Lecture 1:
Deontic Logic and Anankastic Conditionals, Overview

ESSLLI'16: Deontic modality: linguistic and logical perspectives on oughts and ends

Cleo Condoravdi & Leon van der Torre

Abstract

We first discuss traditional deontic logic introduced at the end of the sixties by Hansson-Lewis and based on dyadic operators and preference based semantics, and contrast it with the use of preference in decision theory. Then we discuss the history of deontic logic as a debate between these traditional and alternative semantics. Finally, we introduce anankastic conditionals as a central challenge for deontic reasoning.

Hansson-Lewis preference based deontic logic vs. preference in decision theory

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.

Modal language

Traditional or 'standard' deontic logic, typically referred to as SDL, was introduced by Von Wright in 1951.

Given a set Φ of propositional letters. The language of traditional deontic logic $\mathfrak{L}_D$ is given by the following BNF:

$$\varphi := \bot \mid p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid \bigcirc\varphi \mid \Box\varphi$$

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

Modal semantics

Definition 1. *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 $\mathfrak{L}_D$ by deontic relational model.*

- *$M, s \vDash p$ iff $s \in V(p)$.*
- *$M, s \vDash \neg\varphi$ iff not $M, s \vDash \varphi$.*
- *$M, s \vDash (\varphi \wedge \psi)$ iff $M, s \vDash \varphi$ and $M, s \vDash \psi$.*
- *$M, s \vDash \bigcirc\varphi$ iff for all t, if Rst then $M, t \vDash \varphi$.*
- *$M, s \vDash \Box\varphi$ iff for all $t \in W$, $M, t \vDash \varphi$.*

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

Limitations of a monadic deontic operator

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

[1] In the original paradox, 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 later.

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 $\bigcirc_1 m \wedge \bigcirc_2 \neg m$, but not a contradiction.

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

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 solution of the paradox that **B** and **C** are represented in the same way (as in A_1-D_1). Finally, the distinction between $\bigcirc_1$ and $\bigcirc_2$ is insufficient for extensions of the paradox that seem to need also operators like $\bigcirc_3$, $\bigcirc_4$, 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 ...

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 paradox can be represented by a preference ordering, for example:

$$\neg f \wedge \neg m > \neg f \wedge m > f \wedge m > f \wedge \neg m$$

Extensions like **E** and **F** can be incorporated by further refining the preference relation. The language is extended with dyadic operators $\bigcirc(p|q)$, 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.

Dyadic language

Given a set Φ of propositional letters. The language of DSDL $\mathfrak{L}_D$ is given by the following BNF:

$$\varphi := \bot \mid p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid \Box\varphi \mid \bigcirc(\varphi/\varphi)$$

[2] A variant of Forrester's paradox, also known as the gentle murderer paradox: You should not kill, but if you kill, you should do it gently.

The intended reading of $\Box\varphi$ is "necessarily φ", $\bigcirc(\varphi/\psi)$ is "It ought to be φ, given ψ". Moreover we use $P(\varphi/\psi)$, read as "φ is permitted, given ψ", as an abbreviation of $\neg \bigcirc (\neg\varphi/\psi)$, and $\Diamond\varphi$, read as "possibly φ", as an abbreviation of $\neg\Box\neg\varphi$.

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

Preference based semantics

Definition 3. *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: if $||\varphi|| \neq \emptyset$ then $\{x \in ||\varphi|| : (\forall y \in ||\varphi||) x \geq y\} \neq \emptyset$. Here $||\varphi|| = \{x \in W : M, x \vDash \varphi\}$.*
- *V is a standard propositional valuation such that for every propositional letter p, $V(p) \subseteq W$.*

Definition 4. *We interpret formulas of $\mathfrak{L}_D$ by deontic relational model.*

- *$M, s \vDash p$ iff $s \in V(p)$.*
- *$M, s \vDash \neg\varphi$ iff not $M, s \vDash \varphi$.*
- *$M, s \vDash (\varphi \wedge \psi)$ iff $M, s \vDash \varphi$ and $M, s \vDash \psi$.*
- *$M, s \vDash \Box\varphi$ iff for all $t \in W$, $M, t \vDash \varphi$.*
- *$M, s \vDash \bigcirc(\psi/\varphi)$ iff $\forall t(((M, t \vDash \varphi) \& \forall u(M, u \vDash \varphi) \Rightarrow t \geq u) \Rightarrow M, t \vDash \psi)$.*

