SEMANTIC ANALYSIS OF CHISHOLM'S PARADOX

Jan Broersen [a] Leendert van der Torre [b]

[a] *Universiteit Utecht, P.O. Box 80.089, 3508 TB, Utrecht*
[b] *CWI, P.O. Box 94079, 1090 GB, Amsterdam*

Abstract

Violation handling is a crucial problem in many applications. therefore its paradoxes have been studied in, amongst others, artificial intelligence, agent theory and computer science. The standard way to study these paradoxes is to model them using a formal language, and use formal logic to consider whether the set of sentences is inconsistent, the sentences logically follow from others, or some other anomaly occurs. During the past decades, developments in temporal, action and non-monotonic logics have contributed to a better understanding of the paradoxes and thus of violation handling. In this paper we propose an alternative way to analyze Chisholm's notorious contrary-to-duty paradox in deontic logic. We model the paradox using semantic models, using insights from conceptual modelling. We aim to gain insight in the open question whether the paradoxes are in some sense logical contradictions, or only *apparent* contradictions. If a paradox is only an apparent contradiction, then there has to be a model interpreting all sentences.

1 Introduction

Violation handling is a crucial problem in many applications, and its paradoxes have therefore been studied in, amongst others, artificial intelligence, agent theory and computer science. The standard way to study these paradoxes is to model them using a formal language, and use formal logic to consider whether the set of sentences is inconsistent, some sentence follows from another one, or some other anomaly occurs. Chisholm's paradox [6] consists of the following four sentences.

it ought to be that a certain man go to the assistence of his neighbours (1)
it ought to be that if he does go he tell them he is comming (2)
if he does not go then he ought not to tell them he is comming (3)
he does not go (4)

In traditional approaches to formalizing this example, see [15] for an overview, each sentence is represented by a logical formula, and then either the set of sentences is inconsistent, one sentence follows from another one, or some other anomaly occurs. Therefore it is traditionally called a paradox. During the past decades, developments in temporal, action or non-monotonic logics have contributed to a better understanding of the paradoxes and thus of violation handling (we take violation handling here to be interpreted in a broad sense, encompassing conflict handling, exception handling, recovery from anomalous states, etc.). A disadvantage of starting with a *logic* to tackle the problem is that one cannot know beforehand whether the formalization of the example is going to be consistent. Another problem with the logical approach is that we cannot use modeling methodologies developed in conceptual modeling.

Therefore, we propose to start by looking for a model on which all sentences can be interpreted. If it is possible to interpret the sentences of the scenario consistently, in the sense that we deal with an apparent paradox, then it should be possible to find such models. Then, if we study the meaning of the sentences in terms of logics interpreted on these models, we never end up in the situation in which our formalization is inconsistent. However, capturing the meaning of the sentences in an intuitive way is not the only criterion for a satisfactory formalization in terms of logical formulas. In addition we have to face the following two problems: on the one hand, we have to avoid unwanted consequences, and on the other hand, we have to ensure desired ones, specific for the example. Of course these two requirements are not completely independent of the consistency issue: formulas specified in stronger logics are more likely to be inconsistent.

Our modeling approach is in line with conceptual modeling approaches in computer science and elsewhere, typically supported by modeling languages like the Unified Modeling Language (UML). We take the following modeling steps:

Temporal structure. The first step of our formalization is to model the temporal choice structures we are considering. We model that the agent can first tell his neighbors, that he can go to the assistance, etc. We also introduce a deadline. It was argued in [3] that temporal achievement obligations need a deadline, since otherwise they cannot be violated. Though there are several alternatives which all look acceptable, this first step does not seem to raise any substantial problems.

Facts. Having fixed the temporal structures, we consider the facts of the example, represented by the fourth sentence. Modeling this step, however, leads already to substantial problems. Has the agent not gone to the assistance? Has it become impossible to go to the assistance? Does he intend not to go to the assistance? Can the man still tell that he will go to the assistance, or is this fact already settled, but unknown to us?

Obligations. Once we have made a choice at the first and second step, we have to decide how to model obligations. In general, the semantics of obligations can be suitably defined in terms of preference relations distinguishing between deontically ideal, deontically sub-ideal and deontically bad circumstances. In this paper we use a function selecting choices that are deontically ideal.

Detachment. Having represented the aspects described in the first three steps, in the fourth step we consider two classical reasoning patterns in the example, called ‘deontic detachment’ (Ideally the man ought to tell...) and ‘factual detachment’ (... but in the actual situation he should not tell.).

One source for the research question of this paper is a dispute among the authors. The former [2] believes that Chisholm’s four sentences are closer to ‘consistent’, whereas the latter believes that they are closer to ‘apparently inconsistent’. The intuition of the first of the authors of this paper is that Chisholm’s paradox is not a real paradox in language or reasoning, and that a consistent interpretation should be possible. He claims that the problems discussed in the literature are only due to insufficient formalizations of the example. The first author thus considers the problem of Chisholm’s paradox a modelling problem.

The layout of this paper is as follows. In Section 2 we repeat the traditional symbolic analysis of Chisholm’s paradox. In Section 3 we introduce our semantic structures, and we show how Chisholm’s paradox can be modeled in them. In Section 4 we relate our semantic analysis to the deontic logic literature on Chisholm’s paradox.

2 Symbolic analysis of Chisholm’s paradox

Chisholm modeled his paradox using normal monadic modal logic of type KD, often called Standard Deontic Logic (SDL, Von Wright [17]), a normal modal system containing the propositional theorems, axioms $K : O(p \rightarrow q) \rightarrow (O(p) \rightarrow O(q))$ and $D : \neg(O(p) \wedge O(\neg p))$, and closed under modus ponens and necessitation. The sentences were originally represented as $S_1 = \{O(h), h \rightarrow O(t), \neg h \rightarrow O(\neg t), \neg h\}$ or $S_2 = \{O(h), O(h \rightarrow t), \neg h \rightarrow O(\neg t), \neg h\}$. Both formalizations are unsatisfactory. The sentences of S_1 are not logically independent (the second formula is entailed by the fourth), and there is no form of deontic detachment ensuring that given that there is an obligation to help, there is also an obligation to tell. Deontic detachment is present in the sentences of S_2. However, S_2 is inconsistent.

Roughly, two reactions are possible to the inadequacy of SDL for modelling Chisholm’s scenario: either SDL is too ‘poor’, where a logic is ‘poor’ (‘rich’) if it’s language is ‘poor’ (‘rich’), or it is too ‘strong’, where a logic is ‘strong’ (‘weak’) if a relatively ‘high’ (‘low’) percentage of the formulas of its language are valid. If we consider it to be too poor, we may add temporal expressiveness, conditionality, other notions of obligation, etc. If we consider it too strong, we may weaken it by eliminating some standard inference rules (aggregation, closure under material implication, etc). Typically, logical approaches to the problem combine these two directions of solution, which means that they combine an extension of the language with weakening of the inference capacities.

A third possible reaction to the inadequacy of SDL is that the solution of the paradox requires a nonmonotonic logic. But this approach can also be seen as a combination of the other two reactions: making

SDL non-monotonic requires a syntactical distinction between formulas (or formula parts) with a defeasible and strict interpretation, which makes the language richer. On the other hand, non-monotonicity implies that inference becomes 'context dependent', which can be considered a form of weakening. In this paper, we do not discuss non-monotonic solutions, see [12] for a discussion.

An example of the approach where the language is extended, is the one where the set is represented using an additional normal modal operator $\Box$ (for "necessity"): $S_3 = \{\Box(O(h)), \Box(h \rightarrow O(t)), \Box(\neg h \rightarrow O(\neg t)), \neg h\}$. This can be rewritten to the following set of equivalent formulas using a strict implication $>$, defined as usual by $p > q = \Box(p \rightarrow q)$: $S_3 = \{\top > O(h), h > O(t), \neg h > O(\neg t)), \neg h\}$. S_3 is closely related to the SDL formalization S_1. Like S_1 it is consistent and shows 'factual detachment' as a pattern of inference. But it has as an advantage over S_1 that the second sentence is not a logical consequence of the fourth. Yet, still it is not satisfactory, since it does not show the important inference pattern of deontic detachment.

Factual detachment. From the latter two sentences we should derive, in some sense, that it is obligatory not to tell, given that the agent is making the best out of the sad circumstances that he is violating the primary obligation.

Deontic detachment. From the first two sentences we should derive, in some sense, that it is obligatory for the agent to tell his neighbors that he will come. The reason is that ideally, the agent tells his neighbors that he will come, and he will go there. This derivation has been called a moral cue for action.

