

# Strategic Deontic Temporal Logic as a Reduction to ATL, with an Application to Chisholm’s Scenario

Jan Broersen

Department of Information and Computing Sciences  
Utrecht University  
Utrecht, The Netherlands  
broersen@cs.uu.nl

**Abstract.** In this paper we extend earlier work on deontic deadlines in CTL to the framework of alternating time temporal logic (ATL). The resulting setting enables us to model several concepts discussed in the deontic logic literature. Among the issues discussed are: conditionality, ought implies can, deliberateness, settledness, achievement obligations versus maintenance obligations and deontic detachment. We motivate our framework by arguing for the importance of temporal order obligations, from the standpoint of agent theory as studied in computer science. In particular we will argue that in general *achievement* obligations cannot do without a deadline condition saying the achievement has to take place before it. Then we define our logic as a reduction to ATL. We demonstrate the applicability of the logic by discussing a possible solution to Chisholm’s paradox. The solution differs considerably from other known temporal approaches to the paradox.

## 1 Introduction

In agent theory, as studied in computer science, we are interested in designing logical models that describe how agents can reason about and decide what to do, given their obligations, permissions, abilities, desires, intentions, beliefs, etc. Decisions have a temporal aspect, namely, they are about what to do in the *future*, and they deal with conditional information, namely, they have to result from considering and reasoning about *hypothetical* circumstances. The deontic ATL operators we consider in this paper are both *conditional* and *temporal*. Their syntactical form is  $O_A(\rho \leq \delta : \xi_A)$ . The intuitive interpretation of the operator is that *if* the agents in the set  $A$  achieve  $\delta$ , they are obliged to achieve  $\rho$  at the same time or before that, under penalty of suffering the negative condition  $\xi_A$ . A good example of such an obligation is the following: according to Dutch traffic regulations one has to indicate direction before one turns off. In this example,  $\delta$  is ‘turning off’,  $\rho$  is ‘indicating direction’ and  $\xi$  can be the circumstance of being vulnerable for being fined by a police officer. Obligations  $O_A(\rho \leq \delta : \xi_A)$  are thus *conditional* on the actual occurrence of  $\delta$  and are *temporal* in the sense that the achievement  $\rho$  has to precede the condition  $\delta$ . Readers familiar with the deontic logic literature will recognize that another example is the second sentence of Chisholm’s original paradoxical scenario: ‘if one helps, first one has to tell’. In section 9 we discuss formalizations of Chisholm’s scenario in our formalism.

One might wonder why we think obligations expressed as  $O_A(\rho \leq \delta : \xi_A)$  are so important. Let us explain. Obligations that guide agents in the actions they select for

performing are always *about* the future. Obligations about the past are not interesting for agents having to decide what to do, because we may safely assume that agents do not *control* the past. We adopt terminology from BDI-theory, and say that the *reference* time of the obligations we are interested in is the future, while the *validity* time is the present<sup>1</sup>. The latter emphasizes that we want our logic to model the reasoning of an agent that has to decide what to do ‘now’, considering the obligations he has about the future. A rough classification of obligations whose reference time is the future is the division in *achievement* obligations and *maintenance* obligations. Similar terminology is used by Cohen and Levesque [8] who distinguish between *achievement goals* and *maintenance goals*. In an achievement obligation, the objective is to achieve something in the future that is not already (necessarily) true now. For a ‘maintenance obligation’ the objective is to preserve the truth of a condition that *is* already true now. Our main interest in this paper will be with achievement obligations, since as we will see in section 6, maintenance obligations can be rewritten into equivalent achievement obligations. So, for agent theory as studied in computer science achievement obligations are the most interesting type of norms. Now we will argue in section 2 that achievement obligations are close to meaningless without a condition  $\delta$  before whose occurrence the achievement  $\rho$  has to be realized, which explains why obligations of the form  $O_A(\rho \leq \delta : \xi_A)$  are central to our investigations.

In some of the previous work on this subject [4], we referred to the condition  $\delta$  as a ‘deadline’ of an obligation  $O_A(\rho \leq \delta : \xi_A)$ . That was partly because there we studied this type of modality in the purely temporal framework of CTL. Here we use ATL for the temporal component. ATL has elements of logics of Agency. In [5] we showed how to embed Coalition Logic (CL), which is a subset of ATL, in the STIT framework of Horty [11]<sup>2</sup>. Since ATL can be seen as a logic of (strategic) ability it enables us to define different notions of control over conditions. And adding information about control over the condition  $\delta$  (or, to be more precise, absence of control over  $\neg\delta$ , which is something else) is actually what can turn a conditional temporal order obligation into a real deadline obligation, as we explain in section 8. So obligations  $O_A(\rho \leq \delta : \xi_A)$  as such should not be referred to as ‘deadline’ obligations. They are conditional temporal order obligations, which can be made into deadline obligations, by adding that agents  $A$  do not control avoidance of the deadline *condition*  $\delta$ .

## 2 Why Achievement Obligations Need a ‘Deadline’ Condition

Dignum et al [9, 17] stress the importance of providing deadlines (which they do *not* view as particular kinds of *conditionals*, like we do) for obligations from practical considerations. And indeed, in the environments where software agents are envisioned

<sup>1</sup> The distinction between validity time and reference time for logics that contain a temporal modality as one of the logic components, was, for instance, formulated by Lindström and Rabinowicz [13] in the context of temporal epistemic logics. But it equally applies to temporal motivational logics. And we believe that a failure to distinguish the two concepts is the source of a lot of confusion.

<sup>2</sup> And a paper on embedding ATL as a whole into the *strategic* version of Horty’s STIT formalism is under review.

to operate, they will have to deal with deadlines. Think about agents negotiating certain deadlines, to coordinate their behavior. For instance, agents may engage in mutual agreements by means of contracts, which typically contain deadlines.

However, there are more fundamental reasons to equip achievement obligations with deadline conditions. A fundamental assumption in deontic logic is that norms (obligations) can be violated, and that we should be able to reason about this circumstance. Now, if we think about the situation where an agent has an achievement obligation  $\varphi$  for, say, ‘stop smoking’, and we want to make the temporal component of  $\varphi$  explicit, then we cannot model this by writing for  $\varphi$  a formula like  $OF\ stop\_smoking$  ( $O\psi$  for ‘it is obliged that  $\psi$ ’ and  $F\chi$  for ‘some time in the future  $\chi$ ’), because we cannot *violate* this obligation. At any future point, the agent can claim that although he has not stopped smoking yet, he will eventually, at some point even further in the future. Most deontic logicians would agree that obligations that cannot be violated are no obligations at all. So what it takes for an achievement obligation to really be an obligation is reference to a *condition* under which it is violated. And this is exactly what a deadline condition is: a condition giving rise to a violation if the achievement has not been accomplished before. If this condition corresponds to a given point in some time metric, we have a genuine deadline obligation. But if this condition is an abstract proposition  $\delta$ , we have a special kind of conditional obligation, namely, a temporal order obligation.

