Faithful Semantical Embedding of a Dyadic Deontic Logic in HOL

**Christoph Benzmüller · Ali Farjami ·
Xavier Parent**

March 5, 2018

Abstract A shallow semantical embedding of a dyadic deontic logic by Carmo and Jones in classical higher-order logic is presented. This embedding is proven sound and complete, that is, faithful.

The work presented here provides the theoretical foundation for the implementation and automation of dyadic deontic logic within off-the-shelf higher-order theorem provers and proof assistants.

Keywords Dyadic deontic logic; Classical higher-order logic; Semantic embedding; Faithfulness

1 Introduction

Dyadic deontic logic is the logic for reasoning with dyadic obligations ("it ought to be the case that ... if it is the case that ..."). A particular dyadic deontic logic, tailored to so-called contrary-to-duty conditionals, has been proposed by Carmo and Jones [11]. We shall refer to it as DDL in the remainder. DDL comes with a neighborhood semantics and a weakly complete axiomatization over the class of finite models. The framework is immune to the well-known contrary-to-duty paradoxes, like Chisholm's paradox, and other related puzzles. However, the question of how to mechanise and automate reasoning tasks in DDL has not been studied yet.

Christoph Benzmüller
University of Luxembourg, Luxembourg, and Freie Universität Berlin, Germany
E-mail: c.benzmueller@gmail.com

Ali Farjami, Xavier Parent
University of Luxembourg, Luxembourg
E-mail: ali.farjami@uni.lu, xavier.parent@uni.lu

This article adresses this challenge. We essentially devise a faithfull semantical embedding of DDL in classical higher-order logic (HOL). The latter logic thereby serves as an universal meta-logic. Analogous to successful, recent work in the area of computational metaphysics (cf. [6] and the references therein), the key motivation is to mechanise and automate DDL on the computer by reusing existing theorem proving technology for meta-logic HOL. The embedding of DDL in HOL as devised in this article enables just this.

Meta-logic HOL [4], as employed in this article, was originally devised by Church [14], and further developed by Henkin [15], Andrews [1,3,2]. It bases both terms and formulas on simply typed λ-terms. The use of the λ-calculus has some major advantages. For example, λ-abstractions over formulas allow the explicit naming of sets and predicates, something that is achieved in set theory via the comprehension axioms. Another advantage is, that the complex rules for quantifier instantiation at first-order and higher-order types is completely explained via the rules of λ-conversion (the so-called rules of α-, β-, and η-conversion) which were proposed earlier by Church [12,13]. These two advantages are exploited in our embedding of DDL in HOL.

Different notions of semantics for HOL have been thoroughly studied in the literature [7,16]. In this article we assume HOL with Henkin semantics (cf. the detailed description by Benzmüller et. al. [7]). For this notion of HOL, which does not suffer from Gödel's incompleteness results, several sound and complete theorem provers have been developed in the past decades [8]. We propose to reuse these systems for the automation of DDL. The semantical embedding as devised in this article provides both the theoretical foundation for the approach and the practical bridging technology that is enabling DDL applications within existing HOL theorem provers.

The article is structured as follows: Section 2 outlines the syntax and semantics of DDL, as far as needed for this article. Section 3 provides a comparably detailed introduction into HOL; this is needed to keep the article sufficiently self-contained. The semantical embedding of DDL in HOL is then devised and studied in Sec. 4. This section also presents the respective soundness and completeness proofs. Section 5 concludes the paper.

2 The Dyadic Deontic Logic of Carmo and Jones

This section provides a concise introduction of DDL, the dyadic deontic logic proposed by Carmo and Jones. Definitions as required for the remainder are presented. For further details we refer to the literature [11,10].

To define the formulas of DDL we start with an countable set of propositional symbols P, and we choose $\neg$ and $\vee$ as the only primitive connectives.

The set of *DDL formulas* is given as the smallest set of formulas obeying the following conditions:

– Each $p^i \in P$ is an (atomic) DDL formula.

– Given two arbitrary DDL formulas φ and ψ, then

$\neg\varphi$	—	*classical negation*,
$\varphi \vee \psi$	—	*classical disjunction*,
$\bigcirc(\psi/\varphi)$	—	*dyadic deontic obligation: "it ought to be ψ, given φ"*,
$\Box\varphi$	—	*in all worlds*,
$\Box_a\varphi$	—	*in all actual versions of the current world*,
$\Box_p\varphi$	—	*in all potential versions of the current world*,
$\bigcirc_a(\varphi)$	—	*monadic deontic operator for actual obligation*, and
$\bigcirc_p(\varphi)$	—	*monadic deontic operator for primary obligation*

are also DDL formulas.

Further logical connectives can be defined as usual. For example, we may define $\varphi \wedge \psi := \neg(\neg\varphi \vee \neg\psi)$, $\varphi \rightarrow \psi := \neg\varphi \vee \psi$, $\varphi \longleftrightarrow \psi := (\varphi \rightarrow \psi) \wedge (\psi \rightarrow \varphi)$, $\Diamond\varphi := \neg\Box\neg\varphi$, $\Diamond_a\varphi := \neg\Box_a\neg\varphi$, $\Diamond_p\varphi := \neg\Box_p\neg\varphi$, $\top := \neg q \vee q$, for some propositional symbol q, and $\bot := \neg\top$.

A DDL *model* is a structure $M = \langle S, av, pv, ob, V\rangle$, where S is a non empty set of items called possible worlds, V is a function assigning a set of worlds to each atomic formula, that is, $V(p^i) \subseteq S$. av: $S \rightarrow P(S)$, where $P(S)$ denotes the power set of S, is a function mapping worlds to sets of worlds such that $av(s) \neq \emptyset$. $av(s)$ denotes the set of actual versions of the world s. pv: $S \rightarrow P(S)$ is another, similar mapping such that $av(s) \subseteq pv(s)$ and $s \in pv(s)$. $pv(s)$ denotes the set of potential versions of the world s. ob: $P(S) \rightarrow P(P(S))$, which denotes the set of propositions that are obligatory in context $\bar{X} \subseteq S$, is a function mapping set of worlds to sets of sets of worlds. The following conditions hold for ob (where $\bar{X}, \bar{Y}, \bar{Z}$ designate arbitrary subsets of S):

1. $\emptyset \notin ob(\bar{X})$.
2. If $\bar{Y} \cap \bar{X} = \bar{Z} \cap \bar{X}$, then $\bar{Y} \in ob(\bar{X})$ if and only if $\bar{Z} \in ob(\bar{X})$.
3. Let $\bar{\beta} \subseteq ob(\bar{X})$ and $\bar{\beta} \neq \emptyset$. If $(\cap\bar{\beta})\cap\bar{X} \neq \emptyset$ (where $\cap\bar{\beta} = \{s \in S \mid \text{for all } \bar{Z} \in \beta \text{ we have } s \in \bar{Z}\}$), then $(\cap\bar{\beta}) \in ob(\bar{X})$.
4. If $\bar{Y} \subseteq \bar{X}$ and $\bar{Y} \in ob(\bar{X})$ and $\bar{X} \subseteq \bar{Z}$, then $(\bar{Z} \setminus \bar{X}) \cup \bar{Y} \in ob(\bar{Z})$.
5. If $\bar{Y} \subseteq \bar{X}$ and $\bar{Z} \in ob(\bar{X})$ and $\bar{Y} \cap \bar{Z} \neq \emptyset$, then $\bar{Z} \in ob(\bar{Y})$.

Satisfiability of a formula φ for a model $M = \langle S, av, pv, ob, V\rangle$ and a world $s \in S$ is denoted by $M, s \models \varphi$ and we define $V^M(\varphi) = \{s \in S \mid M, s \models \varphi\}$. In order to simplify the presentation, whenever the model M is obvious from context, we write $V(\varphi)$ instead of $V^M(\varphi)$. Moreover, we often use "iff" as shorthand for “if and only if”.

$$\begin{array}{lll}
M,s \models p & \text{iff} & s \in V(p) \\
M,s \models \neg\varphi & \text{iff} & M,s \not\models \varphi \,\text{(that is, not } M,s \models \varphi) \\
M,s \models \varphi \vee \psi & \text{iff} & M,s \models \varphi \text{ or } M,s \models \psi \\
M,s \models \Box\varphi & \text{iff} & V(\varphi) = S \\
M,s \models \Box_a\varphi & \text{iff} & av(s) \subseteq V(\varphi) \\
M,s \models \Box_p\varphi & \text{iff} & pv(s) \subseteq V(\varphi) \\
M,s \models \bigcirc(\psi/\varphi) & \text{iff} & V(\psi) \in ob(V(\varphi)) \\
M,s \models \bigcirc_a\varphi & \text{iff} & V(\varphi) \in ob(av(s)) \text{ and } av(s) \cap V(\neg\varphi) \neq \emptyset \\
M,s \models \bigcirc_p\varphi & \text{iff} & V(\varphi) \in ob(pv(s)) \text{ and } pv(s) \cap V(\neg\varphi) \neq \emptyset
\end{array}$$

Our evaluation rule for $\bigcirc(\_/\_)$ is a simplified version of the one used by Carmo and Jones. Given the constraints placed on *ob*, both rules are equivalent (cf. [5, result II-2-2]).

As usual, a DDL formula φ is *valid in a DDL model* $M = \langle S, av, pv, ob, V\rangle$, denoted as $M \models^{DDL} \varphi$, if and only if for all worlds $s \in S$ holds $M, s \models \varphi$. A formula φ is *valid*, denoted $\models^{DDL} \varphi$, if and only if it is valid in every DDL model.

3 Classical Higher-order Logic

In this section we introduce classical higher-order logic (HOL). The presentation, which has partly been adapted from [5], is rather detailed in order to keep the article sufficiently self-contained.

3.1 Syntax of HOL

For defining the syntax of HOL, we first introduce the set T of *simple types.* We assume that T is freely generated from a set of *basic types* $BT \supseteq \{o, i\}$ using the function type constructor $\rightarrow$. o denotes the (bivalent) set of Booleans, and i a non-empty set of individuals.

For the definition of HOL, we start out with a family of denumerable sets of typed constant symbols $(C_\alpha)_{\alpha\in T}$, called *signature*, and a family of denumerable sets of typed variable symbols $(V_\alpha)_{\alpha\in T}$.[1] We employ Church-style typing, where each term t_α explicitly encodes its type information in subscript α.

The *language of HOL* is given as the smallest set of terms obeying the following conditions.

[1] For example in Section 4 we will assume constant symbols *av*, *pv* and *ob* with types $i \rightarrow i \rightarrow o$, $i \rightarrow i \rightarrow o$ and $(i \rightarrow o) \rightarrow (i \rightarrow o) \rightarrow o$ as part of the signature.

- Every typed constant symbol $c_\alpha \in C_\alpha$ is a HOL term of type α.
- Every typed variable symbol $X_\alpha \in V_\alpha$ is a HOL term of type α.
- If $s_{\alpha\to\beta}$ and t_α are HOL terms of types $\alpha \to \beta$ and α, respectively, then $(s_{\alpha\to\beta}\, t_\alpha)_\beta$, called *application*, is an HOL term of type β.
- If $X_\alpha \in V_\alpha$ is a typed variable symbol and s_β is an HOL term of type β, then $(\lambda X_\alpha s_\beta)_{\alpha\to\beta}$, called *abstraction*, is an HOL term of type $\alpha \to \beta$.

The above definition encompasses the simply typed λ-calculus. In order to extend this base framework into logic HOL we simply ensure that the signature $(C_\alpha)_{\alpha\in T}$ provides a sufficient selection of primitive logical connectives. Without loss of generality, we here assume the following *primitive logical connectives* to be part of the signature: $\neg_{o\to o} \in C_{o\to o}$, $\vee_{o\to o\to o} \in C_{o\to o\to o}$, $\mathit{\Pi}_{(\alpha\to o)\to o} \in C_{(\alpha\to o)\to o}$ and $=_{\alpha\to\alpha\to\alpha}\in C_{\alpha\to\alpha\to\alpha}$, abbreviated as $=^\alpha$. The symbols $\mathit{\Pi}_{(\alpha\to o)\to o}$ and $=_{\alpha\to\alpha\to\alpha}$ are generally assumed for each type $\alpha \in T$. The denotation of the primitive logical connectives is fixed below according to their intended meaning. *Binder notation* $\forall X_\alpha\, s_o$ is used as an abbreviation for $\mathit{\Pi}_{(\alpha\to o)\to o}\lambda X_\alpha s_o$. Universal quantification in HOL is thus modeled with the help of the logical constants $\mathit{\Pi}_{(\alpha\to o)\to o}$ to be used in combination with lambda-abstraction. That is, the only binding mechanism provided in HOL is lambda-abstraction.