Both patterns are intuitively valid. But at the same time, the conflicting conclusions of both patterns are the main cause of the inconsistencies of naive formalizations.

3 Semantic models interpreting Chisholm's paradox

We now define the structures in terms of which all four sentences of Chisholm's scenario can be given intuitive interpretations. We make our models as rich as deemed necessary to give convincing interpretations of the sentences. Since we only look at models, we are not vulnerable to such claims as the one by Prakken and Sergot [10], who argue that temporal logic formalizations do not touch upon the core of the problem. We do not define logics. Of course, future research may reveal that a consistent logic formalization in terms of logics based on our models does not hinge on the temporal dimension. But from a modelling perspective, it is wise to include any possibly relevant aspect in the models. In particular, we want to model the following concepts: *agency*, *choice*, *time*, *fact*, *intention* and *obligation*.

3.1 Agency, Choice and Time

We do not need to justify that we include agents in our models. We model choices because the sentences describe a decision problem for the agent. Given the obligations in the first three sentences and the partial decision (not helping) in the fourth, the decision to make is whether or not to tell. Because choices always concern the future, we also want to model time. To model agency, choice and time, we use models similar to those of alternating time temporal logic (ATL) [1]. This suggests that for future research we might use ATL itself. But we might also use logics richer than ATL, for instance, logics that make ATL's strategies explicit in the object language. Or we might use logics weaker than ATL, for instance CTL, or even plain propositional logic to formalize the agency, choice and time aspects of the sentences.

In our models, 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. Also seeing-to-it-that or STIT models popular in philosophical logic (e.g. [7]) are based on choices, but typically do not have the assumption on non-determinism. A semantic structure contains for each agent a in state s a set of choices (informally: 'actions'), written as $C(a, s)$, under the condition that the intersection of the choices returns a unique follow-up state.

Definition 3.1 *A semantic structure $\mathcal{M}^{AT} = (S, s, \mathcal{C}, \pi)$, consists of a non-empty set S of states, an actual state $s \in S$, a function $\mathcal{C} : A \times S \mapsto 2^{2^S}$ such that for each s the set $\{\bigcap_{a \in A} C \mid C \in \mathcal{C}(a, s)\}$ is a set of singletons, and, finally, an interpretation function π for propositional atoms in states.*

We assume two propositional atoms t and h, for *just* having told and *just* having helped respectively, and a single agent a. We can model the actions in the example using seven states:

- $s_0 : \neg t, \neg h$ is the initial state in which nothing has been told and the neighbors have not been helped. $\mathcal{C}(a, s_0) = \{\{s_1\}, \{s_2\}\}$: the agent has two actions, telling the neighbors that he will come or not telling them, leading to respectively $s_1 : t, \neg h$ and $s_2 : \neg t, \neg h$.
- $s_1 : t, \neg h$ is the state in which the agent has told his neighbors that he will come. $\mathcal{C}(a, s_1) = \{\{s_3\}, \{s_4\}\}$: he can either go to the assistance of his neighbors ($s_3 : \neg t, h$) or not ($s_4 : \neg t, \neg h$)
- $s_2 : \neg t, \neg h$ is the state in which the agent has not told his neighbors that he will come. $\mathcal{C}(a, s_2) = \{\{s_5\}, \{s_6\}\}$: he can again either go to the assistance of his neighbors ($s_5 : \neg t, h$) or not ($s_6 : \neg t, \neg h$)
- the states s_3, s_4, s_5 and s_6 have a single action which brings them to themselves $C(a, s_3) = \{\{s_3\}\}$, etc., to ensure there is always a follow-up state (seriality).

Of course, there are many more possible models that explain the story.

3.2 Facts or Intentions?

The modelling of the fourth sentence, saying that the agent does not help, appears to be the most problematic. Using standard terminology of conceptual modelling, we believe that there are at least two (subjective) interpretations of the story. The most obvious interpretation is that the actual state w of our model is one in which the agent does not help.

4. agent a does not help. The actual state w is one for which $\neg h$ (typically s_4 or s_6).

However, we can also interpret ‘not helping’ as a decision about what (not) to do in the future. To model this aspect we need to model *intention*. A set of actions intended at s' by agent a in state s, written as $I(a, s, s')$, is defined as a subset of these actions. Alternatively we might have modelled intentions as subsets of action traces, or even as subsets of action trees. However, at this point we see no need for this. In 3.3, where we use the same type of functions to model obligation, we elaborate on the role of the arguments s and s'.