A possible objection against the above line of reasoning is that sometimes there do seem to be ways in which to violate an achievement obligation without a deadline. Before explaining the objection, we need to point out that obligations play multiple roles in rational agent modelling. There are two abstraction levels on which they play a role. First there are achievement obligations incurred and represented by agents. These obligations should have a deadline, since otherwise they cannot be violated which means that they do not influence the agents decision making. But there are also achievement obligations that function as specifications for the agents behavior as seen by an agent designer. A good example is the formal property of *fairness*. For instance, a designer might specify that his agent is obliged to distribute its deliberation efforts fairly over its set of goals. The design of the agent may violate this fairness obligation. But note that in general this is not something the agent itself is concerned with. The agent is designed as it is, and it cannot choose to have another design. And thus it cannot violate an achievement obligation without a deadline. Or can he? Of course, we can imagine that an agent indeed is able to alter its own design, thereby violating a certain achievement obligation. For instance, we might claim that an agent violates the achievement obligation to shake hands with president Bush someday by cutting off his own hands. A similar objection is that an agent might perform something irrevocable in its environment. For instance, an agent can be said to violate the achievement obligation to bring back the book to the library some day by burning it<sup>3</sup>.

We acknowledge these as valid objections against our claim that achievement obligations need deadlines. However, there are still some good reasons to claim that deadline conditions are crucial for achievement obligations. For instance, an agent cannot destroy

---

<sup>3</sup> What if we are able to reproduce exactly the same book? Then, apparently burning did not *count as* irrevocably destroying the possibility to comply.

the possibility to be able to stop smoking, so at least for some achievement obligations a deadline is necessary. Furthermore, one seems to have a contrived interpretation of an obligation like ‘having to return a book to the library some day’ if one claims that it does not imply that one *actually* has to bring it back but *only* that one should not irrevocably destroy it.

Our suggestion, is that we can study the difference between, for instance the smoking example and the library example as a difference in the interaction with abilities. In section 8 we will mention several ways in which obligations act with ability. One interaction with ability is that we can choose to define that an obligation for achieving  $\varphi$  implies that an agent is obliged to keep open the possibility of reaching  $\varphi$  (we will however not discuss this interaction in section 8). If we choose to include this property, indeed, in the library example we violate the achievement obligation to bring back the book eventually, by burning the book. However, for the smoking example there is no difference: it is hard to imagine how an agent can violate an obligation to keep open the possibility to stop smoking.

### 3 Reduction Using Negative Conditions

As explained, the syntax of the central modality we study in this paper is  $O_A(\rho \leq \delta : \xi_A)$  (although we permit ourselves a brief generalization to a conditional variant  $O_A(\rho \leq \delta : \xi_A \mid \eta)$  in section 9). The  $O$  stands for obligation,  $A$  is a group of agents having the obligation,  $\rho$  is the condition to be achieved,  $\delta$  the condition functioning as the deadline for the achievement, and  $\xi_A$  a condition necessarily true in case the deadline obligation is violated. We think of  $\xi_A$  as a condition that is in some sense negative for the group of agents  $A$ . Negative conditions typically play a role in the semantics of notions like obligation, commitment and intention. Goals, desires, wants, wishes, objectives, etc. are typically associated with positive conditions. Our approach differs from most others working with negative condition [3, 15] in that we make the negative condition explicit in the syntax of the obligation modalities. In standard deontic logic (SDL) [18], the negative conditions are implicit in the modal semantics: modally accessible worlds are optimal worlds where no violations occur. In Anderson’s reduction for SDL [3], the negative conditions are explicit in the object language through a propositional constant *Viol*. In this paper we go one step further by explicitly giving the negative conditions  $\xi_A$  as parameters for the obligation operator  $O_A(\rho \leq \delta : \xi_A)$ . This has many advantages. For instance, it gives us the machinery to specify that certain obligations are incurred as the result of violating other obligations. That is, we can nest obligations by using formulas of the form  $O_A(\rho \leq \delta : \xi_A)$  for  $\xi_A$ . However, we do not consider such nestings in this paper. The most important reason for making the negative conditions explicit, is the advantage this has in the study of Chisholm’s scenario in section 9. It enables us to view the choice of which *temporal* obligation to comply to as a decision about which *non-temporal* negative condition to prefer. Obviously, such decisions can only be made relative to a preference order over negative conditions. In this paper such an order is not made explicit, since here we are only concerned with a possible definition for temporal obligations *in terms* of negative conditions. Possible logical structures for negative conditions themselves are not explored.

Also, allowing general negative conditions  $\xi_A$  enables us to discuss some principles of deontic logic. For instance, in case  $\xi_A = \perp$ , the norm is ‘regimented’. Actually this means that the norm is no longer a norm, since it is no longer possible to violate it; the norm simply becomes a hard constraint on the behavior of agents. Thus  $O_A(\rho \leq \delta : \perp)$  means that  $A$  can only do  $\rho$  before  $\delta$ , and nothing else. In terms of strategies: the agent has no strategy where eventually he meets a  $\delta$  without having met a  $\rho$  first. The case  $\xi_A = \top$  gives a dual effect: now the penalty is always true, which means that it loses its meaning as a divider between ‘the good’ and ‘the bad’. As a result,  $O_A(\rho \leq \delta : \xi_A)$  loses its normative meaning; any point of achieving  $\rho$  or  $\delta$  becomes as good as any other. This means that  $O_A(\rho \leq \delta : \top)$  has to be generally valid, as is the case for our definitions for the operator.

Our aim is to define the semantics of the modality  $O_A(\rho \leq \delta : \xi_A)$  entirely by constructing reductions to formulas of alternating time temporal logic (ATL) [1, 2] talking about negative conditions. ATL is a temporal logic of agency with a game theoretic semantics in terms of *strategies*. The deadline obligations we define will then be strategic obligations in the sense that they limit the possible strategies of agents by associating negative conditions with courses of agent choices that do not comply to the obligation.

Reducing to ATL has many advantages. We can use the logical machinery of ATL (axiomatization, model checking algorithms), to do reasoning. We can check properties of deontic logics by translating and proving them in ATL. Finally, we can do planning with obligations and deadlines using a satisfiability checker for ATL. We do not have to be too afraid that ATL, as a formal system, might be too weak to encode interesting properties, since it has been shown to have an exponential time complete complexity.