HOL is a logic of terms in the sense that the *formulas of HOL* are given as the terms of type o. In addition to the primitive logical connectives selected above, we could assume *choice operators* $\epsilon_{(\alpha\to o)\to\alpha} \in C_{(\alpha\to o)\to\alpha}$ (for each type α) in the signature. We are not pursuing this here.

Type information as well as brackets may be omitted if obvious from the context, and we may also use infix notion to improve readability. For example, we may write $(s \vee t)$ instead of $((\vee_{o\to o\to o} s_o) t_o)$.

From the selected set of primitive connectives, other logical connectives can be introduced as abbreviations.[2] For example, we may define $s \wedge t := \neg(\neg s \vee \neg t)$, $s \rightarrow t := \neg s \vee t$, $s \longleftrightarrow t := (s \rightarrow t) \wedge (t \rightarrow s)$, $\top := (\lambda X_i X_i) = (\lambda X_i X_i)$, $\bot := \neg\top$ and $\exists X_\alpha s := \neg\forall X_\alpha \neg s$.

Also equality can be defined in HOL by exploiting Leibniz' principle, expressing that two objects are equal if they share the same properties. *Leibniz equality* $\doteq^\alpha$ at type α is thus defined as $s_\alpha \doteq^\alpha t_\alpha := \forall P_{\alpha\to o}(\neg Ps \vee Pt)$.

[2] As demonstrated by Andrews [4], we could in fact start out with only primitive equality in the signature (for all types α) and introduce all other logical connectives as abbreviations based on it. The motivation for the redundant signature as selected here is to stay close to the the choices taken in implemented theorem provers such as LEO-II and Leo-III and also to theory paper [7], which is recommended for further details.

Each occurrence of a variable in a term is either bound by a λ or free. We use $free(s)$ to denote the set of variables with a free occurrence in s. We consider two terms to be *equal* if the terms are the same up to the names of bound variables, that is, we consider α-conversion implicitly.

Substitution of a term s_α for a variable X_α in a term t_β is denoted by $[s/X]t$. Since we consider α-conversion implicitly, we assume the bound variables of t to avoid variable capture.

Well-known operations and relations on HOL terms include *$\beta\eta$-normalization* and *$\beta\eta$-equality*, denoted by $s =_{\beta\eta} t$, *β-reduction* and *η-reduction*. A *β-redex* $(\lambda X s)t$ β-reduces to $[t/X]s$. An *η-redex* $\lambda X(sX)$, where $X \notin free(s)$, η-reduces to s. We write $s =_\beta t$ to mean s can be converted to t by a series of β-reductions and expansions. Similarly, $s =_{\beta\eta} t$ means s can be converted to t using both β and η.

3.2 Semantics of HOL

The semantics of HOL is well understood and thoroughly documented. The introduction provided next focuses on the aspects as needed for this article. For more details we refer to the previously mentioned literature [7].

The semantics of choice for the remainder is Henkin semantics, i.e., we work with Henkin's general models. Henkin models (and standard models) are introduced next. We start out with introducing frame structures.

A *frame* D is a collection $\{D_\alpha\}_{\alpha \in \mathrm{T}}$ of nonempty sets D_α, such that $D_o = \{T, F\}$ (for truth and falsehood). The $D_{\alpha\rightarrow\beta}$ are collections of functions mapping D_α into D_β.

A *model* for HOL is a tuple $M = \langle D, I \rangle$, where D is a frame, and I is a family of typed interpretation functions mapping constant symbols $p_\alpha \in C_\alpha$ to appropriate elements of D_α, called the *denotation of* p_α. The logical connectives $\neg$, $\vee$, Π and $=$ are always given their expected, standard denotations:[3]

- $I(\neg_{o\rightarrow o}) = not \in D_{o\rightarrow o}$ such that $not(T) = F$ and $not(F) = T$.
- $I(\vee_{o\rightarrow o\rightarrow o}) = or \in D_{o\rightarrow o\rightarrow o}$ such that $or(a, b) = T$ iff ($a = T$ or $b = T$).
- $I(=_{\alpha\rightarrow\alpha\rightarrow o}) = id \in D_{\alpha\rightarrow\alpha\rightarrow o}$ such that for all $a, b \in D_\alpha$, $id(a, b) = T$ iff a is identical to b.

[3] Since $=_{\alpha\rightarrow\alpha\rightarrow o}$ (for all types α) is in the signature, it is ensured that the domains $D_{\alpha\rightarrow\alpha\rightarrow o}$ contain the respective identity relations. This addresses an issue discovered by Andrews [2]: if such identity relations were not existing in the $D_{\alpha\rightarrow\alpha\rightarrow o}$, then Leibniz equality in Henkin semantics may not denote as intended.

- $I(\Pi_{(\alpha\to o)\to o}) = all \in D_{(\alpha\to o)\to o}$ such that for all $s \in D_{\alpha\to o}$, $all(s) = T$ iff $s(a) = T$ for all $a \in D_\alpha$; i.e., s is the set of all objects of type α.

Variable assignments are a technical aid for the subsequent definition of an interpretation function $\|.\|^{M,g}$ for HOL terms. This interpretation function is parametric over a model M and a variable assignment g.

A *variable assignment* g maps variables X_α to elements in D_α. $g[d/W]$ denotes the assignment that is identical to g, except for variable W, which is now mapped to d.

The *denotation* $\|s_\alpha\|^{M,g}$ of an HOL term s_α on a model $M = \langle D, I\rangle$ under assignment g is an element $d \in D_\alpha$ defined in the following way:

$$\begin{array}{rcl}
\|p_\alpha\|^{M,g} & = & I(p_\alpha) \\
\|X_\alpha\|^{M,g} & = & g(X_\alpha) \\
\|(s_{\alpha\to\beta}\, t_\alpha)_\beta\|^{M,g} & = & \|s_{\alpha\to\beta}\|^{M,g}(\|t_\alpha\|^{M,g}) \\
\|(\lambda X_\alpha s_\beta)_{\alpha\to\beta}\|^{M,g} & = & \text{the function } f \text{ from } D_\alpha \text{ to } D_\beta \text{ such that} \\
 & & f(d) = \|s_\beta\|^{M,g[d/X_\alpha]} \text{ for all } d \in D_\alpha
\end{array}$$

A model $M = \langle D, I\rangle$ is called a *standard model* if and only if for all $\alpha, \beta \in T$ we have $D_{\alpha\to\beta} = \{f \mid f : D_\alpha \longrightarrow D_\beta\}$. In a *Henkin model (general model)* function spaces are not necessarily full. Instead it is only required that for all $\alpha, \beta \in T$, $D_{\alpha\to\beta} \subseteq \{f \mid f : D_\alpha \longrightarrow D_\beta\}$. However, it is required that the valuation function $\|\cdot\|^{M,g}$ from above is total, so that every term denotes. Note that this requirement, which is called *Denotatpflicht*, ensures that the function domains $D_{\alpha\to\beta}$ never become too sparse, that is, the denotations of the lambda-abstractions as devised above are always contained in them.

Corollary 1 *For any Henkin model $M = \langle D, I\rangle$ and variable assignment g holds:*

1. $\|(\neg_{o\to o}\, s_o)_o\|^{M,g} = T$ *iff* $\|s_o\|^{M,g} = F$.
2. $\|((\vee_{o\to o\to o}\, s_o)\, t_o)_o\|^{M,g} = T$ *iff* $\|s_o\|^{M,g} = T$ *or* $\|t_o\|^{M,g} = T$.
3. $\|((\wedge_{o\to o\to o}\, s_o)\, t_o)_o\|^{M,g} = T$ *iff* $\|s_o\|^{M,g} = T$ *and* $\|t_o\|^{M,g} = T$.
4. $\|((\rightarrow_{o\to o\to o}\, s_o)\, t_o)_o\|^{M,g} = T$ *iff* $\|s_o\|^{M,g} = T$ *then* $\|t_o\|^{M,g} = T$.
5. $\|((\longleftrightarrow_{o\to o\to o}\, s_o)\, t_o)_o\|^{M,g} = T$ *iff* $\|s_o\|^{M,g} = T$ *iff* $\|t_o\|^{M,g} = T$.
6. $\|\top\|^{M,g} = T$.
7. $\|\bot\|^{M,g} = F$.
8. $\|(\forall X_\alpha s_o)_o\|^{M,g} = \|(\forall_{(\alpha\to o)\to o}(\lambda X_\alpha s_o))_o\|^{M,g} = T$ *iff for all $d \in D_\alpha$ we have $\|s_o\|^{M,g[d/X_\alpha]} = T$.*

9. $\|(\exists X_\alpha s_o)_o\|^{M,g} = T$ *iff there exists* $d \in D_\alpha$ *such that* $\|s_o\|^{M,g[d/X_\alpha]} = T$.

Proof We leave the proof as an exercise to the reader.

An HOL formula s_o is *true* in an Henkin model M under assignment g if and only if $\|s_o\|^{M,g} = T$; this is also denoted by $M, g \models^{\mathrm{HOL}} s_o$. An HOL formula s_o is called *valid* in M, which is denoted by $M \models^{\mathrm{HOL}} s_o$, if and only if $M, g \models^{\mathrm{HOL}} s_o$ for all assignments g. Moreover, a formula s_o is called *valid*, denoted by $\models^{\mathrm{HOL}} s_o$, if and only if s_o is valid in all Henkin models M. Finally, we define $\Sigma \models^{\mathrm{HOL}} s_o$ for a set of HOL formulas Σ if and only if $M \models^{\mathrm{HOL}} s_o$ for all Henkin models M with $M \models^{\mathrm{HOL}} t_o$ for all $t_o \in \Sigma$.

Note that any standard model is obviously also a Henkin model. Hence, validity of a HOL formula s_o for all Henkin models, implies validity of s_o for all standard models.

4 Modeling DDL as a Fragment of HOL

This section, as the core contribution of this article, presents a shallow semantical embedding of DDL in HOL and proves its soundness and completeness.

4.1 Semantical Embedding

DDL formulas are identified in our semantical embedding with certain HOL terms (predicates) of type $i \to o$. They can be applied to terms of type i, which are assumed to denote possible worlds. That is, the HOL type i is now identified with a (non-empty) set of worlds. Type $i \to o$ is abbreviated as τ in the remainder. The HOL signature is assumed to contain the constant symbol $av_{i \to \tau}$, $pv_{i \to \tau}$ and $ob_{\tau \to \tau \to o}$. Moreover, for each propositional symbol p^i of DDL, the HOL signature must contain a respective constant symbols p^i_τ. Without loss of generality, we assume that besides those symbols and the primitive logical connectives of HOL, no other constant symbols are given in the signature of HOL.