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

Chisholm scenario represented

The Chisholm scenario in DSDL:

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

A challenge of both the multiple obligation solution using $\bigcirc_1$, $\bigcirc_2$, ... and the preference based semantics is to combine preference orderings, for example combining the Chisholm preferences with preferences originating from the Good Samaritan paradox:

(AB') A man should not be robbed

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

(D') A man is robbed.

$$\neg r \wedge \neg h > r \wedge h > r \wedge \neg h$$

The main drawback of DSDL is that in a monotonic setting, we cannot detach the obligation $\bigcirc m$ 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 $A_3 - D_3$ highlights the dilemma between factual detachment (FD) and deontic detachment (DD). We cannot have both FD and DD, as we derive a dilemma $\bigcirc\neg m \wedge \bigcirc m$.

$$\frac{\bigcirc(m|f), f}{\bigcirc m} FD \qquad \frac{\bigcirc(\neg m|\neg f), \bigcirc\neg f}{\bigcirc\neg m} DD$$

The use of preference in decision theory

Rational choice theory tells us: if C are the best alternatives of A, and $B \cap C$ is nonempty, then $B \cap C$ are the best alternatives of $A \cap B$. This principle is reflected by the S axiom of DSDL:

$$(P(B/A) \wedge \bigcirc((B \rightarrow C)/A)) \rightarrow \bigcirc(C/(A \wedge B))$$

Moreover, we may represent a preference or comparative operator $\succ$ in the language, and define the dyadic operator in terms of the preference logic:

$$O(\psi \mid \phi) =_{def} (\phi \wedge \psi) \succ (\phi \wedge \neg\psi)$$

One may wonder whether the parallel between deontics and rational choice can be extended to utility theory, decision theory, game theory, planning, and so on. First, consider a typical example from Prakken and Sergot's Cottage Regulations: there should be no fence, if there is a fence there should be a white fence, if there is a non-white fence, it should be black, if there is a fence which is neither white nor black, then ....[3] The associated preferences are:

no fence $>$ *white fence* $>$ *black fence* $> \ldots$

However, if this represents a utility ordering over states, then we miss the representation of action. For example, it may be preferred that the sun shines, but we do not say that the sun should shine. As a simple model of action, one might distinguish controllable from uncontrollable propositions, and restrict obligations to controllable propositions. Moreover, we may consider actions instead of states: we should remove the fence if there is one, we may paint the fence white, we may paint it black, etc.

remove $>$ *paint white* $>$ *paint black* $> \ldots$

We may interpret this preference ordering as an ordering of expected utility of actions. Alternatively, the ordering may be generated by another decision rule, such as maximin or minimal regret. Once we are working with a decision theoretic semantics, we may represent probabilities explicitly, and model causality. For example, let n stand for not doing homework and g for getting a good grade for a test. Then we may have the following preference order, which does not reflect that doing homework causes good grades:

$$n \wedge g > \neg n \wedge g > n \wedge \neg g > \neg n \wedge \neg g$$

The use of goals in planning and agent theory

We may interpret $O\phi$ or $O(\phi \mid \psi)$ as goals for ϕ, rather than obligations. This naturally leads to the distinction between maintenance and achievement goals, and to extensions of the logic with beliefs and intentions. Such BDI logics have been developed as formalizations of BDI theory.

BDI theory is developed in theory of mind and has been based on folk psychology. In planning, more efficient alternatives to classical planning have been developed, for example based on hierarchical or graph planning.

[3]This part of the cottage regulations are related to Forrester's paradox. However, note the following difference between Forrester's paradox and the cottage regulations. Once you kill someone, it can no longer be undone, whereas if you build a fence, you can still remove it.

History of deontic logic as a debate between classical vs. 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]

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.

DDL: detachment and constraints

Defeasible deontic logics (DDLs) use techniques developed in non-monotonic logic, such as constrained inference. For example, we can derive $\bigcirc m$ 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 ψ, then ϕ is obligatory. This implies that we no longer have that $O(\phi)$ is represented by $O(\phi|\top)$.

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, 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.

