Mechanised Modal Model Theory

Yiming Xu and Michael Norrish

No Institute Given

1 Introduction

[1p]

Contributions This paper presents the first mechanised proofs of a number of basic results from the first two chapters of Blackburn *et al* [1] (e.g., bounded morphisms, bisimulations and the finite model property *via* selection), as well as

- the saturation of ultraproduct models;
- modal equivalence as bisimilarity between ultrafilter extensions; and
- a close approximation of van Benthem's Characterization Theorem.

We also discuss where HOL's simple type theory lets us down: some standard results (including the best possible statement of van Benthem's Characterization Theorem) seem impossible to prove in our setting.

1.1 Related Work

2 Syntax, Semantics and the Standard Translation

[1.5p] In our formalization, we only consider the basic modal language, in which the only primitive modal operator is the '$\Diamond$'. For a type α, an α-modal formula is either of form $\mathsf{VAR}\ p$, where p is of type α, or a disjunction $\phi \vee \psi$ of two α-modal formulas, or the falsity $\bot$, or a negation $\neg\phi$ of an α-modal formula ϕ, or, finally, of the form $\Diamond\phi$ where ϕ is an α-modal formula. We create a data type called 'form' of the formulas of this modal language.

Definition 11 [1, Definition 1.9]

$$\alpha\ \mathit{form} = \mathsf{VAR}\ \alpha \mid \mathsf{DISJ}\ (\alpha\ \mathit{form})\ (\alpha\ \mathit{form}) \mid \bot \mid (\neg)\ (\alpha\ \mathit{form}) \mid \Diamond\ (\alpha\ \mathit{form})$$

A model where these formulas can be interpreted consists of a *frame* and a *valuation*, where frame is a set with a relation on it.

Definition 12 [1, Definition 1.19]

$$\beta\ \mathit{frame} = \langle\!\langle\ \mathsf{world} : \beta \to \mathit{bool};\ \mathsf{rel} : \beta \to \beta \to \mathit{bool}\ \rangle\!\rangle$$
$$(\alpha,\ \beta)\ \mathit{model} = \langle\!\langle\ \mathsf{frame} : \beta\ \mathit{frame};\ \mathsf{valt} : \alpha \to \beta \to \mathit{bool}\ \rangle\!\rangle$$

In the rest of the paper, the field $\mathfrak{M}.\mathsf{valt}$ of a model $\mathfrak{M}$ will be called the *valuation*, and $\mathfrak{M}^W$, $\mathfrak{M}^R$ and $\mathfrak{M}^V$ are used to denote the world set, the relation, and the valuation of $\mathfrak{M}$ respectively. When we say a (α, β)-model, we mean a model for α-formulas with a β-set as its underlying set. The interpretation of α-modal formulas on an (α, β)-model is given by the predicate *satisfaction*. We read '$\mathfrak{M}, w \Vdash \phi$' as '$\phi$ is satisfied at the world w in $\mathfrak{M}$'.

Definition 13 [1, Definition 1.20]

$$\begin{aligned}
\mathfrak{M}, w \Vdash \mathsf{VAR}\ p &\stackrel{\text{def}}{=} w \in \mathfrak{M}^W \wedge w \in \mathfrak{M}^V\ p \\
\mathfrak{M}, w \Vdash \bot &\stackrel{\text{def}}{=} w \in \mathfrak{M}^W \wedge \mathsf{F} \\
\mathfrak{M}, w \Vdash \neg\phi &\stackrel{\text{def}}{=} w \in \mathfrak{M}^W \wedge \mathfrak{M}, w \nVdash \phi \\
\mathfrak{M}, w \Vdash (\phi_1 \vee \phi_2) &\stackrel{\text{def}}{=} \mathfrak{M}, w \Vdash \phi_1 \vee \mathfrak{M}, w \Vdash \phi_2 \\
\mathfrak{M}, w \Vdash \Diamond\phi &\stackrel{\text{def}}{=} w \in \mathfrak{M}^W \wedge \exists v.\, \mathfrak{M}^R\ w\ v \wedge v \in \mathfrak{M}^W \wedge \mathfrak{M}, v \Vdash \phi
\end{aligned}$$

If ϕ_1, ϕ_2 are α-formulas, we say they are *equivalent* if for every (α, β)-model $\mathfrak{M}$ and every world w in it, we have $\mathfrak{M}, w \Vdash \phi_1 \iff \mathfrak{M}, w \Vdash \phi_2$.

Definition 14 (Equivalence)

$$\begin{aligned}
&(\phi_1 : \alpha\ \mathit{form}) \equiv_{(:\beta)} (\phi_2 : \alpha\ \mathit{form}) \stackrel{\text{def}}{=} \\
&\forall (\mathfrak{M} : (\alpha,\ \beta)\ \mathit{model})\ (w : \beta).\ \mathfrak{M}, w \Vdash \phi_1 \iff \mathfrak{M}, w \Vdash \phi_2
\end{aligned}$$

We cannot omit the type parameter $(:\beta)$ in the definition, since then there will be a type, namely the type of the underlying set of the models we are talking about, that only appears on the right-hand side but not on the left-hand side of the definition, which is not allowed in HOL. Also, we are not allowed to quantify over types, so it is also impossible to define the equivalence to be $\forall \mu.\ \phi_1 \equiv_\mu \phi_2$, where μ denotes a type. Therefore, this definition is not encoding the equivalence in mathematical sense precisely: when we mention equivalence of formulas in usual mathematical language, we are implicitly referring to the class of all models, but the constraint here bans us from talking about all models of all possible types at once.

A modal formula can be translated into a first-order formula via the *standard translation*. A translated formula has at most one free variable. For a modal formula ϕ, $\mathsf{ST}_x\ \phi$ is the standard translation of ϕ using x as the only free variable.

Definition 15 [1, Definition 2.45 (Standard Translation)]

$$\begin{aligned}
\mathsf{ST}_x\ (\mathsf{VAR}\ p) &\stackrel{\text{def}}{=} {}^{\mathsf{f}}\mathsf{P}\ p\ ({}^{\mathsf{f}}\mathsf{VAR}\ x) \\
\mathsf{ST}_x\ \bot &\stackrel{\text{def}}{=} {}^{\mathsf{f}}\bot \\
\mathsf{ST}_x\ (\neg\phi) &\stackrel{\text{def}}{=} {}^{\mathsf{f}}\neg\ (\mathsf{ST}_x\ \phi) \\
\mathsf{ST}_x\ (\phi \vee \psi) &\stackrel{\text{def}}{=} \mathsf{ST}_x\ \phi\ {}^{\mathsf{f}}\vee\ \mathsf{ST}_x\ \psi \\
\mathsf{ST}_x\ (\Diamond\phi) &\stackrel{\text{def}}{=} {}^{\mathsf{f}}\exists\ (x + 1)\ ({}^{\mathsf{f}}\mathsf{R}\ ({}^{\mathsf{f}}\mathsf{VAR}\ x)\ ({}^{\mathsf{f}}\mathsf{VAR}\ (x + 1))\ {}^{\mathsf{f}}\wedge\ \mathsf{ST}_{x+1}\ \phi)
\end{aligned}$$

