VII. Deontic Logic

AS.150.498: Modal Logic and Its Applications
Johns Hopkins University, Spring 2017

Our first application is *deontic logic* which is concerned with obligation, permission, prohibition, and related normative concepts.

1 Syntax, Semantics, and Proof Systems

We will be working with the following language:

Definition 9.1. The **deontic language** $\mathcal{L}_d$ extends the basic sentential language with obligation and permission operators:

$p \mid \bot \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid \mathrm{O}\varphi) \mid \mathrm{P}\varphi$

Read $\mathrm{O}\varphi$ as 'It ought to be the case that φ' and $\mathrm{P}\varphi$ as 'It is permissible that φ.'

Note that O and P are interdefinable: $\mathrm{O} \equiv \neg\mathrm{P}\neg$ and $\mathrm{P} \equiv \neg\mathrm{O}\neg$.

The semantics for $\mathcal{L}_d$ is the standard one based on Kripke models where $w\mathcal{R}v$ just in case v is a **deontically ideal** world relative to w.

Intuitively, $\mathrm{O}\varphi \supset \varphi$ is invalid—that is, $\mathcal{R}$ needn't be reflexive.

Intuitively, $\mathrm{O}\varphi \supset \mathrm{P}\varphi$ is valid—that is, $\mathcal{R}$ should be serial.

At first glance, then, it seems that deontic logic should be at least as strong as **KD** but shouldn't validate **T**.

Definition 9.2. Standard Deontic Logic (SDL) is the logic **KD**:

(PL)	All (substitutions of) tautologies are axioms
(MP)	From φ and $\varphi \supset \psi$ infer ψ
(Nec_d)	From φ infer $\mathrm{O}\varphi$
(K_d)	For any φ, ψ, $\mathrm{O}(\varphi \supset \psi) \supset (\mathrm{O}\varphi \supset \mathrm{O}\psi)$ is an axiom
(D_d)	For any φ, $\mathrm{O}\varphi \supset \mathrm{P}\varphi$ is an axiom
(Duality)	Expressions involving O and P are interchangeable according to the duality $\mathrm{O} \equiv \neg\mathrm{P}\neg$

Intuitively, $\mathrm{O}(\mathrm{O}\varphi \supset \varphi)$ is also valid—while $\mathcal{R}$ needn't be reflexive, this relation should be *shift reflexive*: $\forall w, v(w\mathcal{R}v \supset v\mathcal{R}v)$.

Definition 9.3. SDL^+ is the logic obtained by supplementing **KD** with the axiom schema $\mathrm{O}(\mathrm{O}\varphi \supset \varphi)$.

2 Anderson-Kanger Reduction

In lieu of treating O as a primitive operator, Anderson [1956] and Kanger [1957] proposed reducing deontic logic to ordinary alethic logic as follows:

$\mathrm{O}\varphi \equiv \Box(D \supset \varphi)$ (alternatively: $\mathrm{O}\varphi \equiv \Box(\neg\varphi \supset S)$)

where D designates that all normative requirements have been met (and S designates that a sanction has been imposed).

If the logic of $\Box$ is **K** plus the axiom $\Diamond D$, then the corresponding logic of O is SDL.

Here are some theorems of the combined logic of $\Box$ and O:

$\mathrm{O}D$
$\Box\varphi \supset \mathrm{O}\varphi$
$\Box(\varphi \supset \psi) \supset (\mathrm{O}\varphi \supset \mathrm{O}\psi)$
$\neg\Diamond(\mathrm{O}\varphi \wedge \mathrm{O}\neg\varphi)$
$\mathrm{O}\varphi \supset \Diamond\varphi$ ('ought' implies 'can')

If the logic of $\Box$ is **KT** plus the axiom $\Diamond D$, then the corresponding logic of O is SDL^+.

3 Problems

Conflicting Obligations. SDL rules out conflicting obligations:

1.	$(\mathrm{O}\varphi \wedge \mathrm{O}\neg\varphi) \supset \mathrm{O}(\varphi \wedge \neg\varphi)$	C
2.	$\neg\mathrm{O}(\varphi \wedge \neg\varphi)$	From D_d
3.	$\neg(\mathrm{O}\varphi \wedge \mathrm{O}\neg\varphi)$	PL 1,2

However, such conflicts arguably occur:

(P1) I ought to fight in the war (since I signed a contract to do so).

(P2) I ought not to fight in the war (since the war is unjust).

Some possible responses:

—Abandon **C** and work with neighborhood semantics.

—Abandon $\mathbf{D}_d$.

—Deny the possibility of conflicting obligations. Allow for different kinds of 'ought' (moral, prudential, all-things-considered, etc.) and deny that conflict can arise for any particular 'ought.'

Free Choice Permission. The following inference seems good:

(P1) You may have the whiskey or the gin.
(C) You may have the whiskey and you may have the gin.

However, this inference is invalidated by SDL.

Some possible responses:

—Appeal to Gricean conversational implicature.

—Abandon the standard semantics for $\vee$.

Ross' Paradox. The following inferences seem terrible:

(P1) You ought to mail the letter.
(C) You ought to mail the letter or burn it.

