Lecture 1: Deontic Logic, Overview

ESSLLI’18: Multi-Agent Deontic Logic: Reasoning About Normative Multi-Agent Systems
Gabriella Pigozzi & Leon van der Torre

For more details on the history of deontic logic, see the handbook on deontic logic and normative systems. PDF is freely available at http://deonticlogic.org or College Publications.

1 This course

Monday:
- Deontic logic and normative systems, introduction and overview. History of deontic logic as a debate between classical vs. alternative semantics.
- Multiagent deontic logic. STIT theory addressing the challenges of non-deterministic actions, moral luck, and procrastination.

Tuesday:
- Detachment. Factual versus deontic detachment, detachment with anankastic conditionals.
- Multiagent detachment.

Wednesday:
- Argumentation 1. Formal argumentation theory as a framework to solve conflicts between norms / normative arguments: violations, contrary-to-duty obligations, dilemmas, conflict resolution, constitutive and regulative norms, agents arguing about norms.
- Argumentation 2. Representation results. Hierarchical normative systems and defeasible deontic logic.

Thursday:
- Norm interpretation. Argumentation as a framework to interpret norms.
- Norm revision and merging. Revision of a set of obligations: norm change and AGM contraction and revision of rules.

Friday:
- Norms, BDI agents, and games. Social delegation cycle.
- Multiagent organizations. Collective attitudes as grounds for institutions and norms, constitutive norms in artificial institutions. Wrap up and outlook.
2 Standard deontic logic

Preference-based modal logic for conditionals and counterfactuals from the sixties and seventies is a common root for both the deontic logic community, centered around the biannual conference on deontic logic and normative systems (formerly known as the conference on deontic logic in computer science), and a growing number of researchers in linguistics and philosophy studying deontic modality in language.

2.1 Monadic deontic operators

Traditional or ‘standard’ deontic logic, typically referred to as SDL, was introduced by Von Wright in 1951. His variant of SDL is rarely used nowadays due to the strong limitations discussed in Section 2.2. However, it can still be seen as a fragment of currently used systems, which are introduced in Section 2.3.

Given a set $\Phi$ of propositional letters. The language of traditional deontic logic $L_D$ is given by the following BNF:

$$\varphi ::= \bot \mid p \mid \neg \varphi \mid (\varphi \land \varphi) \mid \Box \varphi \mid \lozenge \varphi$$

where $p \in \Phi$. The intended reading of $\Box \varphi$ is “$\varphi$ is obligatory” and $\lozenge \varphi$ as “$\varphi$ is necessary”. Moreover we use $P \varphi$, read as “$\varphi$ is permitted”, as an abbreviation of $\neg \Box \neg \varphi$ and $F \varphi$, “$\varphi$ is forbidden”, as an abbreviation of $\Box \neg \varphi$.

Definition 1 (Modal semantics) A deontic relational model $M = (W, R, V)$ is a structure where:

- $W$ is a set of worlds.
- $R$ is a serial relation over $W$. That is, $R \subseteq W \times W$ and for all $w \in W$, there exist $v \in W$ such that $Rwv$.
- $V$ is a standard propositional valuation such that for every propositional letter $p$, $V(p) \subseteq W$.

Definition 2 We interpret formulas of $L_D$ by deontic relational model.

- $M, s \models p$ iff $s \in V(p)$.
- $M, s \models \neg \varphi$ iff not $M, s \models \varphi$.
- $M, s \models (\varphi \land \psi)$ iff $M, s \models \varphi$ and $M, s \models \psi$.
- $M, s \models \Box \varphi$ iff for all $t$, if $Rst$ then $M, t \models \varphi$.
- $M, s \models \lozenge \varphi$ iff for all $t \in W$, $M, t \models \varphi$.

For a set of formulas $\Gamma$, we write $M, s \models \Gamma$ iff for all $\varphi \in \Gamma$, $M, s \models \varphi$. For a set of formulas $\Gamma$ and a formula $\varphi$, we say $\varphi$ is a consequence of $\Gamma$ (written as $\Gamma \models \varphi$) if for all models $M$ and worlds $w$, if $M, s \models \Gamma$ then $M, s \models \varphi$. 

2
2.2 Limitations of monadic deontic operators

The following example is a variant of a scenario originally phrased by Chisholm in 1963, who requires a formalisation in which the sentences are mutually consistent and logically independent.¹

(A) It ought to be that Jones does not eat fastfood.