According to the semantic interpretation of the '$\Diamond\phi$', we translate $\Diamond\phi$ into the existential formula. To make sure that we use a fresh variable symbol that is not the same as the variable x which is marking the current state, we use $x+1$ as our new variable symbol. The standard translation gives a first-order reformulation of satisfaction of modal formulas:

Proposition 11 [1, Theorem 2.47 (i)]

$$\vdash \mathfrak{M}, w \Vdash \phi \iff \mathsf{mm2folm}\ \mathfrak{M}, (\lambda\, n.\ w) \vDash \mathsf{ST}_x\ \phi$$

Here mm2folm is the function that turns a modal model into a first-order model, defined as:

$$\begin{array}{l}
\mathsf{mm2folm}\ \mathfrak{M} \stackrel{\mathsf{def}}{=} \\
\ \langle\!\langle \mathsf{Dom} := \mathfrak{M}^W;\ \mathsf{Fun} := (\lambda\, n\ l.\ \mathsf{CHOICE}\ \mathfrak{M}^W);\\
\ \mathsf{Pred} := \\
\ \ (\lambda\, p\ zs. \\
\quad \mathtt{case}\ zs\ \mathtt{of} \\
\quad\ \ [] \Rightarrow \mathsf{F} \\
\quad \mid [w_1] \Rightarrow w_1 \in \mathfrak{M}^W \wedge \mathfrak{M}^V\ p\ w_1 \\
\quad \mid [w_1;\ w_2] \Rightarrow p = 0 \wedge \mathfrak{M}^R\ w_1\ w_2 \wedge w_1 \in \mathfrak{M}^W \wedge w_2 \in \mathfrak{M}^W \\
\quad \mid w_1 :: w_2 :: w_3 :: w_4 \Rightarrow \mathsf{F})\rangle\!\rangle
\end{array}$$

3 Basic Results

We discuss some highlights of formalized results in section 2.1-2.3 in [1] below.

3.1 Tree-like property

A tree-like model is a model whose underlying frame is a tree. If H is a tree with root r, we write $\mathsf{tree}\ H\ r$ in HOL:

Definition 16 [1, Definition 1.7]

$$\begin{array}{l}
\mathsf{tree}\ H\ r \stackrel{\mathsf{def}}{=} \\
\ r \in H.\mathsf{world} \wedge (\forall\, w.\ w \in H.\mathsf{world} \Rightarrow H.\mathsf{rel}\,|_{H.\mathsf{world}}{}^*\ r\ w) \wedge \\
\ (\forall\, w.\ w \in H.\mathsf{world} \Rightarrow \neg H.\mathsf{rel}\ w\ r) \wedge \\
\ \forall\, w.\ w \in H.\mathsf{world} \wedge w \neq r \Rightarrow \exists! w_0.\ w_0 \in H.\mathsf{world} \wedge H.\mathsf{rel}\ w_0\ w
\end{array}$$

The tree-like property says an satisfiable modal formula can be satisfied in a tree-like model. It is proved in HOL as:

Proposition 12 [1, Proposition 2.15]

$$\begin{array}{l}
\vdash (\mathfrak{M}_1 : (\alpha,\ \beta)\ \mathit{model}), (w : \beta) \Vdash\ (\phi : \alpha\ \mathit{form}) \Rightarrow \\
\quad \exists\, (\mathfrak{M} : (\alpha,\ \beta\ \mathtt{list})\ \mathit{model})\ (r : \beta\ \mathtt{list}).\ \mathsf{tree}\ \mathfrak{M}.\mathsf{frame}\ r \wedge \mathfrak{M}, r \Vdash \phi
\end{array}$$

The world set of the tree like model constructed from $\mathfrak{M}$ according to the standard proof of this theorem is a set of lists of worlds in $\mathfrak{M}_1$, that is why passing to a tree-like model does not preserve the type.

3.2 Bisimulation

The definition of bisimulation in HOL is just a direct translation of its mathematical definition:

Definition 17 [1, Definition 2.16 (Bisimulations)]

$$
\begin{array}{l}
\mathfrak{M}_1 \overset{Z}{\leftrightarrows} \mathfrak{M}_2 \overset{\text{def}}{=} \\
\quad \forall\, w_1\ w_2. \\
\qquad w_1 \in \mathfrak{M}_1^W \wedge w_2 \in \mathfrak{M}_2^W \wedge Z\ w_1\ w_2 \Rightarrow \\
\qquad (\forall\, p.\ \mathfrak{M}_1, w_1 \Vdash \mathsf{VAR}\ p \iff \mathfrak{M}_2, w_2 \Vdash \mathsf{VAR}\ p) \wedge \\
\qquad (\forall\, v_1. \\
\qquad\quad v_1 \in \mathfrak{M}_1^W \wedge \mathfrak{M}_1^R\ w_1\ v_1 \Rightarrow \\
\qquad\quad\ \exists\, v_2.\ v_2 \in \mathfrak{M}_2^W \wedge Z\ v_1\ v_2 \wedge \mathfrak{M}_2^R\ w_2\ v_2) \wedge \\
\qquad \forall\, v_2. \\
\qquad\quad v_2 \in \mathfrak{M}_2^W \wedge \mathfrak{M}_2^R\ w_2\ v_2 \Rightarrow \\
\qquad\quad\ \exists\, v_1.\ v_1 \in \mathfrak{M}_1^W \wedge Z\ v_1\ v_2 \wedge \mathfrak{M}_1^R\ w_1\ v_1
\end{array}
$$

It is trivial to prove by induction that bisimilar worlds are modal equivalent. As the most significant theorem on the basic theory of bisimulations, we proved the Hennessy-Milner theorem as, which says modal equivalence and bisimulation on *image finite* models are the same thing. An image-finite model is a model where every world can only be related to finitely many worlds. The Hennessy-Milner Theorem in HOL looks like:

Theorem 13 [1, Theorem 2.24 (Hennessy-Milner Theorem)]

$$
\begin{array}{l}
\vdash \mathsf{image\_finite}\ \mathfrak{M}_1 \wedge \mathsf{image\_finite}\ \mathfrak{M}_2 \wedge w_1 \in \mathfrak{M}_1^W \wedge w_2 \in \mathfrak{M}_2^W \Rightarrow \\
\quad (\mathfrak{M}_1, w_1 \leftrightsquigarrow \mathfrak{M}_2, w_2 \iff \mathfrak{M}_1, w_1 \leftrightarrows \mathfrak{M}_2, w_2)
\end{array}
$$

3.3 Finite model property

There are two classical approach of constructing finite models using model theory, namely via selection and via filtration.

Selection Given $\mathfrak{M}_1, w_1 \Vdash \phi$, where ϕ has degree k, we can construct $\mathfrak{M}_2$, $\mathfrak{M}_3$, $\mathfrak{M}_4$ and $\mathfrak{M}_5$ comsecutively, such that $\mathfrak{M}_5$ is the finite model we want, where:

- $\mathfrak{M}_2$ is the tree-like model obtained from 12 with root w_2 such that $\mathfrak{M}_2, w_2 \Vdash \phi$.
- $\mathfrak{M}_3$ is the restriction of $\mathfrak{M}_2$ to height k.
- $\mathfrak{M}_4$ is obtained from $\mathfrak{M}_3$ by modifying the valuation into $\lambda\, p\ v.$ `if` $p \in$ $\mathsf{prop\_letters}\ \phi$ `then` ${\mathfrak{M}_3}^V\ p\ v$ `else` F.

