Faithful Logic Embeddings in HOL — Deep and Shallow (Isabelle/HOL dataset)

Christoph Benzmüller

June 4, 2025

Abstract

A recipe for the simultaneous deployment of different forms of deep and shallow embeddings of non-classical logics in classical higher-order logic is presented, which enables interactive or even automated faithfulness proofs between the logic embeddings. The approach, which is particularly fruitful for logic education, is explained in detail in an associated CADE conference paper. This paper presents the corresponding Isabelle/HOL dataset (which is only slightly modified to meet AFP requirements).

Contents

1 Introduction

The Isabelle/HOL dataset associated with [1] is presented. Sections 3, 4 and 5 present deep, maximally shallow, and minimally shallow embeddings of propositional modal logic (PML) in classical higher-order logic (HOL). These are connected, as a novel contribution, by automated faithfulness proofs given in Sect. 6. This connection ensures that these deep and shallow embeddings can now be used interchangeably in subsequent applications. Several experiments with the presented embeddings are presented in Sect. 7. The presented work is conceptual in nature and can be adapted to other non-classical logics. For more detailed explanations of the presented material, including a discussion of related works, see [1].

2 Preliminaries

The following preliminaries are are shared between all embeddings introduced in the remainder of this paper.

theory *PMLinHOL-preliminaries*
 imports *Main*
begin

— Type declarations common for both the deep and shallow embedding
typedecl w — Type for possible worlds
typedecl $\mathcal{S}$ — Type for propositional constant symbols
consts p::$\mathcal{S}$ q::$\mathcal{S}$ r::$\mathcal{S}$ — Some propositional constant symbols
type-synonym $\mathcal{W}$ = w⇒*bool* — Type for sets of possible worlds
type-synonym $\mathcal{R}$ = w⇒w⇒*bool* — Type for accessibility relations
type-synonym $\mathcal{V}$ = $\mathcal{S}$⇒w⇒*bool* — Type for valuation functions

— Some useful predicates for accessibility relations
abbreviation(*input*) *reflexive* ≡ λR::$\mathcal{R}$. ∀ x. $R\ x\ x$
abbreviation(*input*) *symmetric* ≡ λR::$\mathcal{R}$. ∀ $x\ y$. $R\ x\ y \longrightarrow R\ y\ x$
abbreviation(*input*) *transitive* ≡ λR::$\mathcal{R}$. ∀ $x\ y\ z$. $(R\ x\ y \wedge R\ y\ z) \longrightarrow R\ x\ z$
abbreviation(*input*) *equivrel* ≡ λR::$\mathcal{R}$. *reflexive* R ∧ *symmetric* R ∧ *transitive* R
abbreviation(*input*) *irreflexive* ≡ λR::$\mathcal{R}$. ∀ x. ¬$R\ x\ x$
abbreviation(*input*) *euclidean* ≡ λR::$\mathcal{R}$. ∀ $x\ y\ z$. $R\ x\ y \wedge R\ x\ z \longrightarrow R\ y\ z$
abbreviation(*input*) *wellfounded* ≡ λR::$\mathcal{R}$. ∀ P::$\mathcal{W}$. (∀ x. (∀ y. $R\ y\ x \longrightarrow P\ y) \longrightarrow P\ x) \longrightarrow$ (∀ x. $P\ x$)
abbreviation(*input*) *converserel* ≡ λR::$\mathcal{R}$. λy::w. λx::w. $R\ x\ y$
abbreviation(*input*) *conversewf* ≡ λR::$\mathcal{R}$. *wellfounded* (*converserel* R)

— Bounded universal quantifier: ∀ x:W. φ stands for ∀ x. $W\ x \longrightarrow \varphi\ x$
abbreviation(*input*) *BoundedAll*::$\mathcal{W}$⇒$\mathcal{W}$⇒*bool* **where** *BoundedAll* W φ ≡ ∀ x. $W\ x \longrightarrow \varphi\ x$
syntax *-BoundedAll*:: *pttrn*⇒$\mathcal{W}$⇒*bool*⇒*bool* ((*3*∀ (-/:-)./ -) [*0*, *0*, *10*] *10*)
translations ∀ x:W. φ ⇌ *CONST BoundedAll* W (λx. φ)

— Backward implication; useful for aesthetic reasons
abbreviation(*input*) *Bimp* (**infixr** ⟵ *50*) **where** $\varphi \longleftarrow \psi \equiv \psi \longrightarrow \varphi$

— Some further settings

declare[[*syntax-ambiguity-warning=false*]]
nitpick-params[*user-axioms*,*expect=genuine*]
end

3 Deep embedding of PML in HOL

theory *PMLinHOL-deep*
imports *PMLinHOL-preliminaries*
begin

— Deep embedding (of propositional modal logic in HOL)
datatype $PML = AtmD\ \mathcal{S}\ (\text{-}^d) \mid NotD\ PML\ (\neg^d) \mid ImpD\ PML\ PML$ (**infixr** $\supset^d$ 93) $\mid BoxD\ PML\ (\Box^d)$