## 4 ATL

We present ATL ([1, 2]) here using a non-standard, but concise and intuitive syntax and semantics.

### 4.1 Core Syntax, Abbreviations and Intuitions

**Definition 1.** *Well-formed formulas of the temporal language  $\mathcal{L}_{ATL}$  are defined by:*

$$\begin{aligned} \varphi, \psi, \dots &:= p \mid \neg\varphi \mid \varphi \wedge \psi \mid \langle [A] \rangle \eta \mid [[A]] \eta \\ \eta, \theta, \dots &:= \eta U^{ee} \theta \end{aligned}$$

where  $\varphi, \psi, \dots$  represent arbitrary well-formed formulas,  $\eta, \theta, \dots$  represent temporal path formulas, the  $p$  are elements from an infinite set of propositional symbols  $\mathcal{P}$ , and  $A$  is a subset of a finite set of agent names  $E$  (we define  $\bar{A} \equiv_{def} E \setminus A$ ). We use the superscript ‘ee’ for the until operator to denote that this is the version of ‘the until’ where  $\varphi$  is not required to hold for the present, nor for the point where  $\psi$ , i.e., the present and the point where  $\phi$  are *both* excluded. Roughly,  $\langle [A] \rangle \eta$  is read as ‘ $A$  can ensure  $\eta$ ’, and the dual  $[[A]] \eta$  is read as ‘ $A$  cannot avoid  $\eta$ ’. A more precise explanation, revealing the existential and universal quantification over strategies in both these operators (which

explains our choice of syntax using a combination of sharp and square brackets for both operators) is as follows:

- $\langle [A] \rangle (\varphi U^{ee} \psi)$ : agents  $A$  have a strategy that, whatever strategy is taken by agents  $\bar{A}$ , ensures that eventually, at some point  $m$ , the condition  $\psi$  will hold, while  $\varphi$  holds from the next moment until the moment before  $m$
- $[\langle A \rangle] (\varphi U^{ee} \psi)$ : for all strategies of agents  $A$  the agents  $\bar{A}$  have a strategy such that eventually, at some point  $m$ , the condition  $\psi$  will hold, while  $\varphi$  holds from the next moment until the moment before  $m$

We use standard propositional abbreviations, and also define the following operators as abbreviations.

### Definition 2

$$\begin{array}{ll}
 \langle [A] \rangle X\varphi \equiv_{def} \langle [A] \rangle (\perp U^{ee} \varphi) & [\langle A \rangle] X\varphi \equiv_{def} [\langle A \rangle] (\perp U^{ee} \varphi) \\
 \langle [A] \rangle F\varphi \equiv_{def} \varphi \vee \langle [A] \rangle (\top U^{ee} \varphi) & [\langle A \rangle] F\varphi \equiv_{def} \varphi \vee [\langle A \rangle] (\top U^{ee} \varphi) \\
 \langle [A] \rangle G\varphi \equiv_{def} \neg [\langle A \rangle] F\neg\varphi & [\langle A \rangle] G\varphi \equiv_{def} \neg [\langle A \rangle] F\neg\varphi \\
 \langle [A] \rangle (\varphi U^e \psi) \equiv_{def} \varphi \wedge \langle [A] \rangle (\varphi U^{ee} \psi) & [\langle A \rangle] (\varphi U^e \psi) \equiv_{def} \varphi \wedge [\langle A \rangle] (\varphi U^{ee} \psi) \\
 \langle [A] \rangle (\varphi U \psi) \equiv_{def} \langle [A] \rangle (\varphi U^e (\varphi \wedge \psi)) & [\langle A \rangle] (\varphi U \psi) \equiv_{def} [\langle A \rangle] (\varphi U^e (\varphi \wedge \psi)) \\
 \langle [A] \rangle (\varphi U_w \psi) \equiv_{def} \neg [\langle A \rangle] (\neg \psi U \neg \varphi) & [\langle A \rangle] (\varphi U_w \psi) \equiv_{def} \neg [\langle A \rangle] (\neg \psi U \neg \varphi)
 \end{array}$$

The informal meanings of the formulas are as follows (the informal meanings in combination with the  $[\langle A \rangle]$  operator follow trivially):

- $\langle [A] \rangle X\varphi$  : agents  $A$  have a strategy to ensure that at any next moment  $\varphi$  will hold
- $\langle [A] \rangle F\varphi$  : agents  $A$  have a strategy to ensure that eventually  $\varphi$  will hold
- $\langle [A] \rangle G\varphi$  : agents  $A$  have a strategy to ensure that holds globally
- $\langle [A] \rangle (\varphi U^e \psi)$ : agents  $A$  have a strategy to ensure that, eventually, at some point  $m$ , the condition  $\psi$  will hold, while  $\varphi$  holds from now until the moment before  $m$
- $\langle [A] \rangle (\varphi U \psi)$ : agents  $A$  have a strategy to ensure that, eventually, at some point the condition  $\psi$  will hold, while  $\varphi$  holds from now until then
- $\langle [A] \rangle (\varphi U_w \psi)$ : agents  $A$  have a strategy to ensure that, if eventually  $\psi$  will hold, then  $\varphi$  holds from now until then, or forever otherwise

## 4.2 Model Theoretic Semantics

The intuition behind ATL models is that agents have choices, such that the non-determinism of each choice is *only* due to the choices other agents have at the same moment. Thus, the simultaneous choice of all agents together, always brings the system to a unique follow-up state. In other words, if an agent would know what the choices of other agents would be, given his own choice, he would know exactly in which state he arrives.

**Definition 3.** An ATL model  $\mathcal{M} = (S, C, \pi)$ , consists of a non-empty set  $S$  of states, a total function  $C : A \times S \mapsto 2^S$  yielding for each agent and each state a set of choices (informally: ‘actions’) under the condition that the intersection of each combination of choices for separate agents gives a unique next system state (i.e., for each  $s$ , the

function  $RX(s) = \{\bigcap_{a \in A} Ch_a \mid Ch_a \in C(a, s)\}$  yields a non-empty set of singleton sets representing the possible follow-up states of  $s$ , and, finally, an interpretation function  $\pi$  for propositional atoms.

Note that from the condition on the function  $C$  it follows that the choices for each individual agent at a certain moment in time are a partitioning of the set of all choices possible for the total system of agents, as embodied by the relation  $\mathcal{R}^{sys} = \{(s, s') \mid s \in S \text{ and } \{s'\} \in RX(s)\}$ . And, also note that this latter condition does not entail the former. That is, there can be partitions of the choices for the total system that do not correspond to the choices of some agent in the system.

