Formalizing Modal Logic in HOL

Yiming Xu

October 2019

A thesis submitted for the degree of Bachelor of Science (Honours)
of the Australian National University

For the following four supervisors/lecturers of mine:
Michael Norrish
Scott Morrison
James Borger
Vigleik Angeltveit

Declaration

The work in this thesis is my own except where otherwise stated.

Yiming Xu

Acknowledgements

Firstly, big thanks to my project main supervisor Michael Norrish and co-supervisor Scott Morrison, as well as the two supervisors/lecturers James Borger and Vigleik Angeltveit who have provided me significant amount of guidance on my studying. The whole list of specific things which I need to thank you about is too long to be displayed here. But overall, I feel that I am the luckiest person in the world for having all of you teaching me all this interesting stuff and helping me developing all these useful skills. I will definitely try my best in my further studying to prove that I deserve all the good education from you. Also, I would thank Joan Licata for being a really helpful and considerate honour convenor, and thank Amnon Neeman for treating us so carefully in his special topic course.

Secondly, thanks to all the honours students. Special thanks to Ameilia Han, Fredrick Yuan, Chris Hone, Mitchell Rowett, Josh Tomlin, Keeley Hoek, Michael Howes, Kelly Maggs, Ben Leedom, Jane Tan, for both collaborating on assignments and chatting.

Moreover, thanks to Robert Culling, for being my tutor from Algebra 1 to Algebra 3, and proofreading several pages. I would also thank Zoey Chen for feeding me.

Finally, thanks to my parents for supporting my studies through all these years.

Contents

Chapter 1

Introduction

There are four questions to answer in order to make sense of our title:

1.1 What is modal logic?

It is hard to find a concise answer to this question. As stated in the textbook 'Modal Logic' by Patrick Blackburn, Maarten de Rijke, and Yde Venema, if you ask three modal logicians, you are likely to get at least three different answers. Therefore, we will begin by asking what is modality. Let us consider first-order logic for a moment. Suppose x is a person. When we say 'x is happy', we are applying the predicate 'is happy' to the person x. But also in our daily conversation, we may say something like 'perhaps x is happy' or 'x must be happy'. Here, 'perhaps' and 'must' are used to describe the 'mode' of the predicate 'is happy', and they are examples of *modalities*. The modalities 'perhaps' and 'must', which are canonically called 'possibly' and 'necessarily' in formal modal logic, are denoted as '$\Diamond$' and '$\Box$' respectively. Let P denote the predicate 'is happy'. In formal language, we can then write 'perhaps x is happy' as '$\Diamond Px$', and write 'x must be happy' as '$\Box Px$'.

In the above discussion, we introduced modalities by considering their semantic meanings. But historically, when logicians start thinking about capturing modalities using formal logic, they enriched propositional logic by adding some extra symbols, called modal operators, together with some axioms describing their behavior, but there was no satisfactory way to define a formal semantics of those modal operators. Before they realized the usefulness of the semantic tools, modal logicians had a hard time attempting to solve problems of distinguishing different systems of axioms. But more than 40 years after the concept of modal logic was established, the usage of Kripke models brought many interesting results to this subject. Work on modal logic using Kripke models is conventionally called ‘modal model theory’, which is exactly what I am formalizing in this project.

Nowadays, modal logic is widely adopted in many disciplines, including, but not limited to, mathematics, philosophy, linguistics, and economics. In particular, the development of modal logic and computer science support each other. With the topics taken from computer science and everywhere else, modal logic is growing rapidly, and we have great chance to see more and more interesting application of this subject to both theoretical research and daily life.

1.2 What is HOL?

A brief overview of HOL can be found in [6]. For a short answer, HOL is an interactive theorem prover: a computer program used to prove mathematical theorems. We stress “interactive”: we do not expect the machine to prove theorems automatically. According to Gödel’s incompleteness theorem, there is no algorithm that can determine the truth value of every mathematical statement. Hence to prove mathematical theorems, we interact with the machine, providing human intelligence and guidance. However, it is also important that interactive theorem provers can and do use automatic techniques. For instance, as there are automatic methods for both first-order logic and Presburger arithmetic, these can be embedded in HOL, making it possible to work at a higher level when interacting with the machine.

There are many theorem provers based on various foundational systems: HOL is based on simple type theory, and there exist other theorem proves based on dependent type theory (e.g. Lean, Coq) and set theory (Mizar, metamath). Each of these systems has their advantages, but there is a trade-off: simple type theory is widely considered as less expressive but rather easy to understand and to be implemented, whereas for more expressive systems, it takes longer for the machine to execute a proof step.

1.3 Why is the combination of the two interesting?

As discussed above, modal model theory has a long history and many theorems in modal model theory have been proved since the usage of Kripke models became popular. Nevertheless, none of these theorems have been machine checked. By formalizing modal model theory in a computer, we will make sure that we have understood every detail of the formalized theorems, find out hidden assumptions, and correct minor errors in their statements. And by formalizing in HOL, we will demonstrate that although simple type theory is a rather weak foundational system and does lack expressiveness, it is still capable of capturing most of the theorems we are interested in. We identify where the lack of expressiveness causes problems.

1.4 What have I done?

My project is to formalize some theorems of modal model theory, based on the first two chapters of the textbook *Modal Logic* [1]. At the beginning of the textbook, the authors give three slogans of this subject:

Slogan 1: Modal languages are simple yet expressive languages for talking about relational structures.

Slogan 2: Modal languages provide an internal, local perspective on relational structures.

Slogan 3: Modal languages are not isolated formal systems.

A reader will see evidence of the three slogans consecutively in this thesis. Chapter 2 and 3 are about formalizing basic properties of modal formulas and their semantic behaviors on models of propositional modal logic. In particular, the locality of a modal language is proved at the end of Chapter 3. From Chapter 4 onwards, modal logic and first-order logic are linked together.

In summary:

- By now, every theorem proved in the book up to section 2.7 that can be captured by the basic modal language and HOL is formalized. The definitions, theorems and proofs are taken from the book, and their statements in HOL are taken to be as close as possible to the original mathematical statements appearing in the book.

- There are some results which are only used but not proved in the textbook. Such results are all formalized, means that they are safe to be quoted for proving things. The most significant part is the work on ultraproducts. This piece of work depends in turn on the work of John Harrison on first-order logic [4], which was done in 1998. The work on ultraproducts is discussed as a interlude in this thesis when it is about to be used.

- Section 2.6 of [1] consists of two parts: characterization and definability. The 'definability' part is not formalized in HOL since it is not possible to capture its statement in HOL. Section 2.7 [1] consists of two parts as well: simulation and safety. The 'safety' part is not formalized in HOL since its statement is not purely about the basic modal language. These two parts are not mentioned in the body of the thesis.

1.5 How to read this thesis

This thesis explains the most interesting parts of the work we have done, and it is as self-contained as possible. We do not assume the reader knows either modal logic or interactive theorem proving, so we will introduce both topics. The approach we have taken to structuring the thesis is explain both topics at the same time.

We explicitly give most of the formal definitions we use, as well as the formal statements of constructions and theorems when necessary. For the sake of length, we only explain the most interesting theorems, and omit those that are less interesting. Though the thesis omits some proofs that are routine or not interesting enough from the theorem-proving aspect, all proofs have been formalized in HOL.

A key role of a human reader is to verify that the formalized statements do actually have the intended meaning. For this reason, for most of the major theorems and definitions, we give both a "human-readable" statement, followed by the pretty-printed statement from the HOL sources, meaning that there is no chance of error in the transfer from checked material to LaTeX. However, the pretty-printing process does turn purely linear computer text into more agreeable printed mathematics, complete with subscripts, superscripts and the like. A reader who wants to fully trust my formalization should carefully compare the English statements and the formalized statements in HOL.

For each definition and each theorem, a clickable link to the HOL sources on Github is provided, where the formal statements can be viewed. We encourage readers to follow at least one hyperlink to see the original text as it was provided to HOL, so the difference between the HOL source and pretty-printed version in the thesis will become clear.

Chapter 2

Getting Started

2.1 HOL syntax

Our theorem prover HOL is based on simple type theory. We are not going to give a convoluted introduction on simple type theory. To read this thesis, the reader only need to know that in simple type theory, whenever we refer to something, it must come with its type. We write $a : \alpha$ to express 'a is a term of the type α'. For a type α, its type universe, which is the set of all the terms of type α, is denoted $\mathcal{U}(:\alpha)$.

In the process of reading this thesis, the reader will get to know how to work with simple type theory. As mentioned in the introduction, the most obvious advantage of simple type theory is its simplicity, which makes most statements in HOL straightforward to read. We can read off the conjunctions, disjunctions and implications in the statement directly. However, there is some special syntax for HOL which is worthwhile to be explained first. While it may be helpful to read this list now, when each instance of this syntax is used for the first time, we will explain it there.

- Inductive types: When defining inductive types, we write bars between the constructors of the type.

- Record types: We put the fields of a record type into '$\langle\!\langle \cdots \rangle\!\rangle$', and separate the fields using ';'. For instance, if we define a type with '$\mathsf{Mytype} = \langle\!\langle\ \mathsf{field1} := \cdots ; \mathsf{field2} := \cdots \rangle\!\rangle$', where the '$\cdots$' will be a type. If A is a term of the type Mytype, we can write $\mathsf{A.field1}$ to get the $\mathsf{field1}$ of A.

- Function application: Unlike what we write in common mathematical textbooks, when we apply a function f of type $\alpha \to \beta$ on a term a of type α, we write $f\ a$ instead of $f(a)$. In turn, this means that function applications can be chained, producing terms such as $f\ a\ b$, which can be read as similar to applying f to two arguments as once, where f's type will be an instance of the pattern $\alpha \ \rightarrow\ \beta \ \rightarrow\ \gamma$. Though it is possible to write functions applied to pairs ($f\ (a, b)$), the "curried" style is more common.

- Predicates as functions: In simple type theory, a predicate *is* a function to the type `bool` consisting of T and F. A predicate P which takes arguments of type α is a term of type $\alpha \ \rightarrow\ \beta$. For a of type α, we write $P\ a$ or $P\ a \iff \mathsf{T}$ to express 'the predicate P is true for a'.

- λ-abstraction: We can use λ-abstraction to define functions. For instance, the function $\lambda\, x.\ x\ +\ 2$ sends x to $x\ +\ 2$. The function $\lambda\, i.\ f\ i$ is the same as the function f, since it means that 'for each i, send i to $f\ i$'.

- Quantification: When using quantifiers in HOL, we put a dot after the thing that we are quantifying over. For example, $\forall\, x.\ P\ x$ reads 'for all x, we have $P\ x$' and $\exists\, x.\ P\ x$ reads 'exists an x such that $P\ x$', where P is a predicate. When quantifying over multiple things, we only write one quantifier at the very beginning. For example, '$\forall\, x\ y.\ R\ x\ y$' reads 'for all x, for all y, we have $R\ x\ y$' and '$\exists\, x\ y.\ R\ x\ y$' reads 'exists x and y such that $R\ x\ y$', where R is a relation.

- Useful functions:
  - CARD: The function CARD takes a set, and gives its cardinality.
  - count: The function count takes a natural number n, and gives the set $\{0, 1, \cdots, n-1\}$.
  - BIJ: The function BIJ takes a term of type $f : \alpha \rightarrow \beta$, an α-set A and a β-set B, and gives the boolean value T if and only if f is a bijection form A to B, similar for the functions INJ and SURJ.
  - CHOICE: The function CHOICE is just the choice function. For a non-empty set X, the only thing we know about $\mathsf{CHOICE}\ X$ is that $\mathsf{CHOICE}\ X\ \in\ X$.
  - $\mathsf{RESTRICT}$: The function $\mathsf{RESTRICT}$ takes a relation on terms of type α and an α-set A, and gives a relation $R\ |_A$ defined as for any term x and y of type α, we have $R\ |_A\ \ x\ y$, which reads 'x and y are related by the relation $R\ |_A$', if and only if $x\ \in\ A$ and $y\ \in\ A$ and $R\ x\ y$.

  - R^* and R^+: For a relation R on α-terms, we use R^* to denote its reflexive and transitive closure, and use R^+ to denote its transitive closure.
  - MAX: The function MAX takes two natural numbers and give the greater one.

- Lists: There are some functions which deal with lists:

  - LENGTH: The function LENGTH takes a list and gives its length, which is a natural number.
  - HD: The function HD takes a list and give the first member of it.
  - EL: The function EL takes a list, a natural number n, and give the n-th member of it (counted from 0).
  - LAST: The function LAST takes a list and gives the last member of it.
  - MAP: The function MAP takes a function of type $\alpha \to \beta$ and an α-list l, gives the β-list such that the n-th member is $f\ a$, where a is the n-th member of l.

2.2 The basic setup

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 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.

In HOL, we create a data type called 'form' of the formulas of this modal language. To define a new inductive type, we give a list of ways to construct terms of the type, separated with the symbol '|'.

Definition 2.1. [1, Definition 1.9] *An α-modal formula as described above is specified formally in HOL as an inductive type:*

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

Note that DISJ is of type $\alpha\ form \to (\alpha\ form \to \alpha\ form)$, which means that it can be regarded as a function that takes two α-modal formulas and gives an α-modal formula. In particular, once DISJ appears, the two arguments after it are always α-modal formula, otherwise it does not make sense. We will write '$\phi_1 \vee \phi_2$' for '$\mathsf{DISJ}\ \phi_1\ \phi_2$' afterwards. We can also regard $\neg$ and $\Diamond$ as functions of type $\alpha\ form \to \alpha\ form$. The functions VAR, DISJ, $(\neg)$, $\Diamond$ together with $\bot$ are called the *constructors* of the type of α-modal formulas. From now on, when we talk about α-modal formula, we will call a term of type α a *propositional letter*. We will just call an α-modal formula an α-formula if no confusion arises.

The non-primitive connectives, the conjunction '$\wedge$', the implication '$\to$', and the truth '$\top$', are defined in a standard way as $\phi_1 \wedge \phi_2 := \neg(\neg\phi_1 \vee \phi_2)$, $\phi_1 \to \phi_2 := \neg\phi_1 \vee \phi_2$ and $\top := \neg\bot$ respectively.

Note that all formulas are of finite size; it is not possible to construct infinite conjunctions or disjunctions.

We have a modal operator that is dual to the diamond: the box $\Box\phi := \neg\Diamond\neg\phi$, as an analogue of the duality between the universal quantifier and the existential quantifier, in the sense that $\exists$ is defined to be $\neg\forall\neg$ in classical logic.

Having defined the syntax of formulas, we can now define their *semantics*. It is easy to come up with a way to interpret formulas which are no more than combinations of propositional letters using the connectives '$\vee$' and '$\neg$'. However, to interpret a modal formula that involves diamonds, we need to assign the syntactical notation '$\Diamond$' an 'actual meaning'.

To interpret modal formulas, we need a *relational structure*. A relational structure consists of a set, which is called the 'set of worlds', and a binary relation on it. Such a relational structure is called a *frame* in the rest of the thesis. If in addition, every world of the frame is equipped with an assignment of truth values on propositional letters, then we will have a *model* of modal formulas. The formula $\Diamond\phi$, where ϕ is a modal formula, is interpreted as 'there exists a world related to the current state where ϕ is true'. Accordingly, '$\Box\phi$' will be interpreted as 'for every point that is related to the current state, ϕ is true'.

For a first example, consider a two-point set $\{a, b\}$, and let the relation be $\{(a, b)\}$. Let the propositional letter p be true on both a and b. Consider the modal formula $\Diamond\mathsf{VAR}\ p$, we say $\Diamond\mathsf{VAR}\ p$ is true at a, since b is a point that is related to a, where the formula $\mathsf{VAR}\ p$ holds. On the other hand, $\Diamond\mathsf{VAR}\ p$ does not hold at b, since there is no world related to b.

Returning to our formalization, we define a frame and a model as follows in HOL:

Definition 2.2. [1, Definition 1.19] *A β-frame consists of a world set and a relation, where the world set has type $\beta \to$ `bool` and the relation has type*

(:$\beta \to \beta \to$ **bool**). *A model for modal logic is a frame together with a function called* valt*. The function* valt *assigns truth values of propositional letters at each world.*

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

Here the $\langle\!\langle\cdots\rangle\!\rangle$ is the notation for defining a structure. When we say a (α, β)-model, we mean a model for α-formulas with a β-set as its underlying set. For a model $\mathfrak{M}$, its field $\mathfrak{M}$.valt will be called the *valuation* in the discussion afterwards. In the rest of the thesis, we use the notations $\mathfrak{M}^W$, $\mathfrak{M}^R$ and $\mathfrak{M}^V$ to denote the world set, the relation, and the valuation of the model $\mathfrak{M}$.

We interpret modal formulas using the function called 'satisfaction'.

Definition 2.3. [1, Definition 1.20] *Satisfaction is a predicate inductively defined on modal formulas, which takes a model $\mathfrak{M}$, a world w in the model, a modal formula, and gives a truth value. We read '$\mathfrak{M}, w \Vdash \phi$' as '$\phi$ is satisfied at the world w in $\mathfrak{M}$'. For $w \in \mathfrak{M}^W$, a propositional letter p is satisfied at w if $\mathfrak{M}^V\ p\ w$ is the boolean value* T*. Falsity is never satisfied, a negation of a formula ϕ is satisfied if ϕ is not satisfied, a disjunction is satisfied if at least one of its disjuncts is satisfied, and $\Diamond\phi$ is satisfied if there exists a world in the model that w is related to where ϕ is satisfied.*

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

Observe that instead of defining the satisfaction of VAR p at w to be $w \in \mathfrak{M}^V\ p$, we include the extra condition that w must live in the underlying set of $\mathfrak{M}$. This is because HOL allows us to write $\mathfrak{M}, w \Vdash \phi$, for every w of the correct type, even if it does not belong to the underlying set of $\mathfrak{M}$. A reader may think that we can define our satisfaction predicate as a function that takes a model $\mathfrak{M}$, and make sure that 'satisfaction on the model $\mathfrak{M}$' is a function from the worlds set of $\mathfrak{M}$ and the set of modal formulas to the set $\{\ \mathsf{T};\ \mathsf{F}\ \}$. We might do this in a dependently typed language, but it is not possible in HOL: we cannot make the domain and the codomain an intrinsic property of a function. The notion of a function from an α-set A to a β-set B is not primitive. Such a function is a term f of type $\alpha \to \beta$, with the additional property that $\forall\, a.\ a \in A \Rightarrow f\ a \in B$.

Though f may satisfy this property, it still has values on elements of α that are not part of the set A.

On each model, the truth value of each modal formula is completely determined by the truth values of the propositional letters appear in it. In HOL, we define a function $\mathsf{prop\_letters}$ that takes a modal formula and gives the set of propositional letters appearing in it, and prove:

Proposition 2.1. [1, Exercise 1.3.1] *If two models $\mathfrak{M}_1$ and $\mathfrak{M}_2$ have the same frame and agree on the valuation on all the propositional letters in ϕ, then ϕ is satisfied at a world w in $\mathfrak{M}_1$ if and only if ϕ is satisfied at w in $\mathfrak{M}_2$.*

$$\begin{aligned}&\vdash \mathfrak{M}_1.\mathsf{frame} \;=\; \mathfrak{M}_2.\mathsf{frame} \;\wedge\; (\forall\, p.\; p \;\in\; \mathsf{prop\_letters}\,\phi \;\Rightarrow\; \mathfrak{M}_1^V\; p \;=\; \mathfrak{M}_2^V\; p) \;\Rightarrow\\ &\quad \forall\, w.\; w \;\in\; \mathfrak{M}_1^W \;\Rightarrow\; (\mathfrak{M}_1, w \Vdash\; \phi \;\iff\; \mathfrak{M}_2, w \Vdash\; \phi)\end{aligned}$$

For two modal formulas using the same type of propositional letters, we have the notion of being *equivalent.*

Definition 2.4 (Equivalence)**.** *If ϕ_1, ϕ_2 are α formulas, $\phi_1 \equiv_{(:\beta)} \phi_2$ means for every (α,β)-model $\mathfrak{M}$ and every world w in it, we have $\mathfrak{M}, w \Vdash\; \phi_1 \iff \mathfrak{M}, w \Vdash\; \phi_2$.*

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

A notable thing is that we need to refer to the type of models when talking about equivalence of formulas. We are not allowed to 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.[1] 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, because of such a specific problem in HOL (actually, in simple type theory), this definition is not encoding the equivalence in mathematical sense precisely, since 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. Such a constraint give rise to some problems in our formalization, as we will see in later chapters.

We can immediately prove that for every type α, if $\phi_1 \equiv_{(:\beta)} \phi_2$ then $\Diamond\phi_1 \equiv_{(:\beta)} \Diamond\phi_2$. If we use set theory as our foundation, then the converse can be proved

[1] See the HOL Logic manual [3] for more details.

very easily: If two diamond formulas $\Diamond\phi_1$ and $\Diamond\phi_2$ are equivalent, then for a contradiction, suppose that ϕ_1 and ϕ_2 are not equivalent, then there exists a model $\mathfrak{M}$ and a world w such that w satisfies f but not g. We can add a world v to the world set of $\mathfrak{M}$ that is only related to w, then v will be a witness of the fact that $\Diamond\phi_1$ and $\Diamond\phi_2$ are not equivalent. But under our definition in HOL, if the $(:\beta)$ is a finite type, the proof is blocked: since we cannot make sure that we can come up with a world v which is not already being used by $\mathfrak{M}$, and hence come up with a fresh world to add to $\mathfrak{M}$ which is only related to w. However, for every model, regardless of its world set is of a finite type or not, we can always create a copy of the model in an infinite type. So it is harmless to only play with equivalence of formulas for models whose underlying set is of an infinite type.