— Further logical connectives as definitions
definition *OrD* (**infixr** $\vee^d$ *92*) **where** $\varphi \vee^d \psi \equiv \neg^d\varphi \supset^d \psi$
definition *AndD* (**infixr** $\wedge^d$ *95*) **where** $\varphi \wedge^d \psi \equiv \neg^d(\varphi \supset^d \neg^d\psi)$
definition *DiaD* ($\Diamond^d$-) **where** $\Diamond^d\varphi \equiv \neg^d(\Box^d(\neg^d\varphi))$
definition *TopD* ($\top^d$) **where** $\top^d \equiv p^d \supset^d p^d$
definition *BotD* ($\bot^d$) **where** $\bot^d \equiv \neg^d \top^d$

— Definition of truth of a formula relative to a model $\langle W, R, V\rangle$ and possible world w
primrec *RelativeTruthD* :: $\mathcal{W} \Rightarrow \mathcal{R} \Rightarrow \mathcal{V} \Rightarrow \text{w} \Rightarrow PML \Rightarrow bool$ ($\langle$-,-,-$\rangle$,- $\models^d$ -) **where**
$\langle W,R,V\rangle,\ w \models^d a^d\ = (V\ a\ w)$
$\mid \langle W,R,V\rangle,\ w \models^d \neg^d\varphi = (\neg\ \langle W,R,V\rangle,\ w \models^d \varphi)$
$\mid \langle W,R,V\rangle,\ w \models^d \varphi \supset^d \psi = (\langle W,R,V\rangle,\ w \models^d \varphi \longrightarrow \langle W,R,V\rangle,\ w \models^d \psi)$
$\mid \langle W,R,V\rangle,\ w \models^d \Box^d\varphi = (\forall v{:}W.\ R\ w\ v \longrightarrow \langle W,R,V\rangle,\ v \models^d \varphi)$

— Definition of validity
definition *ValD* ($\models^d$ -) **where** $(\models^d \varphi) \equiv (\forall\ W\ R\ V.\ \forall w{:}W.\ \langle W,R,V\rangle,w \models^d \varphi)$

— Collection of definitions in a bag called DefD
named-theorems *DefD* **declare** *OrD-def*[*DefD*,*simp*] *AndD-def*[*DefD*,*simp*] *DiaD-def*[*DefD*,*simp*] *TopD-def*[*DefD*,*simp*] *BotD-def*[*DefD*,*simp*] *RelativeTruthD-def*[*DefD*,*simp*] *ValD-def*[*DefD*,*simp*]
end

4 Shallow embedding of PML in HOL (maximal)

theory *PMLinHOL-shallow*
imports *PMLinHOL-preliminaries*
begin

— Shallow embedding (of propositional modal logic in HOL)
type-synonym $\sigma = \mathcal{W} \Rightarrow \mathcal{R} \Rightarrow \mathcal{V} \Rightarrow \text{w} \Rightarrow bool$
definition $AtmS{::}\mathcal{S} \Rightarrow \sigma$ (-^s) **where** $a^s \equiv \lambda W\ R\ V\ w.\ V\ a\ w$
definition $NegS{::}\sigma \Rightarrow \sigma$ ($\neg^s$) **where** $\neg^s\ \varphi \equiv \lambda W\ R\ V\ w.\ \neg(\varphi\ W\ R\ V\ w)$
definition $ImpS{::}\sigma \Rightarrow \sigma \Rightarrow \sigma$ (**infixr** $\supset^s$ *93*) **where** $\varphi \supset^s \psi \equiv \lambda W\ R\ V\ w.\ (\varphi\ W\ R$

$V\ w) \longrightarrow (\psi\ W\ R\ V\ w)$
definition *BoxS*$::\sigma\Rightarrow\sigma$ $(\Box^s)$ **where** $\Box^s\ \varphi \equiv \lambda W\ R\ V\ w.\ \forall v{:}W.\ R\ w\ v \longrightarrow (\varphi\ W\ R\ V\ v)$

— Further logical connectives as definitions
definition *OrS* (**infixr** $\vee^s$ *92*) **where** $\varphi \vee^s \psi \equiv \neg^s\varphi \supset^s \psi$
definition *AndS* (**infixr** $\wedge^s$ *95*) **where** $\varphi \wedge^s \psi \equiv \neg^s(\varphi \supset^s \neg^s\psi)$
definition *DiaS* $(\Diamond^s)$ **where** $\Diamond^s\varphi \equiv \neg^s\ (\Box^s\ (\neg^s\varphi))$
definition *TopS* $(\top^s)$ **where** $\top^s \equiv p^s \supset^s p^s$
definition *BotS* $(\bot^s)$ **where** $\bot^s \equiv \neg^s\ \top^s$

— Definition of truth of a formula relative to a model $\langle W,R,V\rangle$ and possible world w
definition *RelativeTruthS*$::\mathcal{W}\Rightarrow\mathcal{R}\Rightarrow\mathcal{V}\Rightarrow \mathrm{w}\Rightarrow\sigma\Rightarrow bool$ $(\langle\text{-},\text{-},\text{-}\rangle,\text{-}\models^s\text{-})$ **where** $\langle W,R,V\rangle,w \models^s \varphi \equiv \varphi\ W\ R\ V\ w$