**Definition 4.** A strategy  $\alpha_a$  for an agent  $a$ , is a function  $\alpha_a : S \mapsto 2^S$  with  $\forall s \in S : \alpha_a(s) \in C(a, s)$ , assigning choices of the agent  $a$  to states of the ATL model.

Often, strategies are defined as mappings  $\alpha_a : S^+ \mapsto 2^S$ , from finite sequences of states to choices in the final state of a sequence. However, to interpret ATL, this is not necessary, because ATL is not expressive enough to recognize by which sequence of previous states a certain state is reached. More in particular, without affecting truth of any ATL formula, we can always transform an ATL model into one where  $\mathcal{R}^{sys}$  is tree-like. On tree structures it is clear right away that a mapping from states to choices in that state suffices, since any state can only be reached by the actions leading to it.

The strategy function is straightforwardly extended to sets of agents.

**Definition 5.** A full path  $\sigma$  in  $M$  is an infinite sequence<sup>4</sup>  $\sigma = s_0, s_1, s_2, \dots$  such that for every  $i \geq 0$ ,  $s_i \in S$  and  $(s_i, s_{i+1}) \in \mathcal{R}^{sys}$ . We say that the full path  $\sigma$  starts at  $s$  if and only if  $s_0 = s$ . We denote the state  $s_i$  of a full path  $\sigma = s_0, s_1, s_2, \dots$  in  $M$  by  $\sigma[i]$ .

A full path  $\sigma$  complies to a strategy  $\alpha_A$  of a set of agents  $A$  if and only if for every  $n \geq 0$ ,  $\sigma[n+1] \in \alpha_A(\sigma[n])$ . We denote the set of full paths complying to a strategy  $\alpha_A$  by  $\Sigma(\alpha_A)$ .

**Definition 6.** Validity  $M, s \models \varphi$ , of an ATL-formula  $\varphi$  in a world  $s$  of a model  $M = (S, C, \pi)$  is defined as:

$$\begin{aligned}
M, s \models p & \Leftrightarrow s \in \pi(p) \\
M, s \models \neg\varphi & \Leftrightarrow \text{not } M, s \models \varphi \\
M, s \models \varphi \wedge \psi & \Leftrightarrow M, s \models \varphi \text{ and } M, s \models \psi \\
M, s \models \langle [A] \rangle \eta & \Leftrightarrow \exists \alpha_A s. t. \forall \sigma \in \Sigma(\alpha_A) \text{ with } \sigma[0] = s : M, \sigma[0], \sigma \models \eta \\
M, s \models [ \langle A \rangle ] \eta & \Leftrightarrow \forall \alpha_A : \exists \sigma \in \Sigma(\alpha_A) \text{ with } \sigma[0] = s s. t. M, \sigma[0], \sigma \models \eta \\
M, \sigma[0], \sigma \models \varphi U^{ee} \psi & \Leftrightarrow \exists n > 0 s. t. \\
& (1) M, \sigma[n] \models \psi \text{ and} \\
& (2) \forall i \text{ with } 0 < i < n : M, \sigma[i] \models \varphi
\end{aligned}$$

Validity on a ATL model  $M$  is defined as validity in all states of the model. If  $\varphi$  is valid on an ATL model  $M$ , we say that  $M$  is a model for  $\varphi$ . General validity of a formula

<sup>4</sup> Alternatively, we may drop the requirement that  $\mathcal{R}^{sys}$  is serial, and add a maximality condition to the notion of ‘full path’.

$\varphi$  is defined as validity on all ATL models. The logic ATL is the subset of all general validities of  $\mathcal{L}_{\text{ATL}}$  over the class of ATL models.

## 5 Conditional Temporal Order Obligations

In this section we define operators for temporal order obligations as reductions to ATL-formulas talking about negative conditions. We use ATL-formulas indexed with a set of agents, i.e.,  $\xi_A$ , to denote negative conditions. The central observation linking obligations  $O_A(\rho \leq \delta : \xi_A)$  to ATL, is the following:

*$O_A(\rho \leq \delta : \xi_A)$  holds if and only if it is not the case that the group of agents  $A$  has a strategy to achieve  $\delta$ , to avoid  $\rho$  at all moments until  $\delta$  occurs for the first time, and avoid the negative condition  $\xi_A$  at the point where  $\delta$ .*

In other words, if  $A$  want to achieve  $\delta$  at some future point, they have to make sure that before that they achieve  $\rho$ , because otherwise the negative condition  $\xi_A$  will be valid at the point where  $\delta$ . We can rewrite this formally as a truth condition on ATL models:

### Definition 7 (temporal order obligations)

$$\begin{aligned} \mathcal{M}, s \models O_A(\rho \leq \delta : \xi_A) &\Leftrightarrow \nexists \alpha_A, \forall \sigma \in \Sigma(\alpha_A) \text{ with } \sigma[0] = s, \exists j : \\ &\text{such that} \\ &\forall 0 \leq i < j : \mathcal{M}, \sigma[i] \models \neg\rho \wedge \neg\delta \text{ and } \mathcal{M}, \sigma[j] \models \neg\rho \wedge \delta \\ &\text{and} \\ &\mathcal{M}, \sigma[j] \models \neg\xi_A \end{aligned}$$

This says: *if* at some future point  $\delta$  occurs, than  $A$  has no way of ensuring that, if  $\rho$  has not occurred before the point where  $\delta$  occurs for the first time, there is not a negative condition  $\xi_A$  at the point where  $\delta$ . This means that if  $A$  do have a strategy to avoid the negative condition while not doing  $\rho$  before  $\delta$ , they do not have the obligation.

Under the above definition, in case only *some* strategies of agents may lead to negative conditions if they do not ensure that  $\rho$  is achieved before  $\delta$ , the agents are not *obliged* to achieve  $\rho$  before  $\delta$ . This situation actually constitutes a kind of conditionality other than the conditionality with respect to deadline conditions  $\delta$ . Modelling it would require an operator  $O_A(\rho \leq \delta : \xi_A \mid \eta)$ , where  $\eta$  is a temporal formula denoting the subset of paths the obligation holds on. Note that the original obligation reappears as the case where  $\eta$  equals  $\top$ . This kind of conditionality (which is not further explored in this paper) can be modelled using the more expressive variant  $ATL^*$ . We leave this extension and a discussion on the different kinds of conditionality that can be defined for temporal deontic operators for a future paper.

