Handout Lecture 1: Standard Deontic Logic

X. Parent, L. van der Torre, Xin Sun
University of Luxembourg

1 Introduction

The topic of this handout is so-called standard deontic logic (SDL). The key idea of SDL is to assume that there is an analogy between the deontic notions "obligation" and "permission" and the alethic modal notions "necessity" and "possibility." "ϕ is obligatory" is understood as ϕ is true in all morally require situations, while "ϕ is necessary" is understood as ϕ is true in all possible situations. This leads to the development of deontic logic as a branch of modal logic.

⚠ The label "standard" is a misnomer. SDL has been a landmark system until the late 60s, when so-called Dyadic Standard Deontic Logic (see handout DSDL [5]) emerged as a new standard.

2 Language

Definition 1. *Given a set $\mathbb{P}$ of propositional letters, the language of standard deontic logic $\mathfrak{L}$ is the smallest set such that:*

1. $\mathbb{P} \subseteq \mathfrak{L}$.
2. *if* $\phi \in \mathfrak{L}$, *then* $\neg\phi \in \mathfrak{L}$.
3. *if* $\phi \in \mathfrak{L}$ *and* $\psi \in \mathfrak{L}$, *then* $\phi \wedge \psi \in \mathfrak{L}$.
4. *if* $\phi \in \mathfrak{L}$, *then* $\bigcirc\phi \in \mathfrak{L}$.

The boolean connectives $\bot$, $\vee$, $\rightarrow$ and $\leftrightarrow$ are introduced in the usual way. Other abbreviations includes $P\phi ::= \neg \bigcirc (\neg\phi)$ and $F\phi ::= \bigcirc\neg\phi$.

$\mathfrak{L}$ is equivalently represented by the following BNF: for p range over $\mathbb{P}$,

$$\phi = p \mid \neg\phi \mid \phi \wedge \phi \mid \bigcirc\phi$$

The intended reading of $\bigcirc\phi$ is "ϕ is obligatory". $P\phi$ and $F\phi$ are read as "ϕ is permitted" and "ϕ is forbidden" respectively.

⚠ The definition allows for iterated modalities like $\bigcirc(p \wedge \bigcirc p)$. Mixed formulas like $p \wedge \bigcirc q$ are also allowed.

3 Relational semantics

In this section we introduce the relational semantics for SDL.

Definition 2 (Relational model)**.** *A relational model $M = (W, R, V)$ is a tuple where:*

- *W is a (non-empty) set of possible worlds: $s, t, \ldots$*
- *$R \subseteq W \times W$ is a binary relation over W, subject to a constraint of seriality: $(\forall s \in W)(\exists t \in W), (s,t) \in R$.*
- *$V : \mathbb{P} \mapsto 2^W$ is a valuation function for propositional letters such that $V(p) \subseteq W$.*

Here R is known as the relation of deontic alternativeness: $(s,t) \in R$ means that t is an ideal alternative to s, a world in which all the obligations true in s are obeyed.

Intuitively: a sentence of the form $\bigcirc\phi$ is satisfied at a possible world just in case ϕ is satisfied at each of the world's ideal alternatives.

Definition 3 (Satisfaction)**.** *Given a relational model $M = (W, R, V)$ and a world $s \in W$, we define the satisfaction relation $M, s \vDash \phi$ (read as "world s satisfies ϕ in model M") by induction on the structure of ϕ using the following clauses*

- *$M, s \vDash p$ iff $s \in V(p)$.*
- *$M, s \vDash \neg\phi$ iff $M, s \nvDash \phi$.*
- *$M, s \vDash \phi \wedge \psi$ iff $M, s \vDash \phi$ and $M, s \vDash \psi$.*
- *$M, s \vDash \bigcirc\phi$ iff for all $t \in W$, if $(s,t) \in R$ then $M, t \vDash \phi$.*

Exercise 3.1.

(1) Work out the satisfaction rules for $\vee$ and $\rightarrow$ in a relational model.

(2) Same question for P and F.