The mapping $\lfloor\cdot\rfloor$ translates DDL formulas s into HOL terms $\lfloor s\rfloor$ of type τ. The mapping is recursively defined:

$$\begin{array}{ll}
\lfloor p^i\rfloor & = p^i_\tau \\
\lfloor\neg s\rfloor & = \neg_\tau \lfloor s\rfloor \\
\lfloor s\vee t\rfloor & = \vee_{\tau\to\tau\to\tau} \lfloor s\rfloor\lfloor t\rfloor \\
\lfloor\Box s\rfloor & = \Box_{\tau\to\tau} \lfloor s\rfloor \\
\lfloor\bigcirc(t/s)\rfloor & = \bigcirc_{\tau\to\tau\to\tau} \lfloor s\rfloor\lfloor t\rfloor \\
\lfloor\Box_a s\rfloor & = \Box^a_{\tau\to\tau} \lfloor s\rfloor \\
\lfloor\Box_p s\rfloor & = \Box^p_{\tau\to\tau} \lfloor s\rfloor \\
\lfloor\bigcirc_a(s)\rfloor & = \bigcirc^a_{\tau\to\tau} \lfloor s\rfloor \\
\lfloor\bigcirc_p(s)\rfloor & = \bigcirc^p_{\tau\to\tau} \lfloor s\rfloor
\end{array}$$

$\neg_{\tau\to\tau}$, $\vee_{\tau\to\tau\to\tau}$, $\Box_{\tau\to\tau}$, $\bigcirc_{\tau\to\tau\to\tau}$, $\Box^a_{\tau\to\tau}$, $\Box^p_{\tau\to\tau}$, $\bigcirc^a_{\tau\to\tau}$ and $\bigcirc^p_{\tau\to\tau}$ thereby abbreviate the following HOL terms:

$$\begin{array}{ll}
\neg_{\tau\to\tau} & = \lambda A_\tau \lambda X_i \neg(A\,X) \\
\vee_{\tau\to\tau\to\tau} & = \lambda A_\tau \lambda B_\tau \lambda X_i (A\,X \vee B\,X) \\
\Box_{\tau\to\tau} & = \lambda A_\tau \lambda X_i \forall Y_i (A\,Y) \\
\bigcirc_{\tau\to\tau\to\tau} & = \lambda A_\tau \lambda B_\tau \lambda X_i (ob\,A\,B) \\
\Box^a_{\tau\to\tau} & = \lambda A_\tau \lambda X_i \forall Y_i (\neg(av\,X\,Y) \vee A\,Y) \\
\Box^p_{\tau\to\tau} & = \lambda A_\tau \lambda X_i \forall Y_i (\neg(pv\,X\,Y) \vee (A\,Y)) \\
\bigcirc^a_{\tau\to\tau} & = \lambda A_\tau \lambda X_i ((ob\,(av\,X)\,A) \wedge \exists Y_i (av\,X\,Y \wedge \neg(A\,Y))) \\
\bigcirc^p_{\tau\to\tau} & = \lambda A_\tau \lambda X_i ((ob\,(pv\,X)\,A) \wedge \exists Y_i (pv\,X\,Y \wedge \neg(A\,Y)))
\end{array}$$

Analyzing the truth of a translated formula $\lfloor s\rfloor$ in a world represented by term w_i corresponds to evaluating the application $(\lfloor s\rfloor\,w_i)$. In line with previous work [9], we define $\text{vld}_{\tau\to o} = \lambda A_\tau \forall S_i (A\,S)$. With this definition, validity of a DDL formula s in DDL corresponds to the validity of formula $(\text{vld}\,\lfloor s\rfloor)$ in HOL, and vice versa.

4.2 Soundness and Completeness

To prove the soundness and completeness, that is, faithfulness, of the above embedding, a mapping from DDL models into Henkin models is employed.

Definition 1 (Henkin model H^M for DDL model M) For any DDL model $M = \langle S, av, pv, ob, V\rangle$, we define corresponding Henkin models H^M. Thus, let a DDL model $M = \langle S, av, pv, ob, V\rangle$ be given. Moreover, assume that $p^1, ..., p^m \in P$, for $m \geq 1$, are the only propositional symbols of DDL. Remember that our embedding requires the corresponding signature of HOL to provide constant symbols p^j_τ such that $\lfloor p^j\rfloor = p^j_\tau$ for $j = 1, \ldots, m$.

An Henkin model $H^M = \langle \{D_\alpha\}_{\alpha\in T}, I\rangle$ for M is now defined as follows: D_i is chosen as the set of possible worlds S; all other sets $D_{\alpha\to\beta}$ are chosen as (not necessarily full) sets of functions from D_α to D_β. For all $D_{\alpha\to\beta}$ the rule that every term $t_{\alpha\to\beta}$ must have a denotation in $D_{\alpha\to\beta}$ must be obeyed (Denotatpflicht). In particular, it is required that D_τ, $D_{i\to\tau}$ and $D_{\tau\to\tau\to o}$ contain the elements Ip^j_τ, $Iav_{i\to\tau}$, $Ipv_{i\to\tau}$ and $Iob_{\tau\to\tau\to o}$. The interpretation function I of H^M is defined as follows:

1. For $i = 1, \ldots, m$, $Ip^i_\tau \in D_\tau$ is chosen such that $Ip^i_\tau(s) = T$ iff $s \in V(p^j)$ in M.
2. $Iav_{i\to\tau} \in D_{i\to\tau}$ is chosen such that $Iav_{i\to\tau}(s,u) = T$ iff $u \in av(s)$ in M.
3. $Ipv_{i\to\tau} \in D_{i\to\tau}$ is chosen such that $Ipv_{i\to\tau}(s,u) = T$ iff $u \in pv(s)$ in M.
4. $Iob_{\tau\to\tau\to o} \in D_{\tau\to\tau\to o}$ is chosen such that $Iob_{\tau\to\tau\to o}(\bar{X},\bar{Y}) = T$ iff $\bar{Y} \in ob(\bar{X})$ in M.
5. For the logical connectives $\neg$, $\vee$, Π and $=$ of HOL the interpretation function I is defined as usual (see the previous section).

Since we assume that there are no other symbols (besides the p^i_τ, av, pv, ob and $\neg$, $\vee$, Π, and $=$) in the signature of HOL, I is a total function. Moreover, the above construction guarantees that H^M is a Henkin model: $\langle D, I\rangle$ is a frame, and the choice of I in combination with the Denotatpflicht ensures that for arbitrary assignments g, $\|.\|^{H^M,g}$ is an total evaluation function.

Lemma 1 *Let H^M be a Henkin model for a DDL model M. In H^M we have for all $s \in D_i$ and all $\bar{X}, \bar{Y}, \bar{Z} \in D_\tau$ (cf. the conditions DDL models as stated on page 3):*[4]

(av) $Iav_{i\to\tau}(s) \neq \emptyset$.

(pv1) $Iav_{i\to\tau}(s) \subseteq Ipv_{i\to\tau}(s)$.

(pv2) $s \in Ipv_{i\to\tau}(s)$.

(ob1) $\emptyset \notin Iob_{\tau\to\tau\to o}(\bar{X})$.

(ob2) *If* $\bar{Y} \cap \bar{X} = \bar{Z} \cap \bar{X}$, *then* ($\bar{Y} \in Iob_{\tau\to\tau\to o}(\bar{X})$ *iff* $\bar{Z} \in Iob_{\tau\to\tau\to o}(\bar{X})$).

(ob3) *Let* $\bar{\beta} \subseteq Iob_{\tau\to\tau\to o}(\bar{X})$ *and* $\bar{\beta} \neq \emptyset$.
If $(\cap\bar{\beta}) \cap \bar{X} \neq \emptyset$, *where* $\cap\bar{\beta} = \{s \in S \mid$ *for all* $\bar{Z} \in \bar{\beta}$ *we have* $s \in \bar{Z}\}$,
then $(\cap\bar{\beta}) \in Iob_{\tau\to\tau\to o}(\bar{X})$.

(ob4) *If* $\bar{Y} \subseteq \bar{X}$ *and* $\bar{Y} \in Iob_{\tau\to\tau\to o}(\bar{X})$ *and* $\bar{X} \subseteq \bar{Z}$,
then $(\bar{Z} \setminus \bar{X}) \cup \bar{Y} \in Iob_{\tau\to\tau\to o}(\bar{Z})$.

(ob5) *If* $\bar{Y} \subseteq \bar{X}$ *and* $\bar{Z} \in Iob_{\tau\to\tau\to o}(\bar{X})$ *and* $\bar{Y} \cap \bar{Z} \neq \emptyset$,
then $\bar{Z} \in Iob_{\tau\to\tau\to o}(\bar{Y})$.

[4] In the proof we implicitly employ curring and uncurring, and we associate sets with their characteristic functions. This analogously applies to the remainder of this article.

Proof Each statement follows by construction of H^M for M.
(av): By definition of av for $s \in S$ in M, $av(s) \neq \emptyset$; hence, there is $u \in S$ such that $u \in av(s)$. By definition of H^M, $Iav_{i\to\tau}(s,u) = T$, so $u \in Iav_{i\to\tau}(s)$ and hence $Iav_{i\to\tau}(s) \neq \emptyset$ in H^M.
(pv1): By definition of av and pv for $s \in S$ in M, $av(s) \subseteq pv(s)$; hence, for every $u \in av(s)$ we have $u \in pv(s)$. In H^M this means, if $Iav_{i\to\tau}(s,u) = T$, then $Ipv_{i\to\tau}(s,u) = T$. So, $Iav_{i\to\tau}(s) \subseteq Ipv_{i\to\tau}(s)$ in H^M.
(pv2): This case is similar to (av).
(ob1): By definition of ob, we have $\emptyset \notin ob(\bar{X})$; hence, in H^M, $Iob_{\tau\to\tau\to o}(\bar{X}, \emptyset) = F$, that is $\emptyset \notin Iob_{\tau\to\tau\to o}(\bar{X})$.
(ob2): Suppose $\bar{Y} \cap \bar{X} = \bar{Z} \cap \bar{X}$. In M we have $\bar{Y} \in ob(\bar{X})$ iff $\bar{Z} \in ob(\bar{X})$. By definition of H^M we have $Iob_{\tau\to\tau\to o}(\bar{X}, \bar{Y}) = T$ iff $Iob_{\tau\to\tau\to o}(\bar{X}, \bar{Z}) = T$. Hence, $\bar{Y} \in Iob_{\tau\to\tau\to o}(\bar{X})$ iff $\bar{Z} \in Iob_{\tau\to\tau\to o}(\bar{X})$ in H^M.
(ob3): Suppose $\bar{\beta} \subseteq Iob_{\tau\to\tau\to o}(\bar{X})$ and $\bar{\beta} \neq \emptyset$. If $(\cap\bar{\beta}) \cap \bar{X} \neq \emptyset$, by definition of ob in M we have $(\cap\bar{\beta}) \in ob(\bar{X})$. Hence, in H^M, $Iob_{\tau\to\tau\to o}(\bar{X}, (\cap\bar{\beta})) = T$ and then $(\cap\bar{\beta}) \in Iob_{\tau\to\tau\to o}(\bar{X})$.
(ob4) and (ob5) are similar to (ob2). □

Lemma 2 *Let $H^M = \langle\{D_\alpha\}_{\alpha\in T}, I\rangle$ be a Henkin model for a DDL model M. We have $H^M \models^{HOL} \Sigma$ for all $\Sigma \in \{AV, PV1, PV2, OB1, ..., OB5\}$, where*

$$
\begin{array}{lll}
AV & \textit{is} & \forall W_i \exists V_i (av_{i\to\tau} W_i V_i) \\
PV1 & \textit{is} & \forall W_i \forall V_i (av_{i\to\tau} W_i V_i \to pv_{i\to\tau} W_i V_i) \\
PV2 & \textit{is} & \forall W_i (pv_{i\to\tau} W_i W_i) \\
OB1 & \textit{is} & \forall X_\tau \neg ob_{\tau\to\tau\to o} X_\tau (\lambda X_\tau \bot) \\
OB2 & \textit{is} & \forall X_\tau Y_\tau Z_\tau (\ \ (\forall W_i ((Y_\tau W_i \wedge X_\tau W_i) \longleftrightarrow (Z_\tau W_i \wedge X_\tau W_i))) \\
 & & \qquad\qquad\quad \to (ob_{\tau\to\tau\to o} X_\tau Y_\tau \longleftrightarrow ob_{\tau\to\tau\to o} X_\tau Z_\tau)) \\
OB3 & \textit{is} & \forall \beta_{\tau\to\tau\to o} \forall X_\tau \\
 & & (\ \ ((\forall Z_\tau (\beta_{\tau\to\tau\to o} Z_\tau \to ob_{\tau\to\tau\to o} X_\tau Z_\tau)) \wedge \exists Z_\tau (\beta_{\tau\to\tau\to o} Z_\tau)) \\
 & & \quad \to (\ \ (\exists Y_i (((\lambda W_i \forall Z_\tau (\beta_{\tau\to\tau\to o} Z_\tau \to Z_\tau W_i))\, Y_i) \wedge X_\tau Y_i)) \\
 & & \qquad\quad \to ob_{\tau\to\tau\to o} X_\tau (\lambda W_i \forall Z_\tau (\beta_{\tau\to\tau\to o} Z_\tau \to Z_\tau W_i)))) \\
OB4 & \textit{is} & \forall X_\tau Y_\tau Z_\tau \\
 & & (\ \ (\forall W_i (Y_\tau W_i \to X_\tau W_i) \wedge ob_{\tau\to\tau\to o} X_\tau Y_\tau \wedge \forall X_\tau (X_\tau W_i \to Z_\tau W_i)) \\
 & & \quad \to ob_{\tau\to\tau\to o} Z_\tau (\lambda W_i ((Z_\tau W_i \wedge \neg X_\tau W_i) \vee Y_\tau W_i))) \\
OB5 & \textit{is} & \forall X_\tau Y_\tau Z_\tau \\
 & & (\ \ (\forall W_i (Y_\tau W_i \to X_\tau W_i) \wedge ob_{\tau\to\tau\to o} X_\tau Z_\tau \wedge \exists W_i (Y_\tau W_i \wedge Z_\tau W_i)) \\
 & & \quad \to ob_{\tau\to\tau\to o} Y_\tau Z_\tau)
\end{array}
$$

Proof We present detailed arguments for most cases.