Definition 3.2 *A semantic structure $\mathcal{M}^{ATI} = (S, s, \mathcal{C}, \mathcal{I}, \pi)$, extends a structure $\mathcal{M}^{AT} = (S, s, \mathcal{C}, \pi)$ with a function $\mathcal{I} : A \times S \times S \mapsto 2^{2^S}$ such that $\mathcal{I}(a, s, s') \subseteq \mathcal{C}(a, s')$.*

Then, in our second (subjective) interpretation of the story told in the paradox, the actual state is s_0. In this state, it does not make sense to say that the agent does not go to the assistance of his neighbors, but it does make sense to state that he intends not to go to their assistance. We interpret the last sentence by the intentions in the initial state.

4’. a intends not to help. In the initial state we have $\mathcal{I}(a, s_0, s_1) = \{\{s_4\}\}$ and $\mathcal{I}(a, s_0, s_2) = \{\{s_6\}\}$.

3.3 Obligations

Having modelled the facts, the modelling of the obligations is relatively straightforward. A set of obligatory actions for agent a in state s at all possible states s', written as $O(a, s, s')$, is defined analogously to intended actions.

Definition 3.3 *A semantic structure $\mathcal{M}^{ATIO} = (S, s, \mathcal{C}, \mathcal{O}, \mathcal{I}, \pi)$ extends a structure $\mathcal{M}^{ATI} = (S, s, \mathcal{C}, \mathcal{I}, \pi)$ with a function $\mathcal{O} : A \times S \times S \mapsto 2^{2^S}$ such that for all $a \in A$ and $s, s' \in S$ we have $\mathcal{O}(a, s, s') \subseteq C(a, s')$.*

A noteworthy property of our semantic structures is thus that the obligations and intentions refer to two states, the first is the point of reference of the obligation formula and the second is the point of reference for the obliged formula. So, for example, when today it is obligatory that tomorrow the man goes to the assistance of his neighbors, then the first state is today and the second state is tomorrow. The explicit representation of the first state is used below to represent conditional obligations.

- agent a ought to help his neighbor. In every state $s_i \in S$, we have that $\mathcal{O}(a, s_i, s_1) = \{\{s_3\}\}$ and $\mathcal{O}(a, s_i, s_2) = \{\{s_5\}\}$;

- if agent a helps, then agent a ought to tell he will do so. We have that $\mathcal{O}(a, s_3, s_0) = \{\{s_1\}\}$ and $\mathcal{O}(a, s_5, s_0) = \{\{s_1\}\}$;
- if agent a does not help, then agent a ought not to tell he will. We have that $\mathcal{O}(a, s_4, s_0) = \{\{s_2\}\}$ and $\mathcal{O}(a, s_6, s_0) = \{\{s_2\}\}$;

3.4 The consequences of factual and deontic detachment

Now we consider the analogues of deontic and factual detachment. In the traditional symbolic analysis of Chisholm's paradox they are associated with reasoning patterns, but in our present modelling approach, we only consider whether our models do not exclude these patterns. That is, our models have to represent that, in some sense, in the actual situation we have to tell, but ideally we do not tell. Following the discussion in Section 3.2, we distinguish between the interpretation in which the fourth sentence refers to the actual state itself, and the interpretation in which the fourth sentence refers to the intentions in the actual state.

We start with the actual state interpretation. First, in the actual state, we have that it is obligatory not to tell. So $\mathcal{O}(a, s_4, s_0) = \{\{s_2\}\}$ or $\mathcal{O}(a, s_6, s_0) = \{\{s_2\}\}$ (see above), depending on which of the states s_4 or s_6 is the actual state. Second, in the initial state s_0 we have the obligation to tell, i.e., $\mathcal{O}(a, s_0, s_0) = \{\{s_1\}\}$, which results from $\mathcal{O}(a, s_0, s_1) = \{\{s_3\}\}$ (obligation to help) and $\mathcal{O}(a, s_3, s_0) = \{\{s_1\}\}$ (obligation to tell, in case of helping). Thus, in this temporal analysis of Chisholm's paradox, we have in the initial state s_0 that it is obligatory to tell, but once we are in a state in which the man has not gone to the assistance, then we have an obligation not to tell.

