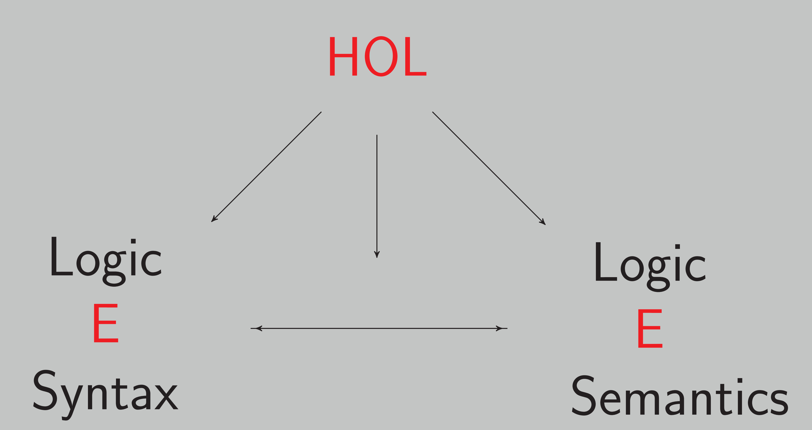
# A Faithful Semantic Embedding of the Dyadic Deontic Logic E in HOL

Christoph Benzmüller, Ali Farjami and Xavier Parent
University of Luxembourg
c.benzmueller@gmail.com, ali.farjami@uni.lu, xavier.parent@uni.lu

## Shallow Semantical Embedding

A semantic embedding of a target logical system defines the syntactic elements of the target language in a background logic (HOL) [2].




**Comprehension axiom**:

$\neg\varphi = \{x|\neg_{o\to o}(\varphi x)\} = \lambda x.\neg_{o\to o}(\varphi x)$

$M, s \vDash \neg\varphi$ if and only if $M, s \nvDash \varphi$ (that is, not $M, s \vDash \varphi$)

## System E: Syntax

Åqvist defined dyadic deontic logic system **E** [1] by the following axioms and rules: ($\Box$ (S5-schemata for necessity) and $\bigcirc(-/-)$ (for conditional obligation) )

| | |
|---|---|
| $\bigcirc(\psi_1 \to \psi_2/\varphi) \to (\bigcirc(\psi_1/\varphi) \to \bigcirc(\psi_2/\varphi))$ | COK |
| $\bigcirc(\psi/\varphi) \to \Box\bigcirc(\psi/\varphi)$ | Abs |
| $\Box\psi \to \bigcirc(\varphi/\psi)$ | Nec |
| $\Box(\varphi_1 \leftrightarrow \varphi_2) \to (\bigcirc(\psi/\varphi_1) \leftrightarrow \bigcirc(\psi/\varphi_2))$ | Ext |
| $\bigcirc(\varphi/\varphi)$ | Id |
| $\bigcirc(\psi/\varphi_1 \wedge \varphi_2) \to \bigcirc(\varphi_2 \to \psi/\varphi_1)$ | Sh |

## System E: Semantics

- A preference model is a structure $M = \langle S, \succeq, V\rangle$ where
  - $S$ is a non-empty set of items called possible worlds;
  - $\succeq \subseteq S \times S$ (intuitively, $\succeq$ is a betterness or comparative goodness relation);
  - $V$ is a function assigning to each atomic sentence a set of worlds (i.e $V(q) \subseteq S$).
- (Satisfaction) Given $opt_{\succeq}(V(\varphi)) = \{s \in V(\varphi) | \forall t(t \vDash \varphi \to s \succeq t)\}$

  $M, s \vDash \bigcirc(\psi/\varphi)$ if and only if $opt_{\succeq}(V(\varphi)) \subseteq V(\psi)$
- (Soundness and Completeness) System **E** is (strongly) sound and complete with respect to the class of all preference models [1].