Proposition 2.2 (`equiv0_DIAM`). *For two modal formulas ϕ_1 and ϕ_2, ϕ_1 and ϕ_2 are equivalent on models with β-world sets where β is an infinite type if and only if $\Diamond\phi_1$ and $\Diamond\phi_2$ are equivalent on models with α-world sets.*

$$\vdash \mathsf{INFINITE}\,\mathcal{U}(:\beta) \;\Rightarrow\; (\Diamond\phi_1 \equiv_{(:\beta)} \Diamond\phi_2 \iff \phi_1 \equiv_{(:\beta)} \phi_2)$$

Chapter 3

Invariant Results and Finite Model Property

In this chapter, we talk about some basic results about modal logic. First, we prove some theorems about when modal satisfaction is invariant under operations and relations. And in the second section, we prove the *finite model property* of modal formulas.

3.1 Invariant results

The key concept we are interested in this section is *modal equivalence.*

Proposition 3.1. [1, Definition 2.1 (Modal Equivalent)] *Two worlds $w \in \mathfrak{M}^W$ and $w' \in \mathfrak{M}'^W$ are called to be 'modal equivalent' (notation: $\mathfrak{M}, w \leftrightsquigarrow \mathfrak{M}', w'$) if they satisfy the same set of modal formulas.*

$$\mathfrak{M}, w \leftrightsquigarrow \mathfrak{M}', w' \overset{\mathsf{def}}{=} \forall\, \phi.\ \mathfrak{M}, w \Vdash \phi \iff \mathfrak{M}', w' \Vdash \phi$$

The three parts in this section are about three ways to get modal equivalence, namely via generated submodels, bounded morphisms, and bisimulation. The first two constructions will be proved to be special cases of the third one.

3.1.1 Generated submodels

Given a model, there is an operation that allows us to restrict our scope to a smaller model without changing satisfaction of modal formulas, this is called the 'generated submodel' construction. When we say '$\mathfrak{M}_1$ is a submodel of $\mathfrak{M}_2$', we mean all the information of $\mathfrak{M}_1$ is inherited from that of $\mathfrak{M}_2$.

Definition 3.1. [1, Definition 2.5, Submodels] *By* $\mathsf{submodel}\ \mathfrak{M}_1\ \mathfrak{M}_2$*, we mean:*

- *The world set of* $\mathfrak{M}_1$ *is a subset of the world set of* $\mathfrak{M}_2$*.*
- *For two worlds* w_1, w_2 *in* $\mathfrak{M}_1$*, we have* $\mathfrak{M}_1^R\ w_1\ w_2$ *iff* $\mathfrak{M}_2^R\ w_1\ w_2$*.*
- *For every world of* $\mathfrak{M}_1$*, its valuation of propositional letters is exactly the same as that in* $\mathfrak{M}_2$*.*

$$\begin{array}{l}\mathsf{submodel}\ \mathfrak{M}_1\ \mathfrak{M}_2 \overset{\mathsf{def}}{=} \\ \quad \mathfrak{M}_1^W \subseteq \mathfrak{M}_2^W\ \wedge \\ \quad (\forall\, w_1\ w_2.\ w_1 \in \mathfrak{M}_1^W \wedge w_2 \in \mathfrak{M}_1^W \Rightarrow (\mathfrak{M}_1^R\ w_1\ w_2 \iff \mathfrak{M}_2^R\ w_1\ w_2))\ \wedge \\ \quad \forall\, w_1.\ w_1 \in \mathfrak{M}_1^W \Rightarrow \forall\, v.\ \mathfrak{M}_1^V\ v\ w_1 \iff \mathfrak{M}_2^V\ v\ w_1\end{array}$$

It is not necessary that submodel construction preserves modal satisfaction. Although the clause about relation says that for every pair of worlds w_1, w_2 in $\mathfrak{M}_1$, they are related in $\mathfrak{M}_1$ iff they are related in $\mathfrak{M}_2$, it can be the case that w_1, w_2 are worlds in $\mathfrak{M}_2$ such that $\mathfrak{M}_2^R\ w_1\ w_2$, where w_2 is the only world that w_1 is related to, but $w_1 \in \mathfrak{M}_1^W$ whereas $w_2 \notin \mathfrak{M}_1^W$. As a consequence, if we have $\mathfrak{M}_2, w_2 \Vdash \phi$, then we will have $\mathfrak{M}_2, w_1 \Vdash \Diamond\phi$ but not $\mathfrak{M}_1, w_1 \Vdash \Diamond\phi$ since there is no world in $\mathfrak{M}_1$ such that w_1 is related to. To avoid such situation, we can add an extra constraint to make sure that for each world w in $\mathfrak{M}_2$, if it is included in the world set of $\mathfrak{M}_1$, then every world $w' \in \mathfrak{M}_2^W$ such that $\mathfrak{M}_2^R\ w\ w'$ must also be included to the world set of $\mathfrak{M}_1$. A submodel which satisfies this extra condition is called a generated submodel (notation: '$\mathfrak{M}_1 \rightarrowtail \mathfrak{M}_2$' reads '$\mathfrak{M}_1$ is a generated submodel of $\mathfrak{M}_2$').

Definition 3.2. [1, Definition 2.5, Generated Submodels]

$$\begin{array}{l}\mathfrak{M}_1 \rightarrowtail \mathfrak{M}_2 \overset{\mathsf{def}}{=} \\ \quad \mathsf{submodel}\ \mathfrak{M}_1\ \mathfrak{M}_2\ \wedge \\ \quad \forall\, w_1\ w_2.\ w_1 \in \mathfrak{M}_1^W \wedge w_2 \in \mathfrak{M}_2^W \wedge \mathfrak{M}_2^R\ w_1\ w_2 \Rightarrow w_2 \in \mathfrak{M}_1^W\end{array}$$

Note that for a generated submodel $\mathfrak{M}_1$ of $\mathfrak{M}_2$, for worlds w_1 and w_2 of w_2, if w_1 is included to the world set of $\mathfrak{M}_1$ and $\mathfrak{M}_2^R\ w_1\ w_2$, we must include w_2 to the world set of $\mathfrak{M}_1$ as well. But if $\mathfrak{M}_2^R\ w_1\ w_2$ and w_2 is included to $\mathfrak{M}_1$, we are allowed not to include w_1 to $\mathfrak{M}_1$. This is because the '$\Diamond$' operator in modal formulas cannot 'look back', in the sense that adding extra connections or discard connections *towards* a world w does not change the satisfaction of modal formulas at w.

Generated submodels do preserve modal satisfaction:

Proposition 3.2. [1, Proposition 2.6]

$$\vdash \mathfrak{M}_1 \rightarrowtail \mathfrak{M}_2 \ \wedge \ w \ \in \ \mathfrak{M}_1^W \ \Rightarrow \ (\mathfrak{M}_1, w \Vdash \ \phi \ \iff \ \mathfrak{M}_2, w \Vdash \ \phi)$$

3.1.2 Bounded morphisms

Just as in algebra, it is natural to investigate 'morphisms' between our structures of interest. Here, these structures are the models. For instance, 'homomorphism' is the weakest notion of 'structure-preserving map':

Definition 3.3. [1, Definition 2.7 (Homomorphisms)] *A homomorphism from a model $\mathfrak{M}_1$ to a model $\mathfrak{M}_2$ (notation: $\mathfrak{M}_1 \xrightarrow{f} \mathfrak{M}_2$) is a function from the world set of $\mathfrak{M}_1$ to the world set of $\mathfrak{M}_2$ that preserves relation and valuation.*

$$\begin{aligned}&\mathfrak{M}_1 \xrightarrow{f} \mathfrak{M}_2 \ \stackrel{\text{def}}{=}\\ &\quad (\forall\, w.\ w \ \in \ \mathfrak{M}_1^W \ \Rightarrow \ f \ w \ \in \ \mathfrak{M}_2^W \ \wedge \ \forall\, p.\ w \ \in \ \mathfrak{M}_1^V \ p \ \Rightarrow \ f \ w \ \in \ \mathfrak{M}_2^V \ p) \ \wedge\\ &\quad \forall\, w \ v.\ w \ \in \ \mathfrak{M}_1^W \ \wedge \ v \ \in \ \mathfrak{M}_1^W \ \wedge \ \mathfrak{M}_1^R \ w \ v \ \Rightarrow \ \mathfrak{M}_2^R \ (f \ w) \ (f \ v)\end{aligned}$$

The second clauses only says 'propositional letters are preserved', and the last clause only says 'relations in the source model are preserved by a homomorphism'. We are allowed to have propositional letters satisfied at the target but not at the source, and we can have relations in the target which are not from a relation in the source. Because of these reasons, we cannot guarantee every world and its image in the target satisfy exactly the same set of modal formulas. Actually, there exists more than one notion of morphisms which gives equivalences, but most of these notions are too strong to be interesting. The only one among these notions that we are interested in here is *bounded morphism.*

Definition 3.4. [1, Definition 2.10 (Bounded Morphisms)] *A bounded morphism between two models $\mathfrak{M}_1$ and $\mathfrak{M}_2$ is a function f between their world sets such that:*

- *For every world w of $\mathfrak{M}_1$, it satisfies the same propositional letters as $f \ w$.*
- *If w, v are worlds in $\mathfrak{M}_1$ such that $\mathfrak{M}_1^R \ w \ v$, then we have $\mathfrak{M}_2^R \ (f \ w) \ (f \ v)$ in $\mathfrak{M}_2$.*
- *If $w \ \in \ \mathfrak{M}_1^W$ and we have $\mathfrak{M}_2^R \ (f \ w) \ v'$ for some $v' \ \in \ \mathfrak{M}_2^W$, then we can find a world v in $\mathfrak{M}_1$ such that $\mathfrak{M}_1^R \ w \ v$ and $f \ v \ = \ v'$.*

In HOL:

$$\begin{array}{l}
\mathsf{bounded\_mor}\ f\ \mathfrak{M}_1\ \mathfrak{M}_2 \overset{\mathsf{def}}{=} \\
\quad \forall\, w. \\
\qquad w\ \in\ \mathfrak{M}_1^W\ \Rightarrow \\
\qquad\quad f\ w\ \in\ \mathfrak{M}_2^W\ \wedge\ (\forall\, a.\, \mathfrak{M}_1, w \Vdash\ \mathsf{VAR}\ a\ \iff\ \mathfrak{M}_2, f\ w \Vdash\ \mathsf{VAR}\ a)\ \wedge \\
\qquad\quad (\forall\, v.\ v\ \in\ \mathfrak{M}_1^W\ \wedge\ \mathfrak{M}_1^R\ w\ v\ \Rightarrow\ \mathfrak{M}_2^R\ (f\ w)\ (f\ v))\ \wedge \\
\qquad\quad \forall\, v'.\ v'\ \in\ \mathfrak{M}_2^W\ \wedge\ \mathfrak{M}_2^R\ (f\ w)\ v'\ \Rightarrow\ \exists\, v.\ v\ \in\ \mathfrak{M}_1^W\ \wedge\ \mathfrak{M}_1^R\ w\ v\ \wedge\ f\ v\ =\ v'
\end{array}$$

We read '$\mathsf{bounded\_mor}\ f\ \mathfrak{M}_1\ \mathfrak{M}_2$' as '$f$ is a bounded morphism from $\mathfrak{M}_1$ to $\mathfrak{M}_2$'. From above, the notion of bounded morphism is a strengthen of homomorphism. For a homomorphism, we only need propositional letters to be preserved, but for bounded morphism, we strengthen the condition on propositional letters to be an 'if and only if'. Moreover, we added a 'backward condition' on relations.

Bounded morphism gives modal equivalences, in the following sense:

Proposition 3.3. [1, Proposition 2.14] *If f is a bounded morphism from $\mathfrak{M}_1$ to $\mathfrak{M}_2$, then for each modal formula ϕ and each world w in $\mathfrak{M}_1$, we have $\mathfrak{M}_1, w \Vdash \phi \iff \mathfrak{M}_2, f\ w \Vdash \phi$.*

$$\vdash \mathsf{bounded\_mor}\ f\ \mathfrak{M}_1\ \mathfrak{M}_2\ \wedge\ w\ \in\ \mathfrak{M}_1^W\ \Rightarrow\ (\mathfrak{M}_1, w \Vdash\ \phi\ \iff\ \mathfrak{M}_2, f\ w \Vdash\ \phi)$$

The above result is very useful. As an application, now we use it to prove the *tree-like property* of the basic modal language. The tree-like property says that for each formula ϕ satisfied on some point in some model, there exists a tree-like model such that ϕ is satisfied at the root of the tree. As the name indicates, a tree-like model is a model such that its underlying frame is a tree.

Definition 3.5. [1, Definition 1.7] *The predicate* tree *takes a frame H and a point r, and* tree *H r means that H is a tree with root r. A frame H is a tree with root r if:*

- *We have r is a world in of H.*
- *For any world $w\ \in\ H$.*world*, we have r is related to w via the reflexive and transitive closure of H.*rel*.*
- *For any world $w\ \in\ H$.*world*, it cannot be linked back to the root r via the reflexive and transitive closure of H.*rel*.*
- *For any world $w\ \in\ H$.*world*, it has a unique predecessor.*

In HOL:

$$
\begin{aligned}
&\mathsf{tree}\ H\ r \overset{\mathsf{def}}{=} \\
&\quad r \in H.\mathsf{world} \land (\forall w.\ w \in H.\mathsf{world} \Rightarrow H.\mathsf{rel}\ |_{H.\mathsf{world}}{}^{*}\ r\ w) \land \\
&\quad (\forall w.\ w \in H.\mathsf{world} \Rightarrow \neg H.\mathsf{rel}\ w\ r) \land \\
&\quad \forall w.\ w \in H.\mathsf{world} \land w \neq r \Rightarrow \exists! w_0.\ w_0 \in H.\mathsf{world} \land H.\mathsf{rel}\ w_0\ w
\end{aligned}
$$

In above, for a relation R on terms of type α and an α-set A, we have $R\ |_A\ \ a\ b$ for terms a, b of type α if and only if both a and b are in A, and $R\ a\ b$. We write R^* to denote the reflective and transitive closure of the relation R.

Every tree-like model is *rooted*, where rooted models are just submodel generated by a singleton set. As an instance of generated models, a rooted model needs to be sitting in an ambient model.

Definition 3.6. [1, Definition 2.5 (Rooted Models)] *The predicate '*rooted_model*' takes three parameters: The model itself, the root r, and the ambient model that it is sitting in. We read '*rooted_model $\mathfrak{M}_1$ r $\mathfrak{M}_2$*' as $\mathfrak{M}_1$ is a rooted model with root r sitting in the ambient model $\mathfrak{M}_2$.*

$$
\begin{aligned}
&\mathsf{rooted\_model}\ \mathfrak{M}_1\ r\ \mathfrak{M}_2 \overset{\mathsf{def}}{=} \\
&\ r \in \mathfrak{M}_2^W \land (\forall a.\ a \in \mathfrak{M}_1^W \iff a \in \mathfrak{M}_2^W \land \mathfrak{M}_2^R\ |_{\mathfrak{M}_2^W}{}^{*}\ r\ a) \land \\
&\ (\forall w_1\ w_2.\ w_1 \in \mathfrak{M}_1^W \land w_2 \in \mathfrak{M}_1^W \Rightarrow (\mathfrak{M}_1^R\ w_1\ w_2 \iff \mathfrak{M}_2^R\ |_{\mathfrak{M}_2^W}\ w_1\ w_2)) \land \\
&\ \forall p\ w.\ \mathfrak{M}_1^V\ p\ w \iff \mathfrak{M}_2^V\ p\ w
\end{aligned}
$$

We now prove the tree-like property of modal formulas:

Proposition 3.4. [1, Proposition 2.15]

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

Proof. Suppose $\mathfrak{M}_1, w \Vdash \phi$. Let $\mathfrak{M}_2$ be the rooted model generated by w, then $\mathfrak{M}_2, w \Vdash \phi$ by Proposition 3.2. To find a tree-like model satisfying ϕ, by Proposition 3.3, it suffices to prove $\mathfrak{M}_2$ is the image of some bounded morphism from some tree-like model $\mathfrak{M}_3$ where the root of the tree is mapped to w. Then $\mathfrak{M}_3$ will be the $\mathfrak{M}$ we want. We construct $\mathfrak{M}_3$ as follows: Take the set of worlds to be the finite sequences $[w; u_1; \cdots; u_n]$ such that $\mathfrak{M}_1^R\ u_i\ u_{i+1}$ for all i. Define ${\mathfrak{M}_3}^R\ [w; u_1; \cdots; u_n]\ [w; v_1; \cdots; v_m]$ iff $m = n+1$, $u_i = v_i$ for $1 \leq i \leq n$, and $\mathfrak{M}_2^R$ $u_n\ v_m$. The valuation is given by $[w; u_1; \cdots; u_n] \in {\mathfrak{M}_3}^V\ p$ iff $u_n \in \mathfrak{M}_2^V\ p$. Such a model in HOL looks like:

$$
\begin{array}{l}
\mathfrak{M}_3 \overset{\mathsf{def}}{=} \\
\quad \langle\!\langle \mathsf{frame} \ := \\
\quad\quad \langle\!\langle \mathsf{world} \ := \\
\quad\quad\quad \{\, l \mid \\
\quad\quad\quad\ \ \mathsf{HD}\ l \ = \ w \ \wedge \ \mathsf{LENGTH}\ l \ > \ 0 \ \wedge \\
\quad\quad\quad\ \ \forall\, m.\ m \ < \ \mathsf{LENGTH}\ l \ - \ 1 \ \Rightarrow \ \mathfrak{M}_2^R \mid_{\mathfrak{M}_2^W} \ (\mathsf{EL}\ m\ l)\ (\mathsf{EL}\ (m \ + \ 1)\ l) \,\}\,; \\
\quad\quad\ \ \mathsf{rel} \ := \\
\quad\quad\quad (\lambda\, l_1\ l_2. \\
\quad\quad\quad\quad \mathsf{LENGTH}\ l_1 \ + \ 1 \ = \ \mathsf{LENGTH}\ l_2 \ \wedge \ \mathfrak{M}_2^R \mid_{\mathfrak{M}_2^W} \ (\mathsf{LAST}\ l_1)\ (\mathsf{LAST}\ l_2) \ \wedge \\
\quad\quad\quad\quad \forall\, m.\ m \ < \ \mathsf{LENGTH}\ l_1 \ \Rightarrow \ \mathsf{EL}\ m\ l_1 \ = \ \mathsf{EL}\ m\ l_2) \rangle\!\rangle; \\
\quad \mathsf{valt} \ := \ (\lambda\, v\ n.\ \mathfrak{M}_2^V\ v\ (\mathsf{LAST}\ n)) \rangle\!\rangle
\end{array}
$$

Here the functions HD and LAST give the first and last member of a list, respectively. The function LENGTH gives the length of a list. The function EL takes a natural number n and a list l, and gives the n-th member of l.

The map LAST will sends a list $[w; u_1; \cdots; u_n]$ in ${\mathfrak{M}_3}^W$ to its last member u_n. We can easily check that LAST is a bounded morphism. Also, the root $[w]$ of $\mathfrak{M}_3$ is sent to w in $\mathfrak{M}_2$, as desired.

□

3.1.3 Bisimulation

The two approaches to obtain modal equivalences have a common feature: both of them lead to a relation on worlds in models such that related worlds satisfy exactly the same set of propositional letters, and once we can make a transition in one model, we can make a corresponding transition in the other. This observation leads us to the concept of *bisimulation*:

Definition 3.7. [1, Definition 2.16 (Bisimulations)] *A bisimulation Z between models $\mathfrak{M}_1$ and $\mathfrak{M}_2$ (notation: $\mathfrak{M}_1 \overset{Z}{\Leftrightarrow} \mathfrak{M}_2$) is a relation between their worlds, such that for worlds $w_1 \in \mathfrak{M}_1^W$ and $w_2 \in \mathfrak{M}_2^W$ which are related by Z, we have:*

- *For every propositional letter p, it is satisfied at w_1 if and only if it is satisfied at w_2.*

- *If we have a world $v_1 \in \mathfrak{M}_1^W$ such that w_1 is related to v_1 by the relation in $\mathfrak{M}_1$, then we can find a world v_2 in $\mathfrak{M}_2$ such that v_1 and v_2 are related by Z where w_2 is related to v_2 in $\mathfrak{M}_2$.*

- *If we have a world $v_2 \in \mathfrak{M}_2^W$ such that w_2 is related to v_2 by the relation in $\mathfrak{M}_2$, then we can find a world v_1 in $\mathfrak{M}_1$ such that v_1 and v_2 are related by Z where w_1 is related to v_1 in $\mathfrak{M}_1$.*

In HOL:

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

When there exists a bisimulation relating two worlds $w \in \mathfrak{M}^W$ and $v \in \mathfrak{N}^W$, we say w and v are *bisimilar*, and write '$\mathfrak{M}, w \leftrightarrows \mathfrak{N}, v$'. Both generated submodels and bounded morphic image give rise to bisimulations:

Proposition 3.5. [1, Proposition 2.19, (iii) and (iv)]

$$\vdash \mathfrak{M}_1 \rightarrowtail \mathfrak{M}_2 \ \Rightarrow\ \forall\, w.\ w \in \mathfrak{M}_1^W \ \Rightarrow\ \mathfrak{M}_1, w \leftrightarrows \mathfrak{M}_2, w$$

$$\vdash \mathfrak{M}_1 \overset{f}{\twoheadrightarrow} \mathfrak{M}_2 \ \Rightarrow\ \forall\, w.\ w \in \mathfrak{M}_1^W \ \Rightarrow\ \mathfrak{M}_1, w \leftrightarrows \mathfrak{M}_2, f\ w$$

Proof. The bisimulation relations are given by relating a world in $\mathfrak{M}_1$ to its copy in $\mathfrak{M}_2$ and relating a world in $\mathfrak{M}_1$ to its image in $\mathfrak{M}_2$ respectively. □

Bisimilar worlds are always modal equivalent.

