atascadero: Ridgview.

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

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

Horty, J. 1996. Agency and obligation. *Synthese* 108:269–307.

Horty, J. 1997. *Agency and Deontic Logic.* Preliminary draft.

Kimbrough, S., and Lee, R. 1997. Formal aspects of electronic commerce: research issues and challenges. *International Journal of Electronic Commerce* 1.

Loewer, B., and Belzer, M. 1983. Dyadic deontic detachment. *Synthese* 54:295–318.

Pearl, J. 1993. A calculus of pragmatic obligation. In *Proceedings of Uncertainty in Artificial Intelligence (UAI'93)*, 12–20.

Smith, T. 1994. *Legal expert systems: discussion of theoretical assumptions.* Ph.D. Dissertation, University of Utrecht.

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

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

Tomberlin, J. 1981. Contrary-to-duty imperatives and conditional obligation. *Noûs* 16:357–375.

van der Torre, L., and Tan, Y. 1995. Cancelling and overshadowing: two types of defeasibility in defeasible deontic logic. In *Proceedings of the IJCAI'95*, 1525–1532.

van der Torre, L., and Tan, Y. 1997. The many faces of defeasibility in defeasible deontic logic. In Nute, D., ed., *Defeasible Deontic Logic.* Kluwer. 79–121.

van der Torre, L., and Tan, Y.-H. 1998a. Diagnosis and decision making in normative reasoning. *Artificial Intelligence and Law.* To appear.

van der Torre, L., and Tan, Y. 1998b. An update semantics for prima facie obligations. In *Proceedings of the ECAI'98.* To appear.

van der Torre, L. 1997. *Reasoning About Obligations: Defeasibility in Preference-based Deontic Logics.* Ph.D. Dissertation, Erasmus University Rotterdam.

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

Vorobej, M. 1986. Conditional obligaton and detachment. *Canadian Journal of Philosophy* 16:11–26.

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

Yu, X. 1995. *Deontic Logic with Defeasible Detachment.* Ph.D. Dissertation, University of Georgia.