A second aspect of definition 7 that has to be explained is that it by no means implies that an obligation requires that  $\delta$  becomes true eventually (which is why it is conditional on  $\delta$ ). However, we do have that if  $A$  cannot avoid that  $\delta$  might never become true, they cannot have a strategy that ensures that at some point  $\delta$  will hold (and where if  $\rho$  has not been done before, there is not a negative condition), which means that they are obliged



every  $\rho$  before  $\delta$  (validity 10 in proposition 2 reflects this). This seems rather counter intuitive. However, in section 8 we define a deliberate version of the operator for which this property is eliminated.

The third thing to discuss is that intuitively, a strategic notion of obligation should distinguish between the strategies that are good and the strategies that are bad. However, our definition suggests that we can define the obligation in terms of what strategies agents *have*. The link between these two views is the use of the negative conditions and the conditionality with respect to occurrence of the condition  $\delta$ . Actually we can view the definition as distinguishing between good and bad strategies in the following sense: the strategies in which an agent eventually meets the condition  $\delta$  without having achieved the condition  $\rho$  before, are the bad strategies, all the others are the good ones.

We can circumscribe the truth condition of definition 7 as an ATL formula. We have the following proposition:

**Proposition 1.** *A formula  $O_A(\rho \leq \delta : \xi_A)$  is true at some point of an ATL model if and only if the point satisfies the ATL formula  $(\delta \wedge (\neg\rho \rightarrow \xi_A)) \vee \neg\langle[A]\rangle((\neg\rho \wedge \neg\delta)U^e(\delta \wedge \neg\rho \wedge \neg\xi_A))$ .*

*Proof.* We only give an impression of the proof. The present is not controlled by any strategy. If  $\delta$  holds presently, and  $\rho$  does not hold presently, there is a violation presently. In the truth condition this corresponds to the case  $j = 0$ , and in the formula to  $\delta \wedge (\neg\rho \rightarrow \xi_A)$ . Equivalence is easy to see. For moments other than the present, the equivalence follows almost directly from the semantics of the ATL operators involved.

## 6 Maintenance Obligations with a Relief Condition

In the introduction we explained what maintenance obligations are. Where achievement obligations for a property  $\rho$  naturally come with a property  $\delta$  functioning as a *deadline condition*, maintenance properties  $\varphi$  come with a property  $\psi$  functioning as a *relief condition*: if the relief condition occurs, the obligation to maintain  $\varphi$  no longer holds. We can define maintenance obligations  $O_A(\varphi \leftarrow \diamond \psi : \xi_A)$  in terms of achievement obligations as follows:

### Definition 8

$$O_A(\varphi \leftarrow \diamond \psi : \xi_A) \equiv_{def} O_A(\psi \leq \neg\varphi : \xi_A)$$

The rationale for the definition is as follows. An agent can comply to obligations  $O_A(\rho \leq \delta : \xi_A)$  in two different ways: (1) he can look at it as having to do  $\rho$  before he does  $\delta$ , but he can also (2) look at it as having to preserve  $\neg\delta$  as long as he has not achieved  $\rho$ . Note that for a maintenance obligation  $O_A(\varphi \leftarrow \diamond \psi : \xi_A)$ , the negative condition occurs at the first point where  $\varphi$  is no longer maintained, provided this point is before  $\psi$ . In section 9 we will use a maintenance obligation to model one of the sentences of Chisholm's scenario.

## 7 More Logical Properties

In this section we mention some logical properties of the defined obligation operator.

**Proposition 2.** *The following schemas are valid:*

$$\models O_A((\rho \wedge \chi) \leq \delta : \xi_A) \rightarrow O_A(\rho \leq \delta : \xi_A) \quad (1)$$

$$\models O_A(\rho \leq \top : \xi_A) \wedge O_A(\chi \leq \top : \zeta_A) \rightarrow O_A((\rho \wedge \chi) \leq \top : \xi_A \vee \zeta_A) \quad (2)$$

$$\models O_A(\top \leq \delta : \xi_A) \quad (3)$$

$$\models O_A(\gamma \leq \gamma : \xi_A) \quad (4)$$

$$\models O_A(\rho \leq \perp : \xi_A) \quad (5)$$

$$\models \neg O_A(\perp \leq \top : \perp_A) \quad (6)$$

$$\models \neg(O_A(\rho \leq \top : \xi_A) \wedge O_A(\neg\rho \leq \top : \xi_A)) \quad (7)$$

$$\models O_A(\rho \leq \delta : \top) \quad (8)$$

$$\models O_A(\rho \leq \delta : \perp) \rightarrow O_A(\rho \leq \delta : \xi_A) \quad (9)$$

$$\models [\langle A \rangle]G\neg\delta \rightarrow O_A(\rho \leq \delta : \xi_A) \quad (10)$$

$$\models O_A(\rho \leq \delta : \xi_A) \rightarrow [\langle A \rangle](O_A(\rho \leq \delta : \xi_A)U_w(\rho \vee \delta)) \quad (11)$$

$$\models O_A(\rho \leq \delta : \xi_A) \rightarrow O_A(O_A(\rho \leq \delta : \xi_A) \leftarrow \diamond (\rho \vee \delta) : \xi_A) \quad (12)$$

$$\models \xi_A \rightarrow O_A(\rho \leq \top : \xi_A) \quad (13)$$

$$\models O_A(\rho \leq \xi_A : \xi_A) \quad (14)$$

**Proposition 3.** *The following schemas are not valid:*

$$\not\models O_A(\rho \leq \delta : \xi_A) \wedge O_A(\delta \leq \gamma : \zeta_A) \rightarrow O_A(\rho \leq \gamma : \xi_A \vee \zeta_A) \quad (15)$$

$$\not\models O_A(\rho \leq \delta : \xi_A) \rightarrow O_A(\rho \leq (\delta \wedge \gamma) : \xi_A) \quad (16)$$

$$\not\models O_A(\rho \leq \delta : \xi_A) \wedge O_A(\rho \leq \gamma : \xi_A) \rightarrow O_A(\rho \leq (\delta \vee \gamma) : \xi_A) \quad (17)$$

$$\not\models O_A(\rho \leq \delta : \xi_A) \wedge O_A(\rho \leq \gamma : \xi_A) \rightarrow O_A(\rho \leq (\delta \wedge \gamma) : \xi_A) \quad (18)$$

$$\not\models O_A(\rho \leq \delta : \xi_A) \wedge O_A(\chi \leq \delta : \xi_A) \rightarrow O_A((\rho \wedge \chi) \leq \delta : \xi_A) \quad (19)$$

$$\not\models O_A(\rho \leq \delta : \xi_A) \rightarrow O_A(\delta \leq \rho : \xi_A) \quad (20)$$

$$\not\models O_A(\perp \leq \delta : \xi_A) \quad \not\models \neg O_A(\perp \leq \delta : \xi_A) \quad (21)$$

$$\not\models \neg O_A(\perp \leq \top : \xi_A) \quad \not\models O_A(\perp \leq \top : \xi_A) \quad (22)$$

$$\not\models \neg(O_A(\rho \leq \delta : \xi_A) \wedge O_A(\neg\rho \leq \delta : \xi_A)) \quad (23)$$

$$\not\models O_A(\rho \leq \top : \xi_A) \quad \not\models \neg O_A(\rho \leq \top : \xi_A) \quad (24)$$

We have no opportunity here to discuss these properties. In stead we briefly discuss some more logical issues.