Theorem 3.6. [1, Theorem 2.20]

$$\vdash \mathfrak{M}_1, w_1 \leftrightarrows \mathfrak{M}_2, w_2 \ \Rightarrow\ \mathfrak{M}_1, w_1 \leftrightsquigarrow \mathfrak{M}_2, w_2$$

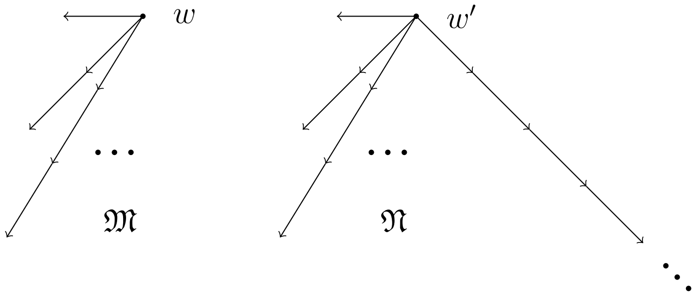
Now we ask if the converse of the above holds: is that the fact that a modal equivalent worlds are always bisimilar? The answer is no, as we have proved in HOL that:

Proposition 3.7. [1, Example 2.23]

$$\vdash \exists\, \mathfrak{M}\ \mathfrak{N}\ w\ v.\ \mathfrak{M}, w \leftrightsquigarrow \mathfrak{N}, v \ \wedge\ \neg(\mathfrak{M}, w \leftrightarrows \mathfrak{N}, v)$$

The models $\mathfrak{M}$ and $\mathfrak{N}$ used in the proof of the theorem above are defined as the picture below, where the arrows denote relations (non-transitive). Both $\mathfrak{M}$ and $\mathfrak{N}$ have infinitely many branches from their roots w, w' respectively. The difference is that $\mathfrak{N}$ has an infinitely long branch, whereas all the branches in $\mathfrak{M}$ are of finite length. In HOL, the worlds of $\mathfrak{M}$ and $\mathfrak{N}$ are captured using pairs of natural numbers. The world $w \in \mathfrak{M}^W$ and $w' \in \mathfrak{N}^W$ are recorded as the pair $(0, 0)$. For a world in $\mathfrak{M}$ or a world in $\mathfrak{N}$ which is the b-th point (counted from the root) on a finite branch of length a, it is recorded as the pair (a, b). The n-th point on the infinite branch in $\mathfrak{N}$ is recorded as the pair $(0, n)$.

Let valuation in $\mathfrak{M}$ and $\mathfrak{N}$ both be such that at every point, there is no propositional letter which is satisfied. The worlds w and w' can be shown to be modal equivalent (using tools which will be introduced in the next section), but they are not bisimilar. Suppose, in order to get a contradiction, that $\mathfrak{M} \overset{Z}{\leftrightarrows} \mathfrak{N}$ and $Z\ w\ w'$, then there exists $v_0 \in \mathfrak{M}^W$ such that $Z\ v_0\ v_0'$, where v_0' is the first world on the infinite branch in $\mathfrak{N}$ such that $\mathfrak{N}^R\ w'\ v_0'$. The branch that v_0 lies on is finitely long, say, the worlds $w, v_0, \cdots, v_n$ are all the worlds on this branch. Then by clause on 'forward condition' in the definition of bisimulation, there are worlds $v_1', \cdots, v_n'$ on the infinite branch of $\mathfrak{N}$ such that $Z\ v_i\ v_i'$ for each $1 \leq i \leq n$. The world v_n' has a successor v_{n+1}' in $\mathfrak{N}$, so the backward clause on relation requires the existence of a world in $\mathfrak{M}$ such that v_n is related to. But such a world does not exist since v_n is at the end of the branch it lies on.



Nonetheless, the converse of Theorem 3.6 does hold on *image finite* models.

Definition 3.8. [1, Page 69, image-finite] *A model $\mathfrak{M}$ is image finite if for every world $w \in \mathfrak{M}^W$, there are only finitely many worlds in $\mathfrak{M}$ that w is related to.*

$$\text{image\_finite}\ \mathfrak{M} \overset{\text{def}}{=} \forall w.\ w \in \mathfrak{M}^W \Rightarrow \text{FINITE}\ \{\ v \mid v \in \mathfrak{M}^W \land \mathfrak{M}^R\ w\ v\ \}$$

Our main theorem is called Hennessy-Milner theorem:

Theorem 3.8. [1, Theorem 2.24 (Hennessy-Milner Theorem)] *For image finite models, modal equivalence and bisimulation are indeed the same thing.*

$$\vdash \mathsf{image\_finite}\ \mathfrak{M}_1 \ \wedge\ \mathsf{image\_finite}\ \mathfrak{M}_2 \ \wedge\ w_1 \in \mathfrak{M}_1^W \ \wedge\ w_2 \in \mathfrak{M}_2^W \ \Rightarrow \\ (\mathfrak{M}_1, w_1 \leftrightsquigarrow \mathfrak{M}_2, w_2 \iff \mathfrak{M}_1, w_1 \rightleftarrows \mathfrak{M}_2, w_2)$$

Proof. We prove the implication from left to right. The other implication is Theorem 3.6. Given that w_1 and w_2 are worlds in $\mathfrak{M}_1$ and $\mathfrak{M}_2$ which are modal equivalent, we prove the relation Z defined as $Z\ v_1\ v_2 \iff \forall\,\phi.\ \mathfrak{M}_1, v_1 \Vdash \phi \iff \mathfrak{M}_2, v_2 \Vdash \phi$ is a bisimulation. The only non-trivial thing to check is that assuming $\mathfrak{M}_1, v_1 \leftrightsquigarrow \mathfrak{M}_2, v_2$ and $\mathfrak{M}_1^R\ v_1\ v_1'$ for some $v_1' \in \mathfrak{M}_1^W$, there exists a world $v_2' \in \mathfrak{M}_2^W$ such that $\mathfrak{M}_2^R\ v_2\ v_2'$ and $\mathfrak{M}_1, v_1' \leftrightsquigarrow \mathfrak{M}_2, v_2'$. Suppose such a v_2' does not exist, we derive a contradiction. Consider the set $S_0 = \{\ u' \mid u' \in \mathfrak{M}_2^W \ \wedge\ \mathfrak{M}_2^R\ v_2\ u'\ \}$ of successors of v_2 , the first claim is that S_0 is finite and nonempty. Finiteness comes from the fact that $\mathfrak{M}_2$ is image finite. Also, and if S_0 is empty, then $\Box\perp$ will be a formula satisfied at v_2 but not at v_1, contradicting the modal equivalence between v_1 and v_2. By assumption, for each world $u' \in S_0$, there is a formula ϕ such that $\mathfrak{M}_1, v_1' \Vdash \phi$ but $\mathfrak{M}_2, u' \nVdash \phi$. As the set S is finite, the set of such ϕs is finite. Then we can take the conjunction of such ϕs to obtain a formula ψ. Then we will have $\mathfrak{M}_1, v_1 \Vdash \Diamond\psi$ but $\mathfrak{M}_1, v_2 \nVdash \Diamond\psi$, contradiction. □

3.2 Finite model property

In this section, we tell the story about Slogan 2 as stated in the introduction: Modal formulas can only capture local information. That is, if a modal formula is satisfied on an arbitrary model, then it can be satisfied on a finite model, where finite model means a model whose world set is finite. Such a result is called the *finite modal property* of modal logic. There are classically two methods of building finite models for satisfiable modal formulas, namely via filtration and selection. Although we have formalized both of them in HOL, the former is almost a direct translation of the mathematical proof and hence is not interesting from the formalizing aspect. We will only talk about finite model property via selection in this section.

In this method, to build a finite model of a satisfiable modal formula ϕ, we start with a model that the formula ϕ is satisfied, delete worlds from the model and only leave finitely many worlds in it. The intuition behind this approach is that every modal formula can only contain finitely many diamonds, each can 'see' one step from the current state. Therefore, each formula can only capture the information of finite depth. To make the notion of 'depth' precise, we define the degree of a modal formula, which counts the number of steps that a modal formula can 'see', as follows:

Definition 3.9. [1, Definition 2.28 (Degree)]

$$\begin{aligned}
\mathsf{DEG}\ (\mathsf{VAR}\ p) &\overset{\text{def}}{=} 0 \\
\mathsf{DEG}\ \bot &\overset{\text{def}}{=} 0 \\
\mathsf{DEG}\ (\neg\phi) &\overset{\text{def}}{=} \mathsf{DEG}\ \phi \\
\mathsf{DEG}\ (\phi_1 \vee \phi_2) &\overset{\text{def}}{=} \mathsf{MAX}\ (\mathsf{DEG}\ \phi_1)\ (\mathsf{DEG}\ \phi_2) \\
\mathsf{DEG}\ (\Diamond\phi) &\overset{\text{def}}{=} \mathsf{DEG}\ \phi\ +\ 1
\end{aligned}$$

In above, the function MAX takes two natural numbers and gives the greater one.

The crucial fact we need about the degree of formulas is that for every finite α-set Φ and every natural number n, there are only finitely many non-equivalent modal formulas of degree up to n which only use the propositional letters in Φ. In the textbook that we are following, the authors prove this fact basically 'by observation', but the proof is long and tedious to formalize (more than 1500 lines in HOL). We will not show the proof, but only show the statement that we have proved in HOL:

Lemma 3.9. [1, Proposition 2.29] *Let Φ be a finite α-set and β be an infinite type. For each natural number n, if we partition the set of α-modal formulas using only propositional letters in Φ of degree up to n using the equivalence relation 'being equivalent on (α, β)-models (models of α-modal formulas with β-world sets)', then we get finitely many equivalence class.*

$$\begin{aligned}
&\vdash \mathsf{FINITE}\ \Phi\ \wedge\ \mathsf{INFINITE}\,\mathcal{U}(:\beta)\ \Rightarrow \\
&\qquad \forall\, n.\ \mathsf{FINITE}\ \{\ \phi \mid \mathsf{DEG}\ \phi\ \le\ n\ \wedge\ \mathsf{prop\_letters}\ \phi\ \subseteq\ \Phi\ \}\ /\!\equiv_{(:\beta)}
\end{aligned}$$

Here $\{\ \phi \mid \mathsf{DEG}\ \phi\ \le\ n\ \wedge\ \mathsf{prop\_letters}\ \phi\ \subseteq\ \Phi\ \}\ /\!\equiv_{(:\beta)}$ is the set of equivalence classes obtained by partitioning the set $\{\ \phi \mid \mathsf{DEG}\ \phi\ \le\ n\ \wedge\ \mathsf{prop\_letters}\ \phi\ \subseteq\ \Phi\ \}$ by the equivalence relation $\equiv_{(:\beta)}$. We require the assumption that the universe of β is infinite since we used Proposition 2.2 when proving the proposition above.

Recall in the last section, we have seen that a bisimulation gives rise to modal equivalence. Modal equivalence means 'satisfying exactly the same formulas', but when we are building a finite model for a formula ϕ, we do not care about the satisfaction of the formulas of degree above DEG ϕ, since such formula cannot affect the satisfaction of ϕ. Therefore, we just need some sort of relations such that related worlds satisfy the same modal formulas up to some degree n. The notion of 'finite approximation of bisimulation', which is called *n-bisimulation*, is used to describe such relations. Let $w_1 \in \mathfrak{M}_1^W$ and $w_2 \in \mathfrak{M}_2^W$, w_1 and w_2 are n-bisimilar if there exists a sequence of relations $Z_n \subseteq \cdots \subseteq Z_0$ such that:

- w_1 and w_2 are related by Z_n
- If $v_1 \in \mathfrak{M}_1^W$ and $v_2 \in \mathfrak{M}_2^W$ are related by Z_0, then v_1 and v_2 satisfy the same propositional letters.
- For $0 \leq i \leq n-1$, if $v_1 \in \mathfrak{M}_1^W$ and $v_2 \in \mathfrak{M}_2^W$ are related by Z_{i+1} and we have $\mathfrak{M}_1^R\ v_1\ u_1$ for $u_1 \in \mathfrak{M}_1^W$, then there exists $u_2 \in \mathfrak{M}_2^W$ such that $\mathfrak{M}_2^R\ v_2\ u_2$ with u_1 and u_2 related by Z_i.
- For $0 \leq i \leq n-1$, if $v_1 \in \mathfrak{M}_1^W$ and $v_2 \in \mathfrak{M}_2^W$ are related by Z_{i+1} and we have $\mathfrak{M}_2^R\ v_2\ u_2$ for $u_2 \in \mathfrak{M}_2^W$, then there exists $u_1 \in \mathfrak{M}_1^W$ such that $\mathfrak{M}_1^R\ v_1\ u_1$ with u_1 and u_2 related by Z_i.

Such a sequence of Z_i is a family of relations indexed by natural numbers from 0 to n. When the world set of $\mathfrak{M}_1$ has type β and the world set of $\mathfrak{M}_2$ has type γ, we encode such a family using a function Z^{s} of type $\texttt{num} \rightarrow \beta \rightarrow \gamma \rightarrow \texttt{bool}$. For each natural number $i \leq n$, applying Z^{s} on i gives us a relation $Z^{\mathsf{s}}\ i$ between terms of type β and γ. In other words, the relation $Z^{\mathsf{s}}\ i$ is the relation Z_i in the usual mathematical notation, and $\mathfrak{M}_1, w_1 \overset{Z^{\mathsf{s}}}{\leftrightarrows}_n \mathfrak{M}_2, w_2$ means w_1 and w_2 are worlds in $\mathfrak{M}_1$ and $\mathfrak{M}_2$ respectively which are n-bisimilar via the family of relations given by Z^{s}, as shown below.

Definition 3.10. [1, Definition 2.30 (n-Bisimulations)]

$$
\begin{array}{l}
\mathfrak{M}_1, w_1 \overset{Z^{\mathsf{s}}}{\underline{\leftrightarrow}}_n \mathfrak{M}_2, w_2 \overset{\mathsf{def}}{=} \\
\quad w_1 \in \mathfrak{M}_1^W \wedge w_2 \in \mathfrak{M}_2^W \wedge \\
\quad (\forall\, m\ a\ b. \\
\qquad a \in \mathfrak{M}_1^W \wedge b \in \mathfrak{M}_2^W \Rightarrow \\
\qquad\quad m + 1 \leq n \Rightarrow Z^{\mathsf{s}}\,(m + 1)\ a\ b \Rightarrow Z^{\mathsf{s}}\ m\ a\ b) \wedge \\
\quad Z^{\mathsf{s}}\ n\ w_1\ w_2 \wedge \\
\quad (\forall\, v_1\ v_2. \\
\qquad v_1 \in \mathfrak{M}_1^W \wedge v_2 \in \mathfrak{M}_2^W \Rightarrow \\
\qquad\quad Z^{\mathsf{s}}\ 0\ v_1\ v_2 \Rightarrow \forall\, p.\, \mathfrak{M}_1, v_1 \Vdash \mathsf{VAR}\ p \iff \mathfrak{M}_2, v_2 \Vdash \mathsf{VAR}\ p) \wedge \\
\quad (\forall\, v_1\ v_2\ u_1\ i. \\
\qquad i + 1 \leq n \wedge v_1 \in \mathfrak{M}_1^W \wedge v_2 \in \mathfrak{M}_2^W \wedge u_1 \in \mathfrak{M}_1^W \wedge \mathfrak{M}_1^R\ v_1\ u_1 \wedge \\
\qquad Z^{\mathsf{s}}\ (i + 1)\ v_1\ v_2 \Rightarrow \\
\qquad\quad \exists\, u_2.\ u_2 \in \mathfrak{M}_2^W \wedge \mathfrak{M}_2^R\ v_2\ u_2 \wedge Z^{\mathsf{s}}\ i\ u_1\ u_2) \wedge \\
\quad \forall\, v_1\ v_2\ u_2\ i. \\
\qquad i + 1 \leq n \wedge v_1 \in \mathfrak{M}_1^W \wedge v_2 \in \mathfrak{M}_2^W \wedge u_2 \in \mathfrak{M}_2^W \wedge \mathfrak{M}_2^R\ v_2\ u_2 \wedge \\
\qquad Z^{\mathsf{s}}\ (i + 1)\ v_1\ v_2 \Rightarrow \\
\qquad\quad \exists\, u_1.\ u_1 \in \mathfrak{M}_1^W \wedge \mathfrak{M}_1^R\ v_1\ u_1 \wedge Z^{\mathsf{s}}\ i\ u_1\ u_2
\end{array}
$$

We will use functions to capture indexed families throughout this thesis. For a family $(A_j)_{j\in J}$ indexed by a set J, where the A_j's are all of the same type, we will capture it using a function A^{s} in HOL such that $A^{\mathsf{s}}\ j$ is A_j. Each function which is used to capture indexing will be decorated with '$^{\mathsf{s}}$' at its right upper corner.

Note that for models $\mathfrak{M}$ and $\mathfrak{N}$, and worlds $w_1 \in \mathfrak{M}_1^W$ and $w_2 \in \mathfrak{M}_2^W$, even if for each natural number n, we have an n-bisimulation between w_1 and w_2, it does not imply that w_1 and w_2 are bisimilar. For the models $\mathfrak{M}$ and $\mathfrak{N}$ used in 3.7, the worlds w and w' are n-bisimilar for all n. Given a natural number n, an n-bisimulation relation Z^{s} can be given as: for each $m \leq n$, $Z^{\mathsf{s}}\ m$ is the relation that relating the points in $\mathfrak{M}$ on the branches of length no more than $n - m$ to the copy of itself in $\mathfrak{N}$. In addition, for each k, the k-th point on the branch in $\mathfrak{M}$ of length n is related to the k-th point on the infinite branch in $\mathfrak{N}$. However, as we have already proved before, the worlds w and w' are not bisimilar.

By induction, we can prove if two worlds are n-bisimilar, then they agree on all the modal formulas up to degree n. The statement in HOL looks like:

Proposition 3.10. [1, Proposition 2.31, one direction]

$$\vdash \mathfrak{M}, w \overset{Z^{\mathsf{s}}}{\leftrightarrows}_n \mathfrak{M}', w' \ \land\ \mathsf{DEG}\,\phi \ \le\ n \ \Rightarrow\ (\mathfrak{M}, w \Vdash\ \phi \quad\Longleftrightarrow\quad \mathfrak{M}', w' \Vdash\ \phi)$$

When we use set theory as foundation, if there are only finitely many propositional letters, then it is true that two worlds in two models agree on all the modal formulas with degree up to n if and only if there exists an n-bisimulation between them. However, we are using simple type theory as foundation in HOL, so this 'if and only if' statement in HOL looks a bit different. The thing we can prove in HOL is that: Let Δ be a finite α-set. If we restrict our scope to the set Σ of α-formulas that only uses propositional letters in Δ. Let $\mathfrak{M}_1$ be an (α, β)-model and $\mathfrak{M}_2$ be a (α, γ)-model, where both β and γ has infinite universe. For each $w_1 \in \mathfrak{M}_1^W$ and $w_2 \in \mathfrak{M}_2^W$, they agree on formulas in Σ up to degree n if and only if there is an n-bisimulation relating them. For the proof of this theorem: One direction is by the theorem shown above. The other direction is similar to the proof of Hennessy-Milner theorem, using the n-bisimulation relation Z^{s} defined by $Z^{\mathsf{s}}\ m\ w_1\ w_2 \iff \forall\,\phi.\ \mathsf{DEG}\,\phi \ \le\ m \ \Rightarrow\ (\mathfrak{M}, w_1 \Vdash\ \phi \iff \mathfrak{M}', w_2 \Vdash\ \phi)$ for each $m \le n$.

We also want a concept that measures the depth of a model. As 'depth' is measuring the distance of from a fixed point to another given point, to talk about the depth of a world $w \in \mathfrak{M}^W$, we need $\mathfrak{M}$ to be naturally equipped with a base point. Hence the 'height' of a world only makes sense for rooted models. To tell HOL about this definition, we start by defining $\mathsf{height}_{\le}$ as an inductive relation:

Definition 3.11. [1, Definition 2.32]

$$\frac{}{\mathsf{height}_{\le}\ \mathfrak{M}\ r\ \mathfrak{M}'\ r\ n}$$

$$\frac{v\ \in\ \mathfrak{M}^W \qquad \exists\, w.\ w\ \in\ \mathfrak{M}^W\ \land\ \mathfrak{M}^R\ w\ v\ \land\ \mathsf{height}_{\le}\ \mathfrak{M}\ r\ \mathfrak{M}'\ w\ n}{\mathsf{height}_{\le}\ \mathfrak{M}\ r\ \mathfrak{M}'\ v\ (n\ +\ 1)}$$

Recall how we defined a rooted model: when we write $\mathsf{rooted\_model}\ \mathfrak{M}\ r\ \mathfrak{M}'$, we mean '$\mathfrak{M}$ is a rooted model generated by the world r in the ambient model $\mathfrak{M}'$'. As $\mathsf{height}_{\le}$ is designed to only make sense for rooted models, we encode the information about the rootedness of the model we are talking about into this definition. Therefore, we read $\mathsf{height}_{\le}\ \mathfrak{M}\ r\ \mathfrak{M}'\ w\ n$ as 'for the rooted model $\mathfrak{M}$ with root r in $\mathfrak{M}'$, the distance from the world w to the root r is less than or equal to n', and we will always have an assumption on rootedness of $\mathfrak{M}$ whence this definition is used. The above rules mean:

- The height of the root for each rooted model is less or equal to every natural number.

