

Automated Reasoning in Deontic Logic*

Ulrich Furbach¹

Claudia Schon¹

Frieder Stolzenburg²

¹ Universität Koblenz-Landau, {uli, schon}@uni-koblenz.de

² Harz University of Applied Sciences, fstolzenburg@hs-harz.de

1 Introduction

Deontic logic is a very well researched branch of mathematical logic and philosophy. Various kinds of deontic logics are discussed for different applications like argumentation theory, legal reasoning and actions in multi-agent systems ([6]). Recently there also is growing interest in modelling human reasoning and testing the models with psychological findings. Deontic logic is an obvious tool to this end, because norms and licenses in human societies can be described easily. In [5] there is a discussion of some of these problems including a solution with the help of deontic logic. This paper concentrates on automated reasoning in deontic logic. We show that deontic logic can be translated into the description logic $\mathcal{ALC}$, for which the first order reasoning system Hyper offers a decision procedure.

2 Deontic Logic as Modal Logic KD

In this section we briefly describe how standard deontic logic can be seen as modal logic K together with the seriality axiom D: $\Box\Phi \rightarrow \Diamond\Phi$. The $\Box$ -operator is interpreted as ‘it is obligatory that’ and the $\Diamond$ as ‘it is permitted that’. Assuming a Kripke-semantics, we have that there are different worlds, in which not all of the norms have to hold, but due to the seriality axiom, there always exists a world in which the norms hold, hence an ideal world.

To demonstrate the use of deontic logic, we consider the well-known problem of *contrary-to-duty obligations* introduced in [4] and the formalization as a normative system $\mathcal{N}'$ given in [10] (where s stands for *steals* and p for *punished*):

a') $\Box\neg s$

b') s

c') $s \rightarrow \Box p$

d') $\Box(\neg s \rightarrow \neg p)$

As shown in [10], the normative system $a')$ - $d')$ is inconsistent. This example is used as a running example throughout this paper.

3 Translating Deontic Logic into Description Logic

Hyper [11] is a theorem prover for first order logic with equality. It is the implementation of the E-hypertableau

calculus [1] which extends the hypertableau calculus with equality handling based on superposition. Hyper has been successfully used in various AI-related applications, like intelligent interactive books or natural language query answering.

Recently the E-hypertableau calculus and its implementation have been extended to deal with knowledge bases given in the description logic $\mathcal{SHIQ}$ [2]. There is a strong connection between modal logic and description logic. As shown in [9], the description logic $\mathcal{ALC}$ is a notational variant of the modal logic $\mathcal{K}_n$. Therefore any formula given in the modal logic $\mathcal{K}_n$ can be translated as usual into an $\mathcal{ALC}$ concept and vice versa.

In addition to that, we have to translate the seriality axiom into description logic. In [7] it is shown, that the seriality axiom can be translated into the following TBox:

$$\mathcal{T} = \{\top \sqsubseteq \exists R.\top\}$$

with R the atomic role introduced by the translation described above.

4 Hypertableau for Deontic Logic

In this paper we used the deontic operator only in a few conditional formulae. In the philosophical literature deontic logic is also used to formulate entire normative systems (e.g. [10]). In practice such normative systems can be rather complex. This makes it difficult for the creator of a normative system to see if a normative system is consistent. We will show that it is helpful to be able to check consistency of normative systems automatically and use the Hyper theorem prover to check the consistency of a given normative system very much the same way as we used it in the previous sections.

The translation of the normative system $\mathcal{N}'$ introduced above into $\mathcal{ALC}$, called $\phi(\mathcal{N}')$ henceforth, is shown in Table 1. Checking the consistency of the normative system $\mathcal{N}'$ corresponds to checking the consistency of $\phi(\mathcal{N}')$ w.r.t. the TBox $\mathcal{T} = \{\top \sqsubseteq \exists R.\top\}$, where $\phi(\mathcal{N}')$ is the conjunction of the concepts given in the right column of Table 1. We transform $\phi(\mathcal{N}')$ into DL-clauses, which is the input language of Hyper. We will not give the result of this transformation here, but refer to [8] for details. Hyper constructs a hypertableau for the resulting set of DL-clauses. This hypertableau is closed and therefore we can conclude, that the set of DL-clauses is unsatisfiable. This tells us, that the normative system $\mathcal{N}'$ formalized above is inconsistent.

*Work supported by DFG grants FU 263/15-1 and STO 421/5-1 ‘Ratiolog’

Deontic Logic	$\mathcal{ALC}$
$\Box\Phi \rightarrow \Diamond\Phi$	$\top \sqsubseteq \exists R.\top$
$\Box\neg s$	$\forall R.\neg S$
s	S
$s \rightarrow \Box p$	$\neg S \sqsubseteq \forall R.P$
$\Box(\neg s \rightarrow \neg p)$	$\forall R.(S \sqsubseteq \neg P)$

Table 1: Translation of the normative system $\mathcal{N}'$ into $\mathcal{ALC}$.