— Definition of validity
definition *ValS* $(\models^s\text{-})$ **where** $\models^s \varphi \equiv \forall\ W\ R\ V.\ \forall w{:}W.\ \langle W,R,V\rangle,w \models^s \varphi$

— Collection of definitions in a bag called DefS
named-theorems *DefS* **declare** *AtmS-def*[*DefS*,*simp*] *NegS-def*[*DefS*,*simp*] *ImpS-def*[*DefS*,*simp*] *BoxS-def*[*DefS*,*simp*] *OrS-def*[*DefS*,*simp*] *AndS-def*[*DefS*,*simp*] *DiaS-def*[*DefS*,*simp*] *TopS-def*[*DefS*,*simp*] *BotS-def*[*DefS*,*simp*] *RelativeTruthS-def*[*DefS*,*simp*] *ValS-def*[*DefS*,*simp*]
end

5 Shallow embedding of PML in HOL (minimal)

theory *PMLinHOL-shallow-minimal*
imports *PMLinHOL-preliminaries*
begin

— The acessibility relation R and the valuation function V are introduced as constants at the meta-level HOL
consts $R::\mathcal{R}$ $V::\mathcal{V}$

— Shallow embedding (of propositional modal logic in HOL)
type-synonym $\sigma = \mathrm{w}\Rightarrow bool$
definition *AtmM*$::\mathcal{S}\Rightarrow\sigma$ (-^m) **where** $a^m \equiv \lambda w.\ V\ a\ w$
definition *NegM*$::\sigma\Rightarrow\sigma$ $(\neg^m)$ **where** $\neg^m\ \varphi \equiv \lambda w.\neg\varphi\ w$
definition *ImpM*$::\sigma\Rightarrow\sigma\Rightarrow\sigma$ (**infixr** $\supset^m$ *93*) **where** $\varphi \supset^m \psi \equiv \lambda w.\ \varphi\ w \longrightarrow \psi\ w$
definition *BoxM*$::\sigma\Rightarrow\sigma$ $(\Box^m)$ **where** $\Box^m\ \varphi \equiv \lambda w.\ \forall v.\ R\ w\ v \longrightarrow \varphi\ v$

— Further logical connectives as definitions
definition *OrM* (**infixr** $\vee^m$ *92*) **where** $\varphi \vee^m \psi \equiv \neg^m\varphi \supset^m \psi$
definition *AndM* (**infixr** $\wedge^m$ *95*) **where** $\varphi \wedge^m \psi \equiv \neg^m(\varphi \supset^m \neg^m\psi)$
definition *DiaM* $(\Diamond^m\text{-})$ **where** $\Diamond^m\varphi \equiv \neg^m\ (\Box^m\ (\neg^m\varphi))$
definition *TopM* $(\top^m)$ **where** $\top^m \equiv p^m \supset^m p^m$
definition *BotM* $(\bot^m)$ **where** $\bot^m \equiv \neg^m\ \top^m$

— Definition of truth of a formula relative to a model $\langle W,R,V\rangle$ and a possible world w
definition *RelativeTruthM*::w⇒σ⇒*bool* (-$\models^m$ -) **where** $w \models^m \varphi \equiv \varphi\ w$

— Definition of validity
definition *ValM* ($\models^m$ -) **where** $\models^m \varphi \equiv \forall\, w$::w. $w \models^m \varphi$

— Collection of definitions in a bag called DefM
named-theorems *DefM* **declare** *AtmM-def*[*DefM*,*simp*] *NegM-def*[*DefM*,*simp*] *ImpM-def*[*DefM*,*simp*] *BoxM-def*[*DefM*,*simp*] *OrM-def*[*DefM*,*simp*] *AndM-def*[*DefM*,*simp*] *DiaM-def*[*DefM*,*simp*] *TopM-def*[*DefM*,*simp*] *BotM-def*[*DefM*,*simp*] *RelativeTruthM-def*[*DefM*,*simp*] *ValM-def*[*DefM*,*simp*]
end

6 Automated faithfulness proofs

theory *PMLinHOL-faithfulness*
imports *PMLinHOL-deep PMLinHOL-shallow PMLinHOL-shallow-minimal*
begin

— Mappings: deep to maximal shallow and deep to minimal shallow
primrec *DpToShMax* ((|-|)) **where** $(\!|\varphi^d|\!)= \varphi^s \mid (\!|\neg^d\ \varphi|\!) = \neg^s\ (\!|\varphi|\!) \mid (\!|\varphi \supset^d \psi|\!) = (\!|\varphi|\!) \supset^s (\!|\psi|\!) \mid (\!|\Box^d\ \varphi|\!) = \Box^s\ (\!|\varphi|\!)$
primrec *DpToShMin* ([[-]]) **where** $[\![\varphi^d]\!]= \varphi^m \mid [\![\neg^d\ \varphi]\!] = \neg^m\ [\![\varphi]\!] \mid [\![\varphi \supset^d \psi]\!] = [\![\varphi]\!] \supset^m [\![\psi]\!] \mid [\![\Box^d\ \varphi]\!] = \Box^m\ [\![\varphi]\!]$