- For a world v of $\mathfrak{M}$, if there exists a world w of $\mathfrak{M}$ such that w is related to v in $\mathfrak{M}$, then if the height of w is no more than n, the height of v is no more than $n \;+\; 1$.

We define the actual height of a world w to be the smallest natural number n such that $\mathsf{height}_{\leq}\ \mathfrak{M}\ r\ \mathfrak{M}'\ w\ n$. The height of a model is the maximum height of its worlds.

We are particularly interested in heights of tree-like models.

Lemma 3.11 (`tree_height_rel_lemma`). *When $\mathfrak{M}$ is tree-like, if $w \in \mathfrak{M}^W$ has height n, then every world $v \in \mathfrak{M}^W$ such that $\mathfrak{M}^R\ w\ v$ will have height $n + 1$.*

$$\begin{array}{l} \vdash \mathsf{tree}\ \mathfrak{M}.\mathsf{frame}\ r\ \wedge\ w\ \in\ \mathfrak{M}^W\ \wedge\ \mathsf{height}\ \mathfrak{M}\ r\ \mathfrak{M}\ w\ =\ n\ \wedge\ \mathfrak{M}^R\ w\ v\ \wedge \\ \quad v\ \in\ \mathfrak{M}^W\ \Rightarrow \\ \quad\ \ \mathsf{height}\ \mathfrak{M}\ r\ \mathfrak{M}\ v\ =\ n\ +\ 1 \end{array}$$

The restriction of a rooted model $\mathfrak{M}$ to the height k is the submodel consisting of all the worlds in $\mathfrak{M}$ of height up to k.

Definition 3.12. [1, Definition 2.32 (Restriction)] *We define a function* hrestriction *that takes a rooted model $\mathfrak{M}$, its root r, the an ambient model $\mathfrak{M}'$ that $\mathfrak{M}$ is sitting in, a natural number k, and give the model obtained by restricting $\mathfrak{M}$ to the height k.*

$$\begin{array}{l} \mathsf{hrestriction}\ \mathfrak{M}\ r\ \mathfrak{M}'\ k\ \stackrel{\mathsf{def}}{=} \\ \ \langle\!\langle \mathsf{frame}\ := \\ \quad \langle\!\langle \mathsf{world}\ :=\ \{\ w \mid w\ \in\ \mathfrak{M}^W\ \wedge\ \mathsf{height}\ \mathfrak{M}\ r\ \mathfrak{M}'\ w\ \leq\ k\ \}\ ;\ \mathsf{rel}\ :=\ \mathfrak{M}^R \rangle\!\rangle; \\ \ \ \mathsf{valt}\ :=\ \mathfrak{M}^V \rangle\!\rangle \end{array}$$

A restriction of a tree-like model is always a tree-like model. Moreover, restriction of every rooted model gives rise of n-bisimulation.

Lemma 3.12. [1, Lemma 2.33] *If we restrict a rooted model $\mathfrak{M}$ to height k, then a world w in the restricted model is $k-$*height $\mathfrak{M}$ *r* $\mathfrak{M}'$ *w-bisimilar to itself in the original model.*

$$\begin{array}{c} \vdash \mathsf{rooted\_model}\ \mathfrak{M}\ r\ \mathfrak{M}'\ \wedge\ w\ \in\ (\mathsf{hrestriction}\ \mathfrak{M}\ r\ \mathfrak{M}'\ k)^W\ \Rightarrow \\ \exists\, Z^{\mathsf{s}}.\ \mathsf{hrestriction}\ \mathfrak{M}\ r\ \mathfrak{M}'\ k, w\ \overset{Z^{\mathsf{s}}}{\leftrightarrows}_{k\ -\ \mathsf{height}\ \mathfrak{M}\ r\ \mathfrak{M}'\ w}\ \mathfrak{M}, w \end{array}$$

Proof. The $k-\mathsf{height}\ \mathfrak{M}\ r\ \mathfrak{M}'\ w$-bisimulation is given by Z^{s} which is defined as $Z^{\mathsf{s}}\ n$ relates a world w_1 in the restricted model $\mathsf{hrestriction}\ \mathfrak{M}\ r\ \mathfrak{M}'\ k$ to a world w_2 in $\mathfrak{M}$ iff $w_1 = w_2$ and the height of w_1 is no more than $k - n$. □

Now we can start building a finite model via selection:

Theorem 3.13. [1, Theorem 2.34]

$$\vdash \mathfrak{M}_1, w_1 \Vdash \phi \Rightarrow \exists \mathfrak{M}\ v.\ \mathsf{FINITE}\ \mathfrak{M}^W \wedge v \in \mathfrak{M}^W \wedge \mathfrak{M}, v \Vdash \phi$$

Proof. Suppose $\mathfrak{M}_1, w_1 \Vdash \phi$ where $\mathfrak{M}_1$ is an (α, β)-model and ϕ has degree k. By Proposition 3.4, there exists a tree-like $(\alpha, \beta\ \mathtt{list})$-model $\mathfrak{M}_2$ with ϕ satisfied at its root w_2. Define $\mathfrak{M}_3 := \mathsf{hrestriction}\ \mathfrak{M}_2\ w_2\ \mathfrak{M}_2\ k$ to be the restriction of $\mathfrak{M}_2$ to height k, then $\mathfrak{M}_3$ is rooted and we have a k-bisimulation Z^{s} such that $\mathfrak{M}_3, w_2 \overset{Z^{\mathsf{s}}}{\underline{\leftrightarrow}}_k \mathfrak{M}_2, w_2$ by Lemma 3.12, hence $\mathfrak{M}_3, w_2 \Vdash \phi$. We can discard all the propositional letters in $\mathfrak{M}_3$ which does not occur in ϕ and obtain the model $\mathfrak{M}_3'$, which looks like:

$$\begin{aligned}
&\mathfrak{M}_3' = \\
&\quad \langle\!\langle \mathsf{frame} := \langle\!\langle \mathsf{world} := \mathfrak{M}_3{}^W;\ \mathsf{rel} := \mathfrak{M}_3{}^R \rangle\!\rangle; \\
&\qquad \mathsf{valt} := (\lambda\, p\ v.\ \mathtt{if}\ p \in \mathsf{prop\_letters}\ \phi\ \mathtt{then}\ \mathfrak{M}_3{}^V\ p\ v\ \mathtt{else}\ \mathsf{F}) \rangle\!\rangle.
\end{aligned}$$

By Proposition 2.1, if a propositional letter does not appear in ϕ, then it has no effect on the satisfaction of ϕ. Hence we still have $\mathfrak{M}_3', w_2 \Vdash \phi$. We will select a finite model inductively from $\mathfrak{M}_3'$.

Let $\mathit{\Phi}$ denote the set of propositional letters used by ϕ, so $\mathit{\Phi}$ is finite. By Lemma 3.9, there are only finitely many non-equivalent formulas of degree less or equal to k which only use propositional letters in $\mathit{\Phi}$ (where equivalence is judged with respect to $(\alpha, \beta\ \mathtt{list})$-models). In other words, the set $\mathit{\Delta} = \{\, \psi \mid \mathsf{DEG}\ \psi \le k \wedge \mathsf{prop\_letters}\ \psi \subseteq \mathit{\Phi} \,\} / \equiv_{(:\beta\ \mathtt{list})}$ is finite. We care about the elements in $\mathit{\Delta}$ which are equivalence classes of formulas starting with a $\Diamond$. For such equivalence classes, taking the intersection with the set $\mathit{\Gamma} = \{\, \psi \mid \exists\, \psi_0.\ \psi = \Diamond \psi_0 \,\}$ does not give the empty set. Take the image of $\mathit{\Delta}$ under the function $\lambda\, s.\ s \cap \mathit{\Gamma}$ and delete the empty set from the image. We obtain a set $\mathit{\Sigma}$ of sets of formulas, where for each set $A \in \mathit{\Sigma}$, A consists of equivalent formulas of degree less or equal to k, only use propositional letters in $\mathit{\Phi}$, and start with a diamond. For each $A \in \mathit{\Sigma}$, we choose a representative using the choice function CHOICE, and collect these representatives into the set $R = \{\, \mathsf{CHOICE}\ (A \cap \mathit{\Gamma}) \mid A \in \mathit{\Delta} \,\} \setminus \{\, \emptyset \,\}$. The set R is finite as it is the image of a function over a finite set.

We will construct sets $S_0, \cdots, S_k$ of worlds in $\mathfrak{M}_3'$, where the worlds in S_n have height n. Start with $S_0 := \{w_2\}$, and inductively, assume $S_0, \cdots, S_n$ has been defined, construct S_{n+1} as follows: Consider an element in $v \in S_n$, for each $\Diamond\phi \in R$ such that $\mathfrak{M}_3', w_2 \Vdash \Diamond\phi$, pick a world $u \in {\mathfrak{M}_3'}^W$ such that ${\mathfrak{M}_3'}^R\ v\ u$ and $\mathfrak{M}_3', u \Vdash \phi$. Do the same thing to all the $v \in S_n$, then take S_{n+1} as the set of all the such u's which are selected in this way. The inductive definition of these S's are encoded in HOL as a primitive recursive function S^{s} such that for each $i \leq k$, $S^{\mathsf{s}}\ i$ will be our S_i. By induction on i, we can prove each S_i is finite, so the set $W_4 := \bigcup_{i \leq k} S_i$ is finite. The resultant finite model we select is:

$$\mathfrak{M}_4 = \langle\!\langle \mathsf{frame} := \langle\!\langle \mathsf{world} := W_4;\ \mathsf{rel} := {\mathfrak{M}_3}^R \rangle\!\rangle;\ \mathsf{valt} := {\mathfrak{M}_3'}^V \rangle\!\rangle.$$

To prove $\mathfrak{M}_4, w_2 \Vdash \phi$, it suffices to give a k-bisimulation between $\mathfrak{M}_4$ and $\mathfrak{M}_3'$ relating w_2 to itself. The k-bisimulation Z^{s} is given as for each $n \leq k$, $Z^{\mathsf{s}}\ n$ is the relation such that for $a_1 \in \mathfrak{M}_4^W$ and $a_2 \in {\mathfrak{M}_3}^W$, we have $Z^{\mathsf{s}}\ n\ a_1\ a_2$ iff:

- The worlds a_1 and a_2 have the same height, which is no more than $k - n$.
- If a formula ϕ only contains the propositional letters in Φ and $\mathsf{DEG}\ \phi \leq n$, it is satisfied at a_1 if and only if it is satisfied at a_2.

The rest of the proof amounts to checking the above indeed gives a k-bisimulation. The proof is again an analogue to the proof of Hennessy-Milner theorem. □

As we took a detour through Proposition 3.4, this construction of the finite model changes the type of model. If we start with an (α, β)-model, then the finite model we build by selection will be a $(\alpha, \beta\ \texttt{list})$-model.

Chapter 4

Standard Translation

As claimed by Slogan 3 in the introduction, modal logic is not an isolated formal system. In this chapter, we connect modal logic and first-order logic together using *standard translation.*

For a model $\mathfrak{M}$, the relation $\mathfrak{M}^R$ is a binary predicate on worlds in $\mathfrak{M}$. For each propositional letter p, the valuation $\mathfrak{M}^V$ gives a way of associating to p a unary predicate on the worlds in $\mathfrak{M}$, called P_p. Explicitly, the predicate P_p is defined by $P_p\ w$ if and only if $\mathfrak{M}^V\ p\ w$. We will see a modal model can also be viewed as a model for a *first-order language.* A first-order language is determined by a set of predicate symbols and a set of function symbols. Given a first-order language L such that L_F is the set of its function symbols and L_P is its set of predicate symbols, a first-order formula is called an L-formula if it only contains function symbols in L_F and predicate symbols in L_P. In our case, the first-order language that we can interpret using a modal model is the one such that the set of function symbols is empty, and the set of predicate symbols consists of a binary one corresponds to the relation on a modal model, and the unary ones of the form $\underline{P_p}$, where p is a propositional letter and P_p is the predicate which is associated to p. This language that we are interested in is called $\mathcal{L}^1_\tau$. A model $\mathfrak{M}$ of a first-order language consists of three pieces of information: a domain, where the variable letters in first-order formulas will be sent to, for each n-ary function symbol, an actual function that takes an n-tuple of elements in $\mathfrak{M}$ and gives an element in $\mathfrak{M}$, and for each n-ary predicate symbol, an actual predicate that takes an n-tuple of elements in $\mathfrak{M}$ and give a boolean value, either T or F. In HOL, the variable symbols, the function symbols, and the predicate symbols, are all encoded as natural numbers, hence a first-order model looks like:

Definition 4.1 (From [4], First-order model)**.**

$$\begin{array}{l} \alpha\ \texttt{folmodel} = \langle\!\langle \\ \quad \mathsf{Dom} : \alpha\ \rightarrow\ \texttt{bool}; \\ \quad \mathsf{Fun} : \texttt{num}\ \rightarrow\ \alpha\ \texttt{list}\ \rightarrow\ \alpha; \\ \quad \mathsf{Pred} : \texttt{num}\ \rightarrow\ \alpha\ \texttt{list}\ \rightarrow\ \texttt{bool} \\ \rangle\!\rangle \end{array}$$

As we can see, an α-first-order model means a first-order model with an α-set as its domain. For an α-first-order model $\mathfrak{M}$ and each function symbol f, where f is a natural number, $\mathfrak{M}.\mathsf{Fun}\ f$ is the actual function which is assigned f, and for each predicate symbol p, $\mathfrak{M}.\mathsf{Pred}\ p$ is the actual predicate which is assigned p.

According to the above discussion, if a first-order formula ϕ is purely a combination of some variable letters, which are to be interpreted as worlds in $\mathfrak{M}$, the binary predicate symbol which is to be interpreted as $\mathfrak{M}^R$, and some unary predicate symbols of the form $\underline{P_p}$, using the first-order connectives, then the information contained in a $(\texttt{num}, \beta)$-modal model $\mathfrak{M}$, where β is an arbitrary type, is enough to interpret the formula ϕ. But to formally interpret a first-order formula, we need to formally convert a modal model into a first-order model. The domain of the resulting model is the world set of $\mathfrak{M}$. As we are not going to use the resulting model to interpret formulas with function symbols, we do not really need any interesting information about the function symbol, hence we send every natural number to the constant function at an arbitrary point in $\mathfrak{M}^W$ that we pick. We fix the symbol of the binary predicate given by $\mathfrak{M}^R$ to be 0, and for each propositional letter p, we just use p itself to be the symbol of the predicate P_p associated to p (that is, in our formalization, we use p itself as the symbol $\underline{P_p}$ in the discussion above). For a modal model $\mathfrak{M}$, the function that converts it to a first-order model is called $\mathsf{mm2folm}$. In HOL:

Definition 4.2 (Conversion from a modal model to a first-order model)**.**

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

Conversely, we can view a β-first-order model $\mathfrak{M}$ as a $(\mathtt{num}, \beta)$-modal model. Given a first-order model, the relation on the modal model we get is given by the binary predicate with symbol 0. The valuations of a propositional letter p is given by the unary predicate in $\mathfrak{M}$ with symbol p. The function converting a first-order model into a modal model is folm2mm.

Definition 4.3 (Conversion from a first-order model into a modal model)**.**

$$\begin{array}{l}
\mathsf{folm2mm}\ \mathfrak{M} \overset{\mathsf{def}}{=} \\
\quad \langle\!\langle \mathsf{frame} := \\
\qquad \langle\!\langle \mathsf{world} := \mathfrak{M}.\mathsf{Dom}; \\
\qquad\quad \mathsf{rel} := (\lambda\, w_1\ w_2.\ \mathfrak{M}.\mathsf{Pred}\ 0\ [w_1;\ w_2] \land w_1 \in \mathfrak{M}.\mathsf{Dom} \land w_2 \in \mathfrak{M}.\mathsf{Dom})\rangle\!\rangle; \\
\quad \mathsf{valt} := (\lambda\, v\ w.\ \mathfrak{M}.\mathsf{Pred}\ v\ [w] \land w \in \mathfrak{M}.\mathsf{Dom})\rangle\!\rangle
\end{array}$$

Whereas the conversion from a modal model into a first-order model preserves all the information of the original model, the conversion from a general first-order model into a modal model will omit a lot of information: we will lose all the functions, all except for one binary predicates and all higher-arity predicates.

Now we can spell out how we formally interpret a first-order formula in the language $\mathcal{L}^1_\tau$. To do this, let us just spell out how to interpret a first-order formula on a first-order model in general. In HOL, a *term* of first-order logic is defined inductively. A first-order term is either a variable symbol standing alone, which is of form ${}^{\mathsf{f}}\mathsf{VAR}\ x$, where x is a variable symbol, or a function symbol applied on a list of terms, which is written as ${}^{\mathsf{f}}\mathsf{Fn}\ f\ l$, where f is the natural number serving as a function symbol, and l is a list of terms. A term of first-order logic should not be confused with a term of a type (a reader should better consider them as unrelated, for the sake of reading this thesis). A first-order formula is defined inductively using four primitive connectives:

Definition 4.4 (From [4], First-order formulas)**.** *The primitive logical connectives we are using here are: the falsity* ${}^{\mathsf{f}}\bot$*, a predicate symbol applied on a list of variables, implication* IMP *(which will be written as the infix* ${}^{\mathsf{f}}\!\!\rightarrow$ *from now on), and the universal quantification.*

$$\mathtt{folform} \;=\; {}^{\mathsf{f}}\bot \mid \mathsf{Pred}\ \mathit{num}\,(\mathtt{term\ list}) \mid \mathsf{IMP}\ \mathtt{folform}\ \mathtt{folform} \mid {}^{\mathsf{f}}\forall\ \mathit{num}\ \mathtt{folform}$$

In above, as we use natural numbers for the predicate symbols, function symbols and variable symbols in first-order formulas, the current construction can only capture countable first-order languages. The non-primitive connectives are defined as:

Definition 4.5 (From [4], Non-primitive first-order connectives)**.**

$$\begin{array}{rcl}
{}^{\mathsf{f}}\neg\ \phi & \stackrel{\mathsf{def}}{=} & \phi\ {}^{\mathsf{f}}\!\!\rightarrow {}^{\mathsf{f}}\bot \\
{}^{\mathsf{f}}\top & \stackrel{\mathsf{def}}{=} & {}^{\mathsf{f}}\neg\ {}^{\mathsf{f}}\bot \\
\phi_1\ {}^{\mathsf{f}}\vee\ \phi_2 & \stackrel{\mathsf{def}}{=} & (\phi_1\ {}^{\mathsf{f}}\!\!\rightarrow \phi_2)\ {}^{\mathsf{f}}\!\!\rightarrow \phi_2 \\
\phi_1\ {}^{\mathsf{f}}\wedge\ \phi_2 & \stackrel{\mathsf{def}}{=} & {}^{\mathsf{f}}\neg\ ({}^{\mathsf{f}}\neg\ \phi_1\ {}^{\mathsf{f}}\vee\ {}^{\mathsf{f}}\neg\ \phi_2) \\
{}^{\mathsf{f}}\exists\ x\ \phi & \stackrel{\mathsf{def}}{=} & {}^{\mathsf{f}}\neg\ ({}^{\mathsf{f}}\forall\ x\ ({}^{\mathsf{f}}\neg\ \phi))
\end{array}$$

A quantified variable is called a bounded variable, otherwise, it is called free. For instance, the 1 in ${}^{\mathsf{f}}\mathsf{Pred}\ 4\ [{}^{\mathsf{f}}\mathsf{VAR}\ 1;\ {}^{\mathsf{f}}\mathsf{VAR}\ 2]$ is free, whereas the 1 in ${}^{\mathsf{f}}\forall\ 1\ ({}^{\mathsf{f}}\mathsf{Pred}\ 4\ [{}^{\mathsf{f}}\mathsf{VAR}\ 1;\ {}^{\mathsf{f}}\mathsf{VAR}\ 2])$ is bounded. For a first-order formula ϕ, we write $\mathsf{FV}\ \phi$ for the set of all its free variables and $\mathsf{BV}\ \phi$ for the set of all its bounded variables. We also have functions form_functions and form_predicates which give the set of function and predicate symbols of a first-order formula, respectively. For a function symbol denoted by f, if it is applied on a list of terms of length n, then it will be recorded as the tuple (f, n). For example, we have $\mathsf{form\_functions}\ ({}^{\mathsf{f}}\mathsf{Pred}\ 1\ [{}^{\mathsf{f}}\mathsf{Fn}\ 2\ [{}^{\mathsf{f}}\mathsf{VAR}\ 1;\ {}^{\mathsf{f}}\mathsf{VAR}\ 2];\ {}^{\mathsf{f}}\mathsf{Fn}\ 2\ []]) \;=\; \{\ (2,2);\ (2,0)\ \}$ and $\mathsf{form\_functions}\ ({}^{\mathsf{f}}\mathsf{Pred}\ 1\ [{}^{\mathsf{f}}\mathsf{Fn}\ 0\ [];\ {}^{\mathsf{f}}\mathsf{Fn}\ 1\ []]) \;=\; \{\ (0,0);\ (1,0)\ \}$. Similarly, a predicate symbol denoted by natural number p followed by a list of length n is recorded as a pair (p, n). Hence both the function form_functions and form_predicates take a formula and give a set of pairs of natural numbers. The language $\mathcal{L}^1_\tau$, as introduced before, is defined as a predicate, where '$\mathcal{L}^1_\tau\ \phi$' reads 'the formula ϕ is in the language $\mathcal{L}^1_\tau$', as follows:

Definition 4.6. [1, Definition 2.44 (The Language $\mathcal{L}^1_\tau$)]

$$\begin{array}{l}
\mathcal{L}^1_\tau\ \phi \stackrel{\mathsf{def}}{=} \\
\quad \mathsf{form\_functions}\ \phi \;=\; \emptyset\ \wedge \\
\quad \mathsf{form\_predicates}\ \phi \;\subseteq\; (0,2)\ \mathsf{INSERT}\ \{\ (p,1) \mid p\ \in\ \mathcal{U}(:\mathit{num})\ \}
\end{array}$$

Given a first-order model $\mathfrak{M}$, we interpret formulas or terms by assigning each variable symbol an element in $\mathfrak{M}.\mathsf{Dom}$. As we are using natural numbers as variable symbols, such an assignment is a function that takes a natural number. We are only interested in the case when a function does send each natural number to an element in $\mathfrak{M}.\mathsf{Dom}$, such a function is called a *valuation* of $\mathfrak{M}$. We write $\mathsf{valuation}\ \mathfrak{M}\ \sigma$ in this case, and read it as 'σ is a valuation of the model $\mathfrak{M}$'.