Solution:

(1) $M, s \vDash \phi \vee \psi$ iff $M, s \vDash \phi$ or $M, s \vDash \psi$.
$M, s \vDash \phi \rightarrow \psi$ iff $M, s \nvDash \phi$ or $M, s \vDash \psi$.

(2) $M, s \models P\phi$ iff there is $t \in W$ such that $(s,t) \in R$ and $M, t \models \phi$.
$M, s \models F\phi$ iff for all $t \in W$, if $(s,t) \in R$ then $M, t \not\models \phi$.

Definition 4 (Validity)**.** *A formula ϕ is valid (notation: $\models \phi$) whenever, for all relational models $M = (W, R, V)$, and all $s \in W$, $M, s \models \phi$.*

Definition 5 (Consequence)**.** *Given a set of formulas Γ, a formula ϕ is a consequence of Γ (notation: $\Gamma \models \phi$) whenever, for all relational model $M = (W, R, V)$, all $s \in W$, if $M, s \models \psi$ for all $\psi \in \Gamma$, then $M, s \models \phi$.*

Exercise 3.2. Explain in what sense the notion of validity may be described as a limiting case of the notion of consequence.
Solution: Validity is a limiting case of consequence in the sense that a formula is valid iff it is the consequence of the empty set. More formally, $\models \phi$ iff $\emptyset \models \phi$.

4 Axiomatisation and completeness

4.1 Axiomatisation

Definition 6. *Let $\mathbf{X}$ be an arbitrary axiomatisation with axioms $Ax_1, Ax_2, ..., Ax_n$ and rules $Ru_1, Ru_2, ...Ru_k$, where each rule $Ru_j (j \leq k)$ is of the form "From ϕ_1, . . . $\phi_{j_{ar}}$ infer ϕ_j". We call j_{ar} the arity of the rule. Then, a derivation for ϕ within $\mathbf{X}$ is a finite sequence $\phi_1, \ldots \phi_m$ of formulas such that:*

1. *$\phi_m = \phi$;*
2. *every ϕ_i in the sequence is*
   (a) *either an **instance** of one of the axioms $Ax_1, Ax_2, ...Ax_n$*
   (b) *or else the result of the application of one of the rules $Ru_j (j \leq k)$ to j_{ar} formulas in the sequence that appear before ϕ_i.*

If there is a derivation for ϕ in $\mathbf{X}$ we write $\vdash_{\mathbf{X}} \phi$, or, if the system $\mathbf{X}$ is clear from the context, we just write $\vdash \phi$. We then also say that ϕ is a theorem *of $\mathbf{X}$, or that $\mathbf{X}$ proves ϕ.*

Definition 7. *The axiomatisation of SDL consists the following axioms and rules.*
Axiom schemes

All tautologies of propositional logic	(PL)
$\bigcirc(\phi \to \psi) \to (\bigcirc\phi \to \bigcirc\psi)$	(K)
$\bigcirc\phi \to P\phi$	(D)

Rules

If $\vdash \phi$ *and* $\vdash \phi \to \psi$ *then* $\vdash \psi$	(MP)
If $\vdash \phi$ *then* $\vdash \bigcirc\phi$	(Nec)

Given a set Γ of formulas, we say that ϕ is derivable from Γ (written as $\Gamma \vdash \phi$) if there are formulas $\psi_1, \ldots, \psi_n \in \Gamma$ such that $\vdash (\psi_1 \wedge \ldots \wedge \psi_n) \to \phi$. (In case where $n = 0$, this means that $\vdash \phi$.) A set Γ is inconsistent iff $\Gamma \vdash \bot$. Otherwise Γ is consistent.

Example 4.1. Below is a derivation of $\bigcirc(p \wedge q) \to \bigcirc p$.

