Contrary-to-duty Obligations

Michael Repplinger

January 15, 2010

Standard deontic logic (SDL, also: KD)

- Close to von Wright's original system, but Kripke style semantics introduce some changes
- (Prakken, Sergot 1996) define their version of SDL semantically, but it will be easier to follow their derivations with the list of axioms of KD provided by (Meyer, Wieringa 1993)

KD axioms

0. Tautologies of propositional calculus
1. $O(p \to q) \to (Op \to Oq)$
2. $Op \to Pp$
3. $Pp \equiv \neg O \neg p$
4. $Fp \equiv \neg Pp$
5. $p, p \to q \Rightarrow q$
6. $p \Rightarrow Op$

KD

- NB: KD6 does not say that whenever p holds, Op holds, but rather that if p is a theorem (i.e. holds unconditionally), then Op is a consequence
- KD6 is the "missing link" of the original system; so far, treatment of facts and treatment of obligations were independent (von Wright 1951, p. 15)
- But KD6 also introduces OT: $O(p \vee \neg p)$ (via law of excluded middle)

Two theorems of KD without which (Prakken, Sergot 1996) are difficult to follow

- $O(p \wedge q) \equiv Op \wedge Oq$
- $\neg(Op \wedge O\neg p)$

Intuitively obvious perhaps, but we need to be sure the system enables them

Preliminary remarks

- A formalization of CTD obligations should model: possible violation of obligations, and preservation (or introduction) of appropriate obligations at every point
- $Og \wedge \neg g$ should be consistent, otherwise the system is uninteresting
- CTD obligations cannot naively be assumed to behave like (primary) obligations (example follows)
- Our decisions which inferences between obligations (and facts) hold influences the behavior of the system

It ought to be that Jones goes to assist his neighbors. It ought to be that if Jones goes, then he tells them he is coming. If Jones doesn't go, then he ought not tell them he is coming. Jones doesn't go. A first attempt:

- Og
- $O(g \rightarrow t)$
- $O(\neg g \rightarrow \neg t)$
- $\neg g$

But then *contrary-to-duty* is meaningless, since all obligations are of the same logical form.

Second attempt. We need to distinguish two types of inference:

- Factual detachment: $p \wedge (p \rightarrow Oq) \Rightarrow Oq$
- Deontic detachment: $Op \wedge O(p \rightarrow q) \Rightarrow Oq$

Both types are at least plausible (material implication; intuitively right consequence). But note that allowing both kinds, together with the following formulas, leads to inconsistency.

- Og
- $O(g \rightarrow t)$
- $\neg g \rightarrow (O \neg t)$
- $\neg g$

(Prakken, Sergot 1996. p. 94)

- Keep your promise
- If you haven't kept your promise, apologize
- You haven't kept your promise

- Ok
- $\neg k \Rightarrow Oa$
- $\neg k$

Leads to what the authors call a *pragmatic oddity*: the most ideal situation attainable, given the facts, is one where you ought to keep your promise and ought to apologize (for not keeping your promise). Not yet inconsistent, however.

(Prakken, Sergot 1996. p. 95)

- Woody and Mia should not meet
- If they meet, they should embrace
- They meet
- They can only embrace if they meet (duh!)

- $O\neg m$
- $m \Rightarrow Oe$
- m
- $\neg\Diamond(e \wedge \neg m)$

Leads to *inconsistency*, which intuitively can be seen if you consider that conflicting obligations arise, which, together with the assumption that an obligation entails a possibility, lead to a contradiction.

(Prakken, Sergot 1996. p. 95) (continued)
Formally: We can immediately derive $O\neg m$, Oe. So $O(\neg m \wedge e)$, by Thm. 2 of KD. So $\Diamond(\neg m \wedge e)$, from the axiom of Prakken & Sergot's modally enriched theory. However: $\neg\Diamond(e \wedge \neg m)$.

- Prakken & Sergot go on to discuss the viability of a temporal solution; in the shortest possible version: up to some time point t some obligation p holds, which is violated at t, after which p no longer holds but instead q. They dismiss the possibility on the basis of two arguments: counterexamples to temporal cases (doubtful), and lack of CTD-ness (more plausible).
- They also discuss the relation of their formalization of CTD obligations to *defeasible reasoning*, i.e. reasoning with exceptions (which is typically modeled by some non-monotonic logic). They point out that this approach fails to distinguish between *exceptions*, i.e. exceptional situations in which the default line of reasoning does not apply, and *violation of obligations*, i.e. the failure to comply with some (moral) duty.