AV:

$$
\begin{array}{ll}
 & \text{For all } s \in D_i\text{: } Iav_{i\to\tau}(s) \neq \emptyset \qquad \text{(by Lemma 1 (av))} \\
\Leftrightarrow & \text{For all } s \in D_i\text{, there exists } u \in D_i \text{ such that } Iav_{i\to\tau}(s,u) = T
\end{array}
$$

$\Leftrightarrow$ For all assignments g, for all $s \in D_i$, there exists $u \in D_i$ such that $\|av\, W\, V\|^{H^M, g[s/W_i][u/V_i]} = T$
$\Leftrightarrow$ For all g, all $s \in D_i$ we have $\|\exists V(av\, W\, V)\|^{H^M, g[s/W_i]} = T$
$\Leftrightarrow$ For all g we have $\|\forall W \exists V(av\, W\, V)\|^{H^M, g} = T$
$\Leftrightarrow$ $H^M \models^{\mathrm{HOL}} AV$

PV1:

Given an arbitary assignment g, and arbitary $s, u \in D_i$ such that $\|av\, W\, V\|^{H^M, g[s/W_i][u/V_i]} = T$
$\Leftrightarrow$ $Iav_{i\to\tau}(s,u) = T$
$\Rightarrow$ $Ipv_{i\to\tau}(s,u) = T$ $\quad$ ($Iav_{i\to\tau}(s) \subseteq Ipv_{i\to\tau}(s)$, by Lemma 1 (pv1))
$\Leftrightarrow$ $\|pv\, W\, V\|^{H^M, g[s/W_i][u/V_i]} = T$

Hence by definition of $\|.\|$, for all g, for all $s, u \in D_i$ we have: $\|av\, W\, V\|^{H^M, g[s/W_i][u/V_i]} = T$ implies $\|pv\, W\, V\|^{H^M, g[s/W_i][u/V_i]} = T$
$\Leftrightarrow$ For all g, all $s, u \in D_i$ we have $\|av\, W\, V \to pv\, W\, V\|^{H^M, g[s/W_i][u/V_i]} = T$
$\Leftrightarrow$ For all g, all $s \in D_i$ we have $\|\forall V\ (av\, W\, V \to pv\, W\, V)\|^{H^M, g[s/W_i]} = T$
$\Leftrightarrow$ For all g we have $\|\forall W\, \forall V\ (av\, W\, V \to pv\, W\, V))\|^{H^M, g} = T$
$\Leftrightarrow$ $H^M \models^{\mathrm{HOL}} PV1$

PV2:

This case is analogous to AV.

OB1:

For all $\bar{X} \in D_\tau : \emptyset \notin Iob_{\tau\to\tau\to o}(\bar{X})$ $\quad$ (by Lemma 1 (ob1))
$\Leftrightarrow$ For all g, all $\bar{X} \in D_\tau$ we have $\|\neg ob\, X\, (\lambda X.\bot)\|^{H^M, g[\bar{X}/X_\tau]} = T$
$\Leftrightarrow$ For all g we have $\|\forall X\, \neg(ob\, X\, (\lambda X_\tau \bot))\|^{H^M, g[\bar{X}/X_\tau]} = T$
$\Leftrightarrow$ $H^M \models^{\mathrm{HOL}} OB1$

OB2:

Given an arbitary assignment g, and arbitary $\bar{X}, \bar{Y}, \bar{Z} \in D_\tau$ such that $\|\forall W((Y\, W \wedge X\, W) \longleftrightarrow (Z\, W \wedge X\, W))\|^{H^M, g[\bar{X}/X_\tau][\bar{Y}/Y_\tau][\bar{Z}/Z_\tau]} = T$
$\Leftrightarrow$ For all $s \in D_i$ we have $\|(Y\, W \wedge X\, W) \longleftrightarrow (Z\, W \wedge X\, W)\|^{H^M, g[\bar{X}/X_\tau][\bar{Y}/Y_\tau][\bar{Z}/Z_\tau][s/W_i]} = T$
$\Leftrightarrow$ For all $s \in D_i$ we have $\|Y\, W \wedge X\, W\|^{H^M, g[\bar{X}/X_\tau][\bar{Y}/Y_\tau][\bar{Z}/Z_\tau][s/W_i]} = T$ iff $\|Z\, W \wedge X\, W\|^{H^M, g[\bar{X}/X_\tau][\bar{Y}/Y_\tau][\bar{Z}/Z_\tau][s/W_i]} = T$
$\Leftrightarrow$ For all $s \in D_i$ we have $s \in \bar{Y} \cap \bar{X}$ iff $s \in \bar{Z} \cap \bar{X}$
$\Leftrightarrow$ $\bar{Y} \cap \bar{X} = \bar{Z} \cap \bar{X}$
$\Rightarrow$ $Iob_{\tau\to\tau\to o}(\bar{X}, \bar{Y}) = T$ iff $Iob_{\tau\to\tau\to o}(\bar{X}, \bar{Z}) = T$ $\quad$ (by Lemma 1 (ob2))
$\Leftrightarrow$ $\|ob\, X\, Y)\|^{H^M, g[\bar{X}/X_\tau][\bar{Y}/Y_\tau][\bar{Z}/Z_\tau]} = T$ iff $\|ob\, X\, Z\|^{H^M, g[\bar{X}/X_\tau][\bar{Y}/Y_\tau][\bar{Z}/Z_\tau]} = T$

$\Leftrightarrow \quad \|ob\,X\,Y \longleftrightarrow ob\,X\,Z\|^{H^M,g[\bar{X}/X_\tau][\bar{Y}/Y_\tau][\bar{Z}/Z_\tau]} = T$

Hence, by definition of $\|.\|$, for all g, for all $\bar{X}, \bar{Y}, \bar{Z} \in D_\tau$ we have:

$\|(\forall W(\,((Y\,W \wedge X\,W) \longleftrightarrow (Z\,W \wedge X\,W))$
$\quad \rightarrow (ob\,X\,Y \longleftrightarrow ob\,X\,Z))\|^{H^M,g[\bar{X}/X_\tau][\bar{Y}/Y_\tau][\bar{Z}/Z_\tau]} = T$

$\Leftrightarrow$ For all g we have
$\|\forall XYZ(\forall W(\,((Y\,W \wedge X\,W) \longleftrightarrow (Z\,W \wedge X\,W))$
$\quad \rightarrow (ob\,X\,Y \longleftrightarrow ob\,X\,Z))\|^{H^M,g} = T$

$\Leftrightarrow \quad H^M \models^{\mathrm{HOL}} OB2$

OB3:

Given an arbitary assignment g, and arbitary $\bar{\beta} \in D_{\tau\rightarrow o}, \bar{X} \in D_\tau$ such that
$\|\forall Z(\beta\,Z \rightarrow ob\,X\,Z)\|^{H^M,g[\bar{\beta}/\beta_{\tau\rightarrow o}][\bar{X}/X_\tau]} = T$ and
$\|\exists Z(\beta\,Z)\|^{H^M,g[\bar{\beta}/\beta_{\tau\rightarrow o}]} = T$ and
$\|\exists Y(((\lambda W \forall Z(\beta\,Z \rightarrow Z\,W))\,Y) \wedge X\,Y)\|^{H^M,g[\bar{\beta}/\beta_{\tau\rightarrow o}][\bar{X}/X_\tau]} = T$

$\Leftrightarrow$ For all $\bar{Z} \in D_\tau$ we have
$\|\beta\,Z\|^{H^M,g[\bar{\beta}/\beta_{\tau\rightarrow o}][\bar{X}/X_\tau][\bar{Z}/Z_\tau]} = T$ implies
$\|ob\,X\,Z\|^{H^M,g[\bar{\beta}/\beta_{\tau\rightarrow o}][\bar{X}/X_\tau][\bar{Z}/Z_\tau]} = T$ and
there exists $\bar{Z} \in D_\tau$ such that $\|\beta\,Z\|^{H^M,g[\bar{\beta}/\beta_{\tau\rightarrow o}][\bar{Z}/Z_\tau]} = T$ and
there exists $s \in D_i$ such that
$\|(\lambda W \forall Z(\beta\,Z \rightarrow Z\,W))\,Y \wedge X\,Y\|^{H^M,g[\bar{\beta}/\beta_{\tau\rightarrow o}][\bar{X}/X_\tau][s/Y_i]} = T$

$\Leftrightarrow$ For all $\bar{Z} \in D_\tau$ we have $\bar{Z} \in \beta$ implies $\bar{Z} \in Iob_{\tau\rightarrow\tau\rightarrow o}(\bar{X})$ and
there exists $\bar{Z} \in D_\tau$ such that $\bar{Z} \in \bar{\beta}$ and
there exists $s \in D_i$ such that $s \in \cap\bar{\beta}$ and $s \in \bar{X}$ (**see ***)

$\Leftrightarrow \quad \bar{\beta} \subseteq Iob_{\tau\rightarrow\tau\rightarrow o}(\bar{X})$ and $\bar{\beta} \neq \emptyset$ and $(\cap\bar{\beta}) \cap \bar{X} \neq \emptyset$

$\Rightarrow \quad Iob_{\tau\rightarrow\tau\rightarrow o}(\bar{X}, (\cap\bar{\beta})) = T$ (by Lemma 1 (ob3))

$\Leftrightarrow \quad \|ob\,X\,(\lambda W \forall Z(\beta\,Z \rightarrow Z\,W))\|^{H^M,g[\bar{\beta}/\beta_{\tau\rightarrow o}][\bar{X}/X_\tau]} = T$

Hence by definition of $\|.\|$, for all g, all $\bar{\beta} \in D_{\tau\rightarrow o}$, all $\bar{X} \in D_\tau$ we have:
$\|((\forall Z(\beta\,Z \rightarrow ob\,X\,Z)) \wedge (\exists Z(\beta\,Z)))$
$\rightarrow ((\exists Y(((\lambda W \forall Z(\beta\,Z \rightarrow Z\,W))Y) \wedge X\,Y))$
$\rightarrow ob\,X\,(\lambda W \forall Z(\beta\,Z \rightarrow Z\,W)))\|^{H^M,g[\bar{\beta}/\beta_{\tau\rightarrow o}][\bar{X}/X_\tau]} = T$

$\Leftrightarrow$ For all g, we have
$\|\forall\beta\forall X(((\forall Z(\beta\,Z \rightarrow ob\,X\,Z)) \wedge (\exists Z(\beta\,Z)))$
$\rightarrow ((\exists Y(((\lambda W \forall Z(\beta\,Z \rightarrow Z\,W))Y) \wedge X\,Y))$
$\rightarrow ob\,X\,(\lambda W \forall Z(\beta\,Z \rightarrow Z\,W))))\|^{H^M,g} = T$

$\Leftrightarrow \quad H^M \models^{\mathrm{HOL}} OB3$

Justification *: By definition of $\|.\|$, $\|\lambda W_i \forall Z_\tau(\beta_{\tau\to o} Z_\tau \to Z_\tau W_i)\|^{H^M, g[\bar{\beta}/\beta_{\tau\to o}][\bar{X}/X_\tau][s/Y_i]}$ is denoting the function f from D_i to D_o such that for all $d \in D_i$, $f(d) = \|\forall Z_\tau(\beta_{\tau\to o} Z_\tau \to Z_\tau W_i)\|^{H^M, g[\bar{\beta}/\beta_{\tau\to o}][\bar{X}/X_\tau][s/Y_i][d/W_i]}$. By definition of $\|.\|$, $\|\forall Z_\tau(\beta_{\tau\to o} Z_\tau \to Z_\tau W_i)\|^{H^M, g[\bar{\beta}/\beta_{\tau\to o}][\bar{X}/X_\tau][s/Y_i][d/W_i]} = T$ iff for all $\bar{Z} \in \bar{\beta}$ we have $d \in \bar{Z}$. Thus, f is the characteristic function of the set $\cap\bar{\beta}$. By the Denotatpflicht, which is obeyed in H^M, we know that $f(= \cap\bar{\beta}) \in D_\tau$.