1. $\vdash ((p \wedge q) \to p)$ (PL)
2. $\vdash \bigcirc((p \wedge q) \to p)$ (Nec), 1
3. $\vdash \bigcirc((p \wedge q) \to p) \to (\bigcirc(p \wedge q) \to \bigcirc p)$ (K)
4. $\vdash (\bigcirc(p \wedge q) \to \bigcirc p)$ (MP), 2,3

Exercise 4.1. Show that $\vdash ((\bigcirc(p \wedge q) \wedge Pp) \to Pq)$.
Solution:

1. $\vdash ((p \wedge q) \to (\neg q \to \neg p))$ (PL)
2. $\vdash \bigcirc((p \wedge q) \to (\neg q \to \neg p))$ (Nec) 1
3. $\vdash \bigcirc((p \wedge q) \to (\neg q \to \neg p)) \to (\bigcirc(p \wedge q) \to \bigcirc(\neg q \to \neg p))$ (K)
4. $\vdash \bigcirc(p \wedge q) \to \bigcirc(\neg q \to \neg p)$ (MP) 2,3
5. $\vdash \bigcirc(\neg q \to \neg p) \to (\bigcirc\neg q \to \bigcirc\neg p)$ (K)
6. $\vdash \bigcirc(p \wedge q) \to (\bigcirc\neg q \to \bigcirc\neg p)$ (PL) 4,5
7. $\vdash \bigcirc(p \wedge q) \to (\neg\bigcirc\neg p \to \neg\bigcirc\neg q)$ (PL) 6
8. $\vdash \bigcirc(p \wedge q) \to (Pp \to Pq)$ 7
9. $\vdash ((\bigcirc(p \wedge q) \wedge Pp) \to Pq)$ (PL) 8

4.2 Soundness and completeness theorem

Theorem 1 (Soundness, weak version)**.** *If* $\vdash \phi$ *then* $\models \phi$

Proof. Left as exercise. □

Theorem 2 (Completeness, weak version)**.** *If* $\models \phi$ *then* $\vdash \phi$

Proof. Based on the method of canonical model construction. Details can be found in Chapter 2 of Chellas [2]. □

Theorem 3 (Soundness and completeness, strong version)**.** $\Gamma \models \phi$ *if and only if* $\Gamma \vdash \phi$

5 Chisholm's paradox

The original phrasing of the paradox by Chisholm [3] requires a formalisation of the following scenario in which the sentences are mutually consistent and logically independent.

(A) It ought to be that Jones goes to the assistance of his neighbours.

(B) It ought to be that if Jones goes to the assistance of his neighbours, then he tells them he is coming.

(C) If Jones doesn't go to the assistance of his neighbours, then he ought not to tell them he is coming.

(D) Jones does not go to their assistance.

First attempt is inconsistent.

A_1 $\bigcirc g$

B_1 $\bigcirc(g \to t)$

C_1 $\neg g \to \bigcirc\neg t$

D_1 $\neg g$

Second attempt is not logically independent.

A_2 $\bigcirc g$

B_2 $\bigcirc(g \to t)$

C_2 $\bigcirc(\neg g \to \neg t)$

D_2 $\neg g$

Third attempt is not logically independent either.

A_3 $\bigcirc g$

B_3 $g \to \bigcirc t$

C_3 $\neg g \to \bigcirc\neg t$

D_3 $\neg g$

A fourth attempt based on B_3 and C_2 is missing as we would derive nothing.

Example 5.1. Show that $\Gamma = \{A_1, B_1, C_1, D_1\}$ has no model, and thus it is inconsistent.

Solution. Assume there exists a relational model $M = (W, R, V)$, and a world $w_1 \in W$, that satisfies all the formulas in Γ. Then $M, w_1 \vDash \bigcirc g$, $M, w_1 \vDash \bigcirc(g \to t)$, $M, w_1 \vDash \neg g \to \bigcirc\neg t$ and $M, w_1 \vDash \neg g$.

From $M, w_1 \vDash \neg g$ and $M, w_1 \vDash \neg g \to \bigcirc\neg t$ we deduce $M, w_1 \vDash \bigcirc\neg t$. By seriality we know there is $w_2 \in W$ such that $(w_1, w_2) \in R$. By the definition of $\bigcirc$ we know $M, w_2 \vDash \neg t$, $M, w_2 \vDash g$, $M, w_2 \vDash g \to t$, which is a contradiction. $\dashv$

Exercise 5.1. Show that the second and the third formalisation of Chisholm's paradox are not logically independent.

