

# Deontic logic for obligations that vary in degrees

Johan Gustafsson, Tiago de Lima and  
Martin Peterson

September 16, 2009

# Background

Two fundamentally different motivations:

- ▶ Computer scientists wish to apply deontic logic to systems where states are "forbidden to a certain degree".

# Background

Two fundamentally different motivations:

- ▶ Computer scientists wish to apply deontic logic to systems where states are "forbidden to a certain degree".
- ▶ Philosophers sometimes claim that (moral) obligations vary in degrees.

## Moral Philosophy & Deontic Logic 1(2)

1. “The creed which accepts as the foundation of morals, Utility, or the Greatest Happiness Principle, holds that actions are right *in proportion as* they tend to promote happiness, wrong as they tend to produce the reverse of happiness.” (Mill 1859: 210)
2. “There are not only discrete deontic qualities, such as rightness and wrongness, but within the space of wrong actions there is a continuum ranging from, say, very wrong indeed to nearly right.” (Eriksson 1997:213)
3. We should “talk about *the degree to which* [an act] was acceptable or not rather than talking about whether it was acceptable or not.” (Selgelid 2009:203)
4. Ross’s theory of prima facie duties.
5. Moral dilemmas?

## Moral Philosophy & Deontic Logic 2(2)

- ▶ A group of people are about to drown aboard a sinking cruise liner. You are morally obliged to rescue them – this is easy and safe.

## Moral Philosophy & Deontic Logic 2(2)

- ▶ A group of people are about to drown aboard a sinking cruise liner. You are morally obliged to rescue them – this is easy and safe.
- ▶ Another group of people are about to drown aboard another sinking cruise liner. It is dangerous and difficult to rescue them. But you are still under an obligation to rescue them.

## Moral Philosophy & Deontic Logic 2(2)

- ▶ A group of people are about to drown aboard a sinking cruise liner. You are morally obliged to rescue them – this is easy and safe.
- ▶ Another group of people are about to drown aboard another sinking cruise liner. It is dangerous and difficult to rescue them. But you are still under an obligation to rescue them.
- ▶ In the second example, your obligation to rescue the people is valid to a lower degree.

## A semantic intuition 1(2)

SDL:  $p$  is obligatory if and only if  $p$  is true in all deontically ideal worlds.

- ▶ Our proposal:



## A semantic intuition 1(2)

SDL:  $p$  is obligatory if and only if  $p$  is true in all deontically ideal worlds.

- ▶ Our proposal:
- ▶ Let  $W$  be the set of all accessible possible worlds and let  $d(w)$  be the deontic value of world  $w \in W$ ; we assume that  $d$  is a cardinal function.

## A semantic intuition 1(2)

SDL:  $p$  is obligatory if and only if  $p$  is true in all deontically ideal worlds.

- ▶ Our proposal:
- ▶ Let  $W$  be the set of all accessible possible worlds and let  $d(w)$  be the deontic value of world  $w \in W$ ; we assume that  $d$  is a cardinal function.
- ▶ Let  $N(w)$  be the value of world  $w$  normalized to a scale between 0 and 1 relative to the set of accessible worlds  $W$ , calculated as:

$$N(w) = \begin{cases} 1 & \text{if } d(w_{min}) = d(w_{max}) \\ \frac{d(w) - d(w_{min})}{d(w_{max}) - d(w_{min})} & \text{otherwise} \end{cases}$$

## A semantic intuition 2(2)

- ▶ We propose that the degree to which  $p$  is obligatory is proportional to the value of the best accessible possible world(s) in which it holds that  $\neg p$ .

# Theorem 1

Let

$$O(p) = \begin{cases} 1 & \text{if } W - p = \emptyset \\ 1 - \max(\{N(w) : w \in W - p\}) & \text{otherwise} \end{cases}$$

Then it holds that:

1.  $O(p) \geq 0$
2. If  $O(p) > 0$  then  $O(\neg p) = 0$
3. If  $W \vdash p$  then  $O(p) = 1$
4.  $O(p \wedge q) = \min(\{O(p), O(q)\})$
5.  $O(p \vee q) \geq O(p)$