(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

Dilemmas and aggregative deontic detachment

Another approach to Chisholm's paradox is to detach both obligations of the dilemma $\bigcirc\neg m \wedge \bigcirc m$, and represent them consistently using some kind of minimal deontic logic, for example using techniques from paraconsistent logic.

From a practical reasoning point of view, a drawback of this approach is that a dilemma is not very useful as a moral

[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).

cue for action. Moreover, intuitively it is not clear that the example presents a true dilemma.

A recent representation of Chisholm's paradox is to replace deontic detachment by so-called aggregative deontic detachment (ADD), and to derive from **A-D** the obligation $\bigcirc(\neg f \wedge \neg m)$ and $\bigcirc m$, but not $\bigcirc \neg m$.

$$\frac{\bigcirc(m|f), f}{\bigcirc m} FD \qquad \frac{\bigcirc(\neg m|\neg f), \bigcirc\neg f}{\bigcirc(\neg m \wedge \neg f)} ADD$$

A drawback of this solution is that we can no longer accept the principle of weakening (also known as inheritance).

$$\frac{\bigcirc(\neg m \wedge \neg f|\top)}{\bigcirc(\neg m|\top)} W$$

Anankastics conditionals as a central challenge for deontic reasoning

In this course, we propose the following example as a more challenging variant of Chisholm's scenario.

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

(B) If Jones wants to eat fastfood for dinner, then he ought not to go to McDonalds.

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

(D) Jones eats fastfood for dinner.

Here McDonalds is assumed to be the best restaurant, and the example is similar to:

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

(B) If Jones wants to eat fastfood for dinner, then he ought not to go to a fastfood restaurant.

(C) If Jones wants to eat fastfood for dinner, then he ought to go to a fastfood restaurant.

(D) Jones eats fastfood for dinner.

Another alternative:

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

(B) If Jones wants to eat fastfood for dinner, then he ought not to go to Quick.

(C) If Jones wants to eat fastfood for dinner, then he ought to go to Quick.

(D) Jones eats fastfood for dinner.

And another popular example:

(A) It ought to be that Jones does not buy a new table.

(B) If Jones wants to invite people for dinner, then he ought to buy a new table.

(C) If Jones does not want to buy a new table, then he ought not to invite people for dinner.

Appendix

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

K $\bigcirc(\varphi \to \psi) \to (\bigcirc\varphi \to \bigcirc\psi)$

D $\bigcirc\varphi \to P\varphi$

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

For every $\varphi \in \mathfrak{L}_{\mathfrak{D}}$, if $\varphi \in \Lambda_D$ then we say φ is a theorem and write $\vdash \varphi$. For a set of formulas Γ and formula φ, we say φ is deducible form Γ (write $\Gamma \vdash \varphi$) if $\vdash \varphi$ or there are formulas $\psi_1, \ldots, \psi_n \in \Gamma$ such that $\vdash (\psi_1 \wedge \ldots \wedge \psi_n) \to \varphi$.

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

S5 S5-schemata for $\Box$

COK $\bigcirc(B \to C/A) \to (\bigcirc(B/A) \to \bigcirc(C/A))$

Abs $\bigcirc(B/A) \to \Box \bigcirc (B/A)$

CON $\Box B \to \bigcirc(B/A)$

Ext $\Box(A \leftrightarrow B) \to (\bigcirc(C/A) \leftrightarrow \bigcirc(C/B))$

Id $\bigcirc(A/A)$

C $\bigcirc(C/(A \wedge B)) \to \bigcirc((B \to C)/A)$

D * $\Diamond A \to (\bigcirc(B/A) \to P(B/A))$

S $(P(B/A) \wedge \bigcirc((B \to C)/A)) \to \bigcirc(C/(A \wedge B))$

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

For every $\varphi \in \mathfrak{L}_{\mathfrak{D}}$, if $\varphi \in \Lambda_D$ then we say φ is a theorem and write $\vdash \varphi$. For a set of formulas Γ and formula φ, we say φ is deducible form Γ (write $\Gamma \vdash \varphi$) if $\vdash \varphi$ or there are formulas $\psi_1, \ldots, \psi_n \in \Gamma$ such that $\vdash (\psi_1 \wedge \ldots \wedge \psi_n) \to \varphi$.