Remark 5.1. What is a paradox? In the mathematical or strong sense: a paradox is a contradiction. Russells paradox in set-theory is of this sense. In the linguistic or weak sense: a paradox is something predicted by the theory, which a native speaker would not say. Example from propositional logic: the law of commutativity of $\wedge$ (it does not apply when the conjuncts are interpreted sequentially).

In deontic logic, we had paradoxes of both types. Chisholms paradox is closer to the strong sense.

Are paradoxes good or bad? As for paradoxes of the first type, usually people agree that they are bad, and that they should be resolved/avoided. Once spotted, they give rise to new developments, as is the case with the Chisholm example.

6 Deontic logic via reduction

In SDL, the operator $\bigcirc$ is viewed as primitive. One may also define it in terms of an alethic modal operator and a designated propositional constant. This is called reduction. In this section we introduce the reduction developed by Anderson [1].

6.1 Language

Assume the language of modal logic is supplemented with a distinguished propositional constant v read as a violation has occurred. In more detail, given a set Φ of propositional letters, and a propositional constant v such that $v \notin \Phi$, the language of Anderson's logic $\mathfrak{L}_A$ is the smallest set such that:

1. $\Phi \cup \{v\} \subseteq \mathfrak{L}_A$.
2. if $\phi \in \mathfrak{L}_A$, then $\neg\phi \in \mathfrak{L}_A$.
3. if $\phi \in \mathfrak{L}_A$ and $\psi \in \mathfrak{L}_A$, then $\phi \wedge \psi \in \mathfrak{L}_A$.
4. if $\phi \in \mathfrak{L}_A$, then $\Box\phi, \Diamond\phi \in \mathfrak{L}_A$.

The deontic operators are defined by:

- $\bigcirc\phi ::= \Box(\neg\phi \to v)$
- $P\phi ::= \Diamond(\neg v \wedge \phi)$

6.2 Semantics

A relational model for Anderson's logic is a tuple $M = (W, R, DEM, V)$ such that

- W is a non-empty set

- $R \subseteq W \times W$
- $DEM \subseteq W$ such that for all $s \in W$ there is $t \in W$ such that $(s, t) \in R$ and $t \notin DEM$.
- V is a valutation function for propositional letters.

Let $M = (W, R, DEM, V)$, $s \in W$, satisfaction of formulas of $\mathfrak{L}_A$ is defined as follows:

- $M, s \vDash p$ iff $s \in V(p)$.
- $M, s \vDash v$ iff $s \in DEM$.
- $M, s \vDash \neg\phi$ iff $M, s \nvDash \phi$.
- $M, s \vDash \phi \wedge \psi$ iff $M, s \vDash \phi$ and $M, s \vDash \psi$.
- $M, s \vDash \Box\phi$ iff for all $t \in W$, if $(s, t) \in R$ then $M, t \vDash \phi$.

6.3 Axiomatisation

The axiomatisation of Anderson's logic consists the following axioms and rules. We write $\vdash_A \phi$ for ϕ is a theorem of Anderson's logic.
Axioms schemes

$$\text{All tautologies of propositional logic} \qquad \text{(PL)}$$
$$\Box(\phi \to \psi) \to (\Box\phi \to \Box\psi) \qquad \text{(K}\Box\text{)}$$
$$\Diamond\neg v \qquad (\Diamond\ \text{d})$$

Rules

$$\text{If } \vdash_A \phi \text{ and } \vdash_A \phi \to \psi \text{ then } \vdash_A \psi \qquad \text{(MP)}$$
$$\text{If } \vdash_A \phi \text{ then } \vdash_A \Box\phi \qquad \text{(Nec }\Box\text{)}$$

6.4 Chisholm's scenario via Anderson

Using Anderson's reduction, the Chisholm's scenario is formalized as

A_4 $\bigcirc g$

B_4 $\Box(g \to \bigcirc t)$

C_4 $\Box(\neg g \to \bigcirc\neg t)$

D_4 $\neg g$

Example 6.1. Show that Anderson's formalization of Chisholm's paradox is consistent/satisfiable.
Solution: The following is a relational model satisfies Anderson's formalization of Chisholm's paradox:
$M = (W, R, DEM, V)$, $W = \{w_1, w_2\}$, $R = \{(w_1, w_2), (w_2, w_2)\}$, $DEM = \{w_1\}$, $V(g) = \{w_2\}, V(t) = \emptyset$. It can be verified that $A_4, \ldots, D_4$ are satisfied at M, w_1.