Some notation:

- A, B, C: formulas
- w, v: worlds
- P, Q, R: propositions = sets of worlds (permitted by them)
- $[\![A]\!]$: set of worlds in which A is true
- O (obligatory), P (permitted)
- function $d(w)$: deontic alternatives to w

Truth conditions & definitions:

- $\models_w OA$ iff $d(w) \subseteq [\![A]\!]$
- Serial accessibility relation ($d(w) \neq \emptyset$)

Truth conditions & definitions (continued):

- Adding $f(w)$, $\Box$ and $\Diamond$. $\models_w \Box A$ iff $f(w) \subseteq [\![A]\!]$. It now follows (together with seriality) that $OA \rightarrow \Diamond A$.
- Assume in addition factual detachment $(p \wedge (p \rightarrow Oq) \Rightarrow Oq)$.
- Question: Deontic detachment $(Op \wedge O(p \rightarrow q) \Rightarrow Oq)$ is not mentioned, but it might follow (from KD axioms, from P&S's semantics, or both)

Multi-level CTD modalities.

- $dc([\![B]\!], w)$: alternatives to w given sub-ideal context B
- $\models_w O_B A$ iff $dc([\![B]\!], w) \subseteq [\![A]\!]$
- Simple case OA redefined: $OA =_{df} O_\top A$

Secondary obligations are always introduced as a conditional; $B \Rightarrow O_B A$. According to P\$S, this allows separation of the obligation from its introduction. NB: $\neg B \Rightarrow O_B A$ is meaningless (but permitted).

Restrictions on *dc*

1. dc is a subset of f
2. if the (sub-ideal) context is not empty, the alternatives are not empty
3. a lower bound: introducing a more specific context $Q \cap R$, ideal worlds (alternatives) are only removed unless the new context R implies a violation of the original context Q.
4. an upper bound: obligations are only removed (i.e. alternative worlds can only be added) in a more specific context, unless some obligation becomes unsatisfiable in the new context and this unsatisfiability does not derive from the previous context.

Two important conditions follow. The first one:

- Up: $PB \rightarrow (O_B A \rightarrow OA)$
- $\Leftrightarrow$ Ctd: $\neg OA \rightarrow (O_B A \rightarrow O\neg B)$

In words: secondary obligations are also primary ones, if the context of the secondary one is permitted (i.e. not sub-optimal wrt to the primary one)

The second condition:

- Down: $(\Diamond(A \wedge B) \wedge \neg\Box(\neg A \rightarrow B)) \rightarrow (OA \rightarrow O_B A)$

A primary obligation is also a secondary one if the context B still allows compliance with the original context, and if the violation of the old context does not automatically lead to the new context. In other words: drop an obligation only if it is violated.

- Arbitrary levels of CTD rules are defined in a straightforward way, all results are preserved.

1. The children should not cycle on the street.
2. If they cycle, they should cycle on the left.
3. If they are not cycling on the left, at least they should cycle on the extreme right.
4. They cannot cycle on the left or extreme right unless they are cycling, and cycling on the extreme right means they don't cycle on the left.
5. The children cycle.

1. $O\neg c$
2. $c \Rightarrow O_c l$
3. $(c \wedge \neg l) \Rightarrow O_{(c \wedge \neg l)} e$
4. $\Box(l \rightarrow c) \wedge \Box(e \rightarrow c) \wedge \Box(e \rightarrow \neg l)$
5. $c \wedge \neg l$

- (2) and (5) yield $O_c l$. With (4) we get $O_c c$.
- *Down* is blocked, so (1), $O\neg c$, is not preserved.
- (3) and (5) yield $O_{(c \wedge \neg l)} e$, with (4) we get $O_{(c \wedge \neg l)} \neg l$
- Again, $O_c l$ is not transported downwards.

The result: the set of sentences is not inconsistent, and we preserve the (only) applicable obligation in the end, i.e. $O_{(c \wedge \neg l)} e$.

- Note that we can derive $O_c\neg(c \wedge \neg l)$, but not $O_{(c\wedge\neg l)}\neg c$. So in context c, the context $c \wedge \neg l$ is forbidden, but not vice versa, so $(c \wedge \neg l)$ is a CTD context of c.

Elements of the paper that I did not cover

- A flaw in the definition of *relatedness* between obligations. In short: given multiple primary obligations, irrelevant obligations can be transported downward. Author suggest ad-hoc solution, but state they would prefer a more motivated solution via indices stating the *source* of obligations.
- The distinction and discussion of *ought-to-be* and *ought-to-do*
- The relation to defeasible reasoning and non-monotonic logic (a somewhat shaky argument anyway, I believe)