OB4:

Given an arbitary assignment g, and arbitary $\bar{X}, \bar{Y}, \bar{Z} \in D_\tau$ such that
$\|\forall W(Y\,W \to X\,W) \wedge ob\,X\,Y \wedge$
$\forall W(X\,W \to Z\,W)\|^{H^M, g[\bar{X}/X_\tau][\bar{Y}/Y_\tau][\bar{Z}/Z_\tau]} = T$

$\Leftrightarrow$ $\|\forall W(Y\,W \to X\,W)\|^{H^M, g[\bar{X}/X_\tau][\bar{Y}/Y_\tau][\bar{Z}/Z_\tau]} = T$ and
$\|ob\,X\,Y\|^{H^M, g[\bar{X}/X_\tau][\bar{Y}/Y_\tau][\bar{Z}/Z_\tau]} = T$ and
$\|\forall W(X\,W \to Z\,W)\|^{H^M, g[\bar{X}/X_\tau][\bar{Y}/Y_\tau][\bar{Z}/Z_\tau]} = T$

$\Leftrightarrow$ For all $s \in D_i$ we have
($s \in \bar{Y}$ implies $s \in \bar{X}$) and $\bar{Y} \in Iob_{\tau\to\tau\to o}(\bar{X})$ and ($s \in \bar{X}$ implies $s \in \bar{Z}$)

$\Leftrightarrow$ $\bar{Y} \subseteq \bar{X}$ and $\bar{Y} \in Iob_{\tau\to\tau\to o}(\bar{X})$ and $\bar{X} \subseteq \bar{Z}$

$\Rightarrow$ $(\bar{Z} \setminus \bar{X}) \cup \bar{Y} \in Iob_{\tau\to\tau\to o}(\bar{Z})$ (by Lemma 1 (ob4))

$\Leftrightarrow$ $\|ob\,Z\,(\lambda W((Z\,W \wedge \neg X\,W) \vee Y\,W))\|^{H^M, g[\bar{X}/X_\tau][\bar{Y}/Y_\tau][\bar{Z}/Z_\tau]} = T$ (**see ****)

Hence by definition of $\|.\|$ for all g, all $\bar{X}, \bar{Y}, \bar{Z} \in D_\tau$ we have
$\|(\forall W(Y\,W \to X\,W) \wedge ob\,X\,Y \wedge \forall W(X\,W \to Z\,W))$
$\to ob\,Z\,(\lambda W((Z\,W \wedge \neg X\,W) \vee Y\,W))\|^{H^M, g[\bar{X}/X_\tau][\bar{Y}/Y_\tau][\bar{Z}/Z_\tau]} = T$

$\Leftrightarrow$ For all g we have
$\|\forall XYZ((\forall W(Y\,W \to X\,W) \wedge ob\,X\,Y \wedge \forall W(X\,W \to Z\,W))$
$\to ob\,Z\,(\lambda W((Z\,W \wedge \neg X\,W) \vee Y\,W)))\|^{H^M, g} = T$

$\Leftrightarrow$ $H^M \models^{\mathrm{HOL}} OB4$

Justification **: Similar to justification *, we can convince ourselves that $\|\lambda W((Z\,W \wedge \neg X\,W) \vee Y\,W)\|^{H^M, g[\bar{X}/X_\tau][\bar{Y}/Y_\tau][\bar{Z}/Z_\tau][\bar{Z}/Z_\tau]}$ is denoting the characteristic function f of the set $(\bar{Z} \setminus \bar{X}) \cup \bar{Y}$. By the Denotatpflicht, which is obeyed in H^M, we know that $f(= (\bar{Z} \setminus \bar{X}) \cup \bar{Y}) \in D_\tau$.

OB5:

This case is analogous to OB4.

□

Lemma 3 *Let H^M be a Henkin model for a DDL model M. For all DDL formulas δ, arbitrary variable assignments g and worlds s it holds: $M, s \models \delta$ if and only if $\|\lfloor\delta\rfloor\, S_i\|^{H^M, g[s/S_i]} = T$.*

Proof The proof of the lemma is by induction on the structure of δ.
In the base case we have $\delta = p^j$ for some $p^j \in P$:

$$
\begin{array}{ll}
 & \|\lfloor p^j \rfloor S\|^{H^M,g[s/S_i]} = T \\
\Leftrightarrow & \|p^j_\tau S\|^{H^M,g[s/S_i]} = T \\
\Leftrightarrow & Ip^j_\tau(s) = T \\
\Leftrightarrow & s \in V(p^j) \quad \text{(by definition of } H^M\text{)} \\
\Leftrightarrow & M, s \vDash p^j
\end{array}
$$

For proving the inductive cases we apply the induction hypothesis, which is formulated as follows: For all δ' that are structurally smaller than δ, for all assignment g and all s we have $\|\lfloor \delta' \rfloor S\|^{H^M,g[s/S_i]} = T$ if and only if $M, s \vDash \delta'$.

We consider each inductive case in turn:
$\delta = \neg\varphi$:

$$
\begin{array}{ll}
 & \|\lfloor \neg\varphi \rfloor S\|^{H^M,g[s/S_i]} = T \\
\Leftrightarrow & \|(\neg_{\tau\to\tau}\lfloor\varphi\rfloor)S\|^{H^M,g[s/S_i]} = T \\
\Leftrightarrow & \|\neg(\lfloor\varphi\rfloor S)\|^{H^M,g[s/S_i]} = T \qquad \text{(since } (\neg_{\tau\to\tau}\lfloor\varphi\rfloor)S =_{\beta\eta} \neg(\lfloor\varphi\rfloor S)\text{)} \\
\Leftrightarrow & \|\lfloor\varphi\rfloor S\|^{H^M,g[s/S_i]} = F \\
\Leftrightarrow & M, s \nvDash \varphi \quad \text{(by induction hypothesis)} \\
\Leftrightarrow & M, s \vDash \neg\varphi
\end{array}
$$

$\delta = \varphi \vee \psi$:

$$
\begin{array}{ll}
 & \|\lfloor \varphi \vee \psi \rfloor S\|^{H^M,g[s/S_i]} = T \\
\Leftrightarrow & \|(\lfloor\varphi\rfloor \vee_{\tau\to\tau\to\tau} \lfloor\psi\rfloor)S\|^{H^M,g[s/S_i]} = T \\
\Leftrightarrow & \|(\lfloor\varphi\rfloor S) \vee (\lfloor\psi\rfloor S)\|^{H^M,g[s/S_i]} = T \\
 & \qquad\qquad\qquad \text{(since } (\lfloor\varphi\rfloor \vee_{\tau\to\tau\to\tau} \lfloor\psi\rfloor)S =_{\beta\eta} ((\lfloor\varphi\rfloor S) \vee (\lfloor\psi\rfloor S))\text{)} \\
\Leftrightarrow & \|\lfloor\varphi\rfloor S\|^{H^M,g[s/S_i]} = T \text{ or } \|\lfloor\psi\rfloor S)\|^{H^M,g[s/S_i]} = T \\
\Leftrightarrow & M, s \vDash \varphi \text{ or } M, s \vDash \psi \quad \text{(by induction hypothesis)} \\
\Leftrightarrow & M, s \vDash \varphi \vee \psi
\end{array}
$$

$\delta = \Box\varphi$:

$$
\begin{array}{ll}
 & \|\lfloor \Box\varphi \rfloor S\|^{H^M,g[s/S_i]} = T \\
\Leftrightarrow & \|(\lambda X \forall Y(\lfloor\varphi\rfloor Y))S\|^{H^M,g[s/S_i]} = T \\
\Leftrightarrow & \text{For all } a \in D_i \text{ we have } \|\lfloor\varphi\rfloor Y\|^{H^M,g[s/S_i][a/Y_i]} = T \\
\Leftrightarrow & \text{For all } a \in D_i \text{ we have } \|\lfloor\varphi\rfloor Y\|^{H^M,g[a/Y_i]} = T \quad (S \notin free(\lfloor\varphi\rfloor)) \\
\Leftrightarrow & \text{For all } a \in S \text{ we have } M, a \models \varphi \quad \text{(by induction hypothesis)} \\
\Leftrightarrow & M, s \models \Box\varphi
\end{array}
$$

$\delta = \Box_a\varphi$:

$$
\begin{array}{ll}
 & \|\lfloor \Box_a\varphi \rfloor S\|^{H^M,g[s/S_i]} = T \\
\Leftrightarrow & \|\,(\lambda X \forall Y(\neg av\, X\, Y \vee \lfloor\varphi\rfloor Y))\, S\|^{H^M,g[s/S_i]} = T \\
\Leftrightarrow & \text{For all } a \in D_i \text{ we have } \|\neg av\, S\, Y \vee \lfloor\varphi\rfloor Y\|^{H^M,g[s/S_i][a/Y_i]} = T
\end{array}
$$

$\Leftrightarrow$ For all $a \in D_i$ we have $\|av\,S\,Y\|^{H^M,g[s/S][a/Y]} = F$ or
$\|\lfloor\varphi\rfloor Y\|^{H^M,g[s/S_i][a/Y_i]} = T$

$\Leftrightarrow$ For all $a \in D_i$ we have $Iav_{i\to\tau}(s,a) = F$ or
$\|\lfloor\varphi\rfloor Y\|^{H^M,g[a/Y_i]} = T \qquad (S \notin free(\lfloor\varphi\rfloor))$

$\Leftrightarrow$ For all $a \in S$ we have $a \notin av(s)$ or
$M, a \models \varphi$ (by induction hypothesis)

$\Leftrightarrow$ $M, s \models \Box_a\varphi$

$\delta = \Box_p\varphi$.

The argument is analogous to $\delta = \Box_a\varphi$.

$\delta = \bigcirc(\psi/\varphi)$:

$\|\lfloor\bigcirc(\psi/\varphi)\rfloor S\|^{H^M,g[s/S_i]} = T$

$\Leftrightarrow$ $\|(\lambda X(ob\lfloor\psi\rfloor\lfloor\varphi\rfloor))S\|^{H^M,g[s/S_i]} = T$

$\Leftrightarrow$ $\|ob\lfloor\psi\rfloor\lfloor\varphi\rfloor\|^{H^M,g[s/S_i]} = T$

$\Leftrightarrow$ $Iob_{\tau\to\tau\to o}(\|\lfloor\psi\rfloor\|^{H^M,g[s/S_i]})(\|\lfloor\varphi\rfloor\|^{H^M,g[s/S_i]}) = T$

$\Leftrightarrow$ $\|\lfloor\varphi\rfloor\|^{H^M,g[s/S_i]} \in Iob_{\tau\to\tau\to o}(\|\lfloor\psi\rfloor\|^{H^M,g[s/S_i]})$

$\Leftrightarrow$ $V(\varphi) \in Iob_{\tau\to\tau\to o}(V(\psi))$ **(see ***)**

$\Leftrightarrow$ $V(\varphi) \in ob(V(\psi))$

$\Leftrightarrow$ $M, s \models \bigcirc(\psi/\varphi)$

Justification *:** We need to show that $\|\lfloor\varphi\rfloor\|^{H^M,g[s/S_i]}$ is identified with $V(\varphi) = \{s \in S \mid M, s \models \varphi\}$ (analogous for ψ). By induction hypothesis, for all assignment g and world s, we have $\|\lfloor\varphi\rfloor S\|^{H^M,g[s/S_i]} = T$ if and only if $M, s \vDash \varphi$. We expand the details of this equivalence. For all assignment g and all worlds $s \in D_i$ we have

$s \in \|\lfloor\varphi\rfloor\|^{H^M,g[s/S_i]}$ (charact. functions are associated with sets)

$\Leftrightarrow$ $\|\lfloor\varphi\rfloor\|^{H^M,g[s/S_i]}(s) = T$

$\Leftrightarrow$ $\|\lfloor\varphi\rfloor\|^{H^M,g[s/S_i]}(\|S\|^{H,g[s/S_i]}) = T$

$\Leftrightarrow$ $\|\lfloor\varphi\rfloor S\|^{H^M,g[s/S_i]} = T$

$\Leftrightarrow$ $M, s \vDash \varphi$ (induction hypothesis)

$\Leftrightarrow$ $s \in V(\varphi)$

Hence, $s \in \|\lfloor\varphi\rfloor\|^{H^M,g[s/S_i]}$ if and only if $s \in V(\varphi)$. By extensionality we thus know that $\|\lfloor\varphi\rfloor\|^{H^M,g[s/S_i]} = V(\varphi)$. Moreover, since H^M obeys the Denotatpflicht we know that $V(\varphi) \in D_\tau$.

$\delta = \bigcirc_a(\varphi)$:

$\|\lfloor\bigcirc_a(\varphi)\rfloor S\|^{H^M,g[s/S_i]} = T$

$\Leftrightarrow$ $\|(\lambda X(ob\,(av\,X)\lfloor\varphi\rfloor \wedge \exists Y(av\,X\,Y \wedge \neg(\lfloor\varphi\rfloor Y)))S\|^{H^M,g[s/S_i]} = T$

$\Leftrightarrow \quad \|ob\,(av\,S)\lfloor\varphi\rfloor \wedge \exists Y(av\,S\,Y \wedge \neg(\lfloor\varphi\rfloor Y))\|^{H^M,g[s/S_i]} = T$
$\Leftrightarrow \quad \|ob\,(av\,S)\lfloor\varphi\rfloor\|^{H^M,g[s/S_i]} = T \quad$ and
$\quad\quad \|\exists Y(av\,S\,Y \wedge \neg(\lfloor\varphi\rfloor Y))\|^{H^M,g[s/S_i]} = T$
$\Leftrightarrow \quad \|ob\,(av\,S)\lfloor\varphi\rfloor\|^{H^M,g[s/S_i]} = T \quad$ and
there exists $a \in D_i$ such that $\|av\,S\,Y \wedge \neg(\lfloor\varphi\rfloor Y)\|^{H^M,g[s/S_i][a/Y_i]} = T$
$\Leftrightarrow \quad Iob_{\tau\rightarrow\tau\rightarrow o}(\|av\,S\|^{H^M,g[s/S_i]})(\|\lfloor\varphi\|^{H^M,g[s/S_i]}) = T \quad$ and
there exists $a \in D_i$ such that
$\|av\,X\,Y\|^{H^M,g[s/S_i][a/Y_i]} = T$ and $\|\lfloor\varphi\rfloor Y\|^{H^M,g[s/S_i][a/Y_i]} = F$
$\Leftrightarrow \quad \|\lfloor\varphi\|^{H^M,g[s/S_i]} \in Iob_{\tau\rightarrow\tau\rightarrow o}(\|av\,S\|^{H^M,g[s/S_i]}) \quad$ and
there exists $a \in D_i$ such that
$\|av\,X\,Y\|^{H^M,g[s/S_i][a/Y_i]} = T$ and $\|\lfloor\varphi\rfloor Y\|^{H^M,g[s/S_i][a/Y_i]} = F$
$\Leftrightarrow \quad V(\varphi) \in Iob_{\tau\rightarrow\tau\rightarrow o}(\|av\,S\|^{H^M,g[s/S_i]}) \quad$ and **(similar to ***)**
there exists $a \in D_i$ such that
$\|av\,X\,Y\|^{H^M,g[a/Y_i]} = T$ and $\|\lfloor\varphi\rfloor Y\|^{H^M,g[a/Y_i]} = F$
$\Leftrightarrow \quad V(\varphi) \in Iob_{\tau\rightarrow\tau\rightarrow o}(av(s)) \quad$ and **(similar to ***)**
there exists $a \in D_i$ such that
$\|av\,X\,Y\|^{H^M,g[a/Y_i]} = T$ and $\|\lfloor\varphi\rfloor Y\|^{H^M,g[a/Y_i]} = F \quad (S \notin free(\lfloor\varphi\rfloor))$
$\Leftrightarrow \quad V(\varphi) \in ob(av(s)) \quad$ and
there exists $a \in S$ such that
$a \in av(s)$ and $M, a \not\models \varphi$ (by induction hypothesis)
$\Leftrightarrow \quad V(\varphi) \in ob(av(s)) \quad$ and
there exists $a \in S$ such that $a \in av(s)$ and $a \notin V(\varphi)$
$\Leftrightarrow \quad V(\varphi) \in ob(av(s)) \quad$ and
there exists $a \in S$ such that $a \in av(s) \cap V(\neg\varphi)$
$\Leftrightarrow \quad V(\varphi) \in ob(av(s))$ and $av(s) \cap V(\neg\varphi) \neq \emptyset$
$\Leftrightarrow \quad M, s \models \bigcirc_a(\varphi)$

$\delta = \bigcirc_p(\varphi)$:

The argument is analogous to $\delta = \bigcirc_a(\varphi)$. □

Lemma 4 *For every Henkin model $H = \langle\{D_\alpha\}_{\alpha\in T}, I\rangle$ such that $H \models^{HOL} \Sigma$ for all $\Sigma \in \{AV, PV1, PV2, OB1,..., OB5\}$, there exists a corresponding DDL model M. Corresponding means that for all DDL formulas δ and for all assignment g and worlds s, $\|\lfloor\delta\rfloor S\|^{H,g[s/S]} = T$ if and only if $M, s \vDash \delta$.*

Proof Suppose that $H = \langle\{D_\alpha\}_{\alpha\in T}, I\rangle$ is a Henkin model such that $H \models^{\mathrm{HOL}} \Sigma$ for all $\Sigma \in$ {AV, PV1, PV2, OB1,..,OB5}. Without loss of generality, we can assume that the domains of H are denumerable [15]. We construct the corresponding DDL model M as follows:

- $S = D_i$.

- $s \in av(u)$ for $s, u \in S$ iff $Iav_{i\to\tau}(s, u) = T$.
- $s \in pv(u)$ for $s, u \in S$ iff $Ipv_{i\to\tau}(s, u) = T$.
- $\bar{X} \in ob(\bar{Y})$ for $\bar{X}, \bar{Y} \in D_i \longrightarrow D_o$ iff $Iob_{\tau\to\tau\to o}(\bar{X}, \bar{Y}) = T$.
- $s \in V(p^j)$ iff $Ip^j_\tau(s) = T$ for all p^j.

Since $H \models^{\mathrm{HOL}} \Sigma$ for all $\Sigma \in$ {AV, PV1, PV2, OB1, .., OB5}, it is straightforward (but tedious) to verify that *av*, *pv* and *ob* satisfy the conditions as required for a DDL model.

Moreover, the above construction ensures that H is a Henkin model H^M for DDL model M. Hence, Lemma 3 applies. This ensures that for all DDL formulas δ, for all assignment g and all worlds s we have $\|\lfloor\delta\rfloor S\|^{H,g[s/S]} = T$ if and only if $M, s \vDash \delta$. □

Theorem 1 (Soundness and Completeness of the Embedding)

$$\models^{DDL} \varphi \text{ if and only if } \{AV,\ PV1,\ PV2,\ OB1,..,OB5\} \models^{HOL} vld\,\lfloor\varphi\rfloor$$

Proof (Soundness, ←) The proof is by contraposition. Assume $\not\models^{DDL} \varphi$, that is, there is a DDL model $M = \langle S, av, pv, ob, V\rangle$, and world $s \in S$, such that $M, s \not\models \varphi$. Now let H^M be a Henkin model for DDL model M. By Lemma 3, for an arbitrary assignment g, it holds that $\|\lfloor\varphi\rfloor S_i\|^{H^M,g[s/S_i]} = F$. Thus, by definition of $\|.\|$, it holds that $\|\forall S_i(\lfloor\varphi\rfloor S_i)\|^{H^M,g} = \|\text{vld}\,\lfloor\varphi\rfloor\|^{H^M,g} = F$. Hence, $H^M \not\models^{\mathrm{HOL}} \text{vld}\,\lfloor\varphi\rfloor$. Furthermore, $H^M \models^{\mathrm{HOL}} \Sigma$ for all $\Sigma \in$ {AV, PV1, PV2, OB1,. . . ,OB5} by Lemma 2. Thus, {AV, PV1, PV2, OB1,..,OB5} $\not\models^{\mathrm{HOL}} \text{vld}\,\lfloor\varphi\rfloor$.

(Completeness, →) The proof is again by contraposition. Assume {AV, PV1, PV2, OB1,..,OB5} $\not\models^{\mathrm{HOL}} \text{vld}\,\lfloor\varphi\rfloor$, that is, there is a Henkin model $H = \langle\{D_\alpha\}_{\alpha\in T}, I\rangle$ such that $H \models^{\mathrm{HOL}} \Sigma$ for all $\Sigma \in$ {AV, PV1, PV2, OB1,..,OB5}, but $\|\text{vld}\,\lfloor\varphi\rfloor\|^{H,g} = F$ for some assignment g . By Lemma 4, there is a DDL model M such that $M \nvDash \varphi$. Hence, $\not\models^{DDL} \varphi$. □

Theorem 1 characterises DDL as a natural fragment of HOL.

5 Conclusion

A shallow semantical embedding of Carmo and Jones's logic of contrary-to-duty conditionals in classical higher-order logic has been presented, and shown to be faithfull (sound an complete). In addition, it has meanwhile been implemented in the proof assistant Isabelle/HOL (see the appendix). This implementation constitutes the first theorem prover for the logic by

Carmo and Jones that is available to date. The foundational theory for this implementation has been laid in this article.

There is much room for future work. First, experiments could investigate whether the provided implementation already supports non-trivial applications in practical normative reasoning, or whether further emendations and improvements are required. Second, the introduced framework could also be used to systematically analyse the properties of Carmo and Jones's dyadic deontic logic within Isabelle/HOL. Third, analogous to previous work in modal logic [9], the provided framework could be extended to study and support first-order and higher-order variants of the framework.

References

1. P.B. Andrews. Resolution in type theory. *Journal of Symbolic Logic*, 36(3):414–432, 1971.

2. P.B. Andrews. General models and extensionality. *Journal of Symbolic Logic*, 37(2):395–397, 1972.

3. P.B. Andrews. General models, descriptions, and choice in type theory. *Journal of Symbolic Logic*, 37(2):385–394, 1972.

4. P.B. Andrews. Church's type theory. *In: E.N. Zalta (ed.) The Stanford Encyclopedia of Philosophy*, Spring, 2014.

5. C. Benzmüller. Cut-elimination for quantified conditional logic. *Journal of Philosophical Logic*, 46(3):333–353, 2017.

6. C. Benzmüller. Recent successes with a meta-logical approach to universal logical reasoning (extended abstract). In S.A. da Costa Cavalheiro and J.L. Fiadeiro, editors, *Formal Methods: Foundations and Applications - 20th Brazilian Symposium, SBMF 2017, Recife, Brazil, November 29 - December 1, 2017, Proceedings*, volume 10623 of *Lecture Notes in Computer Science*, pages 7–11. Springer, 2017.

7. C. Benzmüller, C. Brown, and M. Kohlhase. Higher-order semantics and extensionality. *Journal of Symbolic Logic*, 69(4):1027–1088, 2004.

8. C. Benzmüller and D. Miller. Automation of higher-order logic. In D.M. Gabbay, J.H. Siekmann, and J. Woods, editors, *Handbook of the History of Logic, Volume 9 — Computational Logic*, pages 215–254. North Holland, Elsevier, 2014.

9. C. Benzmüller and L.C. Paulson. Quantified multimodal logics in simple type theory. *Logica Universalis (Special Issue on Multimodal Logics)*, 7(1):7–20, 2013.

10. J. Carmo and A.J.I. Jones. Deontic logic and contrary-to-duties. In D. M. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic: Volume 8*, pages 265–343. Springer Netherlands, Dordrecht, 2002.

11. J. Carmo and A.J.I. Jones. Completeness and decidability results for a logic of contrary-to-duty conditionals. *J. Log. Comput.*, 23(3):585–626, 2013.

12. A. Church. A set of postulates for the foundation of logic. *Annals of Mathematics*, 33(3):346–366, 1932.

13. A. Church. An unsolvable problem of elementary number theory. *American Journal of Mathematics*, 58(2):354–363, 1936.

14. A. Church. A formulation of the simple theory of types. *Journal of Symbolic Logic*, 5(2):56–68, 1940.

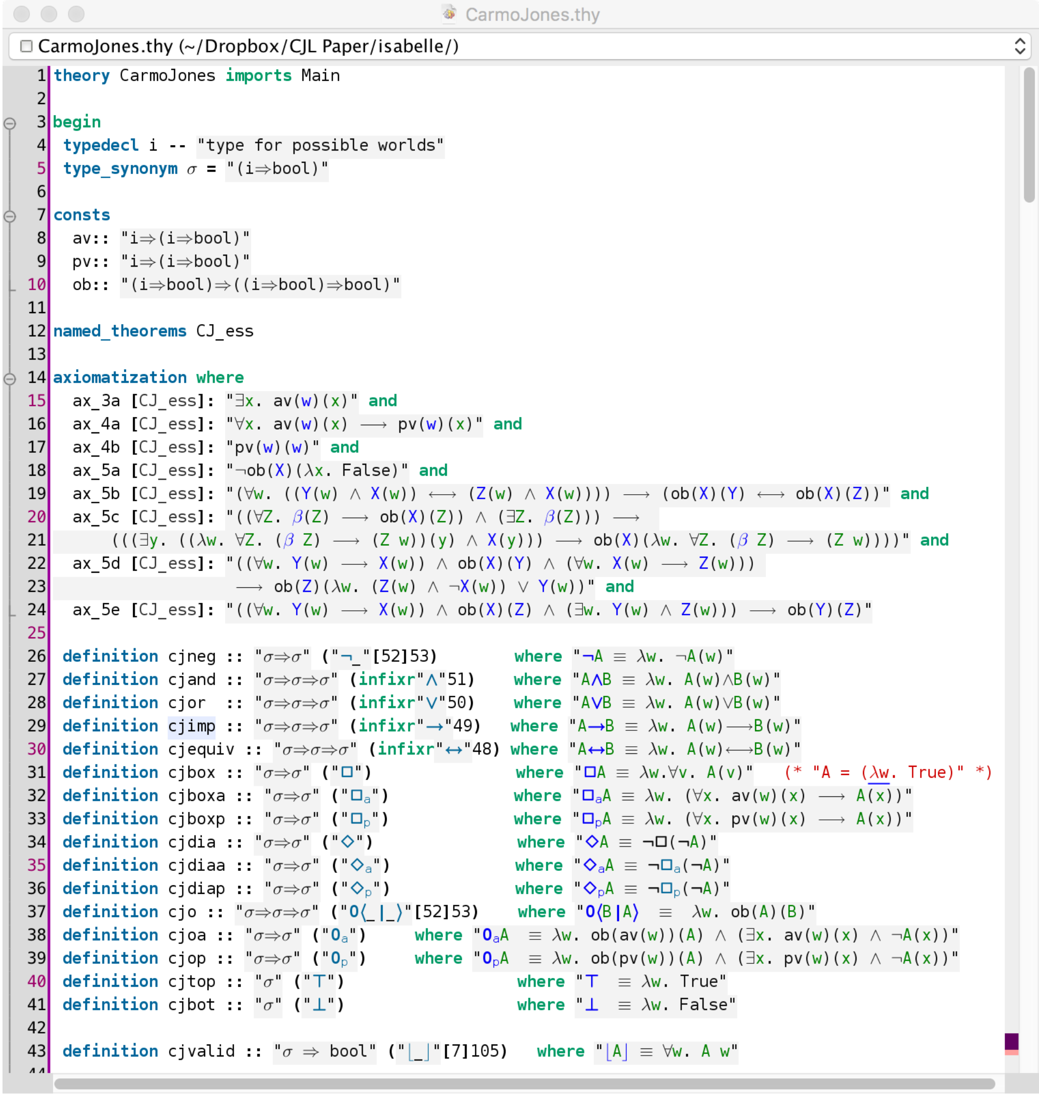
15. L. Henkin. Completeness in the theory of types. *Journal of Symbolic Logic*, 15(2):81–91, 1950.

16. R. Muskens. Intensional models for the theory of types. *Journal of Symbolic Logic*, 75(1):98–118, 2007.

17. T. Nipkow, L.C. Paulson, and M. Wenzel. *Isabelle/HOL — A Proof Assistant for Higher-Order Logic*, volume 2283 of *LNCS*. Springer, 2002.

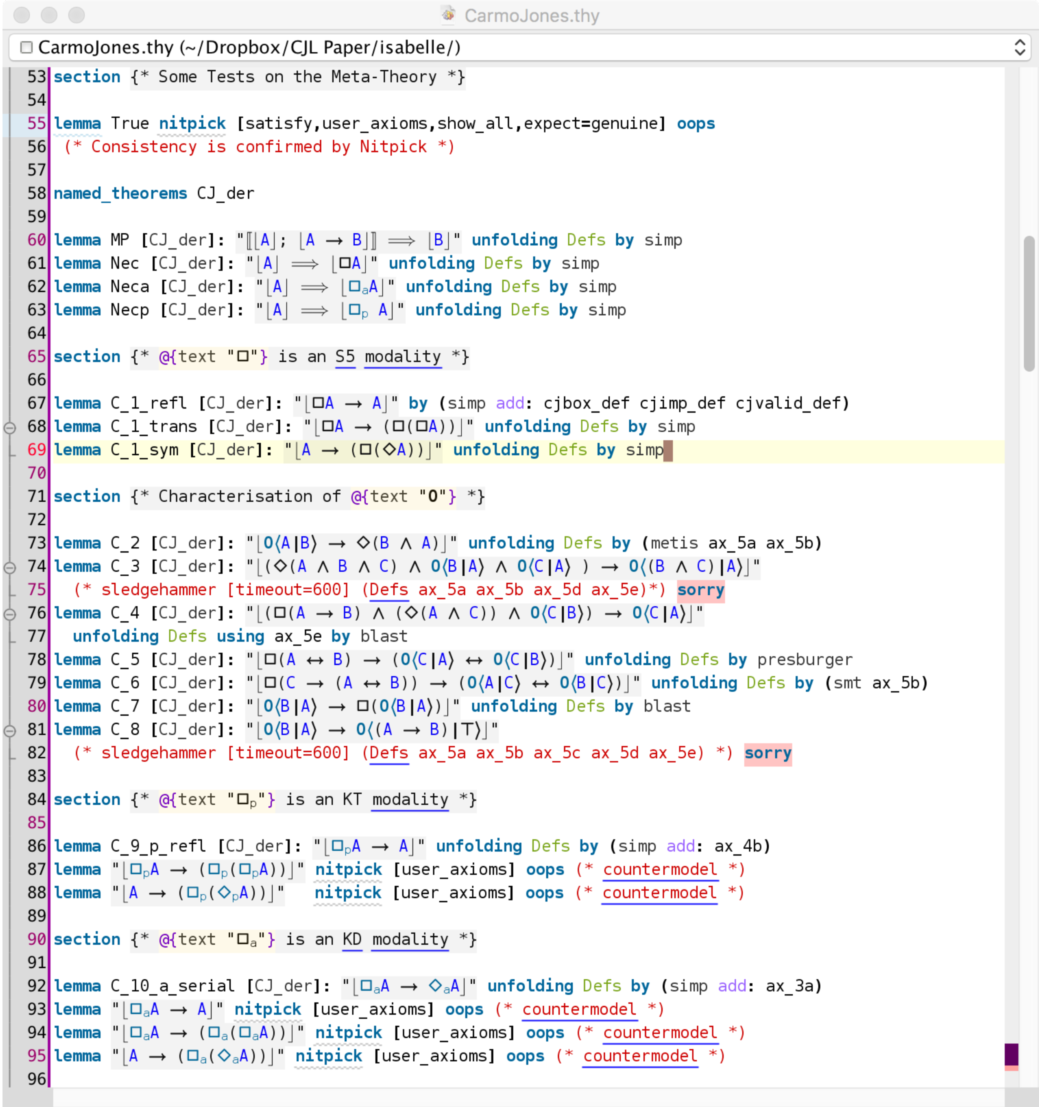
A Implementation in Isabelle/HOL

The semantical embedding as devised in this article has been implemented in the higher-order proof assistant Isabelle/HOL [17]. Figure 1 displays the respective encoding. Figures 2 and 3 report on some experiments.



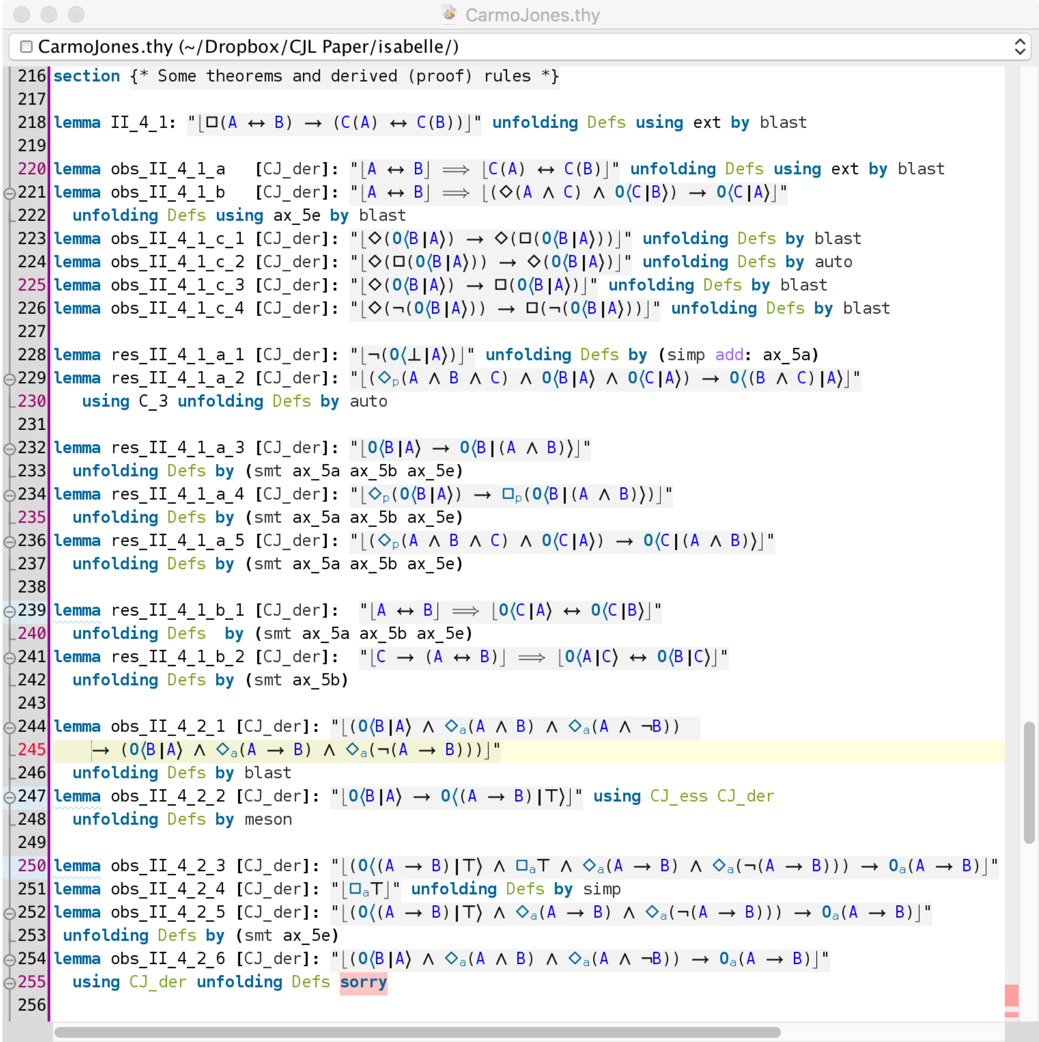
```
theory CarmoJones imports Main

begin
 typedecl i -- "type for possible worlds"
 type_synonym σ = "(i⇒bool)"

consts
  av:: "i⇒(i⇒bool)"
  pv:: "i⇒(i⇒bool)"
  ob:: "(i⇒bool)⇒((i⇒bool)⇒bool)"

named_theorems CJ_ess

axiomatization where
  ax_3a [CJ_ess]: "∃x. av(w)(x)" and
  ax_4a [CJ_ess]: "∀x. av(w)(x) ⟶ pv(w)(x)" and
  ax_4b [CJ_ess]: "pv(w)(w)" and
  ax_5a [CJ_ess]: "¬ob(X)(λx. False)" and
  ax_5b [CJ_ess]: "(∀w. ((Y(w) ∧ X(w)) ⟷ (Z(w) ∧ X(w)))) ⟶ (ob(X)(Y) ⟷ ob(X)(Z))" and
  ax_5c [CJ_ess]: "((∀Z. β(Z) ⟶ ob(X)(Z)) ∧ (∃Z. β(Z))) ⟶
      (((∃y. ((λw. ∀Z. (β Z) ⟶ (Z w))(y) ∧ X(y))) ⟶ ob(X)(λw. ∀Z. (β Z) ⟶ (Z w))))" and
  ax_5d [CJ_ess]: "((∀w. Y(w) ⟶ X(w)) ∧ ob(X)(Y) ∧ (∀w. X(w) ⟶ Z(w)))
                   ⟶ ob(Z)(λw. (Z(w) ∧ ¬X(w)) ∨ Y(w))" and
  ax_5e [CJ_ess]: "((∀w. Y(w) ⟶ X(w)) ∧ ob(X)(Z) ∧ (∃w. Y(w) ∧ Z(w))) ⟶ ob(Y)(Z)"

 definition cjneg :: "σ⇒σ" ("¬_"[52]53)        where "¬A ≡ λw. ¬A(w)"
 definition cjand :: "σ⇒σ⇒σ" (infixr"∧"51)     where "A∧B ≡ λw. A(w)∧B(w)"
 definition cjor  :: "σ⇒σ⇒σ" (infixr"∨"50)     where "A∨B ≡ λw. A(w)∨B(w)"
 definition cjimp :: "σ⇒σ⇒σ" (infixr"→"49)    where "A→B ≡ λw. A(w)⟶B(w)"
 definition cjequiv :: "σ⇒σ⇒σ" (infixr"↔"48) where "A↔B ≡ λw. A(w)⟷B(w)"
 definition cjbox :: "σ⇒σ" ("□")                where "□A ≡ λw.∀v. A(v)"    (* "A = (λw. True)" *)
 definition cjboxa :: "σ⇒σ" ("□a")              where "□aA ≡ λw. (∀x. av(w)(x) ⟶ A(x))"
 definition cjboxp :: "σ⇒σ" ("□p")              where "□pA ≡ λw. (∀x. pv(w)(x) ⟶ A(x))"
 definition cjdia :: "σ⇒σ" ("◇")                 where "◇A ≡ ¬□(¬A)"
 definition cjdiaa :: "σ⇒σ" ("◇a")              where "◇aA ≡ ¬□a(¬A)"
 definition cjdiap :: "σ⇒σ" ("◇p")              where "◇pA ≡ ¬□p(¬A)"
 definition cjo :: "σ⇒σ⇒σ" ("O⟨_|_⟩"[52]53)    where "O⟨B|A⟩  ≡  λw. ob(A)(B)"
 definition cjoa :: "σ⇒σ" ("Oa")      where "OaA  ≡ λw. ob(av(w))(A) ∧ (∃x. av(w)(x) ∧ ¬A(x))"
 definition cjop :: "σ⇒σ" ("Op")      where "OpA  ≡ λw. ob(pv(w))(A) ∧ (∃x. pv(w)(x) ∧ ¬A(x))"
 definition cjtop :: "σ" ("⊤")                   where "⊤  ≡ λw. True"
 definition cjbot :: "σ" ("⊥")                   where "⊥  ≡ λw. False"

 definition cjvalid :: "σ ⇒ bool" ("⌊_⌋"[7]105)   where "⌊A⌋ ≡ ∀w. A w"
```

Fig. 1 Shallow semantical embedding of DDL in Isabelle/HOL.



```
section {* Some Tests on the Meta-Theory *}

lemma True nitpick [satisfy,user_axioms,show_all,expect=genuine] oops
 (* Consistency is confirmed by Nitpick *)

named_theorems CJ_der

lemma MP [CJ_der]: "⟦⌊A⌋; ⌊A → B⌋⟧ ⟹ ⌊B⌋" unfolding Defs by simp
lemma Nec [CJ_der]: "⌊A⌋ ⟹ ⌊□A⌋" unfolding Defs by simp
lemma Neca [CJ_der]: "⌊A⌋ ⟹ ⌊□_aA⌋" unfolding Defs by simp
lemma Necp [CJ_der]: "⌊A⌋ ⟹ ⌊□_p A⌋" unfolding Defs by simp

section {* @{text "□"} is an S5 modality *}

lemma C_1_refl [CJ_der]: "⌊□A → A⌋" by (simp add: cjbox_def cjimp_def cjvalid_def)
lemma C_1_trans [CJ_der]: "⌊□A → (□(□A))⌋" unfolding Defs by simp
lemma C_1_sym [CJ_der]: "⌊A → (□(◇A))⌋" unfolding Defs by simp

section {* Characterisation of @{text "O"} *}

lemma C_2 [CJ_der]: "⌊O⟨A|B⟩ → ◇(B ∧ A)⌋" unfolding Defs by (metis ax_5a ax_5b)
lemma C_3 [CJ_der]: "⌊(◇(A ∧ B ∧ C) ∧ O⟨B|A⟩ ∧ O⟨C|A⟩ ) → O⟨(B ∧ C)|A⟩⌋"
  (* sledgehammer [timeout=600] (Defs ax_5a ax_5b ax_5d ax_5e)*) sorry
lemma C_4 [CJ_der]: "⌊(□(A → B) ∧ (◇(A ∧ C)) ∧ O⟨C|B⟩) → O⟨C|A⟩⌋"
  unfolding Defs using ax_5e by blast
lemma C_5 [CJ_der]: "⌊□(A ↔ B) → (O⟨C|A⟩ ↔ O⟨C|B⟩)⌋" unfolding Defs by presburger
lemma C_6 [CJ_der]: "⌊□(C → (A ↔ B)) → (O⟨A|C⟩ ↔ O⟨B|C⟩)⌋" unfolding Defs by (smt ax_5b)
lemma C_7 [CJ_der]: "⌊O⟨B|A⟩ → □(O⟨B|A⟩)⌋" unfolding Defs by blast
lemma C_8 [CJ_der]: "⌊O⟨B|A⟩ → O⟨(A → B)|⊤⟩⌋"
  (* sledgehammer [timeout=600] (Defs ax_5a ax_5b ax_5c ax_5d ax_5e) *) sorry

section {* @{text "□_p"} is an KT modality *}

lemma C_9_p_refl [CJ_der]: "⌊□_pA → A⌋" unfolding Defs by (simp add: ax_4b)
lemma "⌊□_pA → (□_p(□_pA))⌋" nitpick [user_axioms] oops (* countermodel *)
lemma "⌊A → (□_p(◇_pA))⌋"   nitpick [user_axioms] oops (* countermodel *)

section {* @{text "□_a"} is an KD modality *}

lemma C_10_a_serial [CJ_der]: "⌊□_aA → ◇_aA⌋" unfolding Defs by (simp add: ax_3a)
lemma "⌊□_aA → A⌋" nitpick [user_axioms] oops (* countermodel *)
lemma "⌊□_aA → (□_a(□_aA))⌋" nitpick [user_axioms] oops (* countermodel *)
lemma "⌊A → (□_a(◇_aA))⌋" nitpick [user_axioms] oops (* countermodel *)
```

Fig. 2 Experiments (meta-theory) with the embedding of DDL in Isabelle/HOL. In the "sorry" cases proofs can be automatically found by theorem provers integrated via Sledgehammer, but a reconstruction of these proofs in Isabelle/HOL still fails, since the internal provers are too weak.



```
section {* Some theorems and derived (proof) rules *}

lemma II_4_1: "⌊□(A ↔ B) → (C(A) ↔ C(B))⌋" unfolding Defs using ext by blast

lemma obs_II_4_1_a   [CJ_der]: "⌊A ↔ B⌋ ⟹ ⌊C(A) ↔ C(B)⌋" unfolding Defs using ext by blast
lemma obs_II_4_1_b   [CJ_der]: "⌊A ↔ B⌋ ⟹ ⌊(◇(A ∧ C) ∧ O⟨C|B⟩) → O⟨C|A⟩⌋"
  unfolding Defs using ax_5e by blast
lemma obs_II_4_1_c_1 [CJ_der]: "⌊◇(O⟨B|A⟩) → ◇(□(O⟨B|A⟩))⌋" unfolding Defs by blast
lemma obs_II_4_1_c_2 [CJ_der]: "⌊◇(□(O⟨B|A⟩)) → ◇(O⟨B|A⟩)⌋" unfolding Defs by auto
lemma obs_II_4_1_c_3 [CJ_der]: "⌊◇(O⟨B|A⟩) → □(O⟨B|A⟩)⌋" unfolding Defs by blast
lemma obs_II_4_1_c_4 [CJ_der]: "⌊◇(¬(O⟨B|A⟩)) → □(¬(O⟨B|A⟩))⌋" unfolding Defs by blast

lemma res_II_4_1_a_1 [CJ_der]: "⌊¬(O⟨⊥|A⟩)⌋" unfolding Defs by (simp add: ax_5a)
lemma res_II_4_1_a_2 [CJ_der]: "⌊(◇ₚ(A ∧ B ∧ C) ∧ O⟨B|A⟩ ∧ O⟨C|A⟩) → O⟨(B ∧ C)|A⟩⌋"
   using C_3 unfolding Defs by auto

lemma res_II_4_1_a_3 [CJ_der]: "⌊O⟨B|A⟩ → O⟨B|(A ∧ B)⟩⌋"
  unfolding Defs by (smt ax_5a ax_5b ax_5e)
lemma res_II_4_1_a_4 [CJ_der]: "⌊◇ₚ(O⟨B|A⟩) → □ₚ(O⟨B|(A ∧ B)⟩)⌋"
  unfolding Defs by (smt ax_5a ax_5b ax_5e)
lemma res_II_4_1_a_5 [CJ_der]: "⌊(◇ₚ(A ∧ B ∧ C) ∧ O⟨C|A⟩) → O⟨C|(A ∧ B)⟩⌋"
  unfolding Defs by (smt ax_5a ax_5b ax_5e)

lemma res_II_4_1_b_1 [CJ_der]:  "⌊A ↔ B⌋ ⟹ ⌊O⟨C|A⟩ ↔ O⟨C|B⟩⌋"
  unfolding Defs  by (smt ax_5a ax_5b ax_5e)
lemma res_II_4_1_b_2 [CJ_der]:  "⌊C → (A ↔ B)⌋ ⟹ ⌊O⟨A|C⟩ ↔ O⟨B|C⟩⌋"
  unfolding Defs by (smt ax_5b)

lemma obs_II_4_2_1 [CJ_der]: "⌊(O⟨B|A⟩ ∧ ◇ₐ(A ∧ B) ∧ ◇ₐ(A ∧ ¬B))
    → (O⟨B|A⟩ ∧ ◇ₐ(A → B) ∧ ◇ₐ(¬(A → B)))⌋"
  unfolding Defs by blast
lemma obs_II_4_2_2 [CJ_der]: "⌊O⟨B|A⟩ → O⟨(A → B)|⊤⟩⌋" using CJ_ess CJ_der
  unfolding Defs by meson

lemma obs_II_4_2_3 [CJ_der]: "⌊(O⟨(A → B)|⊤⟩ ∧ □ₐT ∧ ◇ₐ(A → B) ∧ ◇ₐ(¬(A → B))) → Oₐ(A → B)⌋"
lemma obs_II_4_2_4 [CJ_der]: "⌊□ₐT⌋" unfolding Defs by simp
lemma obs_II_4_2_5 [CJ_der]: "⌊(O⟨(A → B)|⊤⟩ ∧ ◇ₐ(A → B) ∧ ◇ₐ(¬(A → B))) → Oₐ(A → B)⌋"
 unfolding Defs by (smt ax_5e)
lemma obs_II_4_2_6 [CJ_der]: "⌊(O⟨B|A⟩ ∧ ◇ₐ(A ∧ B) ∧ ◇ₐ(A ∧ ¬B)) → Oₐ(A → B)⌋"
  using CJ_der unfolding Defs sorry
```

Fig. 3 Further experiments (lemmata and derived rules) with the embedding of DDL in Isabelle/HOL.