Exercise 6.1. Does Anderson's logic validate the D axiom $\bigcirc\phi \to P\phi$? Given a proof to your answer.

References

[1] A. R. Anderson. A reduction of deontic logic to alethic modal logic. *Mind*, 67(265):100–103, 1958.

[2] B. F. Chellas. *Modal Logic: An Introduction*. Cambridge University Press, Cambridge, 1980.

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

[4] W. H. Hanson. Semantics for deontic logic. *Logique Et Analyse*, 8:177–190, 1965.

[5] Leendert van der Torre, Xavier Parent, and Xin Sun. Handout lecture 2 :dyadic standard deontic logic. Unpublished notes.

Appendix

Neighborhood semantics

There is a generalization of the relational semantics, Chellas [2]'s minimal deontic logic.

Definition 8 (Minimal model). *A minimal model $M = (W, N, V)$ is a structure where W and V are as before, and N, so-called neighborhood function, is a mapping from W to sets of subsets of W (i.e. $N(s) \subseteq 2^W$ for each $s \in W$).*

Definition 9 (Satisfaction). *Given a minimal model $M = (W, N, V)$, and a world $s \in W$, we define the satisfaction relation $M, s \vDash \phi$ as before, except for deontic formulas, where:*

$$M, s \vDash \bigcirc\phi \text{ iff } ||\phi|| \in N(s)$$

Here $||\phi|| = \{t \in W : M, t \vDash \phi\}$.

Validity and consequence are defined as in the relational semantics.

The neighborhood approach allows for an extra degree of freedom here. The obligation operator as defined in Definition 9 is very weak. But extra constraints may be put on $N(s)$ in order to make the operator validate more laws, as one thinks fit.

For an illustration, consider Chellas [2]'s minimal deontic logic, which he calls system $\mathcal{D}$.[1] According to Chellas, $\mathcal{D}$ is a more plausible system, because it has the law OD, but not the law OD*:

$$\neg \bigcirc \bot \qquad \text{(OD)}$$

$$\neg \bigcirc (\varphi \wedge \neg\varphi) \qquad \text{(OD}^\star\text{)}$$

OD expresses the seemingly uncontroversial principle "ought" implies "can". OD* rules out the possibility of conflicts between obligations, which seems counter-intuitive. In SDL, we have $\vdash (\neg\bigcirc\bot) \leftrightarrow (\neg\bigcirc(\varphi\wedge\neg\varphi))$. This in fact applies to any normal modal logic of type K, making it impossible to distinguish between OD* and OD, and have one without the other.

Definition 10 (System $\mathcal{D}$). *The axiomatisation of System $\mathcal{D}$ consists the following axioms and rules. We write $\vdash_{\mathcal{D}} \phi$ for ϕ is a theorem in $\mathcal{D}$.*
Axiom scheme

$$\neg \bigcirc \bot \qquad \text{(OD)}$$

Rule

$$\text{If } \vdash_{\mathcal{D}} \phi \to \psi \text{ then } \vdash_{\mathcal{D}} \bigcirc\phi \to \bigcirc\psi \qquad \text{(ROM)}$$

[1] We use calligraphic fonts to avoid confusion with the axiom with the same name.

The relevant constraints to be placed on N are:

$$\text{If } U \subseteq V \text{ and } U \in N(s) \text{ then } V \in N(s) \qquad \text{(closure under superset)}$$

$$\emptyset \notin N(s) \qquad \text{(no-absurd)}$$

Exercise 6.2.

1) Show that ROM holds in view of the property of closure under superset.
2) Work out the evaluation rules for P and F in a minimal model.

