Deontic Logic in Computer Science **ESSLI 2008**
Leon van der Torre & Jörg Hansen 5 August 2008

Solution to homework from lesson 1 (4 August 2008)

Hansson's dyadic deontic system DSDL3 can be defined axiomatically via the PL-instances into the following axiom schemes and rules (plus *modus ponens*):

(DM) $O(A\wedge B/C) \rightarrow (O(A/C) \wedge O(B/C))$
(DC) $(O(A/C) \wedge O(B/C)) \rightarrow O(A\wedge B/C)$
(DN) $O(\top/C)$
(DD-R) If $\nvdash_{PL} \neg C$ then $\vdash_{DSDL3} O(A/C) \rightarrow P(A/C) \in$ DSDL3
(Cond) $O(A\,/C\wedge D) \rightarrow O(D\rightarrow A\,/C)$
(CCMon) $O(A \wedge D\,/C) \rightarrow O(A\,/C\wedge D)$
(RMon) $P(D/C) \rightarrow (O(A\,/C) \rightarrow O(A\,/C\wedge D))$
(CExt) If $\vdash_{PL}$ $C\rightarrow (A \leftrightarrow B)$ is a tautology then $\vdash_{DSDL3}$ $O(A/C) \leftrightarrow O(B/C)$
(ExtC) If $\vdash_{PL}$ $C \leftrightarrow D$ is a tautology then $\vdash_{DSDL3}$ $O(A/C) \leftrightarrow O(A/D)$

The Homework:

Here is an alternative, norm- or imperative-based semantics for dyadic deontic logic:
Let $I=\{A_1, ..., A_n\}$ be a subset of PL-sentences; they are meant to correspond to the termination statements of a set of (unconditional) imperatives $\{!A_1, ..., !A_n\}$, or the 'contents' of a set of unconditonal *O*-norms. *I* is not assumed to be consistent.

Let $I\perp A$ (the *A*-remainders of *I*) be the set of maximal subsets Γ of *I* such that Γ does not derive A, i.e. $I\perp A$ contains all $\Gamma\subseteq I$ such that i) $\Gamma \nvdash_{PL} A$, and ii) there is no $\Delta\subseteq I$ such that $\Gamma\subseteq \Delta$ and $\Gamma \nvdash_{PL} A$.

Consider the following truth definition for $O(A/C)$ (truth definitions for Boolean connectives being as usual):

(a) $I \vDash O(A/C)$ iff for all $\Gamma \in I\perp\neg C$: $\Gamma\cup \{C\} \vdash_{PL} A$

i.e. *A* is obligatory given *C* iff *A* derives – potentially with *C* – from *all* maximal sets of contents of norms that still can be collectively satisfied in the circumstances described by *C*. Thus we 'make the best out of the sad circumstances'.

Then the axiomatic system that is defined just like DSDL3, except without axiom **(RMon)** , is sound and (weakly) complete with respect to this truth definition.

Proof:

... that (RMon) is not valid in the new semantics: for a counterexample consider e.g. the set $I = \{\neg p \wedge s, p \wedge (q \rightarrow s), p \wedge q \wedge \neg r\}$.

$I\perp\neg\top = \{\{\neg p \wedge s\}, \{p \wedge (q \rightarrow s), p \wedge q \wedge \neg r\}\}$, so both $O(s/\top)$ and $P(r/\top)$ are true. However we have

$I\perp\neg r = \{\{\neg p \wedge s\}, \{p \wedge (q \rightarrow s)\}\}$, and since *s* is not a consequence of the right set, we do not have $O(s/r)$.

... for soundness and completeness of the rest: cf. J. Hansen, "Conflicting imperatives and dyadic deontic logic", JAL 3, 2005, 484-511.

New question:

Let's change the above truth definition into

(b) $I \vDash O(A/C)$ iff there exists some $\Gamma \in I\perp\neg C$: $\Gamma\cup \{C\} \vdash_{PL} A$

i.e. something is obligatory if it derives – potentially with *C* – from *some* maximal sets of contents of norms that still can be collectively satisfied in the circumstances described by *C.* Motivation: it is enough if the agent behaves according to one of potentially several, conflicting standards of conduct.

Is then (RMon) valid or invalid? What happens to the other axioms?