## Axiomatization: Axioms (SDL)

- (SDL 1) The tautologies of classical propositional logic.
- (SDL 2) From  $\varphi$  and  $\varphi \rightarrow \psi$  infer  $\psi$
- (SDL 3)  $O(\varphi \rightarrow \psi) \rightarrow (O(\varphi) \rightarrow O(\psi))$
- (SDL 4)  $\neg(O(\perp))$
- (SDL 5) From  $\varphi$  infer  $O(\varphi)$

## Soundness and completeness (SDL)

SDL 1–5 is sound and complete for the class of Kripke models known as KD. A Kripke model is a triple  $\langle W, R, V \rangle$  such that:

- ▶  $W$  is a non-empty set of possible worlds,
- ▶  $R \subseteq W \times W$  and
- ▶  $V : P \rightarrow 2^W$ .

A Kripke model is in the class KD if and only if the relation  $R$  is serial, i.e. if and only if  $R$  satisfies the following constraint:

(CD) For all  $w \in W$  there is a  $v \in W$  such that  
 $(w, v) \in R$ .

## Formal semantics (SDL)

Given a model  $M = \langle W, R, V \rangle$  and a world  $w \in W$ , the semantics for SDL looks as follows.

$$M, w \models \top$$

$$M, w \models p \quad \text{iff} \quad w \in V(p)$$

$$M, w \models \neg\varphi \quad \text{iff} \quad M, w \not\models \varphi$$

$$M, w \models \varphi \wedge \psi \quad \text{iff} \quad M, w \models \varphi \text{ and } M, w \models \psi$$

$$M, w \models O(\varphi) \quad \text{iff} \quad M, v \models \varphi \text{ for every } v \in W \text{ such that } (w, v) \in R$$

## Language (SDLD)

The language  $\mathcal{L}_{SDLD}$  of *SDLD* is the set of formulae  $\varphi$ , defined by the following BNF:

$$\begin{aligned} \varphi &::= \chi \mid O(\chi) > x \mid \neg\varphi \mid \varphi \wedge \varphi \\ \chi &::= \top \mid p \mid \neg\chi \mid \chi \wedge \chi \end{aligned}$$

, where  $x \in [0, 1[$ .

The intended interpretation of  $O(\chi) > x$  is ‘ $\chi$  is obligatory to a degree that is greater than  $x$ ’.



## Models (SDLD)

SDLD models are triples  $\langle W, R, V \rangle$  such that  $W$  and  $V$  are as for SDL models and  $R$  is a function

$$\{(1-x) : x \in [0, 1[ \} \rightarrow (W \times W)\}$$

$R_{1-x}(w)$  denotes the set  $\{v : (w, v) \in R_{1-x}\}$ .

We postulate that  $R$  must satisfy the following two constraints.

(CD)  $R_{1-0}(w) \neq \emptyset$ .

(CI) If  $(1-x) \geq (1-y)$  then  $R_{1-x}(w) \subseteq R_{1-y}(w)$ .

## Formal semantics (SDLD)

Given a SDLD model  $M = \langle W, R, V \rangle$  and a world  $w \in W$  we have:

$$M, w \models \top$$

$$M, w \models p \quad \text{iff} \quad w \in V(p) \text{ and } p \in P$$

$$M, w \models \neg\varphi \quad \text{iff} \quad M, w \not\models \varphi$$

$$M, w \models \varphi \wedge \psi \quad \text{iff} \quad M, w \models \varphi \text{ and } M, w \models \psi$$

$$M, w \models O(\varphi) > x \quad \text{iff} \quad \text{for every } v \in W \text{ if } w \in R_{1-x}(w) \text{ then } M, v \models \varphi$$

## Axiomatization: Axioms (SDLD)

- (SDLD 1) The tautologies of classical propositional logic.
- (SDLD 2) From  $\varphi$  and  $\varphi \rightarrow \psi$  infer  $\psi$
- (SDLD 3)  $(O(\varphi \rightarrow \psi) > x) \rightarrow ((O(\varphi) > x) \rightarrow (O(\psi) > x))$
- (SDLD 4)  $\neg(O(\perp) > 0)$
- (SDLD 5) From  $\varphi$  infer  $O(\varphi) > x$
- (SDLD 6)  $(O(\varphi) > x) \rightarrow (O(\varphi) > y)$  for all  $x \geq y$

## Theorem 2

$\vdash \varphi$  if and only if  $\models \varphi$ ,  
where  $\vdash \varphi$  means that  $\varphi$  is a theorem of SDDL 1–6.