5 An Example from Multi-agent Systems

In multi-agent systems there is a relatively new area of research, namely the formalization of 'robot ethics'. It aims at defining formal rules for the behavior of agents and to prove certain properties. As an example consider Asimov's laws, which aim at regulating the relation between robots and humans. In [3] the authors depict a small example of two surgery robots obeying ethical codes concerning their work. These codes are expressed by means of deontic logic, very much the same as we defined the normative systems in this paper.

We will give a formalization in standard deontic logic of this example together with a description of the proof tasks for Hyper. In our example, the robots *ag1* and *ag2* have two possible actions: *ag1* can terminate a person's life support and *ag1* can delay the delivery of pain medication. We consider two ethical codes O and O^*

- $O \rightarrow \Box\neg action(ag2, delay)$, which means that "If ethical code O holds, then robot *ag2* takes care, that delivery of pain medication is not delayed."
- $O^* \rightarrow O \wedge O^* \rightarrow \Box\neg action(ag1, term)$, which means that "If ethical code O^* holds, then code O holds, and robot *ag1* takes care, that life support is not terminated."

Further we give a slightly modified version of the evaluation of the robot's actions given in [3], where $(+!!)$ describes the most and $(-!!)$ the least desired outcome:

$$\begin{aligned}
& action(ag1, term) \wedge action(ag2, delay) \rightarrow (-!!) \\
& action(ag1, term) \wedge \neg action(ag2, delay) \rightarrow (-!) \\
& \neg action(ag1, term) \wedge action(ag2, delay) \rightarrow (-) \\
& \neg action(ag1, term) \wedge \neg action(ag2, delay) \rightarrow (+!!)
\end{aligned}$$

Further we add formulae stating that the formulae for the evaluation of the robot's actions hold in all reachable worlds. A possible query would be to ask, if the most desirable outcome $(+!!)$ will come to pass, if ethical code O^* is operative. This query can be translated into a satisfiability test: If $O^* \wedge \Diamond\neg(+!!)$ is unsatisfiable, then ethical code O^* ensures outcome $(+!!)$. We have been able to solve this task successfully by translating the above formulae and the query into DL-clauses and use Hyper to test the satisfiability of the resulting set of DL-clauses.

References

- [1] Peter Baumgartner, Ulrich Furbach, and Björn Pelzer. Hyper tableaux with equality. In Frank Pfenning, editor, *Automated Deduction - CADE 21, 21st International Conference on Automated Deduction, Bremen, Germany, July 17-20, 2007, Proceedings*, volume 4603 of *Lecture Notes in Computer Science*, 2007.
- [2] Markus Bender, Björn Pelzer, and Claudia Schon. System description: E-KRHyper 1.4 - extensions for unique names and description logic. In Maria Paola Bonacina, editor, *CADE-24, LNCS*, 2013.
- [3] Selmer Bringsjord, Konstantine Arkoudas, and Paul Bello. Toward a general logicist methodology for engineering ethically correct robots. *IEEE Intelligent Systems*, 21(4):38–44, 2006.
- [4] R. M. Chisolm. Contrary-to-duty imperatives and deontic logic. *Analysis*, 23:33–36, 1963.
- [5] Ulrich Furbach and Claudia Schon. Deontic logic for human reasoning. *CoRR*, abs/1404.6974, 2014.
- [6] John F. Horty. *Agency and Deontic Logic*. Oxford University Press, Oxford, 2001.
- [7] Szymon Klarman and Víctor Gutiérrez-Basulto. Description logics of context. *Journal of Logic and Computation*, 2013.
- [8] Boris Motik, Rob Shearer, and Ian Horrocks. Optimized Reasoning in Description Logics using Hyper-tableaux. In Frank Pfenning, editor, *CADE-21*, volume 4603 of *LNAI*. Springer, 2007.
- [9] Klaus Schild. A correspondence theory for terminological logics: Preliminary report. In *In Proc. of IJCAI-91*, pages 466–471, 1991.
- [10] Frank von Kutschera. *Einführung in die Logik der Normen, Werte und Entscheidungen*. Alber, 1973.
- [11] Christoph Wernhard and Björn Pelzer. System description: E-KRHyper. In Frank Pfenning, editor, *Automated Deduction - CADE 21, 21st International Conference on Automated Deduction, Bremen, Germany, July 17-20, 2007, Proceedings*, volume 4603 of *Lecture Notes in Computer Science*, 2007.