Now we consider the intention variant. The idea of introducing intentions in Chisholm's example is that we can study obligations *given* the intention not to help. So, apart from the semantic notions of obligation ($\mathcal{O}$) and intention ($\mathcal{I}$), we will need a derived notion $\mathcal{X}$ for 'obligation-given-intentions'.

There are again two things we would like to infer from the example, corresponding to intention variants of deontic and factual detachment. On the one hand, we would like to derive that given that the agent intends not to help, ideally he should change his mind and tell his neighbors that he will come, and go there. On the other hand we would like to infer that given the intention not to help, the agent should not tell the neighbors that he will come. Note that both inferences deal with obligations and intentions, and they are thus instances of the interaction between obligations and intentions.

Concerning the issue that the agent should change its mind using deontic detachment, we can still derive from the first two obligations $\mathcal{O}(a, s_0, s_1) = \{\{s_3\}\}$ and $\mathcal{O}(a, s_3, s_0) = \{\{s_1\}\}$ that in the initial state s_0 we have the obligation to tell, i.e., $\mathcal{O}(a, s_0, s_0) = \{\{s_1\}\}$. But, in s_0 we have the intention to choose s_4 or s_6, while in these states s_4 and s_6 we have the obligation that in s_0 the agent has to choose s_2. So, we have $\mathcal{X}(a, s_0, s_0) = \{\{s_2\}\}$ as derived from $\mathcal{I}(a, s_0, s_1) = \{\{s_4\}\}$ and $\mathcal{O}(a, s_4, s_0) = \{\{s_2\}\}$. This represents that if the agent follows his intentions he gets into a conflict - thus he should revise its intentions.

Finally, factual detachment in the intention variant is simply the above mentioned derivation of the 'obligation-given-intentions' $\mathcal{X}(a, s_0, s_0) = \{\{s_2\}\}$; given the intention not to help, the agent should at least not tell that he will.

4 Our semantic analysis and the deontic logic literature

In this paper we cannot give a full survey of the discussions on Chisholm's paradox, see for example [15] for a survey. Numerous other criteria were used to argue that the Chisholm paradox is not sufficiently represented by the set S_3. Unfortunately, these criteria are normally implicit, with the notable exception of the recent article of Jones and Carmo in the handbook of philosophical logic [5]. However, their criteria seem to be tailored to their own logic, and criteria of other authors are not discussed. For example, some authors find that temporal and action aspects should be explicit in the logic, but they are absent in the conditions of Jones and Carmo.

One additional condition is that also a large number of other related so-called contrary-to-duty paradoxes could be formalized analogously, most notably Forrester's paradox: you should not kill, but if you kill you should do it gently. Again, the straightforward formalization of this set is inconsistent in standard deontic logic. Another essential extension is that there can be a fourth obligation in Chisholm's example, that if the agent does not go and he tells, then there is a tertiary obligation, for example to apologize.

In the temporal analysis of Chisholm's paradox, several variants have been distinguished. For example, the original paradox has been called a backward version, because "tell" in the contrary to duty obligation not to tell occurs before the "help" in the primary obligation to help. It was soon discovered that a forward variant of the paradox could easily be formalized in a temporal or action deontic logic, but that the backward version is more complicated. See [14] for a survey of the temporal analysis of the paradox.

The intention variant of Chisholm's paradox clearly is a different example than his original paradox. However, in some discussions of Chisholm's paradox, it seems that people have been discussing this variant. For example, a popular argument why the Chisholm set should derive an obligation to tell and help the neighbors, is that the agent should change its mind from not helping to helping. But in that case, the fourth sentence is not a fact, but only an intention. See the above mentioned survey papers for further discussions and references.

We distinguish in our formalization of Chisholm's example between temporal looking forward and looking back, which may be seen as a formalization of Thomason's context of deliberation and the context of justification [11]. It has been related to decision making and diagnosis in [16], see this paper for a further discussion.

The analysis of the example using intentions is related to a distinction between facts which have been settled, and facts which have not been settled yet, sometimes called necessary facts and normal facts. This can be found in many formalizations of Chisholm's paradox, in particular in the temporal analysis of the paradox, see for example Loewer and Belzer's Dyadic Deontic Detachment or 3D [9]. We again have to refer to the survey papers mentioned earlier, as we are not going to look into the logic of these semantic structures.