The construction of $\mathfrak{M}_5$ requires a lemma:

Lemma 14 [1, Proposition 2.29]

$$\vdash \mathsf{FINITE}\ (\Phi : \alpha \to \mathit{bool}) \wedge \mathsf{INFINITE}\ \mathcal{U}(:\beta) \Rightarrow \forall (n : \mathit{num}).\ \mathsf{FINITE}\ \{\ \phi \mid \mathsf{DEG}\ \phi \leq n \wedge \mathsf{prop\_letters}\ \phi \subseteq \Phi\ \}\ /\equiv_{(:\beta)}$$

We require the assumption that the universe of β is infinite since we relies on the fact that two modal formulas $\Diamond\phi_1$ and $\Diamond\phi_2$ are equivalent if and only if ϕ_1 and ϕ_2 are equivalent. This is easy to prove using set theory. However, in simple type theory, the proof of $\phi_1 \equiv_{(:\beta)} \phi_2$ iff $\Diamond\phi_1 \equiv_{(:\beta)} \Diamond\phi_2$ requires the infiniteness of the type universe of β.

By the lemma above, the set $\Delta = \{\ \psi \mid \mathsf{DEG}\ \psi \leq k \wedge \mathsf{prop\_letters}\ \psi \subseteq \Phi\ \}\ /\ \equiv_{(:\beta\ \mathtt{list})}$ is finite. Let Γ be the set of modal formulas starting with a diamond and let R denote $\{\ \mathsf{CHOICE}\ (A \cap \Gamma) \mid A \in \Delta\ \} \setminus \{\ \emptyset\ \}$, we define S^{s} to be the primitive recursion function:

$$\begin{array}{l}\mathsf{PRIM\_REC}\ \{\ w_2\ \}\\ (\lambda\, s_0\ n.\\ \quad \{\ \mathsf{CHOICE}\ us \mid\\ \quad \exists \phi\ v.\\ \quad\quad \mathfrak{M}_4, v \Vdash \Diamond\phi \wedge \Diamond\phi \in R \wedge v \in s_0 \wedge\\ \quad\quad us = \{\ u \mid \mathfrak{M}_4^R\ v\ u \wedge u \in \mathfrak{M}_4^W \wedge \mathfrak{M}_4, u \Vdash \phi\ \}\ \}\ )\end{array}$$

Then the world set of $\mathfrak{M}_5$ is $\bigcup\ \{\ S^{\mathsf{s}}\ i \mid i \leq k\ \}$. We proved $\mathfrak{M}_5, w_2 \Vdash \phi$ by proving

$$\begin{array}{l}\lambda\, n\ a_1\ a_2.\\ \quad a_1 \in \mathfrak{M}_5^W \wedge a_2 \in \mathfrak{M}_4^W \wedge \mathsf{height}\ \mathfrak{M}_4\ w_2\ \mathfrak{M}_4\ a_1 = \mathsf{height}\ \mathfrak{M}_4\ w_2\ \mathfrak{M}_4\ a_2 \wedge\\ \quad \mathsf{height}\ \mathfrak{M}_4\ w_2\ \mathfrak{M}_4\ a_1 \leq k - n \wedge\\ \quad \forall \phi.\ \mathsf{DEG}\ \phi \leq n \wedge \mathsf{prop\_letters}\ \phi \subseteq s \Rightarrow (\mathfrak{M}_4, a_1 \Vdash \phi \iff \mathfrak{M}_4, a_2 \Vdash \phi)\end{array}$$

is an *n-bisimulation* relating the two copies of w_2 in $\mathfrak{M}_4$ and in $\mathfrak{M}_5$.

As we used 12, we change the type of the model by passing to a finite model via selection:

Theorem 15 [1, Theorem 2.34 (Finite model property, via selection)]

$$\vdash (\mathfrak{M}_1 : (\alpha, \beta)\ \mathit{model}), (w_1 : \beta) \Vdash (\phi : \alpha\ \mathit{form}) \Rightarrow \exists (\mathfrak{M} : (\alpha, \beta\ \mathit{list})\ \mathit{model})\ (v : \beta\ \mathit{list}).\ \mathsf{FINITE}\ \mathfrak{M}^W \wedge v \in \mathfrak{M}^W \wedge \mathfrak{M}, v \Vdash \phi$$

Filtration Given $\mathfrak{M}_1, w \Vdash \phi$, the finite model $\mathfrak{M}_2$ constructed by selection is the filtration of $\mathfrak{M}_1$ by the set of subformulas of ϕ. In HOL, we write $\mathsf{FLT}\ \mathfrak{M}\ \Sigma$ for the filtration of the model $\mathfrak{M}$ via the set Σ. The formalization is almost a direct translation of the mathematical proof in Blackburn *et al* [1]. For the only difference: the world set of the filtrated model using the standard proof is a set of equivalence classes of worlds. To make the type of the finite model same as the model we start with, instead of taking the set of equivalence classes as the world set, we pick only one representative from each equivalence class. Hence we can prove:

Theorem 16 [1, Theorem 2.41 (Finite model property, via filtration)]

$$\vdash (\mathfrak{M}_1 : (\alpha,\ \beta)\ \mathit{model}), (w_1 : \beta) \Vdash\ (\phi : \alpha\ \mathit{form}) \Rightarrow \\ \exists\, (\mathfrak{M}_2 : (\alpha,\ \beta)\ \mathit{model})\ (w_2 : \beta).\ w_2\ \in\ \mathfrak{M}_2^W\ \wedge\ \mathfrak{M}_2, w_2 \Vdash\ \phi\ \wedge\ \mathsf{FINITE}\ \mathfrak{M}_2^W$$

where $\mathfrak{M}_2$ above is taken as $\mathsf{FLT}\ \mathfrak{M}_1\ (\mathsf{subforms}\ \phi)$, and has the same type as $\mathfrak{M}_1$.

4 Mechanizing Ultrafilters and Ultraproducts

[3p] Some results in section 2.5-2.7 relies on some theorems about ultrafilters and ultraproduct.

4.1 ultrafilters

Given a non-empty set J, a set $L \subseteq \mathcal{P}(J)$ is a *filter* if it contains J itself, is closed under binary intersection, and is closed upward.

Definition 18 [1, Definition A.12 (Filters)]

$$\mathsf{filter}\ L\ J\ \stackrel{\mathsf{def}}{=} \\ J\ \neq\ \emptyset\ \wedge\ L\ \subseteq\ \mathsf{POW}\ J\ \wedge\ J\ \in\ L\ \wedge \\ (\forall\, X\ Y.\ X\ \in\ L\ \wedge\ Y\ \in\ L \Rightarrow\ X\ \cap\ Y\ \in\ L)\ \wedge \\ \forall\, X\ Z.\ X\ \in\ L\ \wedge\ X\ \subseteq\ Z\ \wedge\ Z\ \subseteq\ J \Rightarrow\ Z\ \in\ L$$

We call L a *proper filter* if L is not the whole power set. An *ultrafilter* is a filter U such that for every $X \subseteq J$, exactly one of X or $J \setminus X$ is in U.

The ultrafilter theorem, which says every proper filter is contained in an ultrafilter, is proved as follows:

Theorem 17 [1, Fact A.14, first half]

$$\vdash\ \mathsf{proper\_filter}\ L\ J\ \Rightarrow\ \exists\, U.\ \mathsf{ultrafilter}\ U\ J\ \wedge\ L\ \subseteq\ U$$

A subset A of the power set on J has *finite intersection property* if once we take the intersection of a finite, nonempty family in A, the resultant set is nonempty.

Definition 19 [1, Definition A.13 (Finite Intersection Property)]

$$\vdash\ \mathsf{FIP}\ A\ J\ \Longleftrightarrow \\ A\ \subseteq\ \mathsf{POW}\ J\ \wedge\ \forall\, B.\ B\ \subseteq\ A\ \wedge\ \mathsf{FINITE}\ B\ \wedge\ B\ \neq\ \emptyset\ \Rightarrow\ \bigcap B\ \neq\ \emptyset$$

As a corollary of ultrafilter theorem, a set with finite intersection property is contained in an ultrafilter.

4.2 Ultraproducts

The notion of ultraproducts are defined for sets, modal models, and first-order models.

Ultraproduct of sets The ultraproduct of a family of sets $(A_j)_{j\in J}$, where J is non-empty and each A_j is non-empty, is defined as a quotient of the cartesian product of the family. The Cartesian product of the family $(A_j)_{j\in J}$ is the set of functions f with domain J such that for all $j \in J$, $f(j) \in A_j$.

Definition 110 [1, Page 495 (Cartesian product)]

$$\mathsf{Cart\_prod}\ J\ A^{\mathsf{s}} \stackrel{\mathsf{def}}{=} \{\, f \mid \forall j.\ j\ \in\ J\ \Rightarrow\ f\ j\ \in\ A^{\mathsf{s}}\ j\ \}$$

If U is an ultrafilter on J, for two functions f, g in the Cartesian product $\mathsf{Cart\_prod}\ J\ A^{\mathsf{s}}$, we say f and g are U-equivalent (notation: $f \sim_U^{A^{\mathsf{s}}} g$) if the set $\{\, j \mid j\ \in\ J\ \wedge\ f\ j\ =\ g\ j\ \}$ (where the values of f and g agree) is in U. The ultraproduct of $(A_j)_{j\in J}$ modulo U is the quotient of $\mathsf{Cart\_prod}\ J\ A^{\mathsf{s}}$ by $\sim_U^{A^{\mathsf{s}}}$.

Definition 111 [1, Definition 2.69 (Ultraproduct of Sets)]

$$\mathsf{ultraproduct}\ U\ J\ A^{\mathsf{s}} \stackrel{\mathsf{def}}{=} \mathsf{Cart\_prod}\ J\ A^{\mathsf{s}} / \sim_U^{A^{\mathsf{s}}}$$

We write f_U to denote the equivalence class that f belongs to. In the case where $A^{\mathsf{s}}\ j\ =\ A$ for all $j \in J$, the ultraproduct is called the ultrapower of A modulo U. As before, in the definition above, the family $(A_j)_{j\in J}$ is encoded as a function, and hence for $j \in J$, $A^{\mathsf{s}}\ j$ is the set A_j indexed by j.

Ultraproduct for modal models Given a family $\mathfrak{M}^{\mathsf{s}}$ of modal models indexed by J and an ultrafilter U on J, the ultraproduct model of $\mathfrak{M}^{\mathsf{s}}$ modulo U (notation : $\Pi_U\ \mathfrak{M}^{\mathsf{s}}$) is described as follows:

- The world set is the ultraproduct of world sets of $\mathfrak{M}^{\mathsf{s}}$ modulo U.
- Two equivalence classes f_U, g_U of functions are related in $\Pi_U\ \mathfrak{M}^{\mathsf{s}}$ iff there exist $f_0 \in f_U, g_0 \in g_U$, such that $\{\, j\ \in\ J \mid (\mathfrak{M}^{\mathsf{s}}\ j)^R\ (f_0\ j)\ (g_0\ j)\ \}$ is in U.
- A propositional letter p is satisfied at f_U in $\Pi_U\ \mathfrak{M}^{\mathsf{s}}$ iff there exists $f_0 \in f_U$ such that $\{\, j \mid j\ \in\ J\ \wedge\ f_0\ j\ \in\ (\mathfrak{M}^{\mathsf{s}}\ j)^V\ p\ \}$ is in U.

Definition 112 [1, Definition 2.70 (Ultraproduct of Modal Models)]

$$\begin{array}{l}
\mathsf{ultraproduct_model}\ U\ J\ \mathfrak{M}^{\mathsf{s}} \stackrel{\mathsf{def}}{=} \\
\quad \langle\!\langle \mathsf{frame}\ := \\
\quad\quad \langle\!\langle \mathsf{world}\ :=\ \mathsf{ultraproduct}\ U\ J\ (\mathsf{worlds}\ \mathfrak{M}^{\mathsf{s}}); \\
\quad\quad\quad \mathsf{rel}\ := \\
\quad\quad\quad\quad (\lambda f_U\ g_U. \\
\quad\quad\quad\quad\quad \exists f_0\ g_0. \\
\quad\quad\quad\quad\quad\quad f_0\ \in\ f_U\ \wedge\ g_0\ \in\ g_U\ \wedge \\
\quad\quad\quad\quad\quad\quad \{\, j \mid j\ \in\ J\ \wedge\ (\mathfrak{M}^{\mathsf{s}}\ j)^R\ (f_0\ j)\ (g_0\ j)\ \}\ \in\ U) \rangle\!\rangle; \\
\quad \mathsf{valt}\ := \\
\quad\quad (\lambda p\ f_U.\ \exists f_0.\ f_0\ \in\ f_U\ \wedge\ \{\, j \mid j\ \in\ J\ \wedge\ f_0\ j\ \in\ (\mathfrak{M}^{\mathsf{s}}\ j)^V\ p\ \}\ \in\ U) \rangle\!\rangle
\end{array}$$

Here worlds is the function that takes a family of models to the family of their world sets:

$$\mathsf{worlds}\ \mathfrak{M}^{\mathsf{s}} \stackrel{\mathsf{def}}{=} (\lambda j.\ (\mathfrak{M}^{\mathsf{s}}\ j)^{W})$$

As $\sim_U^A$ is an equivalence relation, if one element in an equivalence class satisfies the required condition, then all the elements in the equivalence class will satisfy the condition. Therefore, if we replace all the existential quantifiers with universal quantifiers in the above definition, the construction is still valid, and will give the same model as the current definition.

The critical result we will need about ultraproducts of modal models is a modal version of the fundamental theorem of ultraproducts, which is also called Łoś's theorem.

Theorem 18 (`Los_modal_thm`)

$$\begin{array}{l}\vdash \mathsf{ultrafilter}\ U\ J\ \wedge\ f_U\ \in\ (\Pi_U\ \mathfrak{M}^{\mathsf{s}})^W\ \Rightarrow \\ \quad (\Pi_U\ \mathfrak{M}^{\mathsf{s}}, f_U \Vdash\ \phi\ \Longleftrightarrow\ \exists f_0.\ f_0\ \in\ f_U\ \wedge\ \{\ j \mid j\ \in\ J\ \wedge\ \mathfrak{M}^{\mathsf{s}}\ j, f_0\ j \Vdash\ \phi\ \}\ \in\ U)\end{array}$$

Although it is possible to derive the theorem above from Łoś's theorem of first-order models using standard translation. We proved it directly by induction on modal formulas in HOL.

Ultraproduct for first-order models Given a family $\mathfrak{M}^{\mathsf{s}}$ of first-order models indexed by J and an ultrafilter U on J, the ultraproduct model of $\mathfrak{M}^{\mathsf{s}}$ modulo U (notation : ${}^{\mathsf{f}}\Pi_U\ \mathfrak{M}^{\mathsf{s}}$) is given by:

- The domain is the ultraproduct of the domains of $\mathfrak{M}^{\mathsf{s}}$ over U on J.
- A function with its symbol denoted by the natural number n will send a list zs of equivalence classes to the equivalence class of a function that sending $j \in J$ to $(\mathfrak{M}^{\mathsf{s}}\ j).\mathsf{Fun}\ n\ l$, where the k-th member of the list l is a representative of the k-th member (which is an equivalence class) of zs.
- A predicate with its symbol denoted by p will hold for a list zs of equivalence classes if and only if once we have a list zr such that the k-th member is a representative of the k-th member of zs, the set of elements $j\ \in\ J$ such that $(\mathfrak{M}^{\mathsf{s}}\ j).\mathsf{Pred}\ p\ zr$ is in U.

Definition 113 [1, Definition A.18 (Ultraproduct of First-Order Models)]

$$\begin{array}{l}{}^{\mathsf{f}}\Pi_U\ \mathfrak{M}^{\mathsf{s}} \stackrel{\mathsf{def}}{=} \\ \ \langle\!\langle \mathsf{Dom}\ :=\ \mathsf{ultraproduct}\ U\ J\ (\mathsf{Doms}\ \mathfrak{M}^{\mathsf{s}}); \\ \ \mathsf{Fun}\ := \\ \ \ (\lambda n\ zs. \\ \ \ \ \{\ y \mid \\ \ \ \ \ (\forall j.\ j\ \in\ J\ \Rightarrow\ y\ j\ \in\ (\mathfrak{M}^{\mathsf{s}}\ j).\mathsf{Dom})\ \wedge \\ \ \ \ \ \{\ j \mid j\ \in\ J\ \wedge\ y\ j\ =\ (\mathfrak{M}^{\mathsf{s}}\ j).\mathsf{Fun}\ n\ (\mathsf{MAP}\ (\lambda f_U.\ \mathsf{CHOICE}\ f_U\ j)\ zs)\ \}\ \in \\ \ \ \ \ U\ \}); \\ \ \mathsf{Pred}\ := \\ \ \ (\lambda p\ zs.\ \{\ j \mid j\ \in\ J\ \wedge\ (\mathfrak{M}^{\mathsf{s}}\ j).\mathsf{Pred}\ p\ (\mathsf{MAP}\ (\lambda f_U.\ \mathsf{CHOICE}\ f_U\ j)\ zs)\ \}\ \in\ U)\rangle\!\rangle\end{array}$$

Here we fix the representative of each equivalence class f_U to be $\mathsf{CHOICE}\ f_U$. The function Doms takes a family of first-order models to the family of their domains.

The semantic behavior of ultraproduct models are characterized by *Łoś's theorem*, whose proof can be founded in [2].

Theorem 19 [1, Theorem A.19 (Łoś's theorem)]

$$\vdash \mathsf{ultrafilter}\ U\ J\ \wedge\ \mathsf{valuation}\ ({}^{\mathsf{f}}\Pi_U\ \mathfrak{M}^{\mathsf{s}})\ \sigma\ \wedge\ (\forall j.\ j\ \in\ J\ \Rightarrow\ \mathsf{wffm}\ (\mathfrak{M}^{\mathsf{s}}\ j))\ \Rightarrow$$
$$\mathsf{termval}\ ({}^{\mathsf{f}}\Pi_U\ \mathfrak{M}^{\mathsf{s}})\ \sigma\ t\ =$$
$$\{\ f \mid f \sim_U^{\mathsf{Doms}\ \mathfrak{M}^{\mathsf{s}}}\ (\lambda j.\ \mathsf{termval}\ (\mathfrak{M}^{\mathsf{s}}\ j)\ (\lambda v.\ \mathsf{CHOICE}\ (\sigma\ v)\ j)\ t)\ \}$$

$$\vdash \mathsf{ultrafilter}\ U\ J\ \wedge\ \mathsf{valuation}\ ({}^{\mathsf{f}}\Pi_U\ \mathfrak{M}^{\mathsf{s}})\ \sigma\ \wedge$$
$$(\forall j.\ j\ \in\ J\ \Rightarrow\ \mathsf{wffm}\ (\mathfrak{M}^{\mathsf{s}}\ j))\ \Rightarrow$$
$$({}^{\mathsf{f}}\Pi_U\ \mathfrak{M}^{\mathsf{s}}, \sigma \vDash\ \phi\ \Longleftrightarrow$$
$$\{\ j \mid j\ \in\ J\ \wedge\ \mathfrak{M}^{\mathsf{s}}\ j, (\lambda v.\ \mathsf{CHOICE}\ (\sigma\ v)\ j) \vDash\ \phi\ \}\ \in\ U)$$

5 Ultrafilter Extensions

[2p] The first application of the theory of ultrafilters above is to construct the ultrafilter extension of a model, which has the nice property of being M-saturated. To define M-saturation, we give the following three definitions (the first two are called *finitely satisfiable*, *satisfiable*) consecutively:

Definition 114 [1, Definition 2.53]

$$\mathsf{fin\_satisfiable\_in}\ \Sigma\ X\ \mathfrak{M} \stackrel{\mathsf{def}}{=} \forall S.\ S\ \subseteq\ \Sigma\ \wedge\ \mathsf{FINITE}\ S\ \Rightarrow\ \mathsf{satisfiable\_in}\ S\ X\ \mathfrak{M}$$

$$\mathsf{satisfiable\_in}\ \Sigma\ X\ \mathfrak{M} \stackrel{\mathsf{def}}{=} X\ \subseteq\ \mathfrak{M}^W\ \wedge\ \exists w.\ w\ \in\ X\ \wedge\ \forall \phi.\ \phi\ \in\ \Sigma\ \Rightarrow\ \mathfrak{M}, w \Vdash\ \phi$$

$$\mathsf{M\_sat}\ \mathfrak{M} \stackrel{\mathsf{def}}{=}$$
$$\forall w\ \Sigma.$$
$$w\ \in\ \mathfrak{M}^W\ \wedge\ \mathsf{fin\_satisfiable\_in}\ \Sigma\ \{\ v \mid v\ \in\ \mathfrak{M}^W\ \wedge\ \mathfrak{M}^R\ w\ v\ \}\ \mathfrak{M}\ \Rightarrow$$
$$\mathsf{satisfiable\_in}\ \Sigma\ \{\ v \mid v\ \in\ \mathfrak{M}^W\ \wedge\ \mathfrak{M}^R\ w\ v\ \}\ \mathfrak{M}$$

For M-saturated models, bisimulation and modal equivalence coincides:

Proposition 110 [1, Proposition 2.54]

$$\vdash \mathsf{M\_sat}\ \mathfrak{M}_1\ \wedge\ \mathsf{M\_sat}\ \mathfrak{M}_2\ \wedge\ w_1\ \in\ \mathfrak{M}_1^W\ \wedge\ w_2\ \in\ \mathfrak{M}_2^W\ \wedge\ \mathfrak{M}_1, w_1 \leftrightsquigarrow \mathfrak{M}_2, w_2\ \Rightarrow$$
$$\mathfrak{M}_1, w_1 \leftrightarrows \mathfrak{M}_2, w_2$$

Given a model $\mathfrak{M}$, its ultrafilter extension ${}^{ue}\mathfrak{M}$ is defined as:

Definition 115 [1, Definition 2.57 (Ultrafilter Extension)]

$$
\begin{array}{l}
{}^{ue}\mathfrak{M} \stackrel{\text{def}}{=} \\
\quad \langle\!\langle \mathsf{frame} := \\
\quad\quad \langle\!\langle \mathsf{world} := \{\, u \mid \mathsf{ultrafilter}\ u\ \mathfrak{M}^W \,\}; \\
\quad\quad\quad \mathsf{rel} := \\
\quad\quad\quad\quad (\lambda\, u\ v. \\
\quad\quad\quad\quad\quad \mathsf{ultrafilter}\ u\ \mathfrak{M}^W \ \wedge\ \mathsf{ultrafilter}\ v\ \mathfrak{M}^W \ \wedge \\
\quad\quad\quad\quad\quad \forall X.\ X \in v \Rightarrow \mathfrak{M}_{\Diamond}(X) \in u) \rangle\!\rangle; \\
\quad \mathsf{valt} := (\lambda\, p\ v.\ \mathsf{ultrafilter}\ v\ \mathfrak{M}^W \ \wedge\ \{\, w \mid w \in \mathfrak{M}^W \ \wedge\ \mathfrak{M}^V\ p\ w \,\} \in v) \rangle\!\rangle
\end{array}
$$

Using the ultrafilter theorem and some basic properties about ultrafilters, we derive:

Proposition 111 [1, Proposition 2.59 (i)]

$$
\begin{array}{l}
\vdash \mathsf{ultrafilter}\ u\ \mathfrak{M}^W \Rightarrow \\
\quad (\{\, w \mid w \in \mathfrak{M}^W \ \wedge\ \mathfrak{M}, w \Vdash \phi \,\} \in u \iff {}^{ue}\mathfrak{M}, u \Vdash \phi)
\end{array}
$$

In particular, every world $w \in \mathfrak{M}^W$ is embedded as the principal filter $\pi_w^{\mathfrak{M}^W}$ on $\mathfrak{M}^W$ generated by w in the ultrafilter extension or $\mathfrak{M}$. Also, the above leads to the proof of the fact that the ultrafilter extension of every model is M-saturated. The M-saturatedness of ultrafilter extensions together with 110 immediately gives the central result about ultrafilter extension: bisimularity of worlds in a model $\mathfrak{M}$ can be characterized as bisimularity in ${}^{ue}\mathfrak{M}$.

Theorem 112 [1, Proposition 2.62]

$$
\begin{array}{l}
\vdash w_1 \in \mathfrak{M}_1^W \ \wedge\ w_2 \in \mathfrak{M}_2^W \Rightarrow \\
\quad (\mathfrak{M}_1, w_1 \leftrightsquigarrow \mathfrak{M}_2, w_2 \iff {}^{ue}\mathfrak{M}_1, \pi_{w_1}^{\mathfrak{M}_1^W} \leftrightarrows {}^{ue}\mathfrak{M}_2, \pi_{w_2}^{\mathfrak{M}_2^W})
\end{array}
$$

6 Countably Saturatedness of Ultrapower Models

[2p]

Given a first-order model $\mathfrak{M}$ with no information about interpretation of function symbols, we can expand the model $\mathfrak{M}$ by adding the interpretation of some function symbols. For our propose, we are only interested in adding the interpretation of finitely many nullary function symbols, which are also called *constants*. We write $\mathsf{is\_expansion}\ \mathfrak{M}\ A\ \mathfrak{M}'\ f$ to mean that $\mathfrak{M}'$ is the result of adding each element in A to $\mathfrak{M}$ as a new constant. Further, the function f is a bijection between $\{0, \cdots, n-1\}$ and A, which is assumed to be finite, so that each nullary function symbol c will be interpreted as $f\ c$ in $\mathfrak{M}'$.

Definition 116 [1, Definition A.9 (Expansion)]

$$
\begin{array}{l}
\mathsf{is\_expansion}\ \mathfrak{M}\ A\ \mathfrak{M}'\ f \stackrel{\text{def}}{=} \\
\quad \mathfrak{M}'.\mathsf{Dom} = \mathfrak{M}.\mathsf{Dom} \ \wedge\ \mathsf{BIJ}\ f\ (\mathsf{count}\ (\mathsf{CARD}\ A))\ A \ \wedge \\
\quad \mathfrak{M}'.\mathsf{Fun} = \\
\quad\quad (\lambda\, c\ l.\ \texttt{if}\ c < \mathsf{CARD}\ A \ \wedge\ l = []\ \texttt{then}\ f\ c\ \texttt{else}\ \mathsf{CHOICE}\ \mathfrak{M}.\mathsf{Dom}) \ \wedge \\
\quad \mathfrak{M}'.\mathsf{Pred} = \mathfrak{M}.\mathsf{Pred}
\end{array}
$$

If $\mathfrak{M}$, A and f are all fixed, then the model $\mathfrak{M}'$ such that is_expansion $\mathfrak{M}$ A $\mathfrak{M}'$ f is unique. The only difference between a model and an expansion of it is the interpretation of function symbols.

A set Σ of first-order formulas is called *consistent* with a model $\mathfrak{M}$ if for every finite subset $\Sigma_0 \subseteq \Sigma$, there exists a valuation of $\mathfrak{M}$ such that all elements of Σ_0 are satisfied, in this case, we write consistent $\mathfrak{M}$ Σ. A set Γ of first-order formula is an *x-type* if for each formula in Γ, the only free variable that may contain is x. In this case, we write 'ftype x Γ' in HOL. If Γ is an x-type, when evaluating formulas in Γ, the valuations will only control where the only free variable x goes to. We say Γ is *realized* in $\mathfrak{M}$ if there is an element w in the domain of $\mathfrak{M}$ such that $\mathfrak{M}, (\lambda\, v.\ w) \vDash \phi$ for all $\phi \in \Gamma$. In this case, we write 'frealizes $\mathfrak{M}$ x Γ' in HOL. Let $\mathfrak{M}$ be a model and n be a natural number. For every $A \subseteq \mathfrak{M}$.Dom, with $|A| < n$ and for every $f : \mathbb{N} \to \mathfrak{M}$.Dom, there is a unique $\mathfrak{M}'$ such that is_expansion $\mathfrak{M}$ A $\mathfrak{M}'$ f. If every such $\mathfrak{M}'$ realizes every x-type Γ, then we say $\mathfrak{M}$ is n-saturated. In HOL:

Definition 117 [1, Definition 2.63 (n-Saturated)]

n_saturated $\mathfrak{M}$ n $\stackrel{\text{def}}{=}$
$\forall\, A\ \mathfrak{M}'\ \Gamma\ x\ f.$
IMAGE f $\mathcal{U}(:num) \subseteq \mathfrak{M}$.Dom $\wedge$ FINITE A $\wedge$ CARD $A \leq n$ $\wedge$ $A \subseteq \mathfrak{M}$.Dom $\wedge$
is_expansion $\mathfrak{M}$ A $\mathfrak{M}'$ f $\wedge$
$(\forall\, \phi.\ \phi \in \Gamma \Rightarrow$ form_functions $\phi \subseteq \{\, (c, 0) \mid c < \text{CARD } A \,\})$ $\wedge$ ftype x Γ $\wedge$
consistent $\mathfrak{M}'$ Γ $\Rightarrow$
frealizes $\mathfrak{M}'$ x Γ

We say $\mathfrak{M}$ is countably saturated, and write countably_saturated $\mathfrak{M}$ if $\mathfrak{M}$ is n-saturated for every natural number n. The ultimate goal is to prove a lemma to be used in the proof of Van Benthem characterization theorem: For a family of non-empty models, their ultraproduct on a countably incomplete ultrafilter is *countably saturated*.

Lemma 113 [1, Lemma 2.73]

$$\vdash \text{countably\_incomplete } U\ J \wedge (\forall\, j.\ j \in J \Rightarrow (\mathfrak{M}^{\mathfrak{s}}\ j)^W \neq \emptyset) \Rightarrow \text{countably\_saturated } (\text{mm2folm } (\Pi_U\ \mathfrak{M}^{\mathfrak{s}}))$$

Here a countably incomplete ultrafilter is an ultrafilter that contains a countably infinite family that intersects to the empty set. We proved in HOL that such ultrafilters do exists using 17. The above theorem is not simply a direct consequence from the Łoś's theorem: Łoś's theorem is about ultraproducts of first-order models, and it says nothing about expansion. But to prove 113, we are required to prove a statement for a model obtained by expanding a first-order model which is again obtained by viewing an ultraproduct of modal models as a first-order model.

To deal with this issue, the key observation is that constants are nothing more than forcing some symbols to be sent to some points in a model under

every valuation, hence rather than use nullary function symbols, we fixed a set of variable letters, each corresponds to a function symbol, and only consider the valuations that sends these variable letters to fixed certain points. With this idea, we can remove all the constants in a formula, and hence change our scope from an expanded model back to the unexpanded model. To get rid of the constants $\{0, \cdots, n-1\}$, we replace every $\mathsf{VAR}\ m$ with $\mathsf{VAR}\ (m\ +\ n)$, and replace every constant ${}^{\mathsf{f}}\mathsf{Fn}\ c\ []$ by $\mathsf{VAR}\ c$. This operation is done by the function shift_form which takes a natural number (the number of constants we want to remove), and a first-order formula (where the only function symbols may appear are the constants $0, \cdots, n-1$). Since $0, \cdots, n-1$ in a shifted formula are now designed to be sent to fixed places $f\ 0, \cdots, f\ (n-1)$, it does not make sense to assign these variable symbols anywhere else. Therefore, to talk about evaluation of shifted formula, the first thing is to make sure that the valuations we are considering send the variables which actually denotes constants to the right place. Hence we shift the valuations accordingly. The definition of shifting on valuation together with the semantical behaviour are displayed below:

Definition 118 (Shifting on valuations, `expansion_shift_feval`)

$$\mathsf{shift\_valuation}\ n\ \sigma\ f \stackrel{\mathsf{def}}{=} (\lambda v.\ \mathtt{if}\ v\ <\ n\ \mathtt{then}\ f\ v\ \mathtt{else}\ \sigma\ (v\ -\ n))$$

$$\begin{aligned}\vdash\ &\mathsf{is\_expansion}\ (\mathsf{mm2folm}\ \mathfrak{M})\ A\ \mathfrak{M}'\ f\ \wedge\ \mathsf{valuation}\ (\mathsf{mm2folm}\ \mathfrak{M})\ \sigma\ \wedge\\ &\mathsf{form\_functions}\ \phi\ \subseteq\ \{\ (c_1, 0)\mid c_1\ <\ \mathsf{CARD}\ A\ \}\ \Rightarrow\\ &(\mathfrak{M}', \sigma \vDash\ \phi\ \Longleftrightarrow\\ &\quad \mathsf{mm2folm}\ \mathfrak{M}, \mathsf{shift\_valuation}\ (\mathsf{CARD}\ A)\ \sigma\ f \vDash\ \mathsf{shift\_form}\ (\mathsf{CARD}\ A)\ \phi)\end{aligned}$$

The shifting construction gets us out of the expansion, leaving us a model obtained by converting a ultraproduct modal model to a first-order model. To apply Łoś's theorem on such a model, we prove by induction that:

Proposition 114 (`ultraproduct_comm_feval`)

$$\begin{aligned}\vdash\ &\mathsf{ultrafilter}\ U\ J\ \wedge\ \mathsf{form\_functions}\ \phi\ =\ \emptyset\ \wedge\ \mathsf{valuation}\ (\mathsf{mm2folm}\ (\Pi_U\ \mathfrak{M}^{\mathsf{s}}))\ \sigma\ \Rightarrow\\ &(\mathsf{mm2folm}\ (\Pi_U\ \mathfrak{M}^{\mathsf{s}}), \sigma \vDash\ \phi\ \Longleftrightarrow\ {}^{\mathsf{f}}\Pi_U\ (\lambda j.\ \mathsf{mm2folm}\ (\mathfrak{M}^{\mathsf{s}}\ j)), \sigma \vDash\ \phi)\end{aligned}$$

By 118 and 114, proving 113 amounts to prove:

Lemma 115 (Saturation of ultraproduct model, `ultraproduct_sat`)

$$\begin{aligned}\vdash\ &\mathsf{countably\_incomplete}\ U\ J\ \wedge\ \mathsf{valuation}\ ({}^{\mathsf{f}}\Pi_U\ \mathfrak{M}^{\mathsf{s}})\ f\ \wedge\\ &(\forall j.\ j\ \in\ J\ \Rightarrow\ \mathsf{wffm}\ (\mathfrak{M}^{\mathsf{s}}\ j))\ \wedge\\ &(\forall \phi.\ \phi\ \in\ \Delta\ \Rightarrow\ \mathcal{L}^1_\tau\ \phi\ \wedge\ \mathsf{FV}\ \phi \setminus C\ \subseteq\ \{\ x\ \})\ \wedge\\ &(\forall \Delta_0.\\ &\quad \mathsf{FINITE}\ \Delta_0\ \wedge\ \Delta_0\ \subseteq\ \Delta\ \Rightarrow\\ &\quad \exists \sigma.\\ &\quad\quad \mathsf{valuation}\ ({}^{\mathsf{f}}\Pi_U\ \mathfrak{M}^{\mathsf{s}})\ \sigma\ \wedge\\ &\quad\quad (\forall c.\ c\ \in\ C\ \Rightarrow\ \sigma\ c\ =\ f\ c)\ \wedge\\ &\quad\quad \forall \phi.\ \phi\ \in\ \Delta_0\ \Rightarrow\ {}^{\mathsf{f}}\Pi_U\ \mathfrak{M}^{\mathsf{s}}, \sigma \vDash\ \phi)\ \Rightarrow\\ &\exists \sigma.\\ &\quad \mathsf{valuation}\ ({}^{\mathsf{f}}\Pi_U\ \mathfrak{M}^{\mathsf{s}})\ \sigma\ \wedge\ (\forall c.\ c\ \in\ C\ \Rightarrow\ \sigma\ c\ =\ f\ c)\ \wedge\\ &\quad \forall \phi.\ \phi\ \in\ \Delta\ \Rightarrow\ {}^{\mathsf{f}}\Pi_U\ \mathfrak{M}^{\mathsf{s}}, \sigma \vDash\ \phi\end{aligned}$$

The proof of the above lemma requires rather heavy construction and can be found in [2].

7 Van Benthem's Characterization Theorem

[2p] Note that the standard translation of any modal formula only can only contain unary predicate symbols which corresponds to propositional letters, one binary predicate symbol which corresponds to the relation, and no function symbol. A first-order formula which only use these symbols is called an $\mathcal{L}^1_\tau$-formula. An $\mathcal{L}^1_\tau$-formula which contains only one free variable is called *invariant under bisimulation* if for all models $\mathfrak{M}$ and $\mathfrak{N}$ with $w \in \mathfrak{M}^W$ and $v \in \mathfrak{N}^W$, if there exists a bisimulation relation between $\mathfrak{M}$ and $\mathfrak{N}$ relating w and v, then ϕ holds at w if and only if it holds at v when both $\mathfrak{M}$ and $\mathfrak{N}$ are viewed as first-order models.

Definition 119 [1, Definition 2.67 (Invariant for Bisimulations)]

$$
\begin{array}{l}
\mathsf{invar4bisim}\ (x : \mathit{num})\ (:\alpha)\ (:\beta)\ (\phi : \mathit{folform}) \stackrel{\mathsf{def}}{=} \\
\quad \mathsf{FV}\ \phi \subseteq \{\ x\ \} \wedge \mathcal{L}^1_\tau\ \phi \wedge \\
\quad \forall\, (\mathfrak{M} : (\mathit{num},\ \alpha)\ \mathit{model})\ (\mathfrak{N} : (\mathit{num},\ \beta)\ \mathit{model})\ (v : \beta)\ (w : \alpha). \\
\qquad \mathfrak{M}, w \leftrightarrows \mathfrak{N}, v \Rightarrow \\
\qquad\quad (\mathsf{mm2folm}\ \mathfrak{M}, (\lambda\, (x : \mathit{num}).\ w) \vDash \phi \iff \\
\qquad\qquad \mathsf{mm2folm}\ \mathfrak{N}, (\lambda\, (x : \mathit{num}).\ v) \vDash \phi)
\end{array}
$$

By the same reason as ??, the type parameters here is necessasy. Although it is possible to prove theorems for different types α and β in the above definition, we will only consider the case that α and β are the same when proving theorems.

The Van Benthem characterization theorem says an $\mathcal{L}^1_\tau$ formula with at most one free variable x is invariant under bisimulation precisely when it is equivalent to the standard translation of some modal formula at x. It is immediate from 11 that every such formula which is equivalent to a standard translation is invariant for bisimulation. We cannot prove it as an 'if and only if' statement, since according to the proofs in [1], we can only prove the two directions separately as:

Proposition 116 [1, Theorem 2.68, as two separate directions]

$$
\begin{array}{l}
\vdash \mathsf{FV}\ (\delta : \mathit{folform}) \subseteq \{\ (x : \mathit{num})\ \} \wedge \mathcal{L}^1_\tau\ \delta \wedge \delta\ {}^{\mathsf{f}}\!\equiv_{(:\alpha)}\ \mathsf{ST}_x\ (\phi : \mathit{num}\ \mathit{form}) \Rightarrow \\
\quad \mathsf{invar4bisim}\ x\ (:\alpha)\ (:\alpha)\ \delta
\end{array}
$$

$$
\begin{array}{l}
\vdash \mathsf{INFINITE}\ \mathcal{U}(:\alpha) \wedge \\
\quad \mathsf{invar4bisim}\ (x : \mathit{num})\ (:(\mathit{num} \to \alpha) \to \mathit{bool})\ (:(\mathit{num} \to \alpha) \to \mathit{bool}) \\
\qquad (\delta : \mathit{folform}) \Rightarrow \\
\qquad \exists\, (\phi : \mathit{num}\ \mathit{form}).\ \delta\ {}^{\mathsf{f}}\!\equiv_{(:\alpha)}\ \mathsf{ST}_x\ \phi
\end{array}
$$

which cannot be put together into a double implication. To see the reason: given an $\mathcal{L}^1_\tau$-formula ϕ with no more then one free variable, by the second theorem

above, if ϕ is invariant under bisimulation for models with $(\texttt{num} \to \alpha) \to \texttt{bool}$-worlds, then ϕ is equivalent to a standard translation on model with α-worlds. However, if we want to prove the converse of this statement, we need to start with the assumption that ϕ is equivalent to a standard translation on models with α-worlds, and prove that ϕ is invariant for bisimulation for models with $(\texttt{num} \to \alpha) \to \texttt{bool}$-world. But by the first theorem above, we can only conclude ϕ is invariant for bisimulation for models of type α. If the type universe of $(\texttt{num} \to \alpha) \to \texttt{bool}$ is small enough to be embedded into α, then we will also done. However, the cardinality of the universe of $(\texttt{num} \to \alpha) \to \texttt{bool}$ is larger than that of α, and hence we cannot derive ϕ is invariant for bisimulation for models with $(\texttt{num} \to \alpha) \to \texttt{bool}$-worlds from the fact that ϕ is invariant for bisimulation for models with α-worlds. If we could quantify over types (as we could in a theorem prover based on dependent type theory), then we can could define 'invariant under bisimulation for models of every type', and hence prove the original statement of Van Benthem characterization theorem.

For the proof of the two theorems above, the first one is immediate from 11, and the second one requires another critical lemma:

Theorem 117 [1, Theorem 2.74, one direction]

$$\begin{array}{l} \vdash w \in \mathfrak{M}^W \wedge v \in \mathfrak{N}^W \wedge (\forall \phi.\, \mathfrak{M}, w \Vdash \phi \iff \mathfrak{N}, v \Vdash \phi) \Rightarrow \\ \quad \exists U\ J. \\ \qquad \mathsf{ultrafilter}\ U\ J \wedge \\ \qquad \Pi_U\ (\lambda j.\, \mathfrak{M}), \{\ f \mid (\lambda j.\, w) \sim_U^{\mathsf{worlds}\ (\lambda j.\, \mathfrak{M})} f\ \} \leftrightarrows \Pi_U\ (\lambda j.\, \mathfrak{N}), \{\ g \mid \\ \qquad (\lambda j.\, v) \sim_U^{\mathsf{worlds}\ (\lambda j.\, \mathfrak{N})} g\ \} \end{array}$$

The above lemma relies on 113 and its mathematical proof can be found in [1].

8 Conclusion

[¡1p]

References

1. Blackburn, P., de Rijke, M., Venema, Y.: Modal Logic. Cambridge University Press (2001)
2. Chang, C.C., Keisler, H.J.: Model Theory. North Holland (1990)