Tableaux

The so-called method of tableaux [4] provides an efficient decision procedure for checking whether a formula is valid. This is achieved by analyzing what it would take to produce a counter-example. For ϕ to be valid, all attempts at producing a counterexample for it must fail.

Notation:

- $\mathbf{T}\phi, s$: ϕ is true at world s
- $\mathbf{F}\phi, s$: ϕ is false at world s.

Principle Start by assuming that $\mathbf{F}\phi, s$. Break $\mathbf{F}\phi, s$ into its components up to the simplest ones, until a contradiction is reached.

In PL, this creates a tree structure called tableau. In a modal logic setting, we are not dealing with a single tableau, but with sequences of tableaux.

All formulae with the same world label s are written in a rectangle, called a tableau. The rules for the propositional connectives refer to formulae within the same tableau. Example:

$\mathbf{T}(\phi \wedge \psi), s$
$\mathbf{T}\phi, s$
$\mathbf{T}\psi, s$

$\mathbf{F}(\phi \wedge \psi), s$
$\mathbf{F}\phi, s \quad \mathbf{F}\psi, s$

If Rst, we say that t is "auxiliary to" s (or "next to" s).

⚠ By the requirement of seriality of R, a tableau t always has an auxiliary one attached to it– be it t itself.

The rules for $\bigcirc$ are as follows:

$\mathbf{F} \bigcirc \phi, s$
Rst (t new)
$\mathbf{F}\phi, t$

$\mathbf{T} \bigcirc \phi, s$
Rst
$\mathbf{T}\phi, t$

The rule to the left always creates a new auxiliary tableau (t must be new). The rule to the right incorporates the requirement of seriality of R. It says: if an auxiliary tableau

t has already been created, then fill it in with $\mathbf{T}\phi, t$; otherwise, create one such tableau (using a new t).

Some rules of propositional logic create a tree structure in a tableau. This yields to two alternative tableaux-sequences. Example:

$$\boxed{\mathbf{F} \bigcirc (\phi \wedge \psi), s}$$

Rst (t new)

$$\boxed{\begin{array}{c} \mathbf{F}(\phi \wedge \psi), t \\ \mathbf{F}\phi, t \quad \mathbf{F}\psi, t \end{array}}$$

Two alternative tableaux-sequences. In one, t contrains $\mathbf{F}\phi$. In the other, it contains $\mathbf{F}\psi$.

A tableau is closed if it contains a pair of the form

$$\boxed{\begin{array}{c} \vdots \\ \mathbf{T}\psi, s \\ \mathbf{F}\psi, s \\ \times \end{array}}$$

A sequence of tableaux is closed if some term in it is closed. A system of alternative tableaux-sequences is closed if each of the alternative sequences in it are closed. A formula ϕ is valid, if the construction for $\mathbf{F}\phi, s$ gives a closed system of alternative tableaux-sequences. This means that all the attempts at providing a counterexample for ϕ fail.

The method is illustrated with the example of K:

$$\boxed{\begin{array}{c} \mathbf{F} \bigcirc (\phi \to \psi) \to (\bigcirc\phi \to \bigcirc\psi), s \\ \mathbf{T} \bigcirc (\phi \to \psi), s \\ \mathbf{F} \bigcirc \phi \to \bigcirc\psi, s \\ \mathbf{T} \bigcirc \phi, s \\ \mathbf{F} \bigcirc \psi, s \end{array}}$$

Rst (t new)

$$\boxed{\begin{array}{c} \mathbf{F}\psi, t \\ \mathbf{T}\phi, t \\ \mathbf{T}(\phi \to \psi), t \\ \mathbf{F}\phi, t \quad \mathbf{T}\psi, t \\ \times \qquad \times \end{array}}$$

Exercise 6.3.

1) Define suitable rules for $\top$, $\vee$, $\leftrightarrow$, and $\to$.

2) Same question for P and F.

3) Show the validity of the following formula using tableaux:

$$\bigcirc(\phi \wedge \psi) \to (\bigcirc\phi \wedge \bigcirc\psi)$$