— Proving faithfulness between deep and maximal shallow
theorem *Faithful1a*: $\forall\ W\ R\ V.\ \forall\, w{:}W.\ \langle W,R,V\rangle,w \models^d \varphi \longleftrightarrow \langle W,R,V\rangle,w \models^s (\!|\varphi|\!)$ ⟨*proof*⟩
theorem *Faithful1b*: $\models^d \varphi \longleftrightarrow\ \models^s (\!|\varphi|\!)$ ⟨*proof*⟩
theorem *Faithful2*: $\forall\, w.\ \langle(\lambda x$::w. *True*$),R,V\rangle,w \models^d \varphi \longleftrightarrow w \models^m [\![\varphi]\!]$ ⟨*proof*⟩
theorem *Faithful3*: $\forall\, w.\ \langle(\lambda x$::w. *True*$),R,V\rangle,w \models^s (\!|\varphi|\!) \longleftrightarrow w \models^m [\![\varphi]\!]$ ⟨*proof*⟩
lemma *Sound1*: $\models^m \psi \longrightarrow (\exists\, \varphi.\ \psi{=}[\![\varphi]\!] \wedge \models^d \varphi)$ — sledgehammer: Proof found; metis reconstruction timeout ⟨*proof*⟩
lemma *Sound2*: $\models^m \psi \longrightarrow (\exists\, \varphi.\ \psi{=}[\![\varphi]\!] \wedge \models^m [\![\varphi]\!])$ — sledgehammer: Proof found; metis reconstruction timeout ⟨*proof*⟩
end

7 Appendix: proof automation tests

7.1 Tests with the deep embedding

theory *PMLinHOL-deep-tests*
imports *PMLinHOL-deep*
begin

— Hilbert calculus: proving that the schematic axioms and rules implied by the embedding

lemma *H1*: $\models^d \varphi \supset^d (\psi \supset^d \varphi)$ ⟨*proof*⟩
lemma *H2*: $\models^d (\varphi \supset^d (\psi \supset^d \gamma)) \supset^d ((\varphi \supset^d \psi) \supset^d (\varphi \supset^d \gamma))$ ⟨*proof*⟩
lemma *H3*: $\models^d (\neg^d\varphi \supset^d \neg^d\psi) \supset^d (\psi \supset^d \varphi)$ ⟨*proof*⟩
lemma *MP*: $\models^d \varphi \Longrightarrow \models^d (\varphi \supset^d \psi) \Longrightarrow \models^d \psi$ ⟨*proof*⟩
lemma *HCderived1*: $\models^d (\varphi \supset^d \varphi)$ — sledgehammer(HC1 HC2 HC3 MP) returns: by (metis HC1 HC2 MP)
⟨*proof*⟩

lemma *HCderived2*: $\models^d \varphi \supset^d (\neg^d\varphi \supset^d \psi)$ ⟨*proof*⟩
lemma *HCderived3*: $\models^d (\neg^d\varphi \supset^d \varphi) \supset^d \varphi$ ⟨*proof*⟩
lemma *HCderived4*: $\models^d (\varphi \supset^d \psi) \supset^d (\neg^d\psi \supset^d \neg^d\varphi)$ ⟨*proof*⟩
lemma *Nec*: $\models^d \varphi \Longrightarrow \models^d \Box^d\varphi$ ⟨*proof*⟩
lemma *Dist*: $\models^d \Box^d(\varphi \supset^d \psi) \supset^d (\Box^d\varphi \supset^d \Box^d\psi)$ ⟨*proof*⟩
lemma *cM*: *reflexive* $R \longleftrightarrow (\forall \varphi\ W\ V.\ \forall w{:}W.\ \langle W,R,V\rangle,w \models^d \Box^d\varphi \supset^d \varphi)$ — sledgehammer: Proof found ⟨*proof*⟩
lemma *cBa*: *symmetric* $R \longrightarrow (\forall \varphi\ W\ V.\ \forall w{:}W.\ \langle W,R,V\rangle,w \models^d \varphi \supset^d \Box^d(\Diamond^d\varphi))$ ⟨*proof*⟩
lemma *cBb*: *symmetric* $R \longleftarrow (\forall \varphi\ W\ V.\ \forall w{:}W.\ \langle W,R,V\rangle,w \models^d \varphi \supset^d \Box^d(\Diamond^d\varphi))$ — sledgehammer: No proof ⟨*proof*⟩
lemma *c4a*: *transitive* $R \longrightarrow (\forall \varphi\ W\ V.\ \forall w{:}W.\ \langle W,R,V\rangle,w \models^d \Box^d\varphi \supset^d \Box^d(\Box^d\varphi))$ ⟨*proof*⟩
lemma *c4b*: *transitive* $R \longleftarrow (\forall \varphi\ W\ V.\ \forall w{:}W.\ \langle W,R,V\rangle,w \models^d \Box^d\varphi \supset^d \Box^d(\Box^d\varphi))$ — sledgehammer: No proof ⟨*proof*⟩
lemma *reflexive* $R \longrightarrow (\forall \varphi\ W\ V.\ \forall w{:}W.\ \langle W,R,V\rangle,w \models^d \Box^d\varphi \supset^d \Box^d(\Box^d\varphi))$ **nitpick**[*card* w=*3*] ⟨*proof*⟩
lemma $\models^d \varphi \supset^d \Box^d\varphi$ **nitpick**[*card* w=*2*, *card* $\mathcal{S}$= *1*] ⟨*proof*⟩
lemma $\models^d \Box^d(\Box^d\varphi \supset^d \Box^d\psi) \vee^d \Box^d(\Box^d\psi \supset^d \Box^d\varphi)$ **nitpick**[*card* w=*3*] ⟨*proof*⟩
lemma $\models^d (\Diamond^d(\Box^d \varphi)) \supset^d \Box^d(\Diamond^d \varphi)$ **nitpick**[*card* w=*2*] ⟨*proof*⟩
lemma *KB*: *symmetric* $R \longrightarrow (\forall \varphi\ \psi\ W\ V.\ \forall w{:}W.\ \langle W,R,V\rangle,w \models^d (\Diamond^d(\Box^d\varphi)) \supset^d \Box^d(\Diamond^d\varphi))$ ⟨*proof*⟩
lemma *K4B*: *symmetric* R ∧ *transitive* $R \longrightarrow (\forall \varphi\ \psi\ W\ V.\ \forall w{:}W.\ \langle W,R,V\rangle,w \models^d \Box^d(\Box^d\varphi \supset^d \Box^d\psi) \vee^d \Box^d(\Box^d\psi \supset^d \Box^d\varphi))$ ⟨*proof*⟩
end