(P1) You may have the whiskey.
(C) You may have the whiskey or the gin.

However, these inferences are validated by SDL given the axiom **M**.

Some possible responses:

—Abandon **M** and work with neighborhood semantics.

—Explain the oddness of the inferences in pragmatic terms.

Paradox of Epistemic Obligation. Consider the following argument (cf. Aqvist [1967]):

(P1) There is a fire.

(P2) If there is a fire, it ought to be that the firefighter knows that there is a fire.

(C) It ought to be that there is a fire.

This is terrible but comes out valid in SDL:

1.	F	P1
2.	$F \supset \mathrm{OK}_f F$	P2
3.	$\mathrm{OK}_f F$	MP 2,1
4.	$\mathrm{K}_f F \supset F$	Factivity of K_f
5.	$\mathrm{OK}_f F \supset \mathrm{O}F$	Nec_d, K_d, MP 4
6.	$\mathrm{O}F$	MP 5,3

Some possible responses:

—Abandon **K** and work with neighborhood semantics.

—Abandon Nec_d (but note that this is applied only to $\mathrm{K}_f F \supset F$ which is presumably a validity).

Good Samaritan Paradox. Consider the following argument (Prior [1958]):

(P1) It ought to be that Jones helps Smith who has been robbed.
(C) It ought to be that Smith has been robbed.

This is terrible but comes out valid in SDL:

1.	$\mathrm{O}(H \wedge R)$	P1
2.	$\mathrm{O}R$	M, MP

Some possible responses:

—Deny that $\mathrm{O}(H \wedge R)$ is a good translation of P1.

—Abandon **M** and work with neighborhood semantics.

Chisholm's Paradox. The following statements appear consistent and pairwise logically independent (Chisholm [1963]):

(P1) It ought to be that Jones goes to help his neighbors.

(P2) It ought to be that Jones tells his neighbors he is coming if he is going to help them.

(P3) If Jones doesn't go to help, it ought to be that he doesn't tell his neighbors he is coming.

(P4) Jones doesn't go help his neighbors.

However, these are inconsistent in SDL:

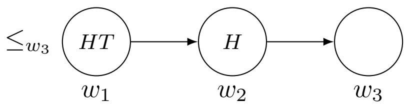
1.	$\mathrm{O}H$	P1
2.	$\mathrm{O}(H \supset T)$	P2
3.	$\neg H \supset \mathrm{O}\neg T$	P3
4.	$\neg H$	P4
5.	$\mathrm{O}H \supset \mathrm{O}T$	K_d, MP 2
6.	$\mathrm{O}T$	MP 5,1
7.	$\mathrm{P}T$	D_d, MP
8.	$\neg\mathrm{O}\neg T$	Duality 7
9.	$\mathrm{O}\neg T$	MP 3,4
10.	$\perp$	From 8,9

Some possible responses:

—Translate P2 and P3 as $H \supset \mathrm{O}T$ and $\neg H \supset \mathrm{O}\neg T$ respectively. But then P2 follows from P4 so the statements are not independent.

—Translate P2 and P3 as $\mathrm{O}(H \supset T)$ and $\mathrm{O}(\neg H \supset \neg T)$ respectively. But then P3 follows from P1 so the statements are not independent.

—Replace the unary obligation and permission operators with the dyadic operators $\mathrm{O}(\psi/\varphi)$ and $\mathrm{P}(\psi/\varphi)$. Read $\mathrm{O}(\psi/\varphi)$ as 'It ought to be the case that ψ *given that* φ' and $\mathrm{P}(\psi/\varphi)$ as 'It is permissible that ψ *given that* φ.' The semantics for these operators is similar to the semantics for counterfactuals in using an ordering on worlds. But now $v \leq_w u$ just in case v is as good as u relative to w. Translating the premises as $\mathrm{O}(H/\neg\bot)$, $\mathrm{O}(T/H)$, $\mathrm{O}(\neg T/\neg H)$, and $\neg H$, these are all true at w_3 in the model below:



—Keep the unary deontic operators but replace the material conditional with a more sophisticated intensional conditional.

Gentle Murderer Paradox. The following statements appear consistent (Forrester [1984]):

(P1) Smith murders Jones.

(P2) Smith ought not murder Jones.

(P3) If Smith murders Jones, he ought to murder Jones gently.

However, these are inconsistent in SDL:

1.	M	P1
2.	$\mathrm{O}\neg M$	P2
3.	$M \supset \mathrm{O}(M \wedge G)$	P3
4.	$\mathrm{O}(M \wedge G)$	MP 3,1
5.	$\mathrm{O}M$	M, MP
6.	$\mathrm{P}M$	D_d, MP
7.	$\neg\mathrm{O}\neg M$	Duality 6
8.	$\bot$	From 2,7

Some possible responses:

—Abandon **M** and work with neighborhood semantics.

—Abandon $\mathbf{D}_d$.

—Work with $\mathrm{O}(\psi/\varphi)$ and $\mathrm{P}(\psi/\varphi)$. Translating the premises as M, $\mathrm{O}(\neg M/\neg\bot)$, and $\mathrm{O}(M \wedge G/M)$, these are all true at w_3 in the model below: