Noname manuscript No.
(will be inserted by the editor)

Implementation of the Dyadic Deontic Logic by Carmo and Jones in Isabelle/HOL

Christoph Benzmüller, Ali Farjami,
Xavier Parent

the date of receipt and acceptance should be inserted later

Abstract A shallow semantical embedding of a dyadic deontic logic (by Carmo and Jones) in classical higher-order logic is presented. This embedding has been encoded in Isabelle/HOL, which turns this system into a proof assistant for deontic logic reasoning. The experiments with this environment provide evidence that this logic *implementation* fruitfully enables interactive and automated reasoning at the meta-level and the object-level.

Keywords Dyadic deontic logic; Carmo and Jones logic; Classical higher-order logic; Semantic embedding.

1 Introduction

Normative notions such as obligation and permission are the subject of deontic logics [23], and conditional obligations are addressed in so called dyadic deontic logic. A particular dyadic deontic logic has been proposed by Carmo and Jones (CJL) [16]. This dyadic deontic logic comes with a neighborhood semantics and a weakly complete axiomatization over the class of finite models. Their framework is immune to some well known contrary-to-duty issues which can still be found in many other, related approaches. It is a complex logic and it's automation is open.

Simple type theory [20], aka classical higher-order logic (HOL), can help to better understand semantical issues. The syntax and semantics of HOL are well understood [7]. Isabelle/HOL [28] is automated proof tool for simple type

Christoph Benzmüller, Ali Farjami, Xavier Parent
University of Luxembourg
E-mail: c.benzmueller@fu-berlin.de, ali.farjami@uni.lu, xavier.parent@uni.lu

theory. Isabelle/HOL is a useful tool for studying automate reasoning and computational aspects of logics. We present an *"embedding"* of CJL in HOL. As mentioned in the Handbook of Deontic Logic and Normative Systems [19] deontic logic is not studied with computational tools very well, so the implementation of CJL logic in computational tools is not obvious. In this article, our implementation utilizes the shallow semantical embedding approach that has been put forward by Benzmüller as a pragmatical solution towards universal logic reasoning (see [14,12]).

The contributions of the paper are twofold. First, we present an *"embedding"* of the Carmo-Jones-Logic (CJL) in HOL. This approach [12] uses classical higher-order logic as (universal) meta-logic to specify, in a shallow way, the syntax and semantics of various object logics, in our case CJL. CJL provides and combines modal and conditional operators and it comes with higher-order relational semantics, which makes it a non-trivial object logic to implement in the shallow semantical embedding approach. We prove that the presented embedding is sound and complete.

Second, this embedding has been encoded in Isabelle/HOL to enable experiment with advised infrastructure for deontic reasoning. A particular focus of our experiments is on nested dyadic obligations and we show that nested dyadic obligations in CJL can be eliminated.

2 Carmo and Jones logic

To define the formulas of CJL we start with an infinite countable set of propositional symbols. We consider $\neg$ and $\vee$ as primitive propositional connectives. The set of *CJL formulas* is given as the smallest set of formulas obeying the following conditions. Every atomic formula of CJL is also a CJL formula. Moreover, given any CJL formulas φ and ψ, then $\neg\varphi$ (negation), $\varphi \vee \psi$ (disjunction), $\bigcirc(\psi/\varphi)$ (dyadic deontic operator), $\Box\varphi$ (in all worlds), $\Box_a\varphi$ (in all actual versions of the current world), $\Box_p\varphi$ (in all potential versions of the current worl),$\bigcirc_a(\varphi)$ (monadic deontic operator for actual obligation) and $\bigcirc_p(\varphi)$ (monadic deontic operator for primary obligation) are also CJL formulas. We consider $\top =_{df} \neg q \vee q$, for some propositional symbol q, and $\bot =_{df} \neg\top$

A *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 truth set to each atomic sentence (i.e $V(q) \subseteq S$). $av : S \rightarrow P(S)$ (where $P(S)$ denotes the power set of S) is a function mapping worlds to sets of worlds such that $av(s) \neq \emptyset$. $pv : S \rightarrow P(S)$ is another similar mapping such that $av(s) \subseteq pv(s)$ and $s \in pv(s)$. $ob : P(S) \mapsto P(P(S))$ is a function mapping from set of worlds to sets of sets of worlds such that (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 $(\bar{Y} \in ob(\bar{X})$ iff $\bar{Z} \in ob(\bar{X}))$.
3. Let $\beta \subseteq ob(\bar{X})$ and $\beta \neq \emptyset$, i.e. let be a non-empty set of elements of $ob(\bar{X})$. If $(\cap\beta) \cap \bar{X} \neq \emptyset$ (where $\cap\beta = \{s \in S : \forall \bar{Z} \in \beta \quad s \in \bar{Z}\}$) then $(\cap\beta) \in ob(\bar{X})$
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 $\bar{Z} \in ob(\bar{Y})$.

$av(s)$ denotes the set of actual versions of the world s; $pv(s)$ denotes the set of potential versions of the world s; and $ob(\bar{X})$ denotes the set of propositions which are obligatory in context $\bar{X} \subseteq S$.

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 $V^M(\varphi) = \{s \in S : M, s \models \varphi\}$. In order to simplify the presentation, whenever the model M is obvious from context, we write $V(\varphi)$ instead of $V^M(\varphi)$.

$M, s \models p$ if and only if $s \in V(p)$

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

$M, s \models \varphi \vee \psi$ if and only if $M, s \models \varphi$ or $M, s \models \psi$

$M, s \models \Box\varphi$ if and only if $V(\varphi) = S$

$M, s \models \Box_a\varphi$ if and only if $av(s) \subseteq V(\varphi)$

$M, s \models \Box_p\varphi$ if and only if $pv(s) \subseteq V(\varphi)$

$M, s \models \bigcirc(\psi/\varphi)$ if and only if $V(\psi) \in ob(V(\varphi))$[1]

$M, s \models \bigcirc_a \varphi$ if and only if $V(\varphi) \in ob(av(s))$ and $av(s) \cap V(\neg\varphi) \neq \emptyset$

$M, s \models \bigcirc_a \varphi$ if and only if $V(\varphi) \in ob(av(s))$ and $av(s) \cap V(\neg\varphi) \neq \emptyset$

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

3 Classical Higher-order Logic

In this section we introduce classical higher-order logic (HOL) in order to keep the article self-continued. The text has been adapted from [**?**].

Predicate logic with higher-order quantification was developed first by Frege in his Begriffsschrift [22] and by Russell in his ramified theory of types [30], which was later simplified by others, including Chwistek and Ramsey [29, 21], Carnap, and finally Church [20] in his simple theory of types, also referred to as classical higher-order logic (HOL).

[1] It can be shown [16], using (1.),(2.),(5.) that this this dyadic deontic operator is equivalent to this alternative dyadic deontic operator $M, s \models \bigcirc 1(\psi/\varphi)$ if and only if $V(\varphi) \cap V(\psi) \neq \varnothing$ and $\forall \bar{X}($if $\bar{X} \subseteq V(\varphi)$ and $\bar{X} \cap V(\psi) \neq \varnothing$ then $V(\psi) \in ob(\bar{X}))$.

HOL bases both terms and formulas on simply typed λ-terms and the equality of terms and formulas is given by equality of such 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 [17, 18]. These two advantages are heavily exploited in our embedding of CJL in HOL.

For defining the language HOL, we first introduce the set T of *simple types*: As usual, we assume that T is freely generated from a set of *basic types* $BT \supseteq \{o, i\}$ using the function type constructor $\to$. 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 $(V_\alpha)_{\alpha \in T}$.[2] 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. Every typed constant symbol $c_\alpha \in C_\alpha$ and every typed variable symbol $X_\alpha \in V_\alpha$ are HOL terms of type α. 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 $\alpha \to \beta$. 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.

The above definition encompasses the simply typed λ-calculus. In order to extend this base framework into 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: $\neg_{o \to o} \in C_{o \to o}$ and $\vee_{o \to o \to o} \in C_{o \to o \to o}$ (for each type α). The denotation of these special constant symbols is fixed below according to their intended meaning. HOL is thus 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) and *primitive equality* $=_{\alpha \to \alpha \to \alpha} \in C_{\alpha \to \alpha \to \alpha}$ (for each type α), abbreviated as $=^\alpha$, in the signature. We are not pursuing this here.

Type information as well as brackets may be omitted if obvious from the context. 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: for example, $\varphi \wedge \psi$, $\varphi \to \psi$, $\varphi \longleftrightarrow \psi$, $\top$

[2] For example in section 4 we will assume constant symbols av, pv, ob with types $i \to i \to o$,$i \to i \to o$ and $(i \to o) \to (i \to o) \to o$ as part of the signature

and $\bot$ abbreviate $\neg(\neg\varphi \vee \neg\psi)$, $\neg\varphi \vee \psi$,$(\varphi \to \psi) \wedge (\psi \to \varphi)$,$\varphi \vee \neg\varphi$ and $\neg\top$ respectively.

Equality can actually 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 $s_\alpha \doteq^{\alpha} t_\alpha := \forall P_{\alpha\to o}(\neg Ps \vee Pt)$.

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 free variables of s (i.e., 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 (i.e., we consider α-conversion implicitly). A term s is *closed* if $free(s)$ is empty.

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 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 variable X is not free in 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.

For each simply typed λ-term s there is a unique *β-normal form* (denoted $s{\downarrow_\beta}$) and a unique $\beta\eta$-normal form (denoted $s{\downarrow_{\beta\eta}}$). From this fact we know $s \equiv_\beta t$ ($s \equiv_{\beta\eta} t$) if and only if $s{\downarrow_\beta} \equiv t{\downarrow_\beta}$ ($s{\downarrow_{\beta\eta}} \equiv t{\downarrow_{\beta\eta}}$).

Remember, that formulas are defined as terms of type o. A *non-atomic formula* is any formula whose β-normal form is of the form (cs) or $((cs)t)$ where c is a primitive logical connective. An *atomic formula* is any other formula.

The semantics of HOL is well understood and thoroughly documented in the literature. Here we briefly recapitulate some essential aspects. A more detailed overview can be found in Benzmüller and Miller [10].

Gödel's incompleteness theorem [24] can be extended directly to HOL since second-order quantification can be used to define Peano arithmetic: that is, there is a "true" formula of HOL (or any extension of it) that is not provable. The notion of truth here, however, is that arising from what is called the *standard model* of HOL in which a functional type, say, $\alpha \to \beta$, contains *all* functions from the type α to the type β. Moreover, the type o is assumed to contain exactly two truth values, namely *truth* and *falsehood*.

Henkin [25] introduced a broader notion of *general model* in which a type contains "enough" functions but not necessarily all functions. Henkin then showed soundness and completeness. More precisely, he showed that provability in HOL coincides with truth in all general models (the standard one as well as the non-standard ones).

Andrews [4] provided an improvement on Henkin's definition of general models by replacing the notion that there be enough functions to provide denotations for all formulas of HOL with a more direct means to define general models based on combinatory logic. Andrews [3] points out that Henkin's definition of general model technically was in error since his definition of general models admitted models in which the axiom of functional extensionality does not hold. Andrews then showed that there is a rather direct way to fix that problem by shifting the underlying logical connectives away from the usual Boolean connectives and quantifiers for a type-indexed family of connectives $\{Q_{\tau\to\tau\to o}\}_\tau$ in which $Q_{\tau\to\tau\to o}$ denotes equality at type τ. An indirect solution, which we also employ here, is to presuppose the presence of the identity relations in all domains $D_{\alpha\to\alpha\to o}$, which ensures the existence of unit sets $\{a\} \in D_{\alpha\to o}$ for all elements $a \in D_\alpha$. The existence of these unit sets in turn ensures that Leibniz equality indeed denotes the intended (fully extensional) identity relation.

Thus, Henkin models with Andrews' correction are fully extensional, i.e., they validate the functional and Boolean extensionality axioms. The Boolean extensionality axiom (abbreviated as $\mathcal{B}_o$) is given as

$$\forall A_o \forall B_o (A \longleftrightarrow B) \to A \doteq^o B$$

The infinitely many functional extensionality axioms (abbreviated as $\mathcal{F}_{\alpha\beta}$) are parameterized over $\alpha, \beta \in T$. They are given as

$$\forall F_{\alpha\to\beta} \forall G_{\alpha\to\beta} (\forall X_\alpha F X \doteq^\beta G X) \to F \doteq^{\alpha\to\beta} G$$

The construction of non-functional models has been pioneered by Andrews [2]. In Andrews's so-called v-complexes, which are based on Schütte's semi-valuation method [31], both the functional and the Boolean extensionality principles fail. Assuming β-equality, functional extensionality splits into two weaker and independent principles η ($F \doteq \lambda X F X$, if X is not free in term F) and ξ (from $\forall X.F \doteq G$ infer $\lambda X F \doteq \lambda X G$, where X may occur free in F and G). Conversely, $\beta\eta$-conversion, which is built-in in many modern implementations of HOL, together with ξ implies functional extensionality. Boolean extensionality, however, is independent of any of these principles. A whole landscape of respective notions of models structures between Andrews's v-complexes and Henkin semantics that further illustrate and clarify the above connections has been developed by Benzmüller, Brown and Kohlhase [7,15, 5], and an alternative development and discussion has been contributed by Muskens [27].

The semantics of choice for the remainder of this work 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 $D_{\alpha\to\beta}$ 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 $\neg$, $\vee$, and $\forall$ are always given the standard denotations, cf. below). Moreover, we assume that the domains $D_{\alpha\to\alpha\to o}$ contain the respective identity relations.

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:

1. $\|p_\alpha\|^{M,g} = I(p_\alpha)$
2. $\|X_\alpha\|^{M,g} = g(X_\alpha)$
3. $\|(s_{\alpha\to\beta}\, t_\alpha)_\beta\|^{M,g} = \|s_{\alpha\to\beta}\|^{M,g}(\|t_\alpha\|^{M,g})$
4. $\|(\lambda X_\alpha . s_\beta)_{\alpha\to\beta}\|^{M,g}$ = the function f from D_α to D_β such that $f(d) = \|s_\beta\|^{M,g[d/X_\alpha]}$ for all $d \in D_\alpha$
5. $\|(\neg_{o\to o}\, s_o)_o\|^{M,g} = T$ if and only if $\|s_o\|^{M,g} = F$
6. $\|((\vee_{o\to o\to o}\, s_o)\, t_o)_o\|^{M,g} = T$ if and only if $\|s_o\|^{M,g} = T$ or $\|t_o\|^{M,g} = T$
7. $\|((\wedge_{o\to o\to o}\, s_o)\, t_o)_o\|^{M,g} = T$ if and only if $\|s_o\|^{M,g} = T$ and $\|t_o\|^{M,g} = T$
8. $\|((\rightarrow_{o\to o\to o}\, s_o)\, t_o)_o\|^{M,g} = T$ if and only if if $\|s_o\|^{M,g} = T$ then $\|t_o\|^{M,g} = T$
9. $\|((\longleftrightarrow_{o\to o\to o}\, s_o)\, t_o)_o\|^{M,g} = T$ if and only if $\|s_o\|^{M,g} = T$ iff $\|t_o\|^{M,g} = T$

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 $D_{\alpha\to\beta} \subseteq \{f \mid f : D_\alpha \longrightarrow D_\beta\}$ (for all $\alpha, \beta \in T$) and that the valuation function $\|\cdot\|^{M,g}$ from above is total (i.e., every term denotes). Any standard model is obviously also a Henkin model.

Truth in a model, validity in a model M and validity are defined as usual: An HOL formula α_o is *true* in model M for world s under assignment g if and only if $\|\alpha_o\|^{M,g} = T$; this is also denoted by $M, g \models^{\mathrm{HOL}} \alpha_o$. An HOL formula α_o is called *valid* in M, which is denoted by $M \models^{\mathrm{HOL}} \alpha_o$, if and only if $M, g \models^{\mathrm{HOL}} \alpha_o$ for all assignments g. Moreover, a formula α_o is called *valid*, which we denote by $\models^{\mathrm{HOL}} \alpha_o$, if and only if α_o is valid for all M. Finally, we define $S \models^{\mathrm{HOL}} \alpha_o$ for a set of HOL formulas Σ if and only if $M \models^{\mathrm{HOL}} \alpha_o$ for all models M with $M \models^{\mathrm{HOL}} \beta_o$ for all $\beta_o \in \Sigma$.

4 Modeling CJL as a fragment of HOL

Regarding the particular choice of HOL, we here assume a set of basic types $BT = \{o, i\}$, where o denotes the type of Booleans as before. Without loss of generality, i is now identified with a (non-empty) set of worlds.

CJL formulas are now identified 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. Type $i \to o$ is abbreviated as τ in the remainder. Moreover, we assume that our signature for HOL contains the symbols $av_{i\to\tau}$, $pv_{i\to\tau}$ and $ob_{\tau\to\tau\to o}$

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

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

$\neg_\tau$, $\vee_{\tau\to\tau\to\tau}$, $\Box_{\tau\to\tau}$, $\bigcirc_{\tau\to\tau\to\tau}$, $\Box^a_{\tau\to\tau}$, $\Box^p_{\tau\to\tau}$, $\bigcirc^a_{\tau\to\tau}$ and $\bigcirc^p_{\tau\to\tau}$ realize the CJL connectives in HOL. They abbreviate the following HOL terms:

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

Analyzing the validity of a translated formula $\lfloor\varphi\rfloor$ for a world represented by term t_i corresponds to evaluating the application $(\lfloor\varphi\rfloor\, t_i)$. In line with previous work [8,9,11,12], we define $\mathrm{vld}_{\tau\to o} = \lambda A_\tau \forall S_i (A\, S)$. With this definition, validity of a CJL formula φ in CJL corresponds to the validity of $(\mathrm{vld}\,\lfloor\varphi\rfloor)$ in HOL, and vice versa.

To prove the soundness and completeness of the above embedding, a mapping from CJL models into Henkin models is employed [6].

Definition 1 (Henkin model H^M) Given a CJL model $M = \langle S, av, pv, ob, V\rangle$. Let $p^1, ..., p^m \in PV$, for $m \geq 1$ be propositional symbols and $\lfloor p^j \rfloor = p^j_\tau$ for $j = 1, ..., m$. An 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 and 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, 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 $Iob_{\tau\to\tau\to o}$. Interpretation I is constructed as follows:

1. For $1 \leq i \leq m$, we choose $I(p^j_\tau) \in D_\tau$ such that $I(p^j_\tau)(s) = T$ for all $s \in D_i$ with $s \in V(p^j)$ and $I(p^j_\tau)(s) = F$ otherwise.
2. We choose $Iav_{i\to\tau} \in D_{i\to\tau}$ such that $Iav_{i\to\tau}(s, u) = T$ if $u \in av(s)$ in M and $Iav_{i\to\tau}(s, u) = F$ otherwise.

3. We choose $Ipv_{i\to\tau} \in D_{i\to\tau}$ such that $Ipv_{i\to\tau}(s,u) = T$ if $u \in pv(s)$ in M and $Ipv_{i\to\tau}(s,u) = F$ otherwise.
4. We choose $Iob_{\tau\to\tau\to o} \in D_{\tau\to\tau\to o}$ such that $Iob_{\tau\to\tau\to o}(\bar{X},\bar{Y}) = T$ iff $\bar{Y} \in ob(\bar{X})$ in M and $Iob_{\tau\to\tau\to o}(\bar{X},\bar{Y}) = F$ otherwise.

It is not hard to verify that H^M is a Henkin model.

Lemma 1 *Let H^M be a Henkin model for a CJL model M. For H^M we have (cf. The conditions on CJL models on page 3) :*[3]

(av) $Iav_{i\to\tau}(s) \neq \emptyset$ *for all* $s \in D_i$.
(pv1) $Iav_{i\to\tau}(s) \subseteq Ipv_{i\to\tau}(s)$ *for all* $s \in D_i$.
(pv2) $s \in Ipv_{i\to\tau}(s)$ *for all* $s \in D_i$.
(ob1) : $\emptyset \notin Iob_{\tau\to\tau\to o}(\bar{X})$ *for all* $\bar{X} \in D_\tau$.
(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* $\bar{Z} \in Iob_{\tau\to\tau\to o}(\bar{X})$ *for all* $\bar{X},\bar{Y},\bar{Z} \in D_\tau$.
(ob3) Let $\beta \subseteq Iob_{\tau\to\tau\to o}(\bar{X})$ *and* $\beta \neq \emptyset$, *i.e. let be a non-empty set of elements of* $Iob_{\tau\to\tau\to o}(\bar{X})$. *If* $(\cap\beta) \cap \bar{X} \neq \emptyset$ *(where* $\cap\beta = \{s \in S : \forall \bar{Z} \in \beta \quad s \in \bar{Z}\}$*) then* $(\cap\beta) \in Iob_{\tau\to\tau\to o}(\bar{X})$ *for all* $\bar{X} \in D_\tau$.
(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* $(\bar{Z}\setminus\bar{X}) \cup \bar{Y} \in Iob_{\tau\to\tau\to o}(\bar{Z})$ *for all* $\bar{X},\bar{Y},\bar{Z} \in D_\tau$.
(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* $\bar{Z} \in Iob_{\tau\to\tau\to o}(\bar{Y})$ *for all* $\bar{X},\bar{Y},\bar{Z} \in D_\tau$.

Proof (av): By definition of av for $s \in S$, $av(s) \neq \emptyset$; hence, there is $u \in S$ such that $u \in av(s)$. By definition $Iav_{i\to\tau}(s,u) = T$, so $u \in Iav_{i\to\tau}(s)$ and hence $Iav_{i\to\tau}(s) \neq \emptyset$.
(pv1): By definition of av and pv for $s \in S$, $av(s) \subseteq pv(s)$; hence, for every $u \in av(s)$ we have $u \in pv(s)$. It means, if $Iav_{i\to\tau}(s,u) = T$ then $Ipv_{i\to\tau}(s,u) = T$. So, $Iav_{i\to\tau}(s) \subseteq Ipv_{i\to\tau}(s)$.
(pv2): similar to (av)
(ob1): By definition of ob, we have $\emptyset \notin ob(\bar{X})$; hence, $Iob_{\tau\to\tau\to o}(\bar{X},\emptyset) = F$, 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 $\bar{Z} \in ob(\bar{X})$. It means $Iob_{\tau\to\tau\to o}(\bar{X},\bar{Y}) = T$ 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 $\bar{Z} \in Iob_{\tau\to\tau\to o}(\bar{X})$.
(ob3): Suppose $\beta \subseteq Iob_{\tau\to\tau\to o}(\bar{X})$ and $\beta \neq \emptyset$. If $(\cap\beta) \cap \bar{X} \neq \emptyset$, by definition of ob in M we have $(\cap\beta) \in ob(\bar{X})$. Hence $Iob_{\tau\to\tau\to o}(\bar{X},(\cap\beta)) = T$ and then $(\cap\beta) \in Iob_{\tau\to\tau\to o}(\bar{X})$.
(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 CJL model M. We have* $H^M \models^{HOL} \varphi$ *for* $\varphi \in \Sigma = \{AV, PV1, PV2, OB1, ..., OB5\}$, *where :*

(AV) $\forall W_i.\exists X_i.av_{i\to\tau}(W_i)(X_i)$
(PV1) $\forall W_i.\forall X_i.av_{i\to\tau}(W_i)(X_i) \to pv_{i\to\tau}(W_i)(X_i)$

[3] In the proof we implicitly employ curring and uncurring, and we associate sets with their characteristic functions. This also applies to other pages in the remainder of this article.

(PV2) $\forall W_i.pv_{i\rightarrow\tau}(W_i)(W_i)$
(OB1) $\forall X_\tau.\neg ob_{\tau\rightarrow\tau\rightarrow o}(X_\tau)(\lambda X_\tau.False)$
(OB2) $\forall X_\tau Y_\tau Z_\tau.(\forall W_i.((Y_\tau(W_i)\wedge X_\tau(W_i))\longleftrightarrow(Z_\tau(W_i)\wedge X_\tau(W_i)))\rightarrow (ob_{\tau\rightarrow\tau\rightarrow o}(X_\tau)(Y_\tau)\longleftrightarrow ob_{\tau\rightarrow\tau\rightarrow o}(X_\tau)(Z_\tau))$
(OB3) $\forall\beta_{\tau\rightarrow\tau\rightarrow o}.\forall X_\tau((\forall Z_\tau.\beta_{\tau\rightarrow\tau\rightarrow o}(Z_\tau)\rightarrow ob_{\tau\rightarrow\tau\rightarrow o}(X_\tau)(Z_\tau))\wedge(\exists Z_\tau.\beta_{\tau\rightarrow\tau\rightarrow o}(Z_\tau)))\rightarrow (((\exists Y_i.((\lambda W_i.\forall Z_\tau.(\beta_{\tau\rightarrow\tau\rightarrow o}Z_\tau)\rightarrow(Z_\tau W_i))(Y_i)\wedge X_\tau(Y_i)))\rightarrow ob_{\tau\rightarrow\tau\rightarrow o}(X_\tau)(\lambda W_i.\forall Z_\tau.(\beta_{\tau\rightarrow\tau\rightarrow o}Z_\tau)\rightarrow(Z_\tau W_i))))$
(OB4) $\forall X_\tau Y_\tau Z_\tau.((\forall W_i.Y_\tau(W_i)\rightarrow X_\tau(W_i))\wedge ob_{\tau\rightarrow\tau\rightarrow o}(X_\tau)(Y_\tau)\wedge(\forall X_\tau(W_i)\rightarrow Z_\tau(W_i)))\rightarrow ob_{\tau\rightarrow\tau\rightarrow o}(Z_\tau)(\lambda W_i.(Z_\tau(W_i)\wedge\neg X_\tau(W_i))\vee Y_\tau(W_i)))$
(OB5) $\forall X_\tau Y_\tau Z_\tau.((\forall W_i.Y_\tau(W_i)\rightarrow X_\tau(W_i))\wedge ob_{\tau\rightarrow\tau\rightarrow o}(X_\tau)(Z_\tau)\wedge(\exists W_i.Y_\tau(W_i)\wedge Z_\tau(W_i))\rightarrow ob_{\tau\rightarrow\tau\rightarrow o}(Y_\tau)(Z_\tau))$

Proof (AV): By lemma1 (av) we have $Iav_{i\rightarrow\tau}(s)\neq\emptyset$ in H^M for all $s\in D_i$. Hence, there is $u\in D_i$ such that $Iav_{i\rightarrow\tau}(w,u)=T$. By definition of $\|.\|$ of HOL we have for all variable assignments g, for all $s\in D_i$, there exists $u\in D_i$ such that $\|av_{i\rightarrow\tau}(W_i)(X_i)\|^{H^M,g[s/S_i][u/X_i]}=T$. Hence, by definition of $\|.\|$ for all variable assignments g, for all $s\in D_i$ we have $\|\exists X_i av_{i\rightarrow\tau}(W_i)(X_i)\|^{H^M,g[s/W_i]}=T$. Hence by definition $\|.\|$, for all all variable assignments g we have:

$$\|\forall W_i.\exists X_i av_{i\rightarrow\tau}(W_i)(X_i)\|^{H^M,g}=T$$

Which by definition of $\models^{\mathrm{HOL}}$ means $H^M\models^{\mathrm{HOL}} AV$.
(PV1):Suppose for all variable assignments g, for all $s\in D_i$, for all $u\in D_i$ that is $\|av_{i\rightarrow\tau}(W_i)(X_i)\|^{H^M,g[s/W_i][u/X_i]}=T$. By definition $\|.\|$ we have $Iav_{i\rightarrow\tau}(s,u)=T$. By lemma1 (pv1) we have $Iav_{i\rightarrow\tau}(s)\subseteq Ipv_{i\rightarrow\tau}(s)$. It means if $Iav_{i\rightarrow\tau}(s,u)=T$ then $Ipv_{i\rightarrow\tau}(s,u)=T$. Hence by definition of $\|.\|$, for all variable assignments g, for all $s\in D_i$ and for all $u\in D_i$ we have :

$$\|pv_{i\rightarrow\tau}(W_i)(X_i)\|^{H^M,g[s/W_i][u/X_i]}=T.$$

Hence by definition of $\|.\|$, for all variable assignments g and for all $s\in D_i$ we have:

$$\|\forall X_i.av_{i\rightarrow\tau}(W_i)(X_i)\rightarrow pv_{i\rightarrow\tau}(W_i)(X_i)\|^{H^M,g[s/W_i]}=T$$

Hence, by definition of $\|.\|$, for all variable assignments g we have:

$$\|\forall W_i.\forall X_i.av_{i\rightarrow\tau}(W_i)(X_i)\rightarrow pv_{i\rightarrow\tau}(W_i)(X_i)\|^{H^M,g}=T$$

Which by definition of $\models^{\mathrm{HOL}}$ means $H^M\models^{\mathrm{HOL}} PV1$.
(PV2): is similar to AV.
(OB1): By lemma1 (OB1) we have $\emptyset\notin Iob_{\tau\rightarrow\tau\rightarrow o}(\bar{X})$ in H^M for all $\bar{X}\in D_\tau$. Hence, by definition of $\|.\|$ of HOL we have for all variable assignments g, for all $\bar{X}\in D_\tau$:

$$\|\neg ob_{\tau\rightarrow\tau\rightarrow o}(X_\tau)(\lambda X_\tau.\bot)\|^{H^M,g[\bar{X}/X_\tau]}=T\ .$$

Hence, by definition of $\|.\|$, for all variable assignments g we have:

$$\|\forall X_\tau.\neg ob_{\tau\rightarrow\tau\rightarrow o}(X_\tau)(\lambda X_\tau.\bot)\|^{H^M,g}=T$$

By definition of $\models^{\text{HOL}}$ means $H^M \models^{\text{HOL}} OB1$.
(OB2): Suppose for all variable assignments g, for all $\bar{X}, \bar{Y}, \bar{Z} \in D_\tau$ we have :

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

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

$$\|(Y_\tau(W_i) \wedge X_\tau(W_i)) \longleftrightarrow (Z_\tau(W_i) \wedge X_\tau(W_i))\|^{H^M, g[\bar{X}/X_\tau][\bar{Y}/Y_\tau][\bar{Z}/Z_\tau][s/W_i]} = T$$

By definition $\|.\|$, this means:

$$\|Y_\tau(W_i) \wedge X_\tau(W_i)\|^{H^M, g[\bar{X}/X_\tau][\bar{Y}/Y_\tau][\bar{Z}/Z_\tau][s/W_i]} = T \text{ iff } \|Z_\tau(W_i) \wedge X_\tau(W_i)\|^{H^M, g[\bar{X}/X_\tau][\bar{Y}/Y_\tau][\bar{Z}/Z_\tau][s/W_i]} = T$$

From this we get that $s \in \bar{Y} \cap \bar{X}$ iff $s \in \bar{Z} \cap \bar{X}$ for all $s \in D_i$. So, $\bar{Y} \cap \bar{X} = \bar{Z} \cap \bar{X}$. By Lemma1 (ob2) we have for all $\bar{X}, \bar{Y}, \bar{Z} \in D_\tau$, if $\bar{Y} \cap \bar{X} = \bar{Z} \cap \bar{X}$, then $Iob_{\tau\to\tau\to o}(\bar{X}, \bar{Y}) = T$ iff $Iob_{\tau\to\tau\to o}(\bar{X}, \bar{Z}) = T$. By definition of $\|.\|$ we have:

$$\|ob_{\tau\to\tau\to o}(X_\tau)(Y_\tau)\|^{H^M, g[\bar{X}/X_\tau][\bar{Y}/Y_\tau][\bar{Z}/Z_\tau]} = T \text{ iff } \|ob_{\tau\to\tau\to o}(X_\tau)(Z_\tau)\|^{H^M, g[\bar{X}/X_\tau][\bar{Y}/Y_\tau][\bar{Z}/Z_\tau]} = T$$

Hence, by definition of $\|.\|$ we have:

$$\|ob_{\tau\to\tau\to o}(X_\tau)(Y_\tau) \longleftrightarrow ob_{\tau\to\tau\to o}(X_\tau)(Z_\tau)\|^{H^M, g[\bar{X}/X_\tau][\bar{Y}/Y_\tau][\bar{Z}/Z_\tau]} = T$$

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

$$\|(\forall W_i.((Y_\tau(W_i) \wedge X_\tau(W_i)) \longleftrightarrow (Z_\tau(W_i) \wedge X_\tau(W_i))) \to (ob_{\tau\to\tau\to o}(X_\tau)(Y_\tau) \longleftrightarrow ob_{\tau\to\tau\to o}(X_\tau)(Z_\tau))\|^{H^M, g[\bar{X}/X_\tau][\bar{Y}/Y_\tau][\bar{Z}/Z_\tau]} = T$$

Hence, by definition of $\|.\|$ for all variable assignments g we have :

$$\|\forall X_\tau Y_\tau Z_\tau (\forall W_i.((Y_\tau(W_i) \wedge X_\tau(W_i)) \longleftrightarrow (Z_\tau(W_i) \wedge X_\tau(W_i))) \to (ob_{\tau\to\tau\to o}(X_\tau)(Y_\tau) \longleftrightarrow ob_{\tau\to\tau\to o}(X_\tau)(Z_\tau))\|^{H^M, g} = T.$$

Which by definition of $\models^{\text{HOL}}$ means $H^M \models^{\text{HOL}} OB2$.
(OB3): Suppose for all variable assignments g, for all $\beta \in \beta_{\tau\to o}$, for all $\bar{X} \in D_\tau$ that is :

$$\|\forall Z_\tau.\beta_{\tau\to\tau\to o}(Z_\tau) \to ob_{\tau\to\tau\to o}(X_\tau)(Z_\tau))\|^{H^M, g[\beta/\beta_{\tau\to o}][\bar{X}/X_\tau]} = T \text{ and}$$
$$\|\exists Z_\tau.\beta_{\tau\to\tau\to o}(Z_\tau)\|^{H^M, g[\beta/\beta_{\tau\to o}]} = T \text{ and}$$
$$\|ob_{\tau\to\tau\to o}(X_\tau)(\lambda W_i.\forall Z_\tau.(\beta_{\tau\to\tau\to o} Z_\tau) \to (Z_\tau W_i)))\|^{H^M, g[\beta/\beta_{\tau\to o}][\bar{X}/X_\tau]} = T$$

First: Hence by definition of $\|.\|$, for all variable assignments g, for all $\beta \in \beta_{\tau\to o}$, for all $\bar{X} \in D_\tau$ and for all $\bar{Z} \in D_\tau$ we have:

$$\|\beta_{\tau\to\tau\to o}(Z_\tau) \to ob_{\tau\to\tau\to o}(X_\tau)(Z_\tau))\|^{H^M, g[\beta/\beta_{\tau\to o}][\bar{X}/X_\tau][\bar{Z}/Z_\tau]} = T$$

Again, by definition of $\|.\|$ we have:

$$\text{if } \|\beta_{\tau\to o}(Z_\tau)\|^{H^M, g[\beta/\beta_{\tau\to o}][\bar{X}/X_\tau][\bar{Z}/Z_\tau]} = T \text{ then}$$
$$\|ob_{\tau\to\tau\to o}(X_\tau)(Z_\tau))\|^{H^M, g[\beta/\beta_{\tau\to o}][\bar{X}/X_\tau][\bar{Z}/Z_\tau]} = T.$$

It means if $\bar{Z} \in \beta$ then $\bar{Z} \in Iob_{\tau\to\tau\to o}(X_\tau)$ for all $\bar{Z} \in D_\tau$. So, $\beta \subseteq Iob_{\tau\to\tau\to o}(\bar{X})$.
Second: Hence by definition of $\|.\|$, for all variable assignments g and at least for one $\bar{Z} \in D_\tau$ we have $\|\beta_{\tau\to\tau\to o}(Z_\tau)\|^{H^M, g[\beta/\beta_{\tau\to o}][\bar{Z}/Z_\tau]} = T$. It means exist $\bar{Z} \in \beta$. So, $\beta \neq \emptyset$
Third : The expression $\lambda W_i.\forall Z_\tau.(\beta_{\tau\to\tau\to o} Z_\tau) \to (Z_\tau W_i))$ is an expression of type τ. By definition of $\|.\|$, $\|\lambda W_i.\forall Z_\tau.(\beta_{\tau\to\tau\to o} Z_\tau) \to (Z_\tau W_i))\|^{H^M,g}$ is equal to function f from D_i to D_o such that $f(s) = \|\lambda W_i.\forall Z_\tau.(\beta_{\tau\to\tau\to o} Z_\tau) \to (Z_\tau W_i))\|^{H^M, g[s/W_i]}$ for all $s \in D_i$. Hence by definition of $\|.\|$, for all variable assignments g, for all $s \in D_i$, for $\bar{Z} \in D_\tau$ we have:

$$\|\lambda W_i.\forall Z_\tau.(\beta_{\tau\to\tau\to o} Z_\tau) \to (Z_\tau W_i))\|^{H^M, g[s/W_i][\bar{Z}/Z_\tau]} = T \text{ iff}$$
$$\forall \bar{Z} \in \beta \to s \in \bar{Z}$$

Hence by definition of $\|.\|$, $f(s) = T$ iff $s \in \cap\beta$ for all $s \in D_i$.
So, $\lambda W_i.\forall Z_\tau.(\beta_{\tau\to\tau\to o} Z_\tau) \to (Z_\tau W_i))$ represents $\cap\beta$.
Hence we have $\beta \subseteq Iob_{\tau\to\tau\to o}(\bar{X})$ and $\beta \neq \emptyset$ and $(\cap\beta) \cap \bar{X} \neq \emptyset$. By Lemma1 (ob3) we have $Iob_{\tau\to\tau\to o}(\bar{X}, (\cap\beta)) = T$. Hence by definition of $\|.\|$, for all variable assignments g, for all $\beta \in \beta_{\tau\to o}$,for all $\bar{X} \in D_\tau$ we have:

$$\|ob_{\tau\to\tau\to o}(X_\tau)(\lambda W_i.\forall Z_\tau.(\beta_{\tau\to\tau\to o} Z_\tau) \to (Z_\tau W_i))))\|^{H^M, g[\beta/\beta_{\tau\to o}][\bar{X}/X_\tau]} = T.$$

Hence by definition of $\|.\|$, for all variable assignments g, for all $\beta \in \beta_{\tau\to o}$, for all $\bar{X} \in D_\tau$ we have:

$$\|((\forall Z_\tau.\beta_{\tau\to\tau\to o}(Z_\tau) \to ob_{\tau\to\tau\to o}(X_\tau)(Z_\tau)) \wedge (\exists Z_\tau.\beta_{\tau\to o}(Z_\tau))) \to$$
$$(((\exists Y_i.((\lambda W_i.\forall Z_\tau.(\beta_{\tau\to\tau\to o} Z_\tau) \to (Z_\tau W_i))(Y_i) \wedge X_\tau(Y_i))) \to$$
$$ob_{\tau\to\tau\to o}(X_\tau)(\lambda W_i.\forall Z_\tau.(\beta_{\tau\to o} Z_\tau) \to (Z_\tau W_i))))\|^{H^M, g[\bar{X}/X_\tau][\beta/\beta_{\tau\to o}]} = T$$

Hence by definition of $\|.\|$, for all variable assignments g we have:

$$\|\forall\beta_{\tau\to o}.\forall X_\tau.((\forall Z_\tau.\beta_{\tau\to\tau\to o}(Z_\tau) \to ob_{\tau\to\tau\to o}(X_\tau)(Z_\tau)) \wedge (\exists Z_\tau.\beta_{\tau\to o}(Z_\tau))) \to$$
$$(((\exists Y_i.((\lambda W_i.\forall Z_\tau.(\beta_{\tau\to\tau\to o} Z_\tau) \to (Z_\tau W_i))(Y_i) \wedge X_\tau(Y_i))) \to$$
$$b_{\tau\to\tau\to o}(X_\tau)(\lambda W_i.\forall Z_\tau.(\beta_{\tau\to o} Z_\tau) \to (Z_\tau W_i))))\|^{H^M,g} = T$$

Which by definition of $\models^{\mathrm{HOL}}$ means $H^M \models^{\mathrm{HOL}} OB3$.
(OB4): Suppose for all variable assignments g, for all $\bar{X}, \bar{Y}, \bar{Z} \in D_\tau$ that is :

$$\|(\forall W_i.Y_\tau(W_i) \to X_\tau(W_i)) \wedge ob_{\tau\to\tau\to o}(X_\tau)(Y_\tau) \wedge (\forall W_i.X_\tau(W_i) \to$$
$$Z_\tau(W_i))\|^{H^M, g[\bar{X}/X_\tau][\bar{Y}/Y_\tau][\bar{Z}/Z_\tau]} = T$$

By definition $\|.\|$, for all variable assignments g and for all $\bar{X}, \bar{Y}, \bar{Z} \in D_\tau$ we have:

$$\|\forall W_i.Y_\tau(W_i) \to X_\tau(W_i)\|^{H^M, g[\bar{X}/X_\tau][\bar{Y}/Y_\tau]} = T \text{ and}$$
$$\|ob_{\tau\to\tau\to o}(X_\tau)(Y_\tau)\|^{H^M, g[\bar{X}/X_\tau][\bar{Y}/Y_\tau]} = T \text{ and}$$
$$\|\forall W_i.X_\tau(W_i) \to Z_\tau(W_i)\|^{H^M, g[\bar{X}/X_\tau][\bar{Y}/Y_\tau][\bar{Z}/Z_\tau]} = T.$$

First: for all variable assignments g, for all $\bar{X}, \bar{Y}, \bar{Z} \in D_\tau$ and for all $s \in D_i$ we have: $\|Y_\tau(W_i) \to X_\tau(W_i)\|^{H^M, g[\bar{X}/X_\tau][\bar{Y}/Y_\tau][s/W_i]} = T$. By definition of $\|.\|$ we have:

$$\text{if } \|Y_\tau(W_i)\|^{H^M, g[\bar{X}/X_\tau][\bar{Y}/Y_\tau][s/W_i]} = T \text{ then} \\ \|X_\tau(W_i)\|^{H^M, g[\bar{X}/X_\tau][\bar{Y}/Y_\tau][s/W_i]} = T.$$

It means if $s \in \bar{Y}$ then $s \in \bar{X}$ for all $s \in D_i$. So, $\bar{Y} \subseteq \bar{X}$.
Second: for all variable assignments g, for all $\bar{X}, \bar{Y} \in D_\tau$ we have :

$$\|ob_{\tau\to\tau\to o}(X_\tau)(Y_\tau)\|^{H^M, g[\bar{X}/X_\tau][\bar{Y}/Y_\tau]} = T$$

Hence by definition of $\|.\|$ we have $\bar{Y} \in Iob_{\tau\to\tau\to o}(\bar{X})$.
Third: similar to first we can derive $\bar{X} \subseteq \bar{Z}$.
Hence we have $\bar{Y} \subseteq \bar{X}$ and $\bar{Y} \in Iob_{\tau\to\tau\to o}(\bar{X})$. By Lemma1 (ob4) we have $(\bar{Z} \setminus \bar{X}) \cup \bar{Y} \in Iob_{\tau\to\tau\to o}(\bar{Z})$.
The expression $\lambda W_i.(Z_\tau(W_i) \wedge \neg X_\tau(W_i)$ is an expression of type τ. By definition of $\|.\|$, $\|\lambda W_i.(Z_\tau(W_i) \wedge \neg X_\tau(W_i)\|^{H^M, g}$ for all $s \in D_i$ is equal to function f from D_i to D_o such that $f(s) = \|\lambda W_i.(Z_\tau(W_i) \wedge \neg X_\tau(W_i)\|^{H^M, g[s/W_i]}$ for all $s \in D_i$. Hence by definition of $\|.\|$, for all variable assignments g, for $\bar{X}, \bar{Y}, \bar{Z} \in D_\tau$ and for all $s \in D_i$ we have:

$$\|\lambda W_i.(Z_\tau(W_i) \wedge \neg X_\tau(W_i)\|^{H^M, g[s/W_i][\bar{X}/X_\tau][\bar{Y}/Y_\tau][\bar{Z}/Z_\tau]} = T \text{ iff} \\ s \in (\bar{Z} \setminus \bar{X}) \cup \bar{Y}.$$

Hence, the expression $\lambda W_i.(Z_\tau(W_i) \wedge \neg X_\tau(W_i)$ represents $(\bar{Z} \setminus \bar{X}) \cup \bar{Y}$.
By definition of $\|.\|$ for all variable assignments g, for all $\bar{X}, \bar{Y}, \bar{Z} \in D_\tau$ we have :

$$\|ob_{\tau\to\tau\to o}(Z_\tau)(\lambda W_i.(Z_\tau(W_i) \wedge \neg X_\tau(W_i)) \vee Y_\tau(W_i)))\|^{H^M, g[\bar{X}/X_\tau][\bar{Y}/Y_\tau][\bar{Z}/Z_\tau]} = T$$

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

$$\|(\forall W_i.Y_\tau(W_i) \to X_\tau(W_i)) \wedge ob_{\tau\to\tau\to o}(X_\tau)(Y_\tau) \wedge (\forall X_\tau(W_i) \to Z_\tau(W_i))) \to \\ ob_{\tau\to\tau\to o}(Z_\tau)(\lambda W_i.(Z_\tau(W_i) \wedge \neg X_\tau(W_i)) \vee Y_\tau(W_i))\|^{H^M, g[\bar{X}/X_\tau][\bar{Y}/Y_\tau][\bar{Z}/Z_\tau]} = T$$

Hence by definition of $\|.\|$ for all variable assignments g we have :

$$\|\forall X_\tau Y_\tau.((\forall W_i.Y_\tau(W_i) \to X_\tau(W_i)) \wedge ob_{\tau\to\tau\to o}(X_\tau)(Y_\tau) \wedge (\forall W_i.X_\tau(W_i) \to \\ Z_\tau(W_i))) \to ob_{\tau\to\tau\to o}(Z_\tau)(\lambda W_i.(Z_\tau(W_i) \wedge \neg X_\tau(W_i)) \vee Y_\tau(W_i)))\|^{H^M, g} = T$$

Which by definition of $\models^{\text{HOL}}$ means $H^M \models^{\text{HOL}} OB4$.
(OB5): is similar to (OB4).

Lemma 3 *Let H^M be a Henkin model for a CJL model M. For all CJL formulas δ, arbitrary variable assignment 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 is by induction on the structure of δ. (We note $\lfloor\delta\rfloor$ as δ_τ). As an inductive hypothesis we make the following assumption:
For all CJL formulas φ of degree of complexity less than δ, such that for every variable assignment g and worlds s we have :

$$M, s \models \varphi \text{ if and only if } \|\lfloor\varphi\rfloor\, S_i\|^{H^M, g[s/S_i]} = T.$$

The cases for $\delta = p^j$: $M, s \vDash p^j$ if and only if $s \in V(p^j)$ in M. Due to the definition of I in H^M, $I(p^j_\tau)(s) = T$. By definition of $\lfloor\cdot\rfloor$, g and H^M holds $\|\lfloor p^j\rfloor\, S_i\|^{H^M, g[s/S_i]} = \|p^j_\tau\, S_i\|^{H^M, g[s/S_i]} = T$.
For the other direction suppose $\|\lfloor p^j\rfloor\, S_i\|^{H^M, g[s/S_i]} = T$. By definition of $\lfloor\cdot\rfloor$, holds $\|p^j_\tau\, S_i\|^{H^M, g[s/S_i]} = T$. Due to the definition of I in H^M this is equivalent to $s \in V(p^j)$ in M, which means $M, s \vDash p^j$.
The case for $\delta = \varphi \vee \psi$: In model M, $M, s \models \varphi \vee \psi$ if and only if $M, s \models \varphi$ or $M, s \models \psi$. Hence, by induction hypothesis we have $\|\lfloor\varphi\rfloor\, S_i\|^{H^M, g[s/S_i]} = T$ or $\|\lfloor\psi\rfloor\, S_i\|^{H^M, \lfloor g\rfloor[s/S_i]} = T$. By definition of $\lfloor\cdot\rfloor$ and since $((\lfloor\varphi \vee \psi\rfloor)S_i) =_{\beta\eta} ((\lfloor\varphi\rfloor S_i) \vee (\lfloor\psi\rfloor)S_i)$ it holds $\|\lfloor\varphi \vee \psi\rfloor\, S_i\|^{H^M, g[s/S_i]} = \|\lfloor\varphi\rfloor \vee \lfloor\psi\rfloor, S_i\|^{H^M, \lfloor g\rfloor[s/S_i]} = T$.
For the other direction suppose $\|\lfloor\varphi \vee \psi\rfloor\, S_i\|^{H^M, g[s/S_i]} = T$. By definition of $\lfloor\cdot\rfloor$ we have $\|\lfloor\varphi \vee \psi\rfloor\, S_i\|^{H^M, g[s/S_i]} = \|\lfloor\varphi\rfloor\ \vee \lfloor\psi\rfloor, S_i\|^{H^M, \lfloor g\rfloor[s/S_i]} = T$. Hence, by definition of $\|.\|$, $\|\lfloor\varphi\rfloor\, S_i\|^{H^M, g[s/S_i]} = T$ or $\|\lfloor\psi\rfloor\, S_i\|^{H^M, \lfloor g\rfloor[s/S_i]} = T$. Hence, by induction hypothesis we must have: $M, s \models \varphi$ or $M, s \models \psi$. So, $M, s \models \varphi \vee \psi$.
The case for $\delta = \neg\varphi$ is similar to case $\delta = \varphi \vee \psi$.
The case for $\delta = \Box\varphi$: We have $M, s \models \Box\varphi$ if and only if for all $u \in S$, $M, u \models \varphi$. The latter condition is equivalent by induction hypothesis to this one: for all $u \in S$ we have $\|\varphi_\tau\, X_i\|^{H^M, g[u/X_i]} = T$. That is equivalent to $\|\forall X_i . \varphi_\tau\, X_i\|^{H^M, g} = T$. That means, by definition of $\Box$, $\lfloor\cdot\rfloor$ and $\lambda-$conversion that is $\|\lfloor\Box\varphi\rfloor S_i\|^{H^M, g[s/S_i]} = T$.
For the other direction suppose $\|\lfloor\Box\varphi\rfloor S_i\|^{H^M, g[s/S_i]} = T$. By definition of $\lfloor\cdot\rfloor$, holds $\|(\lambda X_i . \forall Y_i(\varphi_\tau Y_i))S_i\|^{H^M, g[s/S_i]} = T$. Hence, by definition of $\|.\|$ and $\lambda-$conversion, for all $u \in S$ we have $\|\varphi_\tau\, X_i\|^{H^M, g[s/S_i][u/X_i]} = T$. So, from $\|\varphi_\tau\, X_i\|^{H^M, g^{'}[u/X_i]} = T$ where $g^{'} = g[s/S_i]$ by induction hypothesis for all $u \in S$, $M, u \models \varphi$. Hence, by definition of $\Box$ in M, $M, s \models \Box\varphi$.
The case for $\delta = \Box_a\varphi$: We have $M, s \models \Box_a\varphi$ if and only if for all u with $u \in av(s)$ we must have $M, u \models \varphi$. The latter condition is equivalent by induction hypothesis to this one: for all u with $u \in av(s)$ we have $\|\varphi_\tau\, X_i\|^{H^M, g[u/X_i]} = T$. That is equivalent to

$$\|\neg(av_{i\to\tau}(S_i, X_i)) \vee (\varphi_\tau\, X_i)\|^{H^M, g[u/X_i][s/S_i]} = T$$

and thus to

$$\|\forall Y_i(\neg(av_{i\to\tau}(S_i, Y_i)) \vee (\varphi_\tau\, Y_i))\|^{H^M, g[s/S_i]} =$$
$$\|\Box^a_{\tau\to\tau}\varphi_\tau S_i\|^{H^M, g[s/S_i]} = \|\lfloor\Box_a\varphi\rfloor S_i\|^{H^M, g[s/S_i]} = T.$$

For the other direction suppose $\|\lfloor\Box_a\varphi\rfloor S_i\|^{H^M, g[s/S_i]} = T$. By definition of $\lfloor\cdot\rfloor$, holds $\|(\lambda X_i . \forall Y_i(\neg(av_{i\to\tau}(S_i, Y_i)) \vee (\varphi_\tau\, Y_i)))S_i\|^{H^M, g[s/S_i]} = T$. Hence, by

definition of $\|.\|$ and $\lambda-$conversion, for all $u \in S$ we have $\|\neg(av_{i\to\tau}(S_i, X_i)) \vee (\varphi_\tau\, X_i)\|^{H^M, g[u/X_i][s/S_i]} = T$. Hence, for all $u \in av(s)$ we have $\|\varphi_\tau\, X_i\|^{H^M, g[u/X_i]} = T$. Hence, by induction hypothesis for all $u \in av(s)$ we have $M, u \models \varphi$. Hence, by definition of $\Box_a$ in M, $M, s \models \Box_a\varphi$.
The case for $\delta = \Box_p\varphi$ is similar to case $\delta = \Box_a\varphi$.
The case for $\delta = \bigcirc(\psi/\varphi)$, we have $M, s \models \bigcirc(\psi/\varphi)$ if and only if $V(\psi) \in ob(V(\varphi))$. By definition of I, for $Iob_{\tau\to\tau\to o} \in D_{\tau\to\tau\to o}$ we have $Iob_{\tau\to\tau\to o}(\varphi_\tau, \psi_\tau) = T$. By definition of $\|.\|$, $\|ob_{\tau\to\tau\to o}(\varphi_\tau, \psi_\tau)\|^{H^M, g} = T$. Also, by $\beta-$reduction we have:

$$\|(\lambda X_i.ob_{\tau\to\tau\to o}(\varphi_\tau, \psi_\tau))S_i\|^{H^M, g} =_\beta \|ob_{\tau\to\tau\to o}(\varphi_\tau, \psi_\tau)\|^{H^M, g}$$

That is equivalent to

$$\|(\lambda X_i.ob_{\tau\to\tau\to o}(\varphi_\tau, \psi_\tau))S_i\|^{H^M, g} = T.$$

Hence, for all $s \in D_i$ we have:

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

For the other direction suppose $\|ob_{\tau\to\tau\to o}(\varphi_\tau, \psi_\tau)S_i\|^{H^M, g[s/S_i]} = T$. So, by definition of I, $Iob_{\tau\to\tau\to o}(\varphi_\tau, \psi_\tau) = T$. Hence, by definition of I we have $V(\psi) \in ob(V(\varphi))$. So, $M, s \models \bigcirc(\psi/\varphi)$.
The case for $\delta = \bigcirc_a(\varphi)$, we have $M, s \models \bigcirc_a(\varphi)$ if and only if $V(\varphi) \in ob(av(s))$ and $av(s) \cap V(\neg\varphi) \neq \emptyset$. By induction hypothesis we have $\|ob_{\tau\to\tau\to o}(av_{i\to\tau}(S_i), \varphi_\tau)S_i\|^{H^M, g[s/S_i]} = T$ and for $u \in av(s) \cap V(\neg\varphi)$, $\|av_{i\to\tau}(S_i, X_i)\|^{H^M, g[s/S_i][u/X_i]} = T$ and also $\|\varphi_\tau X_i\|^{H^M, g[u/X_i]} = T$. Since $((\lfloor\varphi \wedge \psi\rfloor)S_i) =_{\beta\eta} ((\lfloor\varphi\rfloor S_i) \wedge (\lfloor r\rfloor)S_i)$ we can derive:

$$\|ob_{\tau\to\tau\to o}(av_{i\to\tau}(S_i), \varphi_\tau)S_i \wedge av_{i\to\tau}(S_i, X_i) \wedge \varphi_\tau X_i\|^{H^M, g[s/S_i][u/X_i]} = T$$

That is equivalent to

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

For the other direction suppose $\|\lfloor\bigcirc_a(\varphi)\rfloor S_i\|^{H^M, g[s/S_i]} = T$. By definition of $\lfloor\cdot\rfloor$, holds

$$\|(\lambda X_i.ob_{\tau\to\tau\to o}(av_{i\to\tau}(X_i), \varphi_\tau) \wedge \exists Y_i(av_{i\to\tau}(X_i, Y_i) \wedge \varphi_\tau Y_i))S_i\|^{H^M, g[s/S_i]} = T$$

So for some $u \in D_i$ we have:

$$\|ob_{\tau\to\tau\to o}(av_{i\to\tau}(S_i), \varphi_\tau)S_i \wedge av_{i\to\tau}(S_i, X_i) \wedge \varphi_\tau X_i\|^{H^M, g[s/S_i][u/X_i]} = T$$

Hence, by induction hypothesis from $\|ob_{\tau\to\tau\to o}(av_{i\to\tau}(S_i), \varphi_\tau)S_i\|^{H^M, g[s/S_i]} = T$ by definition of I we must have $V(\varphi) \in ob(av(s))$, and from $\|av_{i\to\tau}(S_i, X_i)\|^{H^M, g[s/S_i][u/X_i]} = T$ and $\|\varphi_\tau X_i\|^{H^M, g[u/X_i]} = T$ we must have $u \in av(s) \cap V(\neg\varphi)$. Hence, by definition of $\bigcirc_a(\varphi)$ in M, $M, s \models \bigcirc_a(\varphi)$.
The case for $\delta = \bigcirc_p(\varphi)$ is similar to case $\delta = \bigcirc_a(\varphi)$.

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

Proof Throughout the proof whenever possible we omit types in order to avoid making the notation too cumbersome. Suppose that $H = \langle \{D_\alpha\}_{\alpha\in T}, I\rangle$ is a Henkin model such that $H \models^{\mathrm{HOL}} \Sigma \in \{\mathrm{AV, PV1, PV2, OB1,..,OB5}\}$. We construct the corresponding CJL 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 $\bar{X}, \bar{Y} \in D_\tau$ iff $Iob_{\tau\to\tau\to o}(\bar{X}, \bar{Y}) = T$.
- $s \in V(p)$ iff $I(p_\tau)(s) = T$ for all p.

Since $H \models^{\mathrm{HOL}} \Sigma$ where $\Sigma \in \{\mathrm{AV, PV1, PV2, OB1, .., OB5}\}$, the components av, pv and ob all satisfy the required constraints.

The proof of the second half of the lemma is by induction on the structure of δ. We start with the case where δ is p. We have $\|\lfloor p\rfloor S\|^{H,g[s/S]} = T$

$$\begin{aligned}
&\Leftrightarrow \|p_\tau S\|^{H,g[s/S]} = T\\
&\Leftrightarrow I(p_\tau)(s) = T\\
&\Leftrightarrow s \in V(p)
\end{aligned}$$

For the inductive cases we make the hypothesis that the claim holds for sentences δ' shorter than δ:

Inductive hypothesis: For all assignment g and state s, $\|\lfloor\delta'\rfloor S\|^{H,g[s/S]} = T$ if and only if $M, s \vDash \delta'$

We consider each inductive case in turn:

(a) $\delta = \varphi \vee \psi$. In this case:

$$\begin{aligned}
\|\lfloor\varphi\vee\psi\rfloor S\|^{H,g[s/S]} = T &\Leftrightarrow \|(\lfloor\varphi\rfloor \vee_{\tau\to\tau\to\tau} \lfloor\psi\rfloor)S\|^{H,g[s/S]} = T\\
&\Leftrightarrow \|(\lfloor\varphi\rfloor S) \vee (\lfloor\psi\rfloor S)\|^{H,g[s/S]} = T\\
&\Leftrightarrow \|\lfloor\varphi\rfloor S\|^{H,g[s/S]} = T \text{ or } \|\lfloor\psi\rfloor S)\|^{H,g[s/S]} = T\\
&\Leftrightarrow M, s \vDash \varphi \text{ or } M, s \vDash \psi\\
&\Leftrightarrow M, s \vDash \varphi \vee \psi
\end{aligned}$$

(b) $\delta = \neg\varphi$. In this case:

$$\begin{aligned}
\|\lfloor\neg\varphi\rfloor S\|^{H,g[s/S]} = T &\Leftrightarrow \|(\neg_{\tau\to\tau}\lfloor\varphi\rfloor)S\|^{H,g[s/S]} = T\\
&\Leftrightarrow \|\neg(\lfloor\varphi\rfloor)S)\|^{H,g[s/S]} = T\\
&\Leftrightarrow \|\lfloor\varphi\rfloor S\|^{H,g[s/S]} = F
\end{aligned}$$

$$\Leftrightarrow M, s \not\models \varphi$$
$$\Leftrightarrow M, s \models \neg\varphi$$

(c) $\delta = \Box\varphi$. We have the following chain of equivalences:

$$\begin{aligned}
&\|\lfloor\Box\varphi\rfloor S\|^{H,g[s/S]} = T \\
&\quad\Leftrightarrow \|(\lambda X \forall Y.\lfloor\varphi\rfloor Y)S\|^{H,g[s/S]} = T \\
&\quad\Leftrightarrow \text{ For all } a \in D_i : \|\lfloor\varphi\rfloor Y\|^{H,g[s/S][a/Y]} = T \\
&\quad\Leftrightarrow \text{ For all } a \in S : M, a \models \varphi \\
&\quad\Leftrightarrow M, s \models \Box\varphi
\end{aligned}$$

(d) $\delta = \Box_a\varphi$. We have the following chain of equivalences:

$$\begin{aligned}
&\|\lfloor\Box_a\varphi\rfloor S\|^{H,g[s/S]} = T \\
&\quad\Leftrightarrow \|\,(\lambda X \forall Y.\neg av_{i\to\tau}(X)(Y) \vee \lfloor\varphi\rfloor Y)\, S\|^{H,g[s/S]} = T \\
&\quad\Leftrightarrow \text{ For all } a \in D_i : \|\neg av_{i\to\tau}(S)(Y) \vee \lfloor\varphi\rfloor Y\|^{H,g[s/S][a/Y]} = T \\
&\quad\Leftrightarrow \text{ For all } a \in D_i : \|av_{i\to\tau}(S)(Y)\|^{H,g[s/S][a/Y]} = F \text{ or } \|\lfloor\varphi\rfloor Y\|^{H,g[s/S][a/Y]} = T \\
&\quad\Leftrightarrow \text{ For all } a \in D_i : Iav_{i\to\tau}(s,a) = F \text{ or } \|\lfloor\varphi\rfloor Y\|^{H,g[s/S][a/Y]} = T \\
&\quad\Leftrightarrow \text{ For all } a \in S : a \notin av(s) \text{ or } M, a \models \varphi \\
&\quad\Leftrightarrow M, s \models \Box_a\varphi
\end{aligned}$$

(e) $\delta = \Box_p\varphi$. Same argument as with $\Box_a$.

(f) $\delta = \bigcirc(\psi/\varphi)$. We have the following chain of equivalences:

$$\begin{aligned}
&\|\lfloor\bigcirc(\psi/\varphi)\rfloor S\|^{H,g[s/S]} = T \\
&\quad\Leftrightarrow \|(\lambda X ob(\lfloor\varphi\rfloor)(\lfloor\psi\rfloor))S\|^{H,g[s/S]} = T \\
&\quad\Leftrightarrow \|(ob(\lfloor\varphi\rfloor)(\lfloor\psi\rfloor))S\|^{H,g[s/S]} = T \qquad (\beta\text{−reduction}) \\
&\quad\Leftrightarrow Iob_{\tau\to\tau\to o}(\|\lfloor\varphi\rfloor\|^{H,g[s/S]})(\|\lfloor\psi\rfloor\|^{H,g[s/S]}) = T \\
&\quad\Leftrightarrow \|\lfloor\varphi\rfloor\|^{H,g[s/S]} \in Iob_{\tau\to\tau\to o}(\|\lfloor\psi\rfloor\|^{H,g[s/S]}) \\
&\quad\Leftrightarrow V(\varphi) \in Iob_{\tau\to\tau\to o}(V(\psi)) \qquad (\textbf{see the justification *}) \\
&\quad\Leftrightarrow V(\varphi) \in ob(V(\psi)) \\
&\quad\Leftrightarrow M, s \models \bigcirc(\psi/\varphi)
\end{aligned}$$

Justification *: What we need to show is: $\|\lfloor\varphi\rfloor\|^{H,g[s/S]}$ is identified with $V(\varphi)$ (analogously ψ). By induction hypothesis, for all assignment g and state s, we have $\|\lfloor\varphi\rfloor S\|^{H,g[s/S]} = T$ if and only if $M, s \vDash \varphi$. Expanding the details of this equivalence we have: for all assignment g and state s

$\Leftrightarrow s \in \|\lfloor\varphi\rfloor\|^{H,g[s/S]}$ (functions to type o are associated with sets)

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

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

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

$\Leftrightarrow M, s \vDash \varphi$

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

Hence, $s \in \|\lfloor\varphi\rfloor\|^{H,g[s/S]}$ if and only if $s \in V(\varphi)$.
By extensionality we thus know that $\|\lfloor\varphi\rfloor\|^{H,g[s/S]}$ is identified with $V(\varphi)$

(g) $\delta = \bigcirc_a(\varphi)$. We have the following chain of equivalences:

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

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

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

$\Leftrightarrow \|ob(av(S))(\lfloor\varphi\rfloor)\|^{H,g[s/S]} = T$ and $\|\exists Y(av(S)(Y) \wedge \neg(\lfloor\varphi\rfloor Y))\|^{H,g[s/S]} = T$ ($\beta-$reduction)

$\Leftrightarrow \|ob(av(S))(\lfloor\varphi\rfloor)\|^{H,g[s/S]} = T$ and $\|av(S)(Y) \wedge \neg(\lfloor\varphi\rfloor Y)\|^{H,g[s/S][a/Y]} = T$

$\Leftrightarrow Iob_{\tau\to\tau\to o}(av(s))(\lfloor\varphi\rfloor) = T$ and $\|av(S)(Y)\|^{H,g[s/S][a/Y]} = T$ and $\|(\lfloor\varphi\rfloor Y)\|^{H,g[s/S][a/Y]} = F$

$\Leftrightarrow V(\varphi) \in ob(av(s))$ and $a \in av(s)$ and $a \notin V(\varphi)$

$\Leftrightarrow V(\varphi) \in ob(av(s))$ and $a \in av(s) \cap V(\neg\varphi)$

$\Leftrightarrow V(\varphi) \in ob(av(s))$ and $av(s) \cap V(\neg\varphi) \neq \emptyset$

$\Leftrightarrow M, s \models \bigcirc_a(\varphi)$

(h) $\delta = \bigcirc_p(\varphi)$. Same argument as with $\bigcirc_a(\varphi)$.

Theorem 1 (Soundness and Completeness of the Embedding)

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

Proof (Soundness, $\leftarrow$) The proof is by contraposition. Assume $\not\models^{CJL} \varphi$, i.e, there is a CJL model $M = \langle S, av, pv, ob, V\rangle$, and world $s \in S$, such that $M, s \not\models \varphi$. By Lemma 3 for an arbitrary assignment g it holds that $\|\lfloor\varphi\rfloor\ S_i\|^{H^M, g[s/S_i]} = F$ in Henkin model $H^M = \langle\{D_\alpha\}_{\alpha\in T}, I\rangle$ for M. 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, $H^M \not\models^{\text{HOL}}$ vld $\lfloor\varphi\rfloor$. Furthermore, $H^M \models^{\text{HOL}} \Sigma$ where $\Sigma \in$ {AV, PV1, PV2, OB1,..,OB5} by Lemma 2 and thus {AV, PV1, PV2, OB1,..,OB5} $\not\models^{\text{HOL}}$ vld $\lfloor\varphi\rfloor$.

(Completeness, $\rightarrow$) The proof is again by contraposition. Assume {AV, PV1, PV2, OB1,..,OB5} $\not\models^{\text{HOL}}$ vld $\lfloor\varphi\rfloor$, i.e., there is a Henkin model $H = \langle\{D_\alpha\}_{\alpha\in T}, I\rangle$ and an assignment g such that $H \models^{\text{HOL}} \Sigma$ where $\Sigma \in$ {AV, PV1, PV2, OB1,..,OB5} and $\|\text{vld}\,\lfloor\varphi\rfloor\|^{H,g} = F$. By Lemma 4, there is a CJL model M such that $M \not\models \varphi$. Hence, $\not\models^{CJL} \varphi$.

This shows that CJL is a natural fragment of HOL.

5 Implementation of CJL in Isabelle/HOL

In the following we sketch the "implementation" of our embedding in Isabelle/HOL:

```
theory CarmoJones imports Main
begin
  named_theorems CJ_ess
  named_theorems CJ_der

 typedecl i -- "type for possible worlds"
 type_synonym τ = "(i⇒bool)"
 consts
  av:: "i⇒(i⇒bool)"
  pv:: "i⇒(i⇒bool)"
  ob:: "(i⇒bool)⇒((i⇒bool)⇒bool)"
  aw::i (* actual world *)
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)"

```

For the classical CJL connectives we have:

```
 definition cjneg :: "τ⇒τ" ("¬_"[52]53)
  where "¬A ≡ λw. ¬A(w)"
 definition cjand :: "τ⇒τ⇒τ" (infixr"∧"51)
  where "A∧B ≡ λw. A(w)∧B(w)"
 definition cjor  :: "τ⇒τ⇒τ" (infixr"∨"50)
  where "A∨B ≡ λw. A(w)∨B(w)"
 definition cjimp :: "τ⇒τ⇒τ" (infixr"→"49)
  where "A→B ≡ λw. A(w)⟶B(w)"
 definition cjequiv :: "τ⇒τ⇒τ" (infixr"↔"48)
  where "A↔B ≡ λw. A(w)⟷B(w)"
 definition cjtop :: "τ" ("⊤")
  where "⊤  ≡ λw. True"
 definition cjbot :: "τ" ("⊥")
  where "⊥  ≡ λw. False"
```

And for CJL non-classical connectives we have:

```
definition cjbox :: "τ⇒τ" ("□")
 where "□A ≡ λw.∀v. A(v)"    (* same as "A = (λw. True)" *)
definition cjboxa :: "τ⇒τ" ("□a")
 where "□aA ≡ λw. (∀x. av(w)(x) ⟶ A(x))"
definition cjboxp :: "τ⇒τ" ("□p")
 where "□pA ≡ λw. (∀x. pv(w)(x) ⟶ A(x))"
definition cjdia :: "τ⇒τ" ("◇") (* where "◇A ≡ λw. ∃v. A(v)" *)
 where "◇A ≡ ¬□(¬A)"
definition cjdiaa :: "τ⇒τ" ("◇a") (* where "◇aA ≡ λw. (∃x. av(w)(x) ∧ A(x))" *)
 where "◇aA ≡ ¬□a(¬A)"
definition cjdiap :: "τ⇒τ" ("◇p") (* where "◇pA ≡ λw. (∃x. pv(w)(x) ∧ A(x))" *)
 where "◇pA ≡ ¬□p(¬A)"
definition cjo1 :: "τ⇒τ⇒τ" ("O1⟨_|_⟩"[52]53)
 where "O1⟨B|A⟩  ≡ λw. (∃x. A(x) ∧ B(x)) ∧ (∀X. ((∀y. X(y) ⟶ A(y)) ∧ (∃z. X(z) ∧ B(z))) ⟶ ob(X)(B))"
definition cjo :: "τ⇒τ⇒τ" ("O⟨_|_⟩"[52]53)
 where "O⟨B|A⟩  ≡  λw. ob(A)(B)"
definition cjoa :: "τ⇒τ" ("Oa")
 where "OaA  ≡ λw. ob(av(w))(A) ∧ (∃x. av(w)(x) ∧ ¬A(x))"
definition cjop :: "τ⇒τ" ("Op")
 where "OpA  ≡ λw. ob(pv(w))(A) ∧ (∃x. pv(w)(x) ∧ ¬A(x))"
```

Also we define validity and actual validity (of world "aw") as follows:

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

5.1 Nested dyadic deontic operator in CJL

A particular focus of our experiments has been on nested dyadic obligations. We show that nested dyadic obligations in CJL can be eliminated.

Definition 2 (Unnested disjunctive normal form (UDNF)) Let $\mathcal{L}^O$ be the sublanguage of $\mathcal{L}$ without formula containing the $\Box$-operator. We say that a formula $B \in \mathcal{L}^O$ is in Unnested Disjunctive Normal Form (**UDNF**) if it is a disjunction of conjunctions of the form
$\delta = C \wedge \bigcirc(A_1|B_1) \wedge ... \wedge \bigcirc(A_n|B_n) \wedge \bigcirc(A_{n+1}|B_{n+1}) \wedge ... \wedge \bigcirc(A_{n+k}|B_{n+k})$
where $n; k \in \mathbb{N}$ and all of the formula $\alpha, A_m, B_m (m \leq n+k)$ are propositional formula ($\bot$and $\top$are considered propositional formulae). The formula δ is called a canonical conjunction and the formula $\bigcirc(A_m|B_m)$ are called prenex formula [26].

Lemma 5 (Meyer and van der Hoek [26] Lemma 1.7.6.2) *If B is in **UDNF** and contains a prenex formula σ, then B is equivalent to a formula of the form $B = E \vee (D \wedge \sigma)$where $E, D,$ and σ are all in **UDNF**.*

Theorem 2 *For every formula A, B, C, D in CJL we have:*

$\bigcirc(A|(E \vee (D \wedge \bigcirc(B|C)))) \leftrightarrow ((\bigcirc(B|C) \wedge \bigcirc(A|(E \vee D)) \vee ((\neg \bigcirc (B|C) \wedge \bigcirc(A|E))$
$\bigcirc(A|(E \vee (D \wedge \neg \bigcirc (B|C)))) \leftrightarrow ((\neg \bigcirc (B|C) \wedge \bigcirc(A|(E \vee D)) \vee ((\bigcirc(B|C) \wedge \bigcirc(A|E))$
$\bigcirc((E \vee (D \wedge \bigcirc(B|C))|F)) \leftrightarrow ((\bigcirc(B|C) \wedge \bigcirc((E \vee D)|F) \vee ((\neg \bigcirc (B|C) \wedge \bigcirc(E|F))$

$$\bigcirc((E \vee (D \wedge \neg \bigcirc (B|C))|F)) \leftrightarrow ((\neg \bigcirc (B|C) \wedge \bigcirc((E \vee D)|F) \vee ((\bigcirc(B|C) \wedge \bigcirc(E|F))$$

Proof *We proof this lemma in Isabelle/HOL:*

```
lemma "⌊O⟨A |(E ∨ (D ∧ O⟨B|C⟩)) ⟩ ↔ (( O⟨B | C ⟩ ∧ O⟨A |(E ∨ D)⟩ ) ∨ (¬ O⟨B|C⟩∧ O⟨A|E⟩ )) ⌋" unfolding Defs
  by (smt ax_5a ax_5b ax_5e)
lemma "⌊O⟨A |(E ∨ (D ∧ ¬ O⟨B|C⟩)) ⟩ ↔ ((¬ O⟨B | C ⟩ ∧ O⟨A |(E ∨ D)⟩ ) ∨ ( O⟨B|C⟩∧ O⟨A|E⟩ )) ⌋" unfolding Defs
  by (smt ax_5a ax_5b ax_5e)
lemma "⌊O⟨ (E ∨ (D ∧ O⟨B|C⟩)) |F ⟩ ↔ (( O⟨B | C ⟩ ∧ O⟨(E ∨ D) | F⟩ ) ∨ ( ¬ O⟨B|C⟩∧ O⟨E |F⟩ )) ⌋" unfolding Defs
  by (smt ax_5b)
lemma "⌊O⟨ (E ∨ (D ∧ ¬ O⟨B|C⟩)) |F ⟩ ↔ ((¬ O⟨B | C ⟩ ∧ O⟨(E ∨ D) | F⟩ ) ∨ ( O⟨B|C⟩∧ O⟨E |F⟩ )) ⌋" unfolding Defs
  by (smt ax_5b)
```

6 Conclusion

The embeddings approach bridges between the Tarski view of logics (for meta-logic HOL) and the Kripke view (for the embedded source logics) and exploits the fact that well-known translations of logics, respectively, their Kripke-style semantical characterizations, can often be elegantly and directly formalized in HOL. We have presented a straightforward embedding of the dyadic deontic logic has been proposed by Carmo and Jones in simple type theory and we have shown that this embedding is sound and complete. Future work includes the systematic analysis of further properties of dyadic deontic logics in HOL.

Acknowledgements ...

References

1. Andrews, P.: Church's type theory. In: E.N. Zalta (ed.) The Stanford Encyclopedia of Philosophy, Spring 2014 edn., Stanford University (2014)
2. Andrews, P.B.: Resolution in type theory. Journal of Symbolic Logic **36**(3), 414–432 (1971)
3. Andrews, P.B.: General models and extensionality. Journal of Symbolic Logic **37**(2), 395–397 (1972)
4. Andrews, P.B.: General models, descriptions, and choice in type theory. Journal of Symbolic Logic **37**(2), 385–394 (1972)
5. Benzmüller, C.: Equality and extensionality in automated higher-order theorem proving. Ph.D. thesis, Saarland University (1999)
6. Benzmüller, C.: Automating quantified conditional logics in HOL. In: F. Rossi (ed.) 23rd International Joint Conference on Artificial Intelligence (IJCAI-13), pp. 746–753. Beijing, China (2013).
7. Benzmüller, C., Brown, C., Kohlhase, M.: Higher-order semantics and extensionality. Journal of Symbolic Logic **69**(4), 1027–1088 (2004). DOI 10.2178/jsl/1102022211.
8. Benzmüller, C., Gabbay, D., Genovese, V., Rispoli, D.: Embedding and automating conditional logics in classical higher-order logic. Annals of Mathematics and Artificial Intelligence **66**(1-4), 257–271 (2012). DOI 10.1007/s10472-012-9320-z.
9. Benzmüller, C., Paulson, L.: Exploring properties of normal multimodal logics in simple type theory with LEO-II. In: C. Benzmüller, C. Brown, J. Siekmann, R. Statman (eds.) Reasoning in Simple Type Theory — Festschrift in Honor of Peter B. Andrews on His 70th Birthday, Studies in Logic, Mathematical Logic and Foundations, pp. 386–406. College Publications (2008).
10. Benzmüller, C., Miller, D.: Automation of Higher-Order Logic. Chapter in: Gabbay, D.M., Siekmann, J.H., Woods, J. (eds.) Handbook of the History of Logic, Volume 9 — Computational Logic, North Holland, Elsevier, pp. 215-254 (2014).

11. Benzmüller, C., Paulson, L.: Multimodal and intuitionistic logics in simple type theory. The Logic Journal of the IGPL **18**(6), 881–892 (2010). DOI 10.1093/jigpal/jzp080.
12. Benzmüller, C., Paulson, L.: Quantified multimodal logics in simple type theory. Logica Universalis (Special Issue on Multimodal Logics) **7**(1), 7–20 (2013). DOI 10.1007/s11787-012-0052-y.
13. Benzmüller, C., Woltzenlogel Paleo, B.: Automating Gödel's ontological proof of God's existence with higher-order automated theorem provers. In: T. Schaub, G. Friedrich, B. O'Sullivan (eds.) ECAI 2014, *Frontiers in Artificial Intelligence and Applications*, vol. 263, pp. 93 – 98. IOS Press (2014). DOI 10.3233/978-1-61499-419-0-93.
14. Benzmüller, C.: Universal Reasoning, Rational Argumentation and Human-Machine Interaction. arXiv preprint arXiv:1703.09620 (2017).
15. Brown, C.E.: Set comprehension in church's type theory. Ph.D. thesis, Department of Mathematical Sciences, Carnegie Mellon University (2004). See also Chad E. Brown, *Automated Reasoning in Higher-Order Logic*, College Publications, 2007
16. José Carmo and Andrew J. I. Jones, Completeness and decidability results for a logic of contrary-to-duty conditionals, J. Log. Comput.,23 ,3, 585–626, 2013.
17. Church, A.: A set of postulates for the foundation of logic. Annals of Mathematics **33**(2), 346–366 (1932)
18. Church, A.: An unsolvable problem of elementary number theory. American Journal of Mathematics **58**(2), 354–363 (1936)
19. Amit Chopra, Leon van der Torre, Harko Verhagen, Serena Villata, Handbook of Normative Multiagent Systems.(To appear, College publication.)
20. Church, A.: A formulation of the simple theory of types. Journal of Symbolic Logic **5**(2), 56–68 (1940)
21. Chwistek, L.: The Limits of Science: Outline of Logic and of the Methodology of the Exact Sciences. London: Routledge and Kegan Paul (1948)
22. Frege, G.: Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens. Halle (1879). Translated in [**?**]
23. Gabbay, D., Horty, J., Parent, X., van der Meyden, R. and van der Torre, L. : Handbook of deontic logic and normative systems. College Publications, 2013
24. Gödel, K.: Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I. Monatshefte der Mathematischen Physik, **38**, 173–198 (1931)
25. Henkin, L.: Completeness in the theory of types. Journal of Symbolic Logic **15**(2), 81–91 (1950)
26. J-J Ch. Meyer and W van der Hoek. Epistemic Logic for AI and Computer Science (Cambridge Tracts in Theoretical Computer Science). Cambridge University Press, 1995. ISBN 0511569858.
27. Muskens, R.: Intensional models for the theory of types. Journal of Symbolic Logic **72**(1), 98–118 (2007)
28. T. Nipkow, L. C. Paulson, and M. Wenzel, Isabelle/HOL: A Proof Assistant for Higher-Order Logic. Volume 2283 in LNCS. Springer, (2002)
29. Ramsey, F.P.: The foundations of mathematics. In: Proceedings of the London Mathematical Society, *2*, vol. 25, pp. 338–384 (1926)
30. Russell, B.: Mathematical logic as based on the theory of types. American Journal of Mathematics **30**(3), 222–262 (1908)
31. Schütte, K.: Semantical and syntactical properties of simple type theory. Journal of Symbolic Logic **25**(4), 305–326 (1960)