Horty [7] defines obligations relative to paths, and not to strategies, as we have done. This requires that we mark subsets of paths (histories) as optimal, in stead of choices in states. Also Jamroga [8] distinguishes between optimal and non-optimal *paths*, but he does not restrict himself to the subset of paths going through the state of evaluation.

5 Summary

Violation handling is a crucial problem in many applications, and its paradoxes have therefore been studied in, amongst others, artificial intelligence, agent theory and computer science. The standard way to study these paradoxes is to model them using a formal language, and use formal logic to consider whether the set of sentences is inconsistent, the sentences logically follow from other ones, or some other anomaly occurs. In this paper we propose semantic analysis as an alternative. The main result of the semantic analysis in this paper is the crucial role of the interpretation of the fourth sentence of the paradox. We have shown that there are at least two interpretations of this sentence, either as a factual description or as an intentional one. We also show how these two interpretations lead to distinct models of the consequences of factual and deontic detachment. Finally we show how traces of our discussion can be found in the deontic logic literature.

The first results seem to indicate that a model for Chisholm's paradox can be found. We leave it to the reader to decide for him or herself what this result means for deciding the dispute between the two authors. The first author claims that existence of a model demonstrates that the Chisholm set is consistent; if the sentences appear to be consistent (in spite of the problem with inconsistent formalizations in the literature) there should be some model that reflects this. The second author claims that existence of a model shows that it is only an apparent contradiction; it appears contradictory because of the problem with inconsistent formalizations in the literature, but the model shows that it is not.

Chisholm's paradox is relevant for our research on conflicting mental attitudes in the BOID architecture. In future research we intend to use our semantic structures to model the interaction between obligations and intentions. Interactions among desires and obligations have been studied in agent logics. For example, the unfortunate situation has been studied in which an agent desires to do something but is obliged to do otherwise [4]. Interactions between beliefs and obligations have been studied in defeasible deontic logic, for example to decide whether 'normally it should be the case that p' together with the absence of p is a violation or just an exception [13]. The interaction between obligations and intentions seems to have attracted less attention. For example, also obligations may lead to intentions for moral or social agents [4]. Therefore, in this paper we introduce semantic structures to study this interaction. Summarizing, BDI agents operating within an organization, institution or other artificial social system need to deal with norms and social laws of the environment. They give rise to *obligations* for the agent. Agents also act according to their own preferences, desires, wishes and wants. These give rise to *intentions* of the agent. In this paper we are interested in the interaction among obligations and intentions.

Acknowledgements

Thanks to Paul Harrestein and Joris Hulstijn for discussions on the modelling of Chisholm's paradox.

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] J.M. Broersen. *Modal Action Logics for Reasoning about Reactive Systems*. PhD thesis, Faculteit der Exacte Wetenschappen, Vrije Universiteit Amsterdam, februari 2003.

[3] J.M. Broersen. On the logic of 'being motivated to achieve ρ, before δ'. In J. Alferes and J. Leite, editors, *Proceedings Ninth European Conference on Logics in Artificial Intelligence (JELIA'04)*, volume 3229 of *Lecture Notes in Artificial Intelligence*, pages 334–346. Springer, 2004. DOI: 10.1007/b100483.

[4] J.M. Broersen, M. Dastani, J. Hulstijn, and L.W.N. van der Torre. Goal generation in the BOID architecture. *Cognitive Science Quarterly Journal*, 2(3-4):428–447, 2002.

[5] J. Carmo and A. Jones. Deontic logic and contrary-to-duties. In D. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic - Second Edition, Volume 3: Extensions to Classical Systems 2*, pages 203–279. klw, 2005. In press.

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

[7] J.F. Horty. *Agency and Deontic Logic*. Oxford University Press, 2001.

[8] W. Jamroga, W. van der Hoek, and M. Wooldridge. On obligations and abilities. 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 165–181. Springer, 2004.

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

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

[11] R. Thomason. Deontic logic as founded on tense logic. In R. Hilpinen, editor, *New Studies in Deontic Logic*, pages 165–176. D. Reidel Publishing Company, 1981.

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

[13] 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 Academic Publishers, 1997.

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

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

[16] L.W.N. van der Torre and Y.H. Tan. Diagnosis and decision making in normative reasoning. *Artificial Intelligence and Law*, 7:51–67, 1999.

[17] G.H. von Wright. Deontic logic. *Mind*, 60:1–15, 1951.