theory *PMLinHOL-deep-further-tests*
imports *PMLinHOL-deep-tests*
begin

— Implied modal principle
lemma *K-Dia*: $\models^d (\Box^d(\varphi \supset^d \psi)) \supset^d ((\Diamond^d\varphi) \supset^d \Diamond^d\psi)$ ⟨*proof*⟩
lemma *T1a*: $\models^d \Box^d p^d \supset^d ((\Diamond^d q^d) \supset^d \Diamond^d(p^d \wedge^d q^d))$ ⟨*proof*⟩
lemma *T1b*: $\models^d \Box^d p^d \supset^d ((\Diamond^d q^d) \supset^d \Diamond^d(p^d \wedge^d q^d))$ — alternative interactive proof in modal object logic K
⟨*proof*⟩
end

theory *PMLinHOL-deep-Loeb-tests*
imports *PMLinHOL-deep*
begin

— Löb axiom: with the deep embedding automated reasoning tools are not very responsive

lemma *Loeb1*: $\forall\varphi.\ \models^d\ \Box^d(\ \Box^d\varphi\ \supset^d\varphi)\ \supset^d\ \Box^d\varphi$ **nitpick**[*card* w=*1*,*card* $\mathcal{S}$=*1*] ⟨*proof*⟩

lemma *Loeb2*: (*conversewf* R ∧ *transitive* R) $\longrightarrow$ $(\forall\varphi\ W\ V.\ \forall w{:}W.\ \langle W,R,V\rangle,w \models^d \Box^d(\Box^d\varphi \supset^d\varphi) \supset^d \Box^d\varphi)$ — sledgehammer: No Proof ⟨*proof*⟩

lemma *Loeb3*: (*conversewf* R ∧ *transitive* R) $\longleftarrow$ $(\forall\varphi\ W\ V.\ \forall w{:}W.\ \langle W,R,V\rangle,w \models^d \Box^d(\Box^d\varphi \supset^d\varphi) \supset^d \Box^d\varphi)$ — sledgehammer: No Proof ⟨*proof*⟩

lemma *Loeb3a*: *conversewf* R $\longleftarrow$ $(\forall\varphi\ W\ V.\ \forall w{:}W.\ \langle W,R,V\rangle,w \models^d \Box^d(\Box^d\varphi \supset^d\varphi) \supset^d \Box^d\varphi)$ — sledgehammer: No Proof ⟨*proof*⟩

lemma *Loeb3b*: *transitive* R $\longleftarrow$ $(\forall\varphi\ W\ V.\ \forall w{:}W.\ \langle W,R,V\rangle,w \models^d \Box^d(\Box^d\varphi \supset^d\varphi) \supset^d \Box^d\varphi)$ — sledgehammer: No Proof ⟨*proof*⟩

lemma *Loeb3c*: *irreflexive* R $\longleftarrow$ $(\forall\varphi\ W\ V.\ \forall w{:}W.\ \langle W,R,V\rangle,w \models^d \Box^d(\Box^d\varphi \supset^d\varphi) \supset^d \Box^d\varphi)$ — sledgehammer: No Proof ⟨*proof*⟩

end

7.2 Tests with the maximal shallow embedding

theory *PMLinHOL-shallow-tests*
  imports *PMLinHOL-shallow*
begin

— Hilbert calculus: proving that the schematic axioms and rules implied by the embedding