Interpretation of terms and formulas are given as $\mathsf{termval}$ and feval. If we give the function $\mathsf{termval}$ a model $\mathfrak{M}$, a valuation σ and a first-order term t, it will give us the element of the domain of $\mathfrak{M}$ that t is interpreted as. If we give the function feval a model $\mathfrak{M}$, a valuation σ and a first-order formula ϕ, it will give us the truth value of ϕ in $\mathfrak{M}$ under the valuation σ. When ϕ is true in $\mathfrak{M}$ under σ, we write $\mathfrak{M}, \sigma \vDash\ \phi$. We write $\phi_1\ {}^{\mathsf{f}}{\equiv}_{(:\alpha)}\ \phi_2$ to mean the first-order formulas ϕ_1 and ϕ_2 are equivalent on α-first-order models, where equivalence between first-order formulas is defined similarly as that of modal formulas.

If we want to use a first-model $\mathfrak{M}$ to interpret a first-order formula ϕ, the first thing to make sure is that the actual functions assigned to the function symbols appear in ϕ does not send a list of elements in $\mathfrak{M}$ out of the domain of $\mathfrak{M}$. Therefore, a theorem about interpreting a formula ϕ in $\mathfrak{M}$ should start with an assumption

$$\forall f\ n\ l.\ (f, n)\ \in\ \mathsf{form\_functions}\ \phi\ \land\ \mathsf{LENGTH}\ l\ =\ n\ \Rightarrow\ \mathfrak{M}.\mathsf{Fun}\ f\ l\ \in\ \mathfrak{M}.\mathsf{Dom}.$$

But we may want to use the same model to interpret formulas with various function symbols and do not want such an assumption everywhere. Hence for our convenience, unless we have no function symbols at all, we will always assume the models $\mathfrak{M}$ we are working with satisfies $\mathfrak{M}.\mathsf{Fun}\ f\ l\ \in\ \mathfrak{M}.\mathsf{Dom}$ for every function symbol f and list l. For such a model $\mathfrak{M}$, we write $\mathsf{wffm}\ \mathfrak{M}$, meaning '$\mathfrak{M}$ is a well-formed first-order model'.

At noted before, the functions $\mathsf{mm2folm}$ and $\mathsf{folm2mm}$ are not inverses. However, we have:

Proposition 4.1 (`L1tau_mm2folm_folm2mm_comm_feval`). *An $\mathcal{L}^1_\tau$ formula is satisfied in $\mathfrak{M}$ under σ if and only if it is satisfied under σ in the model we obtain by firstly converting $\mathfrak{M}$ to a modal model, and then back to a first-order model.*

$$\vdash \mathcal{L}^1_\tau\ \phi\ \land\ \mathsf{valuation}\ \mathfrak{M}\ \sigma\ \Rightarrow\ (\mathsf{mm2folm}\ (\mathsf{folm2mm}\ \mathfrak{M}), \sigma \vDash\ \phi\ \iff\ \mathfrak{M}, \sigma \vDash\ \phi)$$

Also we note:

Proposition 4.2 (From [4], `holds_valuation`). *For a fixed model, the truth value of a first-order formula only depends on what a valuation sends its free variable to.*

$$\vdash (\forall\, v.\ v\ \in\ \mathsf{FV}\,\phi\ \Rightarrow\ \sigma_1\ v\ =\ \sigma_2\ v)\ \Rightarrow\ (\mathfrak{M}, \sigma_1 \vDash\ \phi\ \iff\ \mathfrak{M}, \sigma_2 \vDash\ \phi)$$

Therefore, although a valuation of $\mathfrak{M}$ assigns every natural number an element in $\mathfrak{M}.\mathsf{Dom}$, what it effectively does is to only control where the free variables in a formula go to. The advantage of using a valuation instead of assigning free variables values one by one is that a valuation can simultaneously control every number of free variables.

With the setup on basics about first-order logic and how it interacts with modal logic, let us build intuition on how modal formulas correspond to first-order formulas. The first thing to note is that as every symbol in a first-order formula is represented by a natural number, without get cumbersome and complicated procedure involved, we can only translate `num`-modal formulas into first-order formulas. Observe that unlike modal formulas which atomic formulas are propositional letters standing alone, even the atomic first-order formulas (except ${}^{\mathsf{f}}\bot$) have variable symbols. Hence to translate a modal formula into a first-order formula, we must introduce some variables as well. For a model $\mathfrak{M}$, it is natural to regard each modal formula as a predicate to be evaluated at worlds of $\mathfrak{M}$, such that this predicate is true at a world w if and only if the formula is satisfied at w. Hence to translate a modal formula into a first-order formula, the only natural thing to do is to get just one variable involved, and this variable will be later assigned to some state in some model when we interpret the translated formula.

Hence for the function ST which translates a modal formula to a first-order formula, the first parameter it takes is a variable symbol x, which is represented by a natural number, that we introduce to mark the world we are looking at, as discussed above. The second parameter is the `num`-modal formula which we want to translate.

Definition 4.7. [1, Definition 2.45 (Standard Translation)] *Here '$\mathsf{ST}_x\ \phi$' reads the standard translation of the modal formula ϕ at x. The translation is defined as:*

- *A propositional letter is translated into the unary predicate symbol represented by p applied on the variable x. Here ${}^{\mathsf{f}}\mathsf{P}\ p\ ({}^{\mathsf{f}}\mathsf{VAR}\ x)$ is the abbreviation of ${}^{\mathsf{f}}\mathsf{Pred}\ p\ [{}^{\mathsf{f}}\mathsf{VAR}\ x]$.*

- *The falsity in modal formula is translated into the falsity in first-order formula.*

- *Inductively, the negation of a modal formula ϕ is translated into the first-order negation of the standard translation of ϕ.*

- *The disjunction of two modal formulas ϕ and ψ is translated into the first-order disjunction of the standard translation of ϕ and the standard translation of ψ.*

- *A modal formula $\Diamond\phi$ is translated to the existential quantifier applied on the variable symbol $x + 1$ and the first-order formula saying ‘$^{\mathsf{f}}\mathsf{VAR}\ x$ is related to $^{\mathsf{f}}\mathsf{VAR}\ (x\ +\ 1)$ and $\mathsf{ST}_{x+1}\ \phi$’. Here $^{\mathsf{f}}\mathsf{R}\ (^{\mathsf{f}}\mathsf{VAR}\ x)\ (^{\mathsf{f}}\mathsf{VAR}\ (x\ +\ 1))$ is the abbreviation of $^{\mathsf{f}}\mathsf{Pred}\ 0\ [(^{\mathsf{f}}\mathsf{VAR}\ x); (^{\mathsf{f}}\mathsf{VAR}\ (x + 1))]$ (Recall that we have fixed 0 as the predicate symbol which corresponds to relation on the modal model). That is:*

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

For the last line, according to the semantic interpretation of the ‘$\Diamond\phi$’, which is ‘there exists a world related to the current state where ϕ is satisfied’, 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, hence the standard translation of $\Diamond\phi$ at x says exactly the same thing as how we interpret it in a modal model.

Some syntactic properties of standard translation are immediate to prove. For instance:

- Every first-order formula obtained by standard translation is $\mathcal{L}^1_\tau$.

- Every first-order formula obtained by standard translation has at most one free variable.

- The negation of a standard translation is a standard translation.

- Conjunctions and disjunctions of standard translations are equivalent to standard translations of big conjunction/disjunction formulas.

On the other hand, standard translations have interesting semantic behavior as well. Their semantic features give a first-order reformulation of modal satisfaction.

Proposition 4.3. [1, Theorem 2.47 (i)] *A modal formula ϕ is satisfied at a world in a modal model $\mathfrak{M}$ iff its standard translation $\mathsf{ST}_x\ \phi$ is satisfied in $\mathfrak{M}$ viewed as a first-order model when the free variable x is assigned this world.*

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

There is a result corresponds to the above using $\mathsf{folm2mm}$.

Proposition 4.4 (`prop_2_47_i0'`). *The standard translation of the formula ϕ using variable symbol x is true in a first-order model $\mathfrak{M}$ under the valuation σ iff ϕ is satisfied at $\sigma\ x$ in $\mathfrak{M}$ viewed as a modal model.*

$$\vdash \mathsf{folm2mm}\ \mathfrak{M}, w \Vdash\ \phi \iff \mathfrak{M}, (\lambda\, n.\ w) \vDash\ \mathsf{ST}_x\ \phi$$

As an interesting consequence of the ‘equivalence’ between a modal formula and its standard translation, we can prove formulas obtained by standard translation are *invariant under bisimulation.*

Definition 4.8. [1, Definition 2.67 (Invariant for Bisimulations)] *An $\mathcal{L}^1_\tau$ formula ϕ with at most one free variable x is invariant for bisimulations 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.*

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

The predicate $\mathsf{invar4bisim}$ takes four parameters, the first one is the name of the only free variable in ϕ and the last one is the formula itself. For the second and third one. Recall the issue we met when defining equivalence of modal formulas as in Chapter 2, it needs to take a type as a parameter since we cannot have type variable which only appears on the right-hand side but not on the left-hand side. For the same reason, here we need to tell HOL explicitly about the type of the world set of the models which can serve as $\mathfrak{M}$ and $\mathfrak{N}$ in our definition. 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 things afterwards.

Proposition 4.5. [1, Theorem 2.68, easy direction] *Every $\mathcal{L}^1_\tau$-formula with only one variable which is equivalent to a standard translation is invariant for bisimulation.*

$$\vdash \mathsf{FV}\,\delta \ \subseteq\ \{\ x\ \}\ \wedge\ \mathcal{L}^1_\tau\ \delta\ \wedge\ \delta\ {}^{\mathsf{f}}\!\!\equiv_{(:\alpha)}\ \mathsf{ST}_x\ \phi\ \Rightarrow\ \mathsf{invar4bisim}\ x\ (:\alpha)\ (:\alpha)\ \delta$$

Proof. By Theorem 3.6 and Proposition 4.4. □

In fact, using set theory as the foundation, we can prove that every formula which is invariant under bisimulation arises as the standard translation of a modal formula, so 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 using the variable symbol x. We can translate the set-theoretic proof into a simple type-theoretic proof. But the proof of the other direction requires more advanced tools, and its statement does not look the same as its mathematical statement as in set theory in HOL. This is because of the lack of expressiveness of simple type theory. We will leave the other direction of the proof to the next chapter.

Chapter 5

Modal Saturation via Ultrafilter Extensions

In Chapter 3, we have seen bisimilarity implies modal equivalence, but only proved the converse for image finite models. In this chapter, we are interested in another particular class of models, called *M-saturated* models, where modal equivalent worlds are bisimilar. We will introduce an operation, called the *ultrafilter extension*, on models, which creates M-saturated models. With the results about M-saturated models, we will conclude this chapter by proving an elegant result about bisimulation: If we have a modal equivalence between worlds w, v in two models $\mathfrak{M}$ and $\mathfrak{N}$, although it may not be the case that w and v are bisimilar, we can find a bisimulation between the ultrafilter extension of $\mathfrak{M}$ and the ultrafilter extension of $\mathfrak{N}$.

Let us explain what is meant by 'M-saturated' first. Being M-saturated is a sort of compactness property, which says 'finite satisfaction implies satisfaction'. We need to give three definitions consecutively to finally get M-saturation to be formally defined in HOL.

Definition 5.1. [1, Definition 2.53 (Satisfiable)] *A set of formulas Σ is called satisfiable in a set X of worlds in a model $\mathfrak{M}$ if there exists an element in X such that all the formulas in Σ are satisfied.*

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

Definition 5.2. [1, Definition 2.53 (Finitely Satisfiable)] *A set of formulas Σ is called finitely satisfiable if every finite subset of Σ is satisfiable.*

$$\textsf{fin\_satisfiable\_in}\ \Sigma\ X\ \mathfrak{M} \overset{\textsf{def}}{=} \forall S.\ S \subseteq \Sigma \land \textsf{FINITE}\ S \Rightarrow \textsf{satisfiable\_in}\ S\ X\ \mathfrak{M}$$

Definition 5.3. [1, Definition 2.53 (M-saturation)] *A model $\mathfrak{M}$ is called M-saturated if for every $w \in \mathfrak{M}^W$, if a set Σ is finitely satisfiable in the set of successors of w, then it is satisfiable in the set of successors of w.*

$$\begin{array}{l} \textsf{M\_sat}\ \mathfrak{M} \overset{\text{def}}{=} \\ \quad \forall\, w\ \Sigma. \\ \qquad w \in \mathfrak{M}^W \wedge \textsf{fin\_satisfiable\_in}\ \Sigma\ \{\ v \mid v \in \mathfrak{M}^W \wedge \mathfrak{M}^R\ w\ v\ \}\ \mathfrak{M} \Rightarrow \\ \qquad\ \ \textsf{satisfiable\_in}\ \Sigma\ \{\ v \mid v \in \mathfrak{M}^W \wedge \mathfrak{M}^R\ w\ v\ \}\ \mathfrak{M} \end{array}$$

For M-saturated models, bisimilarity implies modal equivalence.

Proposition 5.1. [1, Proposition 2.54] *For two worlds w_1 and w_2 living in M-saturated models $\mathfrak{M}_1$ and $\mathfrak{M}_2$ respectively, if w_1 and w_2 are modal equivalent, then they are bisimilar.*

$$\begin{array}{l} \vdash \textsf{M\_sat}\ \mathfrak{M}_1 \wedge \textsf{M\_sat}\ \mathfrak{M}_2 \wedge w_1 \in \mathfrak{M}_1^W \wedge w_2 \in \mathfrak{M}_2^W \wedge \mathfrak{M}_1, w_1 \leftrightsquigarrow \mathfrak{M}_2, w_2 \Rightarrow \\ \quad \mathfrak{M}_1, w_1 \leftrightarrows \mathfrak{M}_2, w_2 \end{array}$$

Proof. Let $\mathfrak{M}_1$ and $\mathfrak{M}_2$ be models. In fact, they can be $(\alpha, \beta), (\alpha, \gamma)$-models respectively, where β and γ are not required to be the same. Under the assumptions, the bisimulation relation Z we need is for $a \in \mathfrak{M}_1^W$ and $b \in \mathfrak{M}_2^W$, we have $Z\ a\ b \iff \forall\,\phi.\ \mathfrak{M}_1, a \Vdash \phi \iff \mathfrak{M}_2, b \Vdash \phi$. To prove this relation is indeed a bisimulation, the only non-trivial clause to check is that for worlds w_1, v_1 of $\mathfrak{M}_1$ and world w_2 of $\mathfrak{M}_2$ such that w_1 and w_2 are modal equivalent, we can find a world v_2 of $\mathfrak{M}_2$ such that $\mathfrak{M}_2^R\ w_2\ v_2$ and v_1 and v_2 are modal equivalent.

Under the assumptions above, let Σ denote the set of formulas satisfied by v_1, we will find a successor w of w_2 where each formula in Σ is satisfied, then the world w will be modal equivalent to v_1. Indeed, if we find such a w, then for a formula ψ which is not satisfied at v_1, we will have $\neg\psi \in \Sigma$ and hence $\neg\psi$ will be satisfied at w, which implies ψ is not satisfied at w.

As $\mathfrak{M}_2$ is M-saturated, it suffices to prove each finite subset $\Delta \subseteq \Sigma$ is satisfied in some successor of w_2. Take such a Δ, then it is satisfied at v_1 by its definition. As Δ is finite, we can conjunct all its elements to obtain a formula ψ. We have $\mathfrak{M}_1, v_1 \Vdash \psi$, and therefore $\mathfrak{M}_1, w_1 \Vdash \Diamond\psi$. By modal equivalence of w_1 and w_2, we then get $\mathfrak{M}_2, w_2 \Vdash \Diamond\psi$, so there exists a successor w' of w_2 that satisfies ψ. Hence w' will satisfy every formula in Δ.

□

Since M-saturated models are nice, here is a natural question: How can we get such models? In the rest of this chapter, we will see the fact the *ultrafilter* *extension* of every model is M-saturated. In order to talk about ultrafilter extensions, we wrote a theory about ultrafilters in HOL. Instead of showing the whole theory here, we will just show what we need for proving the theorems we are interested in.

As its name suggests, an ultrafilter is a special kind of filter.

Definition 5.4. [1, Definition A.12 (Filters)] *Given a non-empty set J, a set L which is a subset of the power set of J (denoted as* $\mathsf{POW}\ J$ *in HOL) is called filter if it contains J itself, is closed under binary intersection, and is closed upward.*

$$\begin{aligned}&\mathsf{filter}\ L\ J \overset{\mathsf{def}}{=}\\ &\quad J \neq \emptyset \land L \subseteq \mathsf{POW}\ J \land J \in L \land\\ &\quad (\forall X\ Y.\ X \in L \land Y \in L \Rightarrow X \cap Y \in L) \land\\ &\quad \forall X\ Z.\ X \in L \land X \subseteq Z \land Z \subseteq J \Rightarrow Z \in L\end{aligned}$$

By induction, closure under binary intersection implies closure under every finite intersection.

The simplest example of a filter is the power set $\mathsf{POW}\ J$ itself. By upward closure, if a filter on J contains the empty set, then the filter must be the whole power set $\mathsf{POW}\ J$. A filter which is not a power set is called a *proper filter.*

For a set J and an element $w \in J$, the filter generated by $\{\ w\ \}$ is the set of subsets of J that contains w, it is trivial to check it is indeed a filter. Such a filter is called a *principal filter.* In HOL, we define a function that takes an element $w \in J$ and a set J, and give the principal filter generated by w, which is denoted as π_w^J. Actually, principal filters are the simplest examples of *ultrafilters.*

Definition 5.5. [1, Definition A.12 (Ultrafilters)] *An ultrafilter on a set J is a proper filter U such that for every $X \subseteq J$, either X or its complement $J \setminus X$ is in U, but not both.*

$$\begin{aligned}&\mathsf{ultrafilter}\ U\ J \overset{\mathsf{def}}{=}\\ &\quad \mathsf{proper\_filter}\ U\ J \land \forall X.\ X \in \mathsf{POW}\ J \Rightarrow (X \in U \iff J \setminus X \notin U)\end{aligned}$$

There are two results about ultrafilter which will be used here, the standard proofs of both of them can be found in Chapter 7 of [5]. The first one is the *ultrafilter theorem.*

Theorem 5.2. [1, Fact A.14, first half] *Every proper filter is contained in an ultrafilter.*

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

The other one is a corollary of the ultrafilter theorem. This corollary says that for every subset of $\mathsf{POW}\ J$ which has *finite intersection property*, it can be extended to an ultrafilter on J. In HOL, the definition of finite intersection property is given as:

Definition 5.6. [1, Definition A.13 (Finite Intersection Property)] *A subset A of $\mathsf{POW}\ J$ has finite intersection property if once we take the intersection of a finite, nonempty family in A, the resultant set is nonempty. We read '$\mathsf{FIP}\ A\ J$' as 'A is a set of subsets of J with finite intersection property'.*

$$\begin{aligned}&\vdash \mathsf{FIP}\ A\ J \iff\\ &\quad A \subseteq \mathsf{POW}\ J \wedge \forall B.\ B \subseteq A \wedge \mathsf{FINITE}\ B \wedge B \neq \emptyset \Rightarrow \bigcap B \neq \emptyset\end{aligned}$$

Note that finite intersection property is a property of subsets of power sets. Therefore, the predicate FIP defined above takes two parameters: a set of subsets of J, and an ambient set J. Every proper filter has finite intersection property.

And the corollary of ultrafilter theorem that we will need is stated as:

Proposition 5.3. [1, Fact A.14, second half] *For every set A of subsets of a non-empty set J with finite intersection property, there exists an ultrafilter on J which contains A.*

$$\vdash \mathsf{FIP}\ A\ J \wedge J \neq \emptyset \Rightarrow \exists\, U.\ \mathsf{ultrafilter}\ U\ J \wedge A \subseteq U$$

The proof of both ultrafilter theorem and its corollary are not technical from the formalization aspect. So we have omitted their proof.

We can now launch on the construction of the ultrafilter extension of a model. For a model $\mathfrak{M}$, the world set of the ultrafilter extension of $\mathfrak{M}$ is simply the set of ultrafilters on the world set of $\mathfrak{M}$, whereas the relation defined on the set of ultrafilters require more explanation.

Fix a model $\mathfrak{M}$ and a subset X of its world set, we can consider two set of worlds determined by X:

Definition 5.7. [1, Definition 2.55 ('Can See' and 'Only See')] *Given a model $\mathfrak{M}$ and a set X of worlds of $\mathfrak{M}$, we define:*

- *The set of worlds that 'can see' X (notation: $\mathfrak{M}_{\Diamond}(X)$) is the set of worlds w of $\mathfrak{M}$ such that there exists some $v \in X$ such that $\mathfrak{M}^R\ w\ v$.*

- *The set of worlds that 'only see' X (notation: $\mathfrak{M}^{\delta}_{\Diamond}(X)$) is the set of worlds w of $\mathfrak{M}$ such that once we have $\mathfrak{M}^R\ w\ v$ for some world $v \in \mathfrak{M}^W$, we must have $v \in X$.*

In HOL:

$$\mathfrak{M}_{\Diamond}(X) \overset{\mathsf{def}}{=} \{ w \mid w \in \mathfrak{M}^W \land \exists v.\, v \in X \land \mathfrak{M}^R\ w\ v \}$$

$$\mathfrak{M}^{\delta}_{\Diamond}(X) \overset{\mathsf{def}}{=} \{ w \mid w \in \mathfrak{M}^W \land \forall v.\, v \in \mathfrak{M}^W \land \mathfrak{M}^R\ w\ v \Rightarrow v \in X \}$$

By definition of satisfaction, for every model formula ϕ, the worlds satisfying $\Diamond\phi$ are exactly the ones that can see a world where ϕ is satisfied, and the worlds that satisfy $\Box\phi$ are exactly the ones that only see the worlds where ϕ is satisfied.

The concept 'can see' and 'only see' are dual to each other.

Proposition 5.4. [1, Proposition 2.56] *A world that can see a world in X is precisely a world that does not only see worlds that are not in X. Similarly, a world that can only see worlds in X is precisely a world which does not see worlds which are not in X.*

$$\vdash X \subseteq \mathfrak{M}^W \Rightarrow \mathfrak{M}_{\Diamond}(X) = \mathfrak{M}^W \setminus \mathfrak{M}^{\delta}_{\Diamond}(\mathfrak{M}^W \setminus X)$$

$$\vdash X \subseteq \mathfrak{M}^W \Rightarrow \mathfrak{M}^{\delta}_{\Diamond}(X) = \mathfrak{M}^W \setminus \mathfrak{M}_{\Diamond}(\mathfrak{M}^W \setminus X)$$

A world can see some world in the union of X and Y if and only if it can see a world in X or a world in Y, hence $\mathfrak{M}_{\Diamond}$ distributes over union. Dually, a world only sees the worlds in the intersection of X and Y if and only if it can only see worlds in X and worlds in Y, therefore, $\mathfrak{M}^{\delta}_{\Diamond}$ distributes over intersections.

Proposition 5.5 (`can_see_UNION`, `only_see_INTER`)**.**

$$\vdash \mathfrak{M}_{\Diamond}(X \cup Y) = \mathfrak{M}_{\Diamond}(X) \cup \mathfrak{M}_{\Diamond}(Y)$$

$$\vdash \mathfrak{M}^{\delta}_{\Diamond}(X \cap Y) = \mathfrak{M}^{\delta}_{\Diamond}(X) \cap \mathfrak{M}^{\delta}_{\Diamond}(Y)$$

Return to the discussion about the definition of relation on the ultrafilter extension of $\mathfrak{M}$. We define:

Definition 5.8. [1, Proposition 2.57 (Relation of Ultrafilter Extension)] *Two ultrafilters u, v on $\mathfrak{M}$ are related in the ultrafilter extension of $\mathfrak{M}$ if for every $X \in v$, the set of worlds that can see X is in u.*

$$^{ue}\mathfrak{M}^R\ u\ v \overset{\mathsf{def}}{=}$$
$$\mathsf{ultrafilter}\ u\ \mathfrak{M}^W \land \mathsf{ultrafilter}\ v\ \mathfrak{M}^W \land \forall X.\, X \in v \Rightarrow \mathfrak{M}_{\Diamond}(X) \in u$$

By the duality between 'can see' and 'only see', this relation has a reformulation:

Proposition 5.6. [1, Exercise 2.5.5] *Two ultrafilters u and v on $\mathfrak{M}^W$ are related if and only if for every subset Y of $\mathfrak{M}^W$, if the set of worlds of $\mathfrak{M}$ that it can only see Y is in u, then Y is in v.*

$$\begin{aligned}&\vdash {}^{ue}\mathfrak{M}^R\ u\ v \iff \\ &\quad \mathsf{ultrafilter}\ u\ \mathfrak{M}^W \ \wedge\ \mathsf{ultrafilter}\ v\ \mathfrak{M}^W\ \wedge \\ &\quad \{\ Y \mid \mathfrak{M}^\delta_\Diamond(Y)\ \in\ u \wedge\ Y\ \subseteq\ \mathfrak{M}^W\ \}\ \subseteq\ v\end{aligned}$$

Proof. Suppose ${}^{ue}\mathfrak{M}^R\ u\ v$ and pick a set Y of worlds such that $\mathfrak{M}^\delta_\Diamond(Y)\ \in\ u$, we will prove $Y\ \in\ v$. We have $\mathfrak{M}^W\ \setminus\ \mathfrak{M}_\Diamond(\mathfrak{M}^W\ \setminus\ Y)\ =\ \mathfrak{M}^\delta_\Diamond(Y)$ by Proposition 5.4, so $\mathfrak{M}^W \setminus \mathfrak{M}_\Diamond(\mathfrak{M}^W \setminus Y)\ \in\ u$. As u is an ultrafilter, this implies $\mathfrak{M}_\Diamond(\mathfrak{M}^W\ \setminus\ Y)\ \notin\ u$. By definition of ${}^{ue}\mathfrak{M}^R$ and the assumption ${}^{ue}\mathfrak{M}^R\ u\ v$, this implies $\mathfrak{M}^W\ \setminus\ Y\ \notin\ v$. Hence $Y\ \in\ v$ as v is an ultrafilter. The other direction is similar.

□

In order to define the ultrafilter extension model, the only remaining issue is to define the valuation. We define a propositional letter p to be satisfied at an ultrafilter v if and only if the set of worlds in $\mathfrak{M}$ which satisfies p is in v. Hence the full definition of ultrafilter extension is:

Definition 5.9. [1, Definition 2.57 (Ultrafilter Extension)] *The ultrafilter extension is defined as a function that takes a model and gives the extended model. We denote the ultrafilter extension of $\mathfrak{M}$ by ${}^{ue}\mathfrak{M}$.*

$$\begin{aligned}&{}^{ue}\mathfrak{M}\ \overset{\mathsf{def}}{=}\\ &\quad \langle\!\langle \mathsf{frame}\ :=\ \langle\!\langle \mathsf{world}\ :=\ \{\ u \mid \mathsf{ultrafilter}\ u\ \mathfrak{M}^W\ \};\ \mathsf{rel}\ :=\ {}^{ue}\mathfrak{M}^R\rangle\!\rangle;\\ &\quad \mathsf{valt}\ :=\ (\lambda\, p\ v.\ \mathsf{ultrafilter}\ v\ \mathfrak{M}^W\ \wedge\ \{\ w \mid w\ \in\ \mathfrak{M}^W\ \wedge\ \mathfrak{M}^V\ p\ w\ \}\ \in\ v)\rangle\!\rangle\end{aligned}$$

The ultrafilter extension also changes the type of the input model, namely, it changes the type of worlds from β to $(\beta\ \rightarrow\ \texttt{bool})\ \rightarrow\ \texttt{bool}$. Ultrafilter extension is indeed an extension, in the sense that $\mathfrak{M}$ is embedded in ${}^{ue}\mathfrak{M}$ as a submodel by the function sending $w\ \in\ \mathfrak{M}^W$ to the principal ultrafilter $\pi_w^{\mathfrak{M}^W}$ generated by w. In general, this embedding does not necessarily give a generated submodel, nevertheless, we have an invariance result for this embedding:

Proposition 5.7. [1, Proposition 2.59 (ii)] *For every model $\mathfrak{M}$ and every world w of $\mathfrak{M}$, w is modal equivalent to the principal filter generated by w, which is a world in the ultrafilter extension of $\mathfrak{M}$.*

$$\vdash w\ \in\ \mathfrak{M}^W\ \Rightarrow\ \mathfrak{M}, w \leftrightsquigarrow {}^{ue}\mathfrak{M}, \pi_w^{\mathfrak{M}^W}$$

This is actually a special case of the following proposition, where u is taken as $\pi_w^{\mathfrak{M}^W}$. The proposition below captures the idea that ultrafilters are used to describe the sense of 'most'. More explicitly, for an ultrafilter U on a set J, we can regard U as the set of subsets of J which can be regarded as 'most of' the elements in J. From this viewpoint, the closure property under intersection can be interpreted as 'if two subsets of J both contain most of the elements in J, then their intersection also contains most of the elements in J'. The upward closure property can be regarded as 'if a subset S of J contains most of the elements in J, then every superset of S also contains most of the elements in J'. Finally, if a subset of J is regarded as 'most of the elements in J', then we are regarding its complement as 'a small part of the elements in J', so its complement cannot also be in the ultrafilter. Given this intuition, the proposition below says that a formula ϕ is satisfied in an ultrafilter u on $\mathfrak{M}^W$ iff ϕ is satisfied at most of the worlds in $\mathfrak{M}$, where the sense of 'most' is measured by u, as described above.

Proposition 5.8. [1, Proposition 2.59 (i)] *A formula ϕ is satisfied at an ultrafilter u in the ultrafilter extension of $\mathfrak{M}$ if and only if in the unextended model, the set of worlds in $\mathfrak{M}$ satisfying ϕ is in u.*

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

Proof. By induction on ϕ. Three cases are straightforward. The diamond case requires some manipulation using Proposition 5.6, Proposition 5.5 (2) and Proposition 5.3.

□

The above proposition leads to a proof of M-saturatedness of ultrafilter extensions.

Proposition 5.9. [1, Proposition 2.61] *The ultrafilter extension of each model is M-saturated.*

$$\vdash \textsf{M\_sat}\ {}^{ue}\mathfrak{M}$$

Proof. Suppose Σ is a set of formulas which is finitely satisfiable in the set of successors of a world $u\ \in\ {}^{ue}\mathfrak{M}^W$, we need to find a world $u'\ \in\ {}^{ue}\mathfrak{M}^W$ such that ${}^{ue}\mathfrak{M}^R\ u\ u'$ and ${}^{ue}\mathfrak{M}, u' \Vdash\ \phi$ for all $\phi\ \in\ \Sigma$. By Proposition 5.6 and Proposition 5.8, it amounts to find an ultrafilter u' on $\mathfrak{M}^W$ such that $\{\ Y \mid \mathfrak{M}^\delta_\Diamond(Y)\ \in\ u\ \}\ \subseteq\ u'$ and $\{\ w \mid w\ \in\ \mathfrak{M}^W\ \wedge\ \mathfrak{M}, w \Vdash\ \phi\ \}\ \in\ u'$ for all $\phi\ \in\ \Sigma$.

Consider the set $\mathit{\Delta}$

$$\{ \{ w \mid w \in \mathfrak{M}^W \wedge \forall \phi.\ \phi \in s \Rightarrow \mathfrak{M}, w \Vdash \phi \} \mid \mathsf{FINITE}\ s \wedge s \subseteq \mathit{\Sigma} \} \cup$$
$$\{ Y \mid \mathfrak{M}^{\delta}_{\Diamond}(Y) \in u \wedge Y \subseteq \mathfrak{M}^W \},$$

we check $\mathit{\Delta}$ has the finite intersection property. The only nontrivial thing to check is that for a in the first set of the union and b in the second set of the union, we have $a \cap b \neq \emptyset$.

Suppose $s \subseteq \Sigma$ is finite, and b is a set of worlds in $\mathfrak{M}$ such that $\mathfrak{M}^{\delta}_{\Diamond}(b) \in u$, we show $\{ w \mid w \in \mathfrak{M}^W \wedge \forall \phi.\ \phi \in s \Rightarrow \mathfrak{M}, w \Vdash \phi \} \cap b \neq \emptyset$. Recall $\mathit{\Sigma}$ is finitely satisfiable in the set of successors of u, we have a world u'' such that ${}^{ue}\mathfrak{M}^R\ u\ u''$ and ${}^{ue}\mathfrak{M}, u'' \Vdash \phi$ for all $\phi \in s$, in other worlds, $\{ w \mid w \in \mathfrak{M}^W \wedge \mathfrak{M}, w \Vdash \phi \} \in u''$ for all $\phi \in s$. Then as s is finite, $\{ w \mid w \in \mathfrak{M}^W \wedge \forall \phi.\ \phi \in s \Rightarrow \mathfrak{M}, w \Vdash \phi \}$ is a big intersection of finitely many sets in u'', and hence is in u''. By Proposition 5.6 again, ${}^{ue}\mathfrak{M}^R\ u\ u''$ gives $\{ Y \mid \mathfrak{M}^{\delta}_{\Diamond}(Y) \in u \wedge Y \subseteq \mathfrak{M}^W \} \subseteq u''$, so $b \in u''$ as well. As two elements in u'' has a nonempty intersection, we are done.

Hence by Proposition 5.3, there exists an ultrafilter u' contains $\mathit{\Delta}$ is routine to check u' is what we want. □

Finally, we arrive at the characterization of modal equivalence as bisimilarity in the ultrafilter extensions:

Theorem 5.10. [1, Proposition 2.62] *Given two models $\mathfrak{M}_1$ and $\mathfrak{M}_2$ with $w_1 \in \mathfrak{M}_1^W$ and $w_2 \in \mathfrak{M}_2^W$, w_1 and w_2 are modal equivalent if and only if the principal filters generated by w_1 in the ultrafilter extension of $\mathfrak{M}_1$ and the principal filter generated by w_2 in the ultrafilter extension of $\mathfrak{M}_2$ are bisimilar.*

$$\vdash w_1 \in \mathfrak{M}_1^W \wedge w_2 \in \mathfrak{M}_2^W \Rightarrow$$
$$(\mathfrak{M}_1, w_1 \leftrightsquigarrow \mathfrak{M}_2, w_2 \iff {}^{ue}\mathfrak{M}_1, \pi_{w_1}^{\mathfrak{M}_1^W} \underline{\leftrightarrow} {}^{ue}\mathfrak{M}_2, \pi_{w_2}^{\mathfrak{M}_2^W})$$

Proof. Bisimulation implies modal equivalence by Theorem 3.6. For the reverse direction, if $w_1 \in \mathfrak{M}_1^W$ and $w_2 \in \mathfrak{M}_2^W$ are modal equivalent, then $\pi_{w_1}^{\mathfrak{M}_1^W} \in {}^{ue}\mathfrak{M}_1{}^W$ is modal equivalent to $\pi_{w_2}^{\mathfrak{M}_2^W} \in {}^{ue}\mathfrak{M}_2{}^W$ by Proposition 5.7. As ${}^{ue}\mathfrak{M}_1$ and ${}^{ue}\mathfrak{M}_2$ are M-saturated by Proposition 5.9, the result follows by Proposition 5.1. □

Chapter 6

Two Characterizing Results

In Chapter 4, we presented the definition of ‘invariant for bisimulation’, and mentioned that if we work with set theory, then we can prove that a first-order formula with no more than one free variable is invariant for bisimulation if and only if it is equivalent to the standard translation of some modal formula. We have translated one direction of the set-theoretic proof of this result there. In the first section of this chapter, we will translate the other half of the proof into HOL, and explain why we cannot get the double implication when working with simple type theory. In the second section, we translate from set theory to simple type theory the proof of a theorem saying a modal formula is *preserved under simulation* if and only if it is *positive existential*. The two new terminologies are to be introduced in the second section of this chapter.

6.1 The ‘modal’ fragment of $\mathcal{L}^1_\tau$ formulas

In Chapter 4, we introduced the standard translation and proved that every modal formula is ‘equivalent to’ its standard translation. We also mentioned that the standard translation of every modal formula is $\mathcal{L}^1_\tau$ and has at most one free variable. However, it is not the fact that every $\mathcal{L}^1_\tau$ formula with at most one free variable is equivalent to the standard translation of a modal formula. For instance, the first-order formula saying ‘there exists a state that is related to the current state’ can never be a standard translation. In HOL:

Proposition 6.1 (`non_ST_exists`). *The formula* ${}^{\mathsf{f}}\exists\ 2\ ({}^{\mathsf{f}}\mathsf{R}\ ({}^{\mathsf{f}}\mathsf{VAR}\ 2)\ ({}^{\mathsf{f}}\mathsf{VAR}\ 1))$ *is not a standard translation.*

$$\vdash \neg\exists\,\phi.\ \mathsf{ST}_1\ \phi\ \ {}^{\mathsf{f}}\!\equiv_{(:\mathit{num})}\ {}^{\mathsf{f}}\exists\ 2\ ({}^{\mathsf{f}}\mathsf{R}\ ({}^{\mathsf{f}}\mathsf{VAR}\ 2)\ ({}^{\mathsf{f}}\mathsf{VAR}\ 1))$$

Here comes a question: which $\mathcal{L}^1_\tau$-formulas with no more than one free variable are equivalent to a standard translation? We have already given a short answer that works for set-theoretic foundation: such formulas are exactly the ones which are invariant under bisimulation. We have proved one direction of this result as Proposition 4.5. In this section, we devote to proving in HOL the other direction, saying 'every formula which is invariant under bisimulation is equivalent to a standard translation'. The tools that this proof will use are centered on saturated models, which we are introducing now.

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 proof in this section, we are only interested in adding the interpretation of finitely many nullary function symbols, which are also called *constants.*

Definition 6.1. [1, Definition A.9 (Expansion)] *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* $\mathfrak{M}'$*.*

$$\begin{aligned}
&\mathsf{is\_expansion}\ \mathfrak{M}\ A\ \mathfrak{M}'\ f \overset{\mathsf{def}}{=} \\
&\quad \mathfrak{M}'.\mathsf{Dom} = \mathfrak{M}.\mathsf{Dom} \land \mathsf{BIJ}\, f\ (\mathsf{count}\ (\mathsf{CARD}\ A))\ A \land \\
&\quad \mathfrak{M}'.\mathsf{Fun} = \\
&\quad\quad (\lambda\, c\ l.\ \mathtt{if}\ c < \mathsf{CARD}\ A \land l = []\ \mathtt{then}\, f\ c\ \mathtt{else}\ \mathsf{CHOICE}\, \mathfrak{M}.\mathsf{Dom}) \land \\
&\quad \mathfrak{M}'.\mathsf{Pred} = \mathfrak{M}.\mathsf{Pred}
\end{aligned}$$

In above, the function CARD gives the cardinality of a finite set. For each natural number n, we have $\mathsf{count}\ n = \{0, \cdots n-1\}$. And $\mathsf{BIJ}\, f\ A\ B$ reads 'f is a bijection from A to B'.

If $\mathfrak{M}$, A and f are all fixed, then the model $\mathfrak{M}'$ such that $\mathsf{is\_expansion}\ \mathfrak{M}\ A\ \mathfrak{M}'\ f$ is unique. We define the expansion as a predicate instead of a function only for the convenience of manipulating the theorem prover when we prove theorems about expansion.

The only difference between a model and an expansion of it is the interpretation of function symbols. Before the expansion, as there is no information contained in $\mathfrak{M}$ about function symbols, once a first-order term contains some function symbol, this term will not make sense to $\mathfrak{M}$, hence $\mathfrak{M}$ can be used to interpret no formula with a function symbol. If A has cardinality m, where m is a natural number, after the expansion, every term $^{\mathsf{f}}\mathsf{Fn}\ c\ []$ for $0 \le c < m$ makes sense to $\mathfrak{M}'$, and is evaluated to the element $f\ c$. Therefore, the formulas which only use these function symbols can be interpreted in the expanded model. The role of f here is to assign each $0 \le c < m$ an element of A, where the term $^{\mathsf{f}}\mathsf{Fn}\ c\ []$ will be evaluated to.

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 $\mathsf{consistent}\ \mathfrak{M}\ \Sigma$. 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 '$\mathsf{ftype}\ x\ \Gamma$' 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 $\phi \in \Gamma$. That is, all the elements in Γ are satisfied at the point w. In this case, we write '$\mathsf{frealizes}\ \mathfrak{M}\ x\ \Gamma$' in HOL.

Definition 6.2. [1, Definition 2.63 (Countably Saturated)] *Let $\mathfrak{M}$ be a model and n be a natural number. For every $A \subseteq \mathfrak{M}.\mathsf{Dom}$, with $|A| < n$ and for every $f : \mathbb{N} \to \mathfrak{M}.\mathsf{Dom}$, there is a unique $\mathfrak{M}'$ such that $\mathsf{is\_expansion}\ \mathfrak{M}\ A\ \mathfrak{M}'\ f$. If every such $\mathfrak{M}'$ realizes every x-type Γ, then we say $\mathfrak{M}$ is n-saturated. In HOL:*

$$
\begin{array}{l}
\mathsf{n\_saturated}\ \mathfrak{M}\ n \overset{\mathsf{def}}{=} \\
\quad \forall\, A\ \mathfrak{M}'\ \Gamma\ x\ f. \\
\qquad \mathsf{IMAGE}\ f\ \mathcal{U}(:\mathit{num}) \subseteq \mathfrak{M}.\mathsf{Dom} \wedge \mathsf{FINITE}\ A \wedge \mathsf{CARD}\ A \le n \wedge A \subseteq \mathfrak{M}.\mathsf{Dom} \wedge \\
\qquad \mathsf{is\_expansion}\ \mathfrak{M}\ A\ \mathfrak{M}'\ f \wedge \\
\qquad (\forall\, \phi.\ \phi \in \Gamma \Rightarrow \mathsf{form\_functions}\ \phi \subseteq \{\, (c, 0) \mid c < \mathsf{CARD}\ A \,\}) \wedge \mathsf{ftype}\ x\ \Gamma \wedge \\
\qquad \mathsf{consistent}\ \mathfrak{M}'\ \Gamma \Rightarrow \\
\qquad\ \ \mathsf{frealizes}\ \mathfrak{M}'\ x\ \Gamma
\end{array}
$$

$$\mathsf{countably\_saturated}\ \mathfrak{M} \overset{\mathsf{def}}{=} \forall\, n.\ \mathsf{n\_saturated}\ \mathfrak{M}\ n$$

As an easy example, we have:

Proposition 6.2. [1, Example 2.64 (iii)] *Let $\mathfrak{M}$ be the model with domain $\mathbb{N}$, no functions, and the only predicates are the unary ones, such that the predicate with symbol n is interpreted as 'greater than n'. Then $\mathfrak{M}$ is not countably saturated.*

$$
\begin{array}{l}
\vdash \neg\mathsf{countably\_saturated} \\
\quad \langle\!\langle \mathsf{Dom} := \mathcal{U}(:\mathit{num});\ \mathsf{Fun} := (\lambda\, f\ l.\ \mathsf{CHOICE}\ \mathcal{U}(:\mathit{num})); \\
\quad \mathsf{Pred} := (\lambda\, n\ v.\ \exists\, x.\ v = [x] \wedge n < x) \rangle\!\rangle
\end{array}
$$

Proof. Consider the set Γ of all formulas of form ${}^{\mathsf{f}}\mathsf{P}\ n\ ({}^{\mathsf{f}}\mathsf{VAR}\ a)$, where n is a natural number. The model $\mathfrak{M}$ is the expansion of itself by adding an empty set of constants. The set Γ is consistent with $\mathfrak{M}$: it is clearly an a-type, and every finite subset of Γ is realized in $\mathfrak{M}$ since there the set of all natural number is not bounded above under the usual ordering. But Γ is not realized at any natural number, since there is no natural number that is greater than every natural number. □

The interest in countably saturated models stems from the fact that if two modal models $\mathfrak{M}_1$ and $\mathfrak{M}_2$ are both countably saturated when viewed as first-order models, then bisimulation and modal equivalence between worlds in $\mathfrak{M}_1$ and in $\mathfrak{M}_2$ coincide.

Theorem 6.3. [1, Proposition 2.65, second half]

$$\begin{aligned}&\vdash \mathsf{countably\_saturated}\ (\mathsf{mm2folm}\ \mathfrak{M}_1)\ \wedge\ \mathsf{countably\_saturated}\ (\mathsf{mm2folm}\ \mathfrak{M}_2)\ \wedge\\ &\quad w_1 \in \mathfrak{M}_1^W \ \wedge\ w_2 \in \mathfrak{M}_2^W \ \Rightarrow\\ &\quad (\mathfrak{M}_1, w_1 \leftrightsquigarrow \mathfrak{M}_2, w_2 \iff \mathfrak{M}_1, w_1 \Leftrightarrow \mathfrak{M}_2, w_2)\end{aligned}$$

By Proposition 5.1, to prove the above, it suffices to prove:

Theorem 6.4. [1, Theorem 2.65, first half] *If a modal model is countably saturated when we view it as a first-order model, then this model is M-saturated.*

$$\vdash \mathsf{countably\_saturated}\ (\mathsf{mm2folm}\ \mathfrak{M})\ \Rightarrow\ \mathsf{M\_sat}\ \mathfrak{M}$$

Proof. Suppose $\mathsf{countably\_saturated}\ (\mathsf{mm2folm}\ \mathfrak{M})$. Let $a \in \mathfrak{M}^W$ and Σ be a set of modal formulas which is finitely satisfiable in the set of successors of a. We find a successor of a in $\mathfrak{M}$ realizing all the formulas in Σ.

Define $\Sigma' = \{\, {}^{\mathsf{f}}\mathsf{R}\ ({}^{\mathsf{f}}\mathsf{Fn}\ 0\ [])\ ({}^{\mathsf{f}}\mathsf{VAR}\ x) \,\} \cup \{\, \mathsf{ST}_x\ \phi \mid \phi \in \Sigma \,\}$ and let $\mathfrak{M}'$ be the model obtained by expanding $\mathsf{mm2folm}\ \mathfrak{M}$ by adding a constant which is represented by 0 and corresponds to the world a. Then the term ${}^{\mathsf{f}}\mathsf{Fn}\ 0\ []$ will be evaluated to a by any valuation of $\mathfrak{M}'$. We claim $\mathsf{consistent}\ \mathfrak{M}'\ \Sigma'$. Take a finite set $\Sigma_0 \subseteq \Sigma'$, we should find an element in $\mathfrak{M}'$ where every formula in Σ_0 is satisfied. For each element in Σ_0 which is a standard translation, choose only one modal formula $p \in \Sigma$ that is translated to it. We do need to choose these formulas using the choice function since the standard translation function is not injective. We call the set of all the formulas chosen by this way A. Then A is a finite subset of Σ. Recall we have assumed Σ is finitely satisfiable in the set of successors of a, hence there exists $b \in \mathfrak{M}^W$ and $\mathfrak{M}^R\ a\ b$ such that $\mathfrak{M}, b \Vdash p$ for every $p \in A$.

It follows by Proposition 4.3 that no matter whether $^{\mathsf{f}}\mathsf{R}\ (^{\mathsf{f}}\mathsf{Fn}\ 0\ [])\ (^{\mathsf{f}}\mathsf{VAR}\ x)$ is in Σ_0 or not, we have Σ_0 is satisfied at b in $\mathfrak{M}'$.

This proves $\mathsf{consistent}\ \mathfrak{M}'\ \Sigma'$. Since $\mathsf{mm2folm}\ \mathfrak{M}$ is countably saturated, the whole set Σ' itself is satisfied in some w in $\mathfrak{M}'$. The fact that $^{\mathsf{f}}\mathsf{R}\ (^{\mathsf{f}}\mathsf{Fn}\ 0\ [])\ (^{\mathsf{f}}\mathsf{VAR}\ x)$ holds at w implies w is a successor of a in $\mathfrak{M}$, and $\{\ \mathsf{ST}_x\ \phi \mid \phi\ \in\ \Sigma\ \}$ holds at w implies that $\mathfrak{M}, w \Vdash\ \phi$ for every $\phi\ \in\ \Sigma$ by Proposition 4.3. □

As a reader may observe, we actually only need that $\mathsf{mm2folm}\ \mathfrak{M}$ is 2-saturated for the proof above.

Knowing the interesting properties of countably saturated models, we now answer the question of where to obtain them. The canonical way to obtain such models involves the usage of ultraproducts, which we will discuss in the following interlude.

6.1.1 Interlude: Countably saturated models via ultraproducts

Although we are ultimately interested in ultraproducts on models, we will begin by introducing the construction of ultraproducts of sets.

Definition 6.3. [1, Page 495 (Cartesian product)] *Suppose J is a non-empty set indexing the family $(A_j)_{j\in J}$, where each A_j is non-empty. The Cartesian product of the family $(A_j)_{j\in J}$ is the set of functions f with domain J such that for all $j \in J$, $f(j) \in A_j$.*

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

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.

Definition 6.4. [1, Definition 2.69 (Ultraproduct of Sets)] *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. For an ultrafilter U on a set J and a family A^{s} indexed by J, $\sim_U^{A^{\mathsf{s}}}$ is an equivalence relation on the Cartesian product of the A^{s}. The ultraproduct of A^{s} modulo U is the set of equivalence classes obtained by partitioning $\mathsf{Cart\_prod}\ J\ A^{\mathsf{s}}$ using the relation $\sim_U^{A^{\mathsf{s}}}$.*

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

We will 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.

We have notions of ultraproduct for both modal and first-order models. For modal models:

Definition 6.5. [1, Definition 2.70 (Ultraproduct of Modal Models)] *Given a family $\mathfrak{M}^{\mathsf{s}}$ of modal models indexed by J and an ultrafilter U on J, where $\mathfrak{M}^{\mathsf{s}}$ is encoded as a function that takes an element of J and gives a model, 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.*
- *For two equivalence classes f_U, g_U of functions in the ultraproduct, they are related 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.*
- *For a propositional letter p and an equivalence class f_U, we have p is satisfied at f_U 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.*

In HOL:

```
$\mathsf{ultraproduct\_model}\ U\ J\ \mathfrak{M}^{\mathsf{s}} \overset{\mathsf{def}}{=}$
  $\langle\!\langle \mathsf{frame} :=$
    $\langle\!\langle \mathsf{world} := \mathsf{ultraproduct}\ U\ J\ (\mathsf{worlds}\ \mathfrak{M}^{\mathsf{s}});$
      $\mathsf{rel} :=$
        $(\lambda f_U\ g_U.$
           $\exists f_0\ g_0.$
              $f_0 \in f_U \wedge g_0 \in g_U \wedge$
              $\{\ j \mid j \in J \wedge (\mathfrak{M}^{\mathsf{s}}\ j)^R\ (f_0\ j)\ (g_0\ j)\ \} \in U)\rangle\!\rangle;$
   $\mathsf{valt} :=$
     $(\lambda p\ f_U.\ \exists f_0.\ f_0 \in f_U \wedge \{\ j \mid j \in J \wedge f_0\ j \in (\mathfrak{M}^{\mathsf{s}}\ j)^V\ p\ \} \in U)\rangle\!\rangle$
```

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

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

In the definition of the relation and valuation of the ultraproduct modal model, the occurrence of the existential quantifier is used to describe the existence of representatives of an equivalence class with a certain additional property. As we expect, since $\sim^A_U$ is an equivalence relation for every ultrafilter, the choice of representative does not matter: 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 6.5 (`Los_modal_thm`). *The modal version of Łoś's theorem states that for U, an ultrafilter on J, and $\mathfrak{M}^{\mathsf{s}}$ a family of models, a modal formula ϕ is satisfied at an equivalence class f_U in the ultraproduct if and only if there exists a function $f_0 \in f_U$ such that the set of elements $j \in J$ such that $\mathfrak{M}^{\mathsf{s}}\ j, f_0\ j \Vdash \phi$ is in U*

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

Proof. Given an ultrafilter U on J and a family $\mathfrak{M}^{\mathsf{s}}$ of modal models, we proceed by induction on ϕ. The base case for $\phi = \mathsf{VAR}\ p$ is directly by definition, and the case for $\phi = \bot$ is by the fact that the empty set is not in the ultrafilter. The cases on disjunction and negation are by basic properties of ultrafilters. We only spell out the proof for diamond case. The induction hypothesis gives for every equivalence f_U in the ultraproduct, we have

$$\Pi_U\ \mathfrak{M}^{\mathsf{s}}, f_U \Vdash \phi\ \iff\ \exists f_0.\ f_0\ \in\ f_U\ \wedge\ \{\, j \mid j\ \in\ J\ \wedge\ \mathfrak{M}^{\mathsf{s}}\ j, f_0\ j \Vdash \phi\,\}\ \in\ U$$

Given a world f_U in $\Pi_U\ \mathfrak{M}^{\mathsf{s}}$, we will prove

$$\Pi_U\ \mathfrak{M}^{\mathsf{s}}, f_U \Vdash \Diamond\phi\ \iff\ \exists f_0.\ f_0\ \in\ f_U\ \wedge\ \{\, j \mid j\ \in\ J\ \wedge\ \mathfrak{M}^{\mathsf{s}}\ j, f_0\ j \Vdash \Diamond\phi\,\}\ \in\ U$$

Left to right: Assume the left-hand side, then there is an equivalence class g_U that is related to f_U and satisfies ϕ. By inductive hypothesis, independence of representatives and definition of the ultraproduct model, for the representative f of f_U and the representative g of g_U, both $\{\, j \mid j\ \in\ J\ \wedge\ \mathfrak{M}^{\mathsf{s}}\ j, g\ j \Vdash \phi\,\}$ and $\{\, j \mid j\ \in\ J\ \wedge\ (\mathfrak{M}^{\mathsf{s}}\ j)^R\ (f\ j)\ (g\ j)\,\}$ are in U, and hence so does their intersection M. Therefore, the set

$$A\ =\ \{\, j \mid j\ \in\ J\ \wedge\ \exists\, v.\ (\mathfrak{M}^{\mathsf{s}}\ i)^R\ (f\ j)\ v\ \wedge\ v\ \in\ (\mathfrak{M}^{\mathsf{s}}\ j)^W\ \wedge\ \mathfrak{M}^{\mathsf{s}}\ j, v \Vdash \phi\,\}$$

is in U, as a superset of M. This proves f can be taken as the f_0 that we require.

Right to left: Suppose there is an $f_0 \in f_U$ such that

$$\{ j \mid j \in J \wedge \exists v.\, (\mathfrak{M}^{\mathsf{s}}\, j)^R\, (f_0\, j)\, v \wedge v \in (\mathfrak{M}^{\mathsf{s}}\, j)^W \wedge \mathfrak{M}^{\mathsf{s}}\, j, v \Vdash \phi \}$$

is in U, we need to find an equivalence class which is related to f_U and satisfies ϕ, which by definition of relation in the ultraproduct model, amounts to find a representative of such an equivalence class. The representative is given by:

- For an element $j \in J$, if there exists a world $v \in (\mathfrak{M}^{\mathsf{s}}\, j)$.world such that $(\mathfrak{M}^{\mathsf{s}}\, j)^R\, (f_0\, j)\, v$ and $\mathfrak{M}^{\mathsf{s}}\, j, v \Vdash \phi$, then we choose such a v to send j to.
- For an element $j \in J$, if such a world v as described above does not exists, we send j to an arbitrary world in $(\mathfrak{M}^{\mathsf{s}}\, j)^W$.

In HOL, the representative described above is defined as:

```
λ j.
  if $\exists v.\, (\mathfrak{M}^{\mathsf{s}}\, j)^R\, (f_0\, j)\, v \wedge v \in (\mathfrak{M}^{\mathsf{s}}\, j)^W \wedge \mathfrak{M}^{\mathsf{s}}\, j, v \Vdash \phi$ then
   CHOICE $\{ v \mid (\mathfrak{M}^{\mathsf{s}}\, j)^R\, (f_0\, j)\, v \wedge v \in (\mathfrak{M}^{\mathsf{s}}\, j)^W \wedge \mathfrak{M}^{\mathsf{s}}\, j, v \Vdash \phi \}$
  else CHOICE $(\mathfrak{M}^{\mathsf{s}}\, j)^W$
```

□

In the case that we are taking the ultraproduct of a constant family of models with $\mathfrak{M}^{\mathsf{s}}\, j = \mathfrak{M}$ for all $j \in J$, we get an ultrapower of $\mathfrak{M}$. Specializing Theorem 6.5 to the case of ultrapowers yields:

Corollary 6.6. [1, Proposition 2.71]

If for every $j \in J$, $\mathfrak{M}^{\mathsf{s}}\, j = \mathfrak{M}$ ($\mathfrak{M}^{\mathsf{s}}$ is a "constant family"), then the equivalence class of the constant function mapping every j to a fixed world w satisfies ϕ in the ultraproduct model iff w satisfies ϕ in the original model $\mathfrak{M}$.

$$\vdash (\forall j.\, j \in J \Rightarrow \mathfrak{M}^{\mathsf{s}}\, j = \mathfrak{M}) \wedge \mathsf{ultrafilter}\ U\ J \Rightarrow$$
$$\forall \phi\ w.\, \Pi_U\, \mathfrak{M}^{\mathsf{s}}, \{ f \mid (\lambda j.\, w) \sim_U^{\mathsf{worlds}\, \mathfrak{M}^{\mathsf{s}}} f \} \Vdash \phi \iff \mathfrak{M}, w \Vdash \phi$$

The construction of ultraproduct of first-order models is similar to the construction for modal models, but a bit more complicated, since we will have predicates and functions to deal with.

Definition 6.6. [1, Definition A.18 (Ultraproduct of 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*.*

In HOL:

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

In above, the function MAP takes a function f and a list l, and gives the list whose n-th member is the image of the n-th member of l under f.

Here we fix the representative of each equivalence class f_U to be CHOICE f_U. The function Doms takes a family of first-order models to the family of their domains. It plays the same role as the function worlds in the definition of ultraproduct of modal models.

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

The first part of this theorem describe how first-order ultraproduct models interpret terms. As for all models, this interpretation is performed by the termval function.

Theorem 6.7. [1, Theorem A.19 (Łoś's theorem) (i)] *For an ultraproduct model of the family* $\mathfrak{M}^{\mathsf{s}}$ *of first-order models, a valuation* σ *assigns each natural number an equivalence class in the ultraproduct of the world sets of the family. A term* t *will be evaluated to the equivalence class of the function that maps a index* $j \in J$

to $\mathsf{termval}\ (\mathfrak{M}^{\mathsf{s}}\ j)\ (\lambda\, v.\ \mathsf{CHOICE}\ (\sigma\ v)\ j)\ t$.

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

In above, for each $j \in J$, the function that sends a variable v (which is a natural number) to $\mathsf{CHOICE}\ (\sigma\ v)\ j$ is indeed a valuation of the model $\mathfrak{M}^{\mathsf{s}}\ j$. As we can see: for each v, the representative $\mathsf{CHOICE}\ (\sigma\ v)$ of the equivalence class $\sigma\ v$ is an element in the Cartesian product $\mathsf{Cart\_prod}\ J\ (\mathsf{Doms}\ \mathfrak{M}^{\mathsf{s}})$. By definition of Cartesian product, this means that for each $j \in J$, we have $\mathsf{CHOICE}\ (\sigma\ v)\ j\ \in\ (\mathfrak{M}^{\mathsf{s}}\ j).\mathsf{Dom}$.

The second part of Łoś's theorem characterizes satisfaction of first-order formulas on ultraproduct models:

Theorem 6.8. [1, Theorem A.19 (Łoś's theorem) (ii)] *For the ultraproduct of a family $\mathfrak{M}^{\mathsf{s}}$ of first-order models over an ultrafilter U on J, a formula ϕ is satisfied under a valuation σ if and only if the set indexing the models $\mathfrak{M}^{\mathsf{s}}\ j$ in the family where ϕ is true under the valuation $\lambda\, v.\ \mathsf{CHOICE}\ (\sigma\ v)\ j$ is in the ultrafilter U.*

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

Proof. By induction on ϕ. The base case for ${}^{\mathsf{f}}\bot$ comes from the fact that the empty set is not in an ultrafilter. The atomic case is a direct translation of its mathematical proof, which uses Theorem 6.7. The implication case is trivial from the inductive hypothesis. We only spell out the proof for the case for universal quantifier.

The implication from right to left is straightforward. From left to right, suppose ${}^{\mathsf{f}}\forall\ x\ \phi$ is satisfied under a valuation σ in the ultraproduct model for $\mathfrak{M}^{\mathsf{s}}$. We need $\{j \mid j \in J \wedge \mathfrak{M}^{\mathsf{s}}\ j, (\lambda\, v.\ \mathsf{CHOICE}\ (\sigma\ v)\ j) \vDash {}^{\mathsf{f}}\forall\ x\ \phi\} \in U$. Suppose not, then as U is an ultrafilter, the complement of the set above, which is:

$$A\ =\ \{\, j \mid j\ \in\ J\ \wedge\ \mathfrak{M}^{\mathsf{s}}\ j, (\lambda\, v.\ \mathsf{CHOICE}\ (\sigma\ v)\ j) \vDash\ {}^{\mathsf{f}}\exists\ x\ ({}^{\mathsf{f}}\neg\ \phi)\,\}$$

is in U. Using choice, we define a function f by sending each $j \in J$ to a chosen point in $(\mathfrak{M}^{\mathsf{s}}\ j).\mathsf{Dom}$ where ϕ is not satisfied if such a point exists, and $\mathsf{CHOICE}\ (\mathfrak{M}^{\mathsf{s}}\ j)$ if such a point does not exist. Using the inductive hypothesis and the fact that $A \in U$, we can show the equivalence class represented by f does not satisfy ϕ, which contradicts our assumption. □

Łoś's theorem gives a classical corollary:

Corollary 6.9. [1, Corollary A.21] *For every ultrafilter U on J, every first-order model $\mathfrak{M}$ is embedded in its ultrapower on U by sending an element in its domain to the equivalence class of the constant function on that element.*

$$\begin{array}{l}\vdash \mathsf{ultrafilter}\ U\ J\ \wedge\ (\forall j.\ j\ \in\ J\ \Rightarrow\ \mathfrak{M}^{\mathsf{s}}\ j\ =\ \mathfrak{M})\ \wedge\ \mathsf{wffm}\ \mathfrak{M}\ \wedge\ \mathsf{valuation}\ \mathfrak{M}\ \sigma\ \Rightarrow\\ \quad(\mathfrak{M}, \sigma \vDash\ \phi\ \iff\ {}^{\mathsf{f}}\Pi_U\ \mathfrak{M}^{\mathsf{s}}, (\lambda\, v.\ \{\ g \mid g \sim_U^{\mathsf{Doms}\ \mathfrak{M}^{\mathsf{s}}}\ (\lambda\, j.\ \sigma\ v)\ \}\,) \vDash\ \phi)\end{array}$$

The above corollary is straightforward to prove once we get the following lemma:

Lemma 6.10 (`ultraproduct_rep_independence_lemma`)**.** *Given a family $\mathfrak{M}^{\mathsf{s}}$ of first-order models indexed over J and an ultrafilter U on J. Let σ be a valuation on the ultraproduct model of $\mathfrak{M}^{\mathsf{s}}$ over U. For a first-order formula ϕ, let σ_{rep} be a function assigning each free variable v of ϕ a representative in the equivalence class $\sigma\ v$. Then the set of $j \in J$ that indexing the models $\mathfrak{M}^{\mathsf{s}}\ j$ where ϕ is satisfied under the valuation $\lambda\, v.\ \mathsf{CHOICE}\ (\sigma\ v)\ j$ is in U if and only if the set of elements $j \in J$ indexing the the models $\mathfrak{M}^{\mathsf{s}}\ j$ where ϕ is satisfied under the valuation $\lambda\, v.\ \sigma_{\mathsf{rep}}\ v\ j$ is in U.*

$$\begin{array}{l}\vdash (\mathsf{ultrafilter}\ U\ J\ \wedge\ \mathsf{valuation}\ ({}^{\mathsf{f}}\Pi_U\ \mathfrak{M}^{\mathsf{s}})\ \sigma)\ \wedge\\ \quad(\forall\, v.\ v\ \in\ \mathsf{FV}\ \phi\ \Rightarrow\ \sigma_{\mathsf{rep}}\ v\ \in\ \sigma\ v)\ \Rightarrow\\ \qquad(\{\, j \mid j\ \in\ J\ \wedge\ \mathfrak{M}^{\mathsf{s}}\ j, (\lambda\, v.\ \mathsf{CHOICE}\ (\sigma\ v)\ j) \vDash\ \phi\,\}\ \in\ U\ \iff\\ \qquad\quad\{\, j \mid j\ \in\ J\ \wedge\ \mathfrak{M}^{\mathsf{s}}\ j, (\lambda\, v.\ \sigma_{\mathsf{rep}}\ v\ j) \vDash\ \phi\,\}\ \in\ U)\end{array}$$

In the theorem above, if the index set J is an α-set and $\mathfrak{M}^{\mathsf{s}}$ is a family of β-first-order models, then σ_{rep} here is of type `num` $\to\ \alpha\ \to\ \beta$. This lemma is very helpful since it enables us to be free of choice of representatives of equivalence classes in the ultraproduct when applying Łoś's theorem.