The logical properties for maintenance obligations with a relief condition follow easily from the properties for achievement obligations with a deadline condition.

Many of the above properties concern properties of single paths within arbitrary strategies. Therefore we were able to give most of the proofs using an LTL theorem prover [12].

An interesting question is whether we can see classical non-temporal obligations (such as the ones of SDL) as limit cases of temporal order obligations. Intuitively it should be the case that if we ‘substitute’ the most common temporal connotations of general obligations in the temporal deontic operators, we get standard deontic operators back. In our opinion, the most likely substitution for this purpose is  $\delta = \top$ . We have the following theorem:

**Theorem 1.** *The logic of  $O_a(\rho \leq \top : \perp)$  is standard deontic logic (the modal logic KD)<sup>5</sup>.*

*Proof.* Substitution in the definition for  $O$  gives  $\neg\langle[A]\rangle((\neg\rho \wedge \neg\top)U^e(\top \wedge \neg\rho \wedge \neg\perp))$ . This reduces to  $\langle\{a\}\rangle X\rho$ . Since there is only one agent, system actions and actions of  $a$  are identical. The seriality condition on system actions ensures modal property D. K follows from the fact that for one agent, the ATL structure is based on a classical Kripke frame. From this it also follows that the logic is exactly KD, since this frame satisfies no additional properties.

## 8 Interactions with Ability: Deadlines and Deliberate Versions

In the previous sections, we did not consider the issue whether or not the conditions  $\rho$ ,  $\delta$  and  $\xi$  were actually ‘under control’ of groups of agents  $A$ . However, as is well known from the deontic literature, issues like ‘ought implies can’, ‘settledness’ and ‘power’ take a central place in it. In this section we study some interactions of obligations and ‘control’.

First we discuss the issue of control over the condition  $\delta$ . We called the obligations ‘temporal order obligations’ exactly because we did not exclude that  $\delta$  was indeed under control of the group of agents  $A$ . In contrast, a deadline obligation can be viewed as a temporal order obligation where the agents  $A$  do *not* control  $\delta$ . However we have to be very careful with what we mean. Actually, not controlling  $\delta$  should not be understood as agents  $A$  not having a strategy for  $F\delta$  (consequently they also would not have a strategy to violate without negative consequences, and thus would be obliged anything before  $\rho$ ). Not controlling  $\delta$  should be understood as not having a strategy for  $G\neg\delta$ . The difference with conditional temporal order obligations is thus that agents  $A$  cannot avoid their duty by pushing a deadline forward indefinitely, that is, they do not control  $\neg\delta$ . We can imagine that a temporal deadline  $D(\delta, n)$  for  $n$  time units is defined as ( $X^n$  represents  $n$  nestings of the next operator.):

### Definition 9

$$D(\delta, n) \equiv_{def} \langle[\emptyset]\rangle X^n(\delta \wedge \langle[\emptyset]\rangle G\neg\delta) \wedge \bigwedge_{0 \leq i < n} \langle[\emptyset]\rangle X^i \neg\delta$$

The ATL formula  $D(\delta, n)$  says that on all paths, after  $n$  steps  $\delta$  is true, while  $\delta$  is never true before or after that. Clearly, in case of a temporal deadline of this kind, no set of agents  $A$  can have a strategy for  $G\neg\delta$ . In this circumstance the temporal order obligation becomes a real deadline obligation:  $\delta$  is sure to happen in  $n$  time units, and agents

<sup>5</sup> Like in conditional deontic logics, the logic of  $O(\varphi \mid \top)$  is often also SDL.

do not have a strategy to avoid  $\rho$  at all points before  $\delta$  and not experience a negative condition at  $\delta$ . We may thus introduce deadline obligations by conjuncting temporal order obligations with formulas  $D(\delta, n)$ . Note that for deadline obligations, there is no longer any conditionality with respect to  $\delta$ , since  $\delta$  is sure to happen at a given point in the future.

Now we return to variants that are due to different possibilities for the control of achievements  $\rho$ . To give content to his concept of ‘categorical imperative’, Kant suggested the principle of ‘ought implies can’. Kant’s principle also makes sense in the present, more profane context. Rational agents are assumed to be realistic, which means that they will not let their decisions be influenced by obligations for conditions  $\rho$  they cannot achieve before  $\delta$  anyway. Obligation variants that incorporate this property can be defined as:

**Definition 11**

$$O_A^{oc}(\rho \leq \delta : \xi_A) \equiv_{def} O_A(\rho \leq \delta : \xi_A) \wedge \langle [A] \rangle (\neg \delta U \rho)$$

For agent theory, Kant’s dictum can be supplemented with a second principle concerning the interaction of obligation and ability. We might call this second principle ‘ought implies can avoid’. This relates to a problem with the definition of  $O$  that has been signaled many times before in deontic logic. It is sometimes called the problem of ‘settledness’ [10, 14]. The issue is that any obligation  $O$  for which compliance is settled, or, in other words, temporally inevitable, is true. In particular we have the property  $\rho \rightarrow O_A(\rho \leq \delta : \xi_A)$ , which is an instance of the more general property 9 of section 7. We avoid the property (and some others that are non-intuitive, such as property 10 of section 7) by defining *deliberate* versions of the obligation operators:

**Definition 12**

$$O_A^{dl}(\rho \leq \delta : \xi_A) \equiv_{def} O_A(\rho \leq \delta : \xi_A) \wedge \neg O_A(\rho \leq \delta : \perp)$$

The formula  $\neg O_A(\rho \leq \delta : \perp)$  says that it is not the case that  $O$  is an obligation for which a violation is impossible (i.e., an obligation for which the negative condition cannot become true). In other words, agents do have a strategy *not* to comply to the obligation. However, if they do so, there will be a negative condition. So, now the obligation is conditional on the possibility not to comply. Thus, agents can only have an obligation to achieve something if they have the *choice* not to do so. i.e., when it is not already settled.