lemma *H1*: $\models^s\ \varphi \supset^s (\psi \supset^s \varphi)$ ⟨*proof*⟩

lemma *H2*: $\models^s\ (\varphi \supset^s (\psi \supset^s \gamma)) \supset^s ((\varphi \supset^s \psi) \supset^s (\varphi \supset^s \gamma))$ ⟨*proof*⟩

lemma *H3*: $\models^s\ (\neg^s\varphi \supset^s \neg^s\psi) \supset^s (\psi \supset^s \varphi)$ ⟨*proof*⟩

lemma *MP*: $\models^s\ \varphi \Longrightarrow\ \models^s (\varphi \supset^s \psi) \Longrightarrow\ \models^s \psi$ ⟨*proof*⟩

lemma *HCderived1*: $\models^s\ (\varphi \supset^s \varphi)$ — sledgehammer(HC1 HC2 HC3 MP) returns: by (metis HC1 HC2 MP)
  ⟨*proof*⟩

lemma *HCderived2*: $\models^s\ \varphi \supset^s (\neg^s\varphi \supset^s \psi)$ ⟨*proof*⟩

lemma *HCderived3*: $\models^s\ (\neg^s\varphi \supset^s \varphi) \supset^s \varphi$ ⟨*proof*⟩

lemma *HCderived4*: $\models^s\ (\varphi \supset^s \psi) \supset^s (\neg^s\psi \supset^s \neg^s\varphi)$ ⟨*proof*⟩

lemma *Nec*: $\models^s\ \varphi \Longrightarrow\ \models^s \Box^s\varphi$ ⟨*proof*⟩

lemma *Dist*: $\models^s\ \Box^s(\varphi \supset^s \psi) \supset^s (\Box^s\varphi \supset^s \Box^s\psi)$ ⟨*proof*⟩

lemma *cM*:*reflexive* R $\longleftrightarrow$ $(\forall\varphi\ W\ V.\ \forall w{:}W.\ \langle W,R,V\rangle,w \models^s \Box^s\varphi \supset^s \varphi)$ — sledgehammer: Proof found ⟨*proof*⟩

lemma *cBa*: *symmetric* R $\longrightarrow$ $(\forall\varphi\ W\ V.\ \forall w{:}W.\ \langle W,R,V\rangle,w \models^s \varphi \supset^s \Box^s(\Diamond^s\varphi))$ ⟨*proof*⟩

lemma *cBb*: *symmetric* R $\longleftarrow$ $(\forall\varphi\ W\ V.\ \forall w{:}W.\ \langle W,R,V\rangle,w \models^s \varphi \supset^s \Box^s(\Diamond^s\varphi))$ — sledgehammer: No proof ⟨*proof*⟩

lemma *c4a*: *transitive* R $\longrightarrow$ $(\forall\varphi\ W\ V.\ \forall w{:}W.\ \langle W,R,V\rangle,w \models^s \Box^s\varphi \supset^s \Box^s(\Box^s\varphi))$ ⟨*proof*⟩

lemma *c4b*: *transitive* R $\longleftarrow$ $(\forall\varphi\ W\ V.\ \forall w{:}W.\ \langle W,R,V\rangle,w \models^s \Box^s\varphi \supset^s \Box^s(\Box^s\varphi))$ — sledgehammer: No proof ⟨*proof*⟩

lemma *reflexive* R $\longrightarrow$ $(\forall\varphi\ W\ V.\ \forall w{:}W.\ \langle W,R,V\rangle,w \models^s \Box^s\varphi \supset^s \Box^s(\Box^s\varphi))$

nitpick[*card* w=*3*] ⟨*proof*⟩
lemma $\models^s \varphi \supset^s \Box^s\varphi$ **nitpick**[*card* w=*2*, *card* $\mathcal{S}$= *1*] ⟨*proof*⟩
lemma $\models^s \Box^s(\Box^s\varphi \supset^s\Box^s\psi) \vee^s \Box^s(\Box^s\psi \supset^s\Box^s\varphi)$ ⟨*proof*⟩
lemma $\models^s (\Diamond^s(\Box^s \varphi)) \supset^s\Box^s(\Diamond^s \varphi)$ **nitpick**[*card* w=*2*] ⟨*proof*⟩
lemma *KB*: *symmetric R* $\longrightarrow (\forall \varphi\ \psi\ W\ V.\ \forall w{:}W.\ \langle W,R,V\rangle,w \models^s (\Diamond^s(\Box^s\varphi)) \supset^s\Box^s(\Diamond^s\varphi))$ ⟨*proof*⟩
lemma *K4B*: *symmetric R* ∧ *transitive R* $\longrightarrow (\forall \varphi\ \psi\ W\ V.\ \forall w{:}W.\ \langle W,R,V\rangle,w \models^s \Box^s(\Box^s\varphi \supset^s\Box^s\psi) \vee^s \Box^s(\Box^s\psi \supset^s\Box^s\varphi))$ ⟨*proof*⟩
end