## Theorem 3 + Observation

What is the relation to the semantic intuition we started with?

Theorem 3:

$M, w \models O(\varphi) > x$

iff

for every  $v \in W$  if  $N(v) \geq 1 - x$  then  $M, v \models \varphi$

Observation The formulae in Theorem 1 are valid in SDDL.

## Moral dilemmas 1(3)

In SDL  $O(p)$  entails  $\neg O(\neg p)$ .

- ▶ What about SDLD? According to Theorem 1.2 it holds that if  $O(p) > 0$  then  $O(\neg p) = 0$

## Moral dilemmas 2(3)

Instead of a single value-based ranking  $N(v)$  we introduce one for each incommensurable value / dimension.

$$N_i(v) = \begin{cases} 1 & \text{if } d_i(v_{min}) = d_i(v_{max}) \\ \frac{d_i(v) - d_i(v_{min})}{d_i(v_{max}) - d_i(v_{min})} & \text{otherwise} \end{cases}$$

$$O(p) = \begin{cases} 1 & \text{if } W - p = \emptyset \\ \max(\{1 - \max(\{N_i(v) : v \in W - p\}) : i \in I\}) & \text{otherwise} \end{cases}$$

## Theorem 4

The condition for  $O(p)$  stated above entails that:

1.  $O(p) \geq 0$
2. If  $W \vdash p$  then  $O(p) = 1$
3.  $O(p) \geq O(p \wedge q)$



## Theorem 5

Definition:  $M, w \models \widehat{O}(\varphi) > x$  iff  
 $\exists i \in I, \forall v \in W$  if  $N_i(v) \geq 1 - x$  then  $M, v \models \varphi$

Theorem 5:

1.  $\models \neg(\widehat{O}(\perp) > 0)$
2.  $\models O_i(\varphi) > x \rightarrow \widehat{O}(\varphi) > x$
3. *From*  $\models \varphi \rightarrow \psi$  *infer*  $\models \widehat{O}(\varphi) > x \rightarrow \widehat{O}(\psi) > x$

### Derivable formulae:

1.  $\models \widehat{O}(T) > x$
2.  $\models \widehat{O}(\varphi \wedge \psi) > x \rightarrow \widehat{O}(\varphi) > x$
3. From  $\models \varphi$  infer  $\widehat{O}(\varphi) > x$

### Non-valid formulae:

1.  $\not\models ((\widehat{O}(\varphi) > x) \wedge (\widehat{O}(\psi) > x)) \rightarrow (\widehat{O}(\varphi \wedge \psi) > x)$
2.  $\not\models ((O_i(\varphi) > x) \wedge (O_j(\psi) > x)) \rightarrow (\widehat{O}(\varphi \wedge \psi) > x)$

## Alternative approaches 1(3): A probabilistic account?

- (P 1)  $O(p) \geq 0$
- (P 2) If  $\vdash p$  then  $O(p) = 1$
- (P 3) Any countable sequence of pairwise incompatible propositions  $p_1, p_2, \dots$  satisfies  $O(p_1 \vee p_2 \vee \dots) = \sum_i O(p_i)$

Observation: SDL 3–5 follow from P 1–3

## Against deontic probabilities

▶  $O(p) + O(\neg p) = 1$

## Against deontic probabilities

- ▶  $O(p) + O(\neg p) = 1$
- ▶  $O(p|q) = \frac{O(p \wedge q)}{O(q)}$

## Against deontic probabilities

- ▶  $O(p) + O(\neg p) = 1$
- ▶  $O(p|q) = \frac{O(p \wedge q)}{O(q)}$
- ▶

$$O(p|q) = \begin{cases} 1 & \text{if } W \cap q - p = \emptyset \\ 1 - \max(\{N_{W \cap q}(w) : w \in W \cap q - p\}) & \text{otherwise} \end{cases}$$

## Some other alternatives

- ▶ Fuzzy logic? No!

## Some other alternatives

- ▶ Fuzzy logic? No!
- ▶ Mcnamara (1996), Torre (1997), Hansson (2001, 2004),