The two principles of ‘ought implies can’ and ‘ought implies can avoid’ come down to the requirement that choices are not empty and have alternatives. Incorporating these principles in the definitions avoids counter intuitive properties like always having the obligation to achieve tautologies ( $O_A(\top \leq \delta : \xi_A)$ ). But in deontic logic, properties like  $O\top$  have actually been defended (it is, for instance, a property of SDL). We think that for the applications of deontic logic in agent theory, they should be excluded. An artificial agent having to deal with obligations is only interested reasoning about obligations that influence his decisions. If there is nothing to choose, either because the set of choices is empty or there is only one alternative, the obligations mean nothing to the agent.

To conclude this section, we want to point out that not all properties of section 7 hold for the variants in this section. To save space, we did not elaborate on the effects of the interaction with abilities on the logical properties. However, we do need to mention the following property that holds for the deliberate variant:

**Proposition 4**

$$\models O_A^{dl}(\rho \leq \delta : \xi_A) \wedge O_A^{dl}(\delta \leq \gamma : \zeta_A) \rightarrow O_A^{dl}(\rho \leq \gamma : \xi_A \vee \zeta_A)$$

This property is crucial in our discussion on the modelling of a Chisholm’s scenario in section 9.

## 9 Modelling Chisholm’s Scenario

The original formulation of Chisholm’s problematic scenario is [7]:

1. it ought to be that a certain man go to the assistance of his neighbors
2. it ought to be that if he does go he tell them he is coming
3. if he does not go then he ought not to tell them he is coming
4. he does not go

The modelling task we pursue in this section is to find a logical formalization that:

- faithfully reflects the natural language meaning, including the temporal aspects (the temporal order in sentence 2, the future directedness of all obligations, the present as the validity time of all obligations, etc.),
- is consistent,
- has no logically redundant sentences,
- derives that A ought not to tell.

As explained in the introduction, we are interested in obligations whose validity time is the present and whose reference time is the future, since these are the obligations an agent has to account for when making a decision about what to do. In particular we will interpret all sentences of Chisholm’s scenario as sentences being valid *presently* while the ‘regulating force’ of the obligations involved refers to the *future*. Note that this differs from many other temporal interpretations of Chisholm’s sentences. For instance, [6] discusses also a backwards looking interpretation that considers a setting where we know for a fact that the man did not help, that the obligation to help has been violated, and whether or not the agent told that he would come. However, most temporal interpretations of the scenario have been particularly aimed at using time to avoid the looming inconsistencies of a-temporal interpretations. For instance, Prakken and Sergot [16] suggest that temporalization can avoid ‘pragmatic oddities’, such as the one consisting of the obligation to help in combination with the obligation not to tell, by stipulating that the validity times of these obligations are disjoint. Following their line of reasoning, the oddity should be solved by interpreting the scenario in such a way that the obligation to help is valid until the moment it is violated, while from that point on the obligation not to tell is valid. We do not regard that as a solution, since we want a solution where the validity time of all obligations is the present.

Let us give our formalization first, before explaining the formulas.

$$O_A^{dl}(\text{help} \leq \text{too\_late} : \xi_A) \wedge D(\text{too\_late}, n) \quad (25)$$

$$O_A^{dl}(\text{tell} \leq \text{help} : \zeta_A) \quad (26)$$

$$\langle [\emptyset] \rangle (\neg \text{help} U \text{too\_late}) \rightarrow O_A^{dl}(\neg \text{tell} \leftarrow \diamond \text{too\_late} : \eta_A) \quad (27)$$

$$\langle [\emptyset] \rangle (\neg \text{help} U \text{too\_late}) \quad (28)$$

In the first sentence we have to model that the man is obliged to go to the assistance of his neighbors. As said, we want to interpret this as an obligation about the future: the man is obliged to help at some future point. However, as explained in section 2, we cannot simply model this as an obligation of the form *OFhelp*. Such obligations are vulnerable for indefinite deferral, and there is no reason for the man to start helping soon. So, if we want to interpret the obligation as an achievement obligation, we have to bring in a condition  $\delta$  before which the helping needs to take place. Since the sentence does not explicitly refer to such a condition, we simply model it as the condition *too\_late* and define that *too\_late* is true in exactly  $n$  time units. One might argue that we are introducing a concept that is not in the natural language description of the obligation. However, we claim that this is the only way we can make sense of the sentence if we interpret it as an achievement obligation. Although from the natural language description we cannot know the exact value of  $n$ , in our opinion it is safe to assume it is a parameter playing a role in the intuitive interpretation of the sentence as an achievement obligation.

As mentioned in the introduction, the second sentence is an outstanding example of the kind of obligations we can model in our formalism. The obligation to tell, with the present as its validity time, is conditional on the condition of helping, while the telling has to precede the helping. We know of no other temporal deontic formalism that can model this sentence as faithful as the present one.

It has been argued that the third sentence should have the same form as the second sentence, since both are conditionals. However, we argue that for our future directed interpretation this is not a sensible requirement. In particular, the second sentence is an achievement obligation, while the third is a maintenance obligation (see section 6): from the present until the moment where it is too late to help, the man has the obligation to preserve the condition of not telling, that is, if he will not help. This conditionality with respect to not helping is simply modelled using a material implication<sup>6</sup> expressing dependency on the condition whether presently it is known for a fact that the man is not going to help.

To interpret the fourth sentence as a fact about the future, we model it as an ATL expression saying that no strategies are possible that possibly result in the man actually helping before it is too late. We acknowledge that this is not necessarily the most intuitive choice. First of all it would contradict (and not violate) the formula modelling the first sentence if this would be a variant that incorporates ‘ought implies can’ (see definition 11). Second, modelling the sentence as a fact about the future is problematic

<sup>6</sup> Actually a conditional obligation of the form  $O_A^{dl}(\neg \text{tell} \leftarrow \diamond \text{too\_late} : \eta_A) \mid (\neg \text{help} U \text{too\_late})$  as briefly mentioned in section 5 would be a better choice here. But this would not affect the main idea behind the solution to Chisholm’s scenario.

as such. Intuitively, one should always keep open the possibility that the man will help. Therefore it would be much better to model the fourth sentence as an *intention*. Actually intentions can be suitably modelled as self-directed obligations, which means we can express them in the present formalism. We leave this for future research.