theory *PMLinHOL-shallow-further-tests*
imports *PMLinHOL-shallow-tests*
begin

— Implied modal principle
lemma *K-Dia*: $\models^s (\Box^s(\varphi \supset^s \psi)) \supset^s ((\Diamond^s\varphi) \supset^s \Diamond^s\psi)$ ⟨*proof*⟩
lemma *T1a*:$\models^s \Box^s p^s \supset^s ((\Diamond^s q^s) \supset^s \Diamond^s (p^s \wedge^s q^s))$ ⟨*proof*⟩
lemma *T1b*: $\models^s \Box^s p^s \supset^s ((\Diamond^s q^s) \supset^s \Diamond^s (p^s \wedge^s q^s))$ — alternative interactive proof in modal object logic K
⟨*proof*⟩
end

theory *PMLinHOL-shallow-Loeb-tests*
imports *PMLinHOL-shallow*
begin

— Löb axiom: with the minimal shallow embedding automated reasoning tools are still partly responsive
lemma *Loeb1*: $\forall \varphi.\ \models^s \Box^s(\Box^s\varphi \supset^s\varphi) \supset^s \Box^s\varphi$ **nitpick**[*card* w=*1*,*card* $\mathcal{S}$=*1*] ⟨*proof*⟩
lemma *Loeb2*: (*conversewf R* ∧ *transitive R*) $\longrightarrow(\forall \varphi\ W\ V.\ \forall w{:}W.\ \langle W,R,V\rangle,w \models^s \Box^s(\Box^s\varphi \supset^s\varphi) \supset^s \Box^s\varphi)$ — sledgehammer: Proof found ⟨*proof*⟩
lemma *Loeb3*: (*conversewf R* ∧ *transitive R*) $\longleftarrow (\forall \varphi\ W\ V.\ \forall w{:}W.\ \langle W,R,V\rangle,w \models^s \Box^s(\Box^s\varphi \supset^s\varphi) \supset^s \Box^s\varphi)$ — sledgehammer: No Proof ⟨*proof*⟩
lemma *Loeb3a*: *conversewf R* $\longleftarrow (\forall \varphi\ W\ V.\ \forall w{:}W.\ \langle W,R,V\rangle,w \models^s \Box^s(\Box^s\varphi \supset^s\varphi) \supset^s \Box^s\varphi)$ — sledgehammer: Proof found ⟨*proof*⟩
lemma *Loeb3b*: *transitive R* $\longleftarrow (\forall \varphi\ W\ V.\ \forall w{:}W.\ \langle W,R,V\rangle,w \models^s \Box^s(\Box^s\varphi \supset^s\varphi) \supset^s \Box^s\varphi)$ — sledgehammer: No Proof ⟨*proof*⟩
lemma *Loeb3c*: *irreflexive R* $\longleftarrow (\forall \varphi\ W\ V.\ \forall w{:}W.\ \langle W,R,V\rangle,w \models^s \Box^s(\Box^s\varphi \supset^s\varphi) \supset^s \Box^s\varphi)$ — sledgehammer: Proof found ⟨*proof*⟩
end

7.3 Tests with the minimal shallow embedding

theory *PMLinHOL-shallow-minimal-tests*
imports *PMLinHOL-shallow-minimal*
begin

— Hilbert calculus: proving that the schematic axioms and rules implied by the

embedding
lemma *H1*: $\models^m \varphi \supset^m (\psi \supset^m \varphi)$ ⟨*proof*⟩
lemma *H2*: $\models^m (\varphi \supset^m (\psi \supset^m \gamma)) \supset^m ((\varphi \supset^m \psi) \supset^m (\varphi \supset^m \gamma))$ ⟨*proof*⟩
lemma *H3*: $\models^m (\neg^m\varphi \supset^m \neg^m\psi) \supset^m (\psi \supset^m \varphi)$ ⟨*proof*⟩
lemma *MP*: $\models^m \varphi \Longrightarrow \models^m (\varphi \supset^m \psi) \Longrightarrow \models^m \psi$ ⟨*proof*⟩
lemma *HCderived1*: $\models^m (\varphi \supset^m \varphi)$ — sledgehammer(HC1 HC2 HC3 MP) returns: by (metis HC1 HC2 MP)
⟨*proof*⟩