## Contrary-To-Duties

- Chisholm's CTD-paradox [4]
  (a) It ought to be that a certain man go to help his neighbours.
  (b) It ought to be that if he goes he tell them he is coming.
  (c) If he does not go, he ought not to tell them he is coming.
  (d) He does not go.

best $s_1 \bullet g \quad t$

– – – – – – – – – – – – – – – – –

2nd best $s_1 \bullet g \quad s_3\bullet$

– – – – – – – – – – – – – – – – –

worst $s_4 \bullet t$

- For example actual world $s_3$ satisfies : $\bigcirc(g)$ $\bigcirc(t/g)$ $\bigcirc(\neg t/\neg g)$ $\neg g$

## Formulas E as Certain HOL Terms

We assume a set of basic types $BT = \{o, i\}$. The mapping $\lfloor\cdot\rfloor$ translates **E** formulas $s$ into HOL terms $\lfloor s\rfloor$ of type $i \to o$. Type $i \to o$ is abbreviated as $\tau$ in the remainder.

$$\begin{aligned}
\lfloor p^j\rfloor &= p^j_\tau \\
\lfloor\neg s\rfloor &= \neg_\tau \lfloor s\rfloor \\
\lfloor s \vee t\rfloor &= \vee_{\tau\to\tau\to\tau} \lfloor s\rfloor\lfloor t\rfloor \\
\lfloor\Box s\rfloor &= \Box_{\tau\to\tau} \lfloor s\rfloor \\
\lfloor\bigcirc(t/s)\rfloor &= \bigcirc_{\tau\to\tau\to\tau} \lfloor s\rfloor\lfloor t\rfloor
\end{aligned}$$

$\neg_{\tau\to\tau}$, $\vee_{\tau\to\tau\to\tau}$, $\Box_{\tau\to\tau}$ and $\bigcirc_{\tau\to\tau\to\tau}$ thereby abbreviate the following HOL terms:

$$\begin{aligned}
\neg_{\tau\to\tau} &= \lambda A_\tau \lambda X_i \neg(A\,X) \\
\vee_{\tau\to\tau\to\tau} &= \lambda A_\tau \lambda B_\tau \lambda X_i (A\,X \vee B\,X) \\
\Box_{\tau\to\tau} &= \lambda A_\tau \lambda X_i \forall Y_i (A\,Y) \\
\bigcirc_{\tau\to\tau\to\tau} &= \lambda A_\tau \lambda B_\tau \lambda X_i \forall W_i ( (\lambda V_i (A\,V \wedge (\forall Y_i (A\,Y \to r_{i\to\tau} V\,Y)))) \, W \to B\,W)
\end{aligned}$$

## Corresponding Henkin Model $H^M$ for Preference Model $M$

Given a preference model $M = \langle S, \succeq, V\rangle$. Let $p^1, ..., p^m \in PV$, for $m \geq 1$ be propositional symbols and $\lfloor p^j\rfloor = p^j_\tau$ for $j = 1, ..., m$. The Henkin model $H^M = \langle\{D_\alpha\}_{\alpha\in T}, I\rangle$ for $M$ is defined as follows:

- $D_i$ is chosen as the set of possible worlds $S$
- $D_{\alpha\to\beta}$ as (not necessarily full) sets of functions from $D_\alpha$ to $D_\beta$.
- For $1 \leq i \leq m$, we choose $Ip^j_\tau \in D_\tau$ such that $Ip^j_\tau(s) = T$ if $s \in V(p^j)$ in $M$ and $Ip^j_\tau(s) = F$ otherwise.
- We choose $Ir_{i\to\tau} \in D_{i\to\tau}$ such that $Ir_{i\to\tau}(u, s) = T$ if $s \succeq u$ in $M$ and $Ir_{i\to\tau}(u, s) = F$ otherwise.

## Corresponding Preference Model $M_H$ for Henkin Model $H$

For every Henkin model $H = \langle\{D_\alpha\}_{\alpha\in T}, I\rangle$ there exists a corresponding preference model $M_H$. Corresponding means that for all **E** formulas $\delta$ and for all assignment $g$ and worlds $s$, $\|\lfloor\delta\rfloor S_i\|^{H,g[s/S_i]} = T$ if and only if $M_H, s \vDash \delta$. We construct the corresponding preference model $M_H$ as follows:

- $S = D_i$.
- $s \succeq u$ for $s, u \in S$ iff $Ir_{i\to\tau}(u, s) = T$.
- $s \in V(p^j_\tau)$ iff $Ip^j_\tau(s) = T$.

## Result: Soundness and Completeness of the Embedding

Given $\text{vld}_{\tau\to o} = \lambda A_\tau \forall S_i (A\,S)$ we have: $\vDash^{E} \varphi$ if and only if $\vDash^{HOL} \text{vld}\lfloor\varphi\rfloor$

## Isabelle/HOL: Propositional Connectives

Isabelle2016-1 - DDLE.thy
File Edit Search Markers Folding View Utilities Macros Plugins Help
DDLE.thy (%USERPROFILE%\Dropbox\thy\Poster\)

```
theory DDLE  imports Main
begin
typedecl i -- "type for possible worlds"
type_synonym σ = "(i⇒bool)"

abbreviation(input) mtrue  :: "σ" ("⊤") where "⊤ ≡ λw. True"
abbreviation(input) mfalse :: "σ" ("⊥")  where "⊥ ≡ λw. False"
abbreviation(input) mnot   :: "σ⇒σ" ("¬_"[52]53)  where "¬φ ≡ λw. ¬φ(w)"
abbreviation(input) mand   :: "σ⇒σ⇒σ" (infixr"∧"51) where "φ∧ψ ≡ λw. φ(w)∧ψ(w)"
abbreviation(input) mor    :: "σ⇒σ⇒σ" (infixr"∨"50) where "φ∨ψ ≡ λw. φ(w)∨ψ(w)"
abbreviation(input) mimp   :: "σ⇒σ⇒σ" (infixr"→"49) where "φ→ψ ≡ λw. φ(w)⟶ψ(w)"
abbreviation(input) mequ   :: "σ⇒σ⇒σ" (infixr"↔"48) where "φ↔ψ ≡ λw. φ(w)⟷ψ(w)"
```

## Isabelle/HOL: Modal Operators

```
abbreviation(input) mbox :: "σ⇒σ" ("□") where "□ ≡ λφ w.  ∀v. φ(v)"
consts r :: "i⇒i⇒bool"   (infixr "r" 70)
  -- "the betterness relation r, used in definition of O"
abbreviation(input) mopt  :: "(i⇒bool)⇒(i⇒bool)"   ("opt<_>")
  where "opt<φ> ≡ (λv. ( (φ)(v) ∧ (∀x. ((φ)(x)   ⟶   v r x) )) )"
abbreviation(input) msubset :: "σ⇒σ⇒bool" (infix "⊆" 53)
  where "φ ⊆ ψ ≡ ∀x. φ x ⟶ ψ x"
abbreviation(input) mcond  :: "σ⇒σ⇒σ" ("○<_|_>")
  where "○<ψ|φ> ≡  λw. opt<φ> ⊆ ψ"
```

## Isabelle/HOL: Validity

```
abbreviation(input) valid :: "σ⇒bool" ("⌊_⌋"[8]109)
  where "⌊p⌋ ≡ ∀w. p w"
```

## Isabelle/HOL: Chisholm Scenario

```
section {* Chisholm Scenario *}
consts g ::  "σ"  t ::  "σ"
context (* Chisholm  Scenario*)
  assumes
  ax1: "⌊ ○<g|⊤> ⌋" and
  ax2: "⌊ ○<t|g> ⌋" and
  ax3: "⌊ ○<¬t|¬g> ⌋" and
  ax4 : "⌊¬g ⌋"
   begin
  lemma True nitpick [satisfy, user_axioms, show_all, expect=genuine]  oops
end
```

## Conclusion

- We have described a faithful semantic embedding of the dyadic deontic logic system **E** in simple type theory.
- This work complements the one reported in [3] where the focus is on neighborhood semantics for dyadic deontic logic.
- Our work provides the theoretical foundation for the implementation and automation of dyadic deontic logics within theorem provers and proof assistants.

## References

[1] Xavier Parent.
Completeness of Åqvist's systems E and F.
*The Review of Symbolic Logic*, 8(1):164–177, 2015.

[2] Christoph Benzmüller.
Recent successes with a meta-logical approach to universal logical reasoning (extended abstract).
volume 10623 of *Lecture Notes in Computer Science*, pages 7–11. Springer, 2017.

[3] C. Benzmüller, A. Farjami, and X. Parent.
Faithful semantical embedding of a dyadic deontic logic in HOL.
Accepted for presentation at DEON 2018. To appear in IfColog.

[4] Roderick M. Chisholm.
Contrary-to-duty imperatives and deontic logic.
*Analysis*, 24:3336, 1963.

## Acknowledgment

This work has been supported by the European Unions Horizon 2020 research and innovation programme under the Marie Skodowska-Curie grant agreement No 690974.