The above formalization is consistent, does not contain logical dependencies, and stays close to the natural language sentences. We now investigate whether it gives rise to the right conclusions. With the formulas modelling the first two sentences, together with the logical principle of proposition 4 for deliberate obligations, we derive  $O_A(\text{tell} \leq \text{too\_late} : \xi_A \vee \zeta_A)$ . Deriving a ‘new’ obligation from the first two sentences has been called ‘deontic detachment’ in the literature. But note that it is a rather special kind of deontic detachment specific for temporal order obligations.

With the formulas modelling the last two sentences, we derive  $O_A(\neg\text{tell} \leftarrow \diamond \text{too\_late} : \eta_A)$ . Obviously, this conflicts with the obligation derived through deontic detachment. But there is no inconsistency, not even when we use one and the same negative condition for all obligations involved (or when  $\xi_A \leftrightarrow \zeta_A \leftrightarrow \eta_A$ ). What the conflicting information tells us is that we cannot avoid one of the negative conditions  $\xi_A \vee \zeta_A$  or  $\eta_A$  becoming true at some point before *too\_late*: we cannot at the same time achieve ‘telling’ and preserve ‘not telling’: a choice has to be made. Of course one of the requirements for a solution to the scenario is that this choice should be ‘not telling’: we should be able to conclude that given the above modelling of the scenario, the obligation not to tell is ‘relevant’, while the obligation to tell is not. This is seen as follows. The agent will want to avoid the negative conditions. And in this case there is a best way to do that. Given the information in sentence 4 that there will be no helping before it is too late, we can derive that negative condition  $\xi_A$  is sure to occur at the point *too\_late*. This means that the derived obligation  $O_A(\text{tell} \leq \text{too\_late} : \xi_A \vee \zeta_A)$  is not interesting for the agent to base its decision on: trying to obey it is pointless, because its negative condition is valid anyway. This leaves the obligation  $O_A(\neg\text{tell} \leftarrow \diamond \text{too\_late} : \eta_A)$  as the relevant one: the agent will want to avoid the negative condition  $\eta_A$ , and thus should not tell.

## 10 Conclusion

In this paper we argued that achievement obligations need a deadline condition that functions as a point where a possible violation of the obligation is payed for. We named the resulting conditional obligations ‘temporal order obligations’. We showed how to define several semantics for temporal order obligations by giving characterizations of these modalities in plain ATL. This has as an advantage that all logic machinery already developed for ATL is applicable. The resulting framework is quite rich: we showed that it enables us to investigate issues like ‘ought implies can’, ‘ought implies can avoid’, deliberateness and deontic detachment. We mentioned logical properties of the defined operators, discussed their conditionality aspect, and demonstrated its applicability by modelling Chisholm’s famous scenario.

Many issues had to be left for future research. In particular the generalization to obligations  $O_A(\rho \leq \delta : \xi_A \mid \eta)$  could not be explained in detail, despite its possible relevance for Chisholm’s scenario. Also intentions, which are also relevant for the scenario, had to be left aside. Among the other issues we are planning to investigate in the

present framework are concepts like ‘power’, ‘responsibility’ and ‘counts as’. For the longer term, we would also like to investigate the relation between deontic semantics and game equilibria.

## References

1. Rajeev Alur, Thomas A. Henzinger, and Orna Kupferman. Alternating-time temporal logic. In *FOCS '97: Proceedings of the 38th Annual Symposium on Foundations of Computer Science (FOCS '97)*, pages 100–109. IEEE Computer Society, 1997.
2. Rajeev Alur, Thomas A. Henzinger, and Orna Kupferman. Alternating-time temporal logic. *Journal of the ACM*, 49(5):672–713, 2002.
3. A.R. Anderson. A reduction of deontic logic to alethic modal logic. *Mind*, 67:100–103, 1958.
4. J.M. Broersen, F. Dignum, V. Dignum, and J.-J. Meyer. Designing a deontic logic of deadlines. In A. Lomuscio and D. Nute, editors, *Proceedings 7th International Workshop on Deontic Logic in Computer Science (DEON'04)*, volume 3065 of *Lecture Notes in Computer Science*, pages 43–56. Springer, 2004. DOI: 10.1007/b98159.
5. J.M. Broersen, A. Herzig, and N. Troquard. From coalition logic to stit. In *Proceedings LCMAS 2005*, Electronic Notes in Theoretical Computer Science. Elsevier, 2005.
6. J.M. Broersen and L.W.N. van der Torre. Semantic analysis of chisholm’s paradox. In K. Verbeeck, K. Tuyls, A. Nowe, B. Manderick, and B. Kuijpers, editors, *Proceedings of the 17th Belgium-Netherlands Artificial Intelligence Conference*, pages 28–34, 2005.
7. R.M. Chisholm. Contrary-to-duty imperatives and deontic logic. *Analysis*, 24:33–36, 1963.
8. P.R. Cohen and H.J. Levesque. Intention is choice with commitment. *Artificial Intelligence*, 42(3):213–261, 1990.
9. F. Dignum and R. Kuiper. Obligations and dense time for specifying deadlines. In *Proceedings of thirty-First HICSS, Hawaii*, 1998.
10. J.A. van Eck. A system of temporally relative modal and deontic predicate logic and its philosophical applications. *Logique et Analyse*, 100:339–381, 1982.
11. J.F. Horty. *Agency and Deontic Logic*. Oxford University Press, 2001.
12. U. Hustadt and B. Konev. Trp++ 2.0: A temporal resolution prover. In F. Baader, editor, *Proceedings of the 19th International Conference on Automated Deduction (CADE-19)*, volume 2741 of *Lecture Notes in Artificial Intelligence*, pages 274–278. Springer, 2003.
13. S. Lindström and W. Rabinowicz. Unlimited doxastic logic for introspective agents. *Erkenntnis*, 50:353–385, 1999.
14. B. Loewer and M. Belzer. Dyadic deontic detachment. *Synthese*, 54:295–318, 1983.
15. J.-J.Ch. Meyer. A different approach to deontic logic: Deontic logic viewed as a variant of dynamic logic. *Notre Dame Journal of Formal Logic*, 29:109–136, 1988.
16. H. Prakken and M.J. Sergot. Contrary-to-duty obligations and defeasible reasoning. *Studia Logica*, 57:91–115, 1996.
17. F. Dignum H. Weigand V. Dignum, J.J. Meyer. Formal specification of interaction in agent societies. In *2nd Goddard Workshop on Formal Approaches to Agent-Based Systems (FAABS)*, Maryland, 2002.
18. G.H. von Wright. Deontic logic. *Mind*, 60:1–15, 1951.