lemma *HCderived2*: $\models^m \varphi \supset^m (\neg^m\varphi \supset^m \psi)$ ⟨*proof*⟩
lemma *HCderived3*: $\models^m (\neg^m\varphi \supset^m \varphi) \supset^m \varphi$ ⟨*proof*⟩
lemma *HCderived4*: $\models^m (\varphi \supset^m \psi) \supset^m (\neg^m\psi \supset^m \neg^m\varphi)$ ⟨*proof*⟩
lemma *Nec*: $\models^m \varphi \Longrightarrow \models^m \Box^m\varphi$ ⟨*proof*⟩
lemma *Dist*: $\models^m \Box^m(\varphi \supset^m \psi) \supset^m (\Box^m\varphi \supset^m \Box^m\psi)$ ⟨*proof*⟩
lemma *cM*: *reflexive R* $\longleftrightarrow (\forall \varphi. \models^m \Box^m\varphi \supset^m \varphi)$ ⟨*proof*⟩
lemma *cBa*: *symmetric R* $\longrightarrow (\forall \varphi. \models^m \varphi \supset^m \Box^m\Diamond^m\varphi)$ ⟨*proof*⟩
lemma *cBb*: *symmetric R* $\longleftarrow (\forall \varphi. \models^m \varphi \supset^m \Box^m\Diamond^m\varphi)$ ⟨*proof*⟩
lemma *c4a*: *transitive R* $\longrightarrow (\forall \varphi. \models^m \Box^m\varphi \supset^m \Box^m(\Box^m\varphi))$ ⟨*proof*⟩
lemma *c4b*: *transitive R* $\longleftarrow (\forall \varphi. \models^m \Box^m\varphi \supset^m \Box^m(\Box^m\varphi))$ ⟨*proof*⟩
lemma *reflexive R* $\longrightarrow (\forall \varphi. \models^m \Box^m\varphi \supset^m \Box^m(\Box^m\varphi))$ **nitpick**[*card* w=*3*,*show-all*]
⟨*proof*⟩
lemma $\models^m \varphi \supset^m \Box^m\varphi$ **nitpick**[*card* w=*2*, *card* $\mathcal{S}$= *1*] ⟨*proof*⟩
lemma $\models^m \Box^m(\Box^m\varphi \supset^m \Box^m\psi) \vee^m \Box^m(\Box^m\psi \supset^m \Box^m\varphi)$ **nitpick**[*card* w=*3*]
⟨*proof*⟩
lemma $\models^m (\Diamond^m(\Box^m \varphi)) \supset^m \Box^m(\Diamond^m \varphi)$ **nitpick**[*card* w=*2*] ⟨*proof*⟩
lemma *KB*: *symmetric R* $\longrightarrow (\forall \varphi\ \psi. \models^m (\Diamond^m(\Box^m\varphi)) \supset^m \Box^m(\Diamond^m\varphi))$ ⟨*proof*⟩
lemma *K4B*: *symmetric R* $\wedge$ *transitive R* $\longrightarrow (\forall \varphi\ \psi. \models^m \Box^m(\Box^m\varphi \supset^m \Box^m\psi) \vee^m \Box^m(\Box^m\psi \supset^m \Box^m\varphi))$ ⟨*proof*⟩
end

theory *PMLinHOL-shallow-minimal-further-tests*
imports *PMLinHOL-shallow-minimal-tests*
begin

— Implied modal principle
lemma *K-Dia*: $\models^m (\Box^m(\varphi \supset^m \psi)) \supset^m ((\Diamond^m\varphi) \supset^m \Diamond^m\psi)$ ⟨*proof*⟩
lemma *T1a*: $\models^m \Box^m p^m \supset^m ((\Diamond^m q^m) \supset^m \Diamond^m(p^m \wedge^m q^m))$ ⟨*proof*⟩
lemma *T1b*: $\models^m \Box^m p^m \supset^m ((\Diamond^m q^m) \supset^m \Diamond^m(p^m \wedge^m q^m))$ — alternative interactive proof in modal object logic K
⟨*proof*⟩
end

theory *PMLinHOL-shallow-minimal-Loeb-tests*
imports *PMLinHOL-shallow-minimal*
begin

— Löb axiom: with the minimal shallow embedding automated reasoning tools are still partly responsive
lemma *Loeb1*: $\forall \varphi. \models^m \Box^m(\Box^m\varphi \supset^m \varphi) \supset^m \Box^m\varphi$ **nitpick**[*card* w=*1*,*card* $\mathcal{S}$=*1*]

$\langle proof \rangle$
lemma *Loeb2*: $(conversewf\ R \wedge transitive\ R) \longrightarrow (\forall \varphi.\ \models^m \Box^m(\Box^m\varphi \supset^m \varphi) \supset^m \Box^m\varphi)$ — sh: Proof found $\langle proof \rangle$
lemma *Loeb3*: $(conversewf\ R \wedge transitive\ R) \longleftarrow (\forall \varphi.\ \models^m \Box^m(\Box^m\varphi \supset^m \varphi) \supset^m \Box^m\varphi)$ — sh: No Proof $\langle proof \rangle$
lemma *Loeb3a*: $conversewf\ R \longleftarrow (\forall \varphi.\ \models^m \Box^m(\Box^m\varphi \supset^m \varphi) \supset^m \Box^m\varphi)$ $\langle proof \rangle$
lemma *Loeb3b*: $transitive\ R \longleftarrow (\forall \varphi.\ \models^m \Box^m(\Box^m\varphi \supset^m \varphi) \supset^m \Box^m\varphi)$ — sledgehammer: No Proof $\langle proof \rangle$
lemma *Loeb3c*: $irreflexive\ R \longleftarrow (\forall \varphi.\ \models^m \Box^m(\Box^m\varphi \supset^m \varphi) \supset^m \Box^m\varphi)$ — sledgehammer: Proof found $\langle proof \rangle$
end

References

[1] C. Benzmüller. Faithful logic embeddings in HOL — deep and shallow. In *Automated Deduction – CADE-30 – 30th International Conference on Automated Deduction, Stuttgart, Germany, July 28–31, Proceedings*, Lecture Notes in Computer Science. Springer, 2025. To appear (preprint: arXiv:2502.19311v1).