(B) It ought to be that if Jones does not eat fastfood for dinner, then he does not go to McDonalds.

(C) If Jones eats fastfood for dinner, then he ought to go to McDonalds.

(D) Jones eats fastfood for dinner.

Consistent representation where sentences are logically independent:

- $(A_1) \bigcirc \neg f$
- $(B_1) \Box (\neg f \rightarrow \bigcirc \neg m)$
- $(C_1) \Box (f \rightarrow \bigcirc m)$
- $(D_1) \neg f$

A drawback of the SDL representation $A_1 - D_1$ is that it does not represent that ideally, the man does not eat fastfood and does not go to McDonalds. Moreover, there does not seem to be a similar solution for the following variant of the scenario:²

(AB) It ought to be that Jones does not eat fastfood and does not go to McDonalds.

(C) If Jones eats fastfood, then he ought to go to McDonalds.

(D) Jones eats fastfood for dinner.

Moreover, SDL only distinguishes between ideal and non-ideal worlds, whereas many ethical dilemmas are based on trade-offs between violations. The challenge is thus how to extend the semantics of SDL in this regard. For example, one can add distinct modal operators for primary and secondary obligations, where a secondary obligation is a kind of repair obligation. From $A_2 - D_2$ we can only derive $\bigcirc_1 m \land \bigcirc_2 \neg m$, but not a contradiction.

- $(A_2) \bigcirc_1 \neg f$
- $(B_2) \bigcirc_1 (\neg f \rightarrow \neg m)$
- $(C_2) f \rightarrow \bigcirc_2 m$
- $(D_2) f$

¹In the original scenario, it ought to be that Jones goes to the assistance of his neighbours, he ought to tell them he is coming, etc. We use the McDonalds example, because going to McDonalds is a means to eat fastfood, and thus we can use a similar scenario for anankastic conditionals tomorrow.

²A variant of Forrester’s scenario, also known as the gentle murderer paradox: You should not kill, but if you kill, you should do it gently.
However, it may not always be easy to distinguish primary from secondary obligations, for example it may depend on the context whether an obligation is primary or secondary. Several authors therefore put as an additional requirement for a representation of the scenario that B and C are represented in the same way (as in A₁-D₁). Finally, the distinction between ⌟₁₁ and ⌟₂₂ is insufficient for extensions of the scenario that seem to need also operators like ⌟₃, ⌟₄, etc, such as the following E and F.

(E) If Jones eats fastfood but does not go to McDonalds, then he should go to Quick.

(F) If Jones eats fastfood but does not go to McDonalds or to Quick, then he should ...  

2.3 Dyadic modal operators

Inspired by rational choice theory in the sixties, preference-based semantics for traditional deontic logic became popular at the end of the sixties (by, for example, Danielsson, Hansson, van Fraassen, Lewis, Spohn). The obligations of Chisholm’s scenario can be represented by a preference ordering, for example:

\[-f \land \neg m > -f \land m > f \land m > f \land \neg m\]

Extensions like E and F can be incorporated by further refining the preference relation. The language is extended with dyadic operators ⌟(p|q), which is true iff the preferred q worlds satisfy p. The class of logics is called Dyadic ‘Standard’ Deontic Logic or DSDL. The notation is inspired by the representation of conditional probability.

Given a set Φ of propositional letters. The language of DSDL L_D is given by the following BNF:

\[\phi := \bot | p | \neg \phi | (\phi \land \phi) | \Box \phi | \Box(\phi/\psi)\]

The intended reading of \(\Box\phi\) is “necessarily \(\phi\)”, \(\Box(\phi/\psi)\) is “It ought to be \(\phi\), given \(\psi\)”. Moreover we use \(P(\phi/\psi)\), read as “\(\phi\) is permitted, given \(\psi\)”, as an abbreviation of \(-\Box(\neg\phi/\psi)\), and \(\Diamond\phi\), read as “possibly \(\phi\)”, as an abbreviation of \(-\Box\neg\phi\).

Unconditional obligations are defined in terms of the conditional ones: \(\Box p = \Box(p|\top)\), where \(\top\) stands for any tautology.

Definition 3 (Preference-based semantics) A dyadic deontic relational model \(M = (W, \geq, V)\) is a structure where:

- \(W\) is a set of worlds.
- \(\geq\) is a reflexive, transitive relation over \(W\) satisfying the following limitedness requirement:³ if \(||\phi|| \neq \emptyset\) then \(\{x \in ||\phi|| : (\forall y \in ||\phi||) x \geq y\} \neq \emptyset\). Here \(||\phi|| = \{x \in W : M, x = \phi\}\).
- \(V\) is a standard propositional valuation such that for every propositional letter \(p\), \(V(p) \subseteq W\).

³It is straightforward to adapt the logic such that the limitedness condition is not required.
Definition 4 We interpret formulas of $\mathcal{L}_D$ by deontic relational model.

- $M, s \models p$ iff $s \in V(p)$.
- $M, s \models \neg \varphi$ iff not $M, s \models \varphi$.
- $M, s \models (\varphi \land \psi)$ iff $M, s \models \varphi$ and $M, s \models \psi$.
- $M, s \models \Box \varphi$ iff for all $t \in W$, $M, t \models \varphi$.
- $M, s \models \Diamond (\psi / \varphi)$ iff $\forall t(((M, t \models \varphi) \land u(M, u \models \varphi) \Rightarrow t \geq u) \Rightarrow M, t \models \psi)$.

Intuitively, $\Diamond (\psi / \varphi)$ holds whenever the best $\varphi$-worlds are $\psi$-worlds.

2.4 Limitations of dyadic deontic operators

The Chisholm scenario in DSDL:

- $(A_3) \Diamond \neg f$
- $(B_3) \Diamond (\neg m | \neg f)$
- $(C_3) \Diamond (m | f)$
- $(D_3) f$

A challenge of both the multiple obligation solution using $\Diamond_1, \Diamond_2, \ldots$ and the preference based semantics is to combine preference orderings, for example combining the Chisholm preferences with preferences originating from the Good Samaritan scenario:

(A'B') A man should not be robbed

(C') If he is robbed, he should be helped

(D') A man is robbed.

$\neg r \land \neg h > r \land h > r \land h$

The main drawback of DSDL is that in a monotonic setting, we cannot detach the obligation $\Diamond m$ from the four sentences. In fact, the preference based solution represents A, B and C, but has little to say about D. So the dyadic representation $A_3D_3$ highlights the dilemma between factual detachment (FD) and deontic detachment (DD). We cannot have both FD and DD, as we derive a dilemma $\Diamond \neg m \land \Diamond m$.

3 Alternative semantics

We consider three main challenges to traditional semantics: normative systems, the use of non-monotonic logic techniques, and the consistent representation of dilemmas.$^4$

$^4$For more details on the history of deontic logic, see the handbook on deontic logic and normative systems (PDF is freely available at http://deonticlogic.org).
3.1 Normative systems

In SDL and DSDL, the logic represents logical relations between deontic operators, but they do not explicitly represent a distinction between norms and obligations. Building on a tradition of Alchourron and Bulygin in the seventies, Makinson argued that norms need to be represented explicitly. This is usually combined with techniques from defeasible deontic logic, discussed below.

Tomorrow in the second lecture we will say more about normative systems and the relation with deontic logic when we zoom in on the problem of detachment.

3.2 DDL: detachment and constraints

Defeasible deontic logics (DDLs) use techniques developed in non-monotonic logic, such as constrained inference. For example, we can derive $\Box m$ from only the first two sentences $A$ and $B$, but not from all four sentences $A$-$D$. Consequently, the inference relation is not monotonic.

For example, we may read $O(\phi|\psi)$ as follows: if the facts are exactly $\psi$, then $\phi$ is obligatory. This implies that we no longer have that $O(\phi)$ is represented by $O(\phi|T)$.

A drawback of the use of non-monotonic techniques is that we often have that violated obligations are no longer derived. This is known as the drowning problem. For example, in the cottage regulations below, if it is no longer derived that there should be no fence once there is a fence, then how do we represent that a violation has occurred?

A second related drawback of this solution is that it does not give the cue for action that the decision maker should change his mind. For example, once there is a fence, it does not represent the obligation to remove the fence.

A third drawback of this approach is that the use of non-monotonic logic techniques like constraints should also be used to represent exceptions, and it thus raises the challenge how to distinguish violations from exceptions. This is highlighted by Prakken and Sergot’s cottage regulations introduced in 1996.

(A”) It ought to be that there is no fence around the cottage
(BC”) If there is a fence around the cottage, then it ought to be white
(G”) If the cottage is close to a cliff, then there ought to be a fence
(D”) There is a fence around the cottage

3.3 Dilemmas

There is a long tradition in deontic logic to represent dilemmas consistently, for example by replacing the Kripke semantics by a neighbourhood semantics, or using techniques from paraconsistent logic.
Using such so-called minimal deontic logics, another approach to represent Chisholm’s scenario is to detach both obligations of the dilemma $\Box \neg m \land \Box m$.

From a practical reasoning point of view, a drawback of this approach is that a dilemma is not very useful as a moral cue for action. Moreover, intuitively it is not clear that the example presents a true dilemma.

4 The search for a unified theory

The main challenge in deontic logic at the moment is a unification of the standard semantics with the alternative ones. The framework of nonmonotonic logic can be used as a general setting for such a unification. In particular, the 2005 textbook of David Makinson “bridges from classical to nonmonotonic logic” can be used as a general framework for such a unification.

Standard semantics has been recast by Kratzer in terms of a modal base and an ordering source, which (roughly) are sets of propositions. The preference ordering is then derived from these two sets by taking all worlds satisfying the modal base, and ordering them by subset relation of satisfaction of formulas of the ordering source. Within this model, normative systems can be introduced by introducing conditionals norms in the ordering source.

A Appendix

A.1 SDL proof system

The proof system of traditional deontic logic $\Lambda_D$ is the smallest set of formulas of $\mathcal{L}_D$ that contains all propositional tautologies, the following axioms:

\begin{align*}
K & \quad \Box (\varphi \rightarrow \psi) \rightarrow (\Box \varphi \rightarrow \Box \psi) \\
D & \quad \Box \varphi \rightarrow P \varphi
\end{align*}

and is closed under modus ponens, and generalization (that is, if $\varphi \in \Lambda_D$, then $\Box \varphi \in \Lambda_D$).

For every $\varphi \in \mathcal{L}_D$, if $\varphi \in \Lambda_D$ then we say $\varphi$ is a theorem and write $\vdash \varphi$. For a set of formulas $\Gamma$ and formula $\varphi$, we say $\varphi$ is deducible form $\Gamma$ (write $\Gamma \vdash \varphi$) if $\vdash \varphi$ or there are formulas $\psi_1, \ldots, \psi_n \in \Gamma$ such that $\vdash (\psi_1 \land \ldots \land \psi_n) \rightarrow \varphi$.

A.2 DSDL Proof system

The proof system of traditional deontic logic $\Lambda_D$, also referred as Aqvist’s system $G$, is the smallest set of formulas of $\mathcal{L}_D$ that contains all propositional tautologies, the following axioms:

\begin{align*}
S5 & \quad \text{S5-schemata for $\Box$} \\
COK & \quad \Box (B \rightarrow C/A) \rightarrow (\Box (B/A) \rightarrow \Box (C/A)) \\
Abs & \quad \Box (B/A) \rightarrow \Box (B/A)
\end{align*}
CON $\square B \rightarrow \Box (B/A)$
Ext $\square (A \leftrightarrow B) \rightarrow (\Box (C/A) \leftrightarrow \Box (C/B))$
Id $\Box (A/A)$
C $\Box (C/(A \land B)) \rightarrow \Box ((B \rightarrow C)/A)$
D $\Diamond A \rightarrow (\Box (B/A) \rightarrow P(B/A))$
S $(P(B/A) \land \Box ((B \rightarrow C)/A)) \rightarrow \Box (C/(A \land B))$

and is closed under *modus ponens*, and *generalization* (that is, if $\varphi \in \Lambda_D$, then $\square \varphi \in \Lambda_D$).

For every $\varphi \in \mathcal{L}_D$, if $\varphi \in \Lambda_D$ then we say $\varphi$ is a theorem and write $\vdash \varphi$. For a set of formulas $\Gamma$ and formula $\varphi$, we say $\varphi$ is deducible form $\Gamma$ (write $\Gamma \vdash \varphi$) if $\vdash \varphi$ or there are formulas $\psi_1, \ldots, \psi_n \in \Gamma$ such that $\vdash (\psi_1 \land \ldots \land \psi_n) \rightarrow \varphi$. 