Proposition 6.11 (`ultraproduct_suffices_rep`)**.** *If we want to find a valuation of a ultraproduct model satisfying a first-order formula ϕ, instead of assigning equivalence classes to natural numbers directly, it suffices to assign representatives.*

$$\begin{array}{l}\vdash \mathsf{ultrafilter}\ U\ J\ \wedge\ (\forall j.\ j\ \in\ J\ \Rightarrow\ \mathsf{wffm}\ (\mathfrak{M}^{\mathsf{s}}\ j))\ \wedge\\ \quad(\forall j.\ \mathsf{valuation}\ (\mathfrak{M}^{\mathsf{s}}\ j)\ (\lambda\, v.\ \sigma_{\mathsf{rep}}\ v\ j))\ \wedge\\ \quad\{\, j \mid j\ \in\ J\ \wedge\ \mathfrak{M}^{\mathsf{s}}\ j, (\lambda\, v.\ \sigma_{\mathsf{rep}}\ v\ j) \vDash\ \phi\,\}\ \in\ U\ \Rightarrow\\ \quad\ {}^{\mathsf{f}}\Pi_U\ \mathfrak{M}^{\mathsf{s}}, (\lambda\, v.\ \{\ g \mid g \sim_U^{\mathsf{Doms}\ \mathfrak{M}^{\mathsf{s}}}\ \sigma_{\mathsf{rep}}\ v\ \}\,) \vDash\ \phi\end{array}$$

All the construction we did above serves to pave a way of getting a countably saturated model. For a family of non-empty models, we will prove that their ultraproduct on a *countably incomplete* ultrafilter is countably saturated. An ultrafilter U on J is countably incomplete if there exists a family S^{s} such that $S^{\mathsf{s}}\ n\ \in\ U$ for each natural number n, and the intersection $\bigcap_{n\in\mathbb{N}} S^{\mathsf{s}}\ n$ is empty. In other words, a countably incomplete ultrafilter is an ultrafilter which is not closed under infinite intersection. Countably incomplete ultrafilters do exist. To see this, first observe that the set A of subsets of $\mathbb{N}$ of form $\mathbb{N} \setminus X$, where X is a finite subset of $\mathbb{N}$, has finite intersection property. Therefore, by Proposition 5.3, there exists an ultrafilter U that contains A. The ultrafilter U will not contain any finite set, otherwise, it will contain both a subset of $\mathbb{N}$ and its complement, and hence contradict the fact that U is an ultrafilter.

Lemma 6.12. [1, Lemma 2.73] *For a family of non-empty models, their ultraproduct on a countably incomplete ultrafilter is countably saturated.*

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

A mathematical proof of the lemma above can be found in Section 6.1 of [2]. It requires some work to translate the mathematical proof into HOL. With all the setup about ultraproduct models, we may expect that the lemma above will be a consequence of Łoś's theorem. But if we take a closer look of the statement, we will find out Łoś's theorem cannot be directly applied here. The obstacles here will become clear when we compare what we want to prove to the statement of Łoś's theorem: Łoś's theorem is about ultraproducts of first-order models, and it says nothing about expansion. But by the definition of countably saturated models, 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. However, as we shall see now, this difference cannot stop us from applying Łoś's theorem.

The first issue is to remove the expansion on the outmost layer. The key observation is that we have an alternative approach to capture the idea of 'constants'. 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 $\mathsf{shift\_form}$ 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 $0, \cdots, n-1$).

As an example, if $\mathfrak{M}'$ is obtained by adding one constant to $\mathfrak{M}$ corresponds to a point $a\ \in\ \mathfrak{M}.\mathsf{Dom}$, then after the expansion, the formulas involves the term $^{\mathsf{f}}\mathsf{Fn}\ 0\ []$ makes sense to $\mathfrak{M}'$. If we do not want to work with expansion, given a formula where the only function symbol that may occur is $(0, 0)$, then we can firstly add 1 to every variable symbol that appears in the formula, and then replace every occurrence of $^{\mathsf{f}}\mathsf{Fn}\ 0\ []$ by $^{\mathsf{f}}\mathsf{VAR}\ 0$. The formula $^{\mathsf{f}}\mathsf{R}\ (^{\mathsf{f}}\mathsf{Fn}\ 0\ [])\ (^{\mathsf{f}}\mathsf{VAR}\ 0)$ will become $^{\mathsf{f}}\mathsf{R}\ (^{\mathsf{f}}\mathsf{VAR}\ 0)\ (^{\mathsf{f}}\mathsf{VAR}\ 1)$, and the formula $^{\mathsf{f}}\mathsf{P}\ p\ (^{\mathsf{f}}\mathsf{VAR}\ 1)\ ^{\mathsf{f}}\!\vee\ ^{\mathsf{f}}\mathsf{P}\ q\ (^{\mathsf{f}}\mathsf{VAR}\ 2)$ will become $^{\mathsf{f}}\mathsf{P}\ p\ (^{\mathsf{f}}\mathsf{VAR}\ 2)\ ^{\mathsf{f}}\!\vee\ ^{\mathsf{f}}\mathsf{P}\ q\ (^{\mathsf{f}}\mathsf{VAR}\ 3)$. Therefore, after applying the shifting construction to a formula, there will be no function symbol remaining. Also, if s is the set of free variables in the formula we start with, then a free variable in the resultant formula is either of form $x\ +\ \mathsf{CARD}\ A$ for some $x \in s$, or an element in $\{0, \cdots, (\mathsf{CARD}\ A - 1)\}$ that is used to capture a constant.

Now if we still want to use an arbitrary valuation to evaluate a shifted formula, something may go wrong. Since $0, \cdots, n-1$ in the 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. Hence 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:

Definition 6.7 (Shifting on valuations)**.**

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

Continue with the previous example. Formerly, we can use the valuation $\lambda\, n.\ b$ where $b\ \in\ \mathfrak{M}.\mathsf{Dom}$ and $b\ \neq\ a$ to evaluate the formula $^{\mathsf{f}}\mathsf{R}\ (^{\mathsf{f}}\mathsf{Fn}\ 0\ [])\ (^{\mathsf{f}}\mathsf{VAR}\ 0)$. But after the shifting, it does not make sense to use the same valuation to evaluate $^{\mathsf{f}}\mathsf{R}\ (^{\mathsf{f}}\mathsf{VAR}\ 0)\ (^{\mathsf{f}}\mathsf{VAR}\ 1)$. To turn this valuation into a valuation that makes sense to the shifted formula, we need to let $^{\mathsf{f}}\mathsf{VAR}\ 0$ in the shifted formula to be evaluated to the correct place a, and let the variable symbol which is formerly sent to b to be also sent to b. Formerly, the variable symbol 0 is sent to b, but now the variable which plays the same role as the 0 after the shifting is the variable symbol 1, hence we need the 1 in the shifted formula to be sent to b, as we can check, according to our definition, $\mathsf{shift\_valuation}\ 1\ (\lambda\, v.\ b)\ (\lambda\, v.\ a)\ =\ (\lambda\, v.\ \texttt{if}\ v\ =\ 0\ \texttt{then}\ a\ \texttt{else}\ b)$ does the correct thing.

The shifting construction gives the desired semantic behavior on first-order formulas. If $\mathfrak{M}'$ is a model we get by adding a bunch of constants corresponds to elements in a set A to a model $\mathfrak{M}$, then for a first-order formula ϕ such that the function symbols appear in ϕ can only be the constants that corresponds to element in A, the formula ϕ is true in $\mathfrak{M}$ under valuation σ if and only if when we 'shift away' all the constants in ϕ and shift the valuation σ accordingly, then the resultant formula will be true on $\mathfrak{M}$ under the shifted valuation. For our aim here, we are interested in expanding a model that is obtained by converting a modal model as a first-order model, so the result we need is:

Proposition 6.13 (`expansion_shift_feval`).

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

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:

Proposition 6.14 (`ultraproduct_comm_feval`). *For the ultraproduct of a family of modal models, if we view the resultant modal ultraproduct model as a first-order model, this first-order model will satisfy the same first-order formulas without function symbols as the model we obtain by firstly view each modal model in the family as a first-order model, then take their ultraproduct as first-order models.*

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

Proof. By induction on ϕ. □

Actually, we also have:

Proposition 6.15 (`ultraproduct_comm_feval'`).

$$\begin{array}{l} \vdash \mathsf{ultrafilter}\ U\ J\ \wedge\ \mathcal{L}^1_\tau\ \phi\ \wedge\ (\forall\, j.\ j\ \in\ J\ \Rightarrow\ \mathsf{wffm}\ (\mathfrak{M}^{\mathsf{s}}\ j))\ \wedge \\ \quad \mathsf{IMAGE}\ \sigma\ \mathcal{U}(:\mathit{num})\ \subseteq\ \mathsf{ultraproduct}\ U\ J\ (\mathsf{Doms}\ \mathfrak{M}^{\mathsf{s}})\ \Rightarrow \\ \quad\ ({}^{\mathsf{f}}\Pi_U\ \mathfrak{M}^{\mathsf{s}}, \sigma \vDash\ \phi\ \iff\ \mathsf{mm2folm}\ (\Pi_U\ (\lambda\, j.\ \mathsf{folm2mm}\ (\mathfrak{M}^{\mathsf{s}}\ j))), \sigma \vDash\ \phi) \end{array}$$

In summary, the above two propositions express the fact that the order of taking ultraproduct and converting between modal and first-order models do not matter if we only consider the satisfaction of $\mathcal{L}^1_\tau$-formulas.

According to the discussion above, Proposition 6.14 and Proposition 6.13 reduce our task to the following:

Lemma 6.16 (Saturation of ultraproduct model, `ultraproduct_sat`). *Let $\mathfrak{M}^{\mathsf{s}}$ be a family of well-formed first-order models indexed by J, a countably incomplete ultrafilter U on J, a set Δ of $\mathcal{L}^1_\tau$-formulas which contain no free variables other than the ones in the set $\{x\} \cup C$, and a function f from C into the domain of ${}^{\mathsf{f}}\Pi_U\ \mathfrak{M}^{\mathsf{s}}$ (the function f serves to give meaning to the free variables in C, treating them as constants). If for every finite subset Δ_0 of Δ, there exists a valuation σ that agrees with f on the elements of C (i.e., it sends the 'constants' to the correct places), and all the formulas in Δ_0 are satisfied in ${}^{\mathsf{f}}\Pi_U\ \mathfrak{M}^{\mathsf{s}}$ under σ, then there exists a valuation σ sending the constants to the correct places that makes every formula in Δ satisfied in ${}^{\mathsf{f}}\Pi_U\ \mathfrak{M}^{\mathsf{s}}$ (which just means that σ assigns the only 'real' free variable x to a point in ${}^{\mathsf{f}}\Pi_U\ \mathfrak{M}^{\mathsf{s}}$ such that all the elements in Δ are satisfied).*

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

The above is a classical theorem on ultraproduct models. To prove it, we need another lemma about countably incomplete ultrafilters:

Proposition 6.17 (`countably_incomplete_chain`). *In a countably incomplete ultrafilter U on J, we can find a chain $J = J_0 \supseteq J_1 \supseteq J_2 \supseteq \cdots$ with each J_i in U, such that $\bigcap_{n\in\mathbb{N}} J_n = \emptyset$. The J^{s} below is a function that takes a index n, here* *a natural number, to the set that n is indexing.*

$$\begin{array}{l} \vdash \mathsf{countably\_incomplete}\ U\ J\ \Rightarrow \\ \quad \exists\, J^{\mathsf{s}}. \\ \qquad J^{\mathsf{s}}\,0\ =\ J\ \wedge\ (\forall\, n.\ J^{\mathsf{s}}\ n\ \in\ U\ \wedge\ J^{\mathsf{s}}\ (n\ +\ 1)\ \subseteq\ J^{\mathsf{s}}\ n)\ \wedge \\ \qquad \bigcap\ \{\ J^{\mathsf{s}}\ n \mid n\ \in\ \mathcal{U}(:\boldsymbol{num})\ \}\ =\ \emptyset \end{array}$$

Proof. By definition of countable incompleteness, there exists a family X_n in U indexed by natural numbers such that $\bigcap_{n \in \mathbb{N}} X_n = \emptyset$. Define $K_n := \bigcap_{m \leq n} X_n$. In HOL, the family K_n is defined as a recursive function K^{s} such that $K^{\mathsf{s}}\ 0$ is X_0 and $K^{\mathsf{s}}\ (n\ +\ 1)$ is the intersection of $K^{\mathsf{s}}\ n$ and X_{n+1}. We get the desired chain J_n by inserting J at the beginning of K_n. □

Now we can prove the saturation of ultraproducts:

Proof. Under the given assumptions, if $\varDelta$ is finite, there is nothing to prove. Hence we assume $\varDelta$ is infinite. As we are using a countable first-order language, every infinite set of first-order formula is countable, and hence there exists a bijection *enum* from the set of all natural numbers to the set $\varDelta$. It suffices to prove the existence of a valuation σ such that σ agree with f on C and moreover, ${}^{\mathsf{f}}\Pi_U\ \mathfrak{M}^{\mathsf{s}}, \sigma \vDash\ \mathit{enum}\ n$ for all natural number n. The σ we want is an assignment of variables to equivalence classes. But by Proposition 6.11 and Łoś's theorem, instead of assigning equivalence classes, it suffices to find out a function σ_{rep} that assigning each natural number a representative of some equivalence class satisfying the following conditions.

- $\forall\, v\ j.\ j\ \in\ J\ \Rightarrow\ \sigma_{\mathsf{rep}}\ v\ j\ \in\ (\mathfrak{M}^{\mathsf{s}}\ j).\mathsf{Dom}$
- $\forall\, c.\ c\ \in\ C\ \Rightarrow\ \{\ g \mid g \sim_U^{\mathsf{Doms}\ \mathfrak{M}^{\mathsf{s}}}\ \sigma_{\mathsf{rep}}\ c\ \}\ =\ f\ c$
- $\forall\, k.\ \{\ j \mid j\ \in\ J\ \wedge\ \mathfrak{M}^{\mathsf{s}}\ j, (\lambda\, v.\ \sigma_{\mathsf{rep}}\ v\ j) \vDash\ \mathit{conj}\ k\ \}\ \in\ U$

The first item says for each free variable v, the function that σ_{rep} assigns v must be an element in the Cartesian product. The second item says that the equivalence class assigned free variables in C has already been fixed by f. Both of these two are easy to be satisfied. We devote to finding a σ_{rep} satisfying the third condition.

By Proposition 6.17, we have a chain I^{s} where $I^{\mathsf{s}}\ n\ \in\ U$ and $I^{\mathsf{s}}\ (n\ +\ 1)\ \subseteq\ I^{\mathsf{s}}\ n$ for each n, which start with $I^{\mathsf{s}}\ 0\ =\ J$. Moreover, the intersection of this chain is empty. Let *conj* be the recursive function that $\mathit{conj}\ 0\ =\ {}^{\mathsf{f}}\top$, and $\mathit{conj}\ n$

is the conjunction of first-order formulas in s from *enum* 0 to *enum* $(n - 1)$. We define J^{s} to be the function that takes a natural number n and gives the set:

$$
\begin{array}{l}
J^{\mathsf{s}}\ k\ = \\
\quad \{\, j \mid \\
\quad\quad j \in J \wedge \\
\quad\quad \forall \sigma.\ (\forall c.\ c \in C \Rightarrow \sigma\ c = \mathsf{CHOICE}\ (f\ c)\ j) \Rightarrow \mathfrak{M}^{\mathsf{s}}\ j, \sigma \vDash {}^{\mathsf{f}}\exists\ x\ (\mathit{conj}\ k)\ \}
\end{array}
$$

Then $J^{\mathsf{s}}\ 0 = J$, and for $n > 0$, $J^{\mathsf{s}}\ n$ is the subset of J indexing the set of models $\mathfrak{M}^{\mathsf{s}}\ j$ with a point in its domain such that the conjunction from *enum* 0 to *enum* $(n - 1)$ are satisfied. Therefore, J^{s} is a descending chain. Since every finite subset of Δ is satisfied in ${}^{\mathsf{f}}\Pi_U\ \mathfrak{M}^{\mathsf{s}}$ by assumption, Łoś's theorem implies that $J^{\mathsf{s}}\ n \in U$ for every n. Define for each natural number n, $X^{\mathsf{s}}\ n$ is the intersection $I^{\mathsf{s}}\ n \cap J^{\mathsf{s}}\ n$, then X^{s} is a descending chain in U starting with J and the intersection of all $X^{\mathsf{s}}\ n$ is the empty set. For such a chain, each element $j \in J$ can only belong to finitely many of the sets in the family X^{s}. Hence there exists a function N that send an element j to smallest set in the chain X^{s} that j belongs to. That is, for all $j \in J$, we have $j \in X^{\mathsf{s}}\ (N\ j)$ and $j \notin X^{\mathsf{s}}\ a$ for every $a > N\ j$.

The σ_{rep} we are looking for can be taken as the function that takes a free variable v and an element $j \in J$ to an element in the domain of $\mathfrak{M}^{\mathsf{s}}\ j$, defined as:

- If $v \in C$ (v is a free variable which is actually used to capture a constant), then v it sent to $\mathsf{CHOICE}\ (f\ v)\ j$.

- If $v \notin C$ (which means that v is the x in our assumption), then choose an element $a \in (\mathfrak{M}^{\mathsf{s}}\ j).\mathsf{Dom}$ such that the formula *conj* $(N\ j)$ is satisfied in $\mathfrak{M}^{\mathsf{s}}\ j$ under the valuation that sends a free variable n in C to $\mathsf{CHOICE}\ (f\ n)\ j$ and sends the free variable x to a. We can choose such an element since we can easily prove its existence from the fact that $J^{\mathsf{s}}\ n$ is in U for every natural number n.

The first two conditions are immediate to check. It remains to show $\{\, j \mid j \in J \wedge \mathfrak{M}^{\mathsf{s}}\ j, (\lambda\, v.\ \sigma_{\mathsf{rep}}\ v\ j) \vDash \mathit{conj}\ k\,\} \in U$ for each k. Fix an arbitrary k, as we have known that $X^{\mathsf{s}}\ k$ is in U, it suffices to check $X^{\mathsf{s}}\ k \subseteq \{\, j \mid j \in J \wedge \mathfrak{M}^{\mathsf{s}}\ j, (\lambda\, v.\ \sigma_{\mathsf{rep}}\ v\ j) \vDash \mathit{conj}\ k\,\}$. For every $j \in X^{\mathsf{s}}\ k$, by definition of the function N, we have $k \leq N\ j$. As $j \in X^{\mathsf{s}}\ (N\ j)$, in particular, $j \in J^{\mathsf{s}}\ (N\ j)$. From here, we can deduce $\mathfrak{M}^{\mathsf{s}}\ j, (\lambda\, v.\ \sigma_{\mathsf{rep}}\ v\ j) \vDash \mathit{conj}\ (N\ j)$ by the definition of σ_{rep} and the definition of J^{s}. As *conj* m implies *conj* n for $n \leq m$, we are done. □

This is the end of the interlude.

With the help of Lemma 6.12, we yield another theorem about 'modal equivalence between two worlds implies bisimilarity of the two worlds when embedded in some other models'.

Theorem 6.18. [1, Theorem 2.74, one direction] *If two worlds $w \in \mathfrak{M}^W$ and $v \in \mathfrak{N}^W$ are modal equivalent, then we can find an ultrafilter U on J such that in ultrapower models of $\mathfrak{M}$ and $\mathfrak{N}$ on U respectively, there is a bisimulation between the worlds corresponding to w and v.*

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

Proof. The U we require here can be an arbitrary countably incomplete ultrafilter. Then by Lemma 6.12, the models $\mathsf{mm2folm}\ (\Pi_U\ (\lambda j.\, \mathfrak{M}))$, $\mathsf{mm2folm}\ (\Pi_U\ (\lambda j.\, \mathfrak{N}))$ are countably incomplete. Hence we are done by Proposition 6.3 and Corollary 6.6. □

The last ingredient we need for the main theorem we are proving is the compactness theorem of first-order logic. The standard statement of compactness theorem says that for a set Σ of modal formulas, if for each finite subset $\Sigma_0 \subseteq \Sigma$, there exists a model such that all the formulas in Σ_0 are satisfied, then there exists a model such that all the formulas in Σ are satisfied. This standard version of compactness theorem is formalized in 1998 in HOL by John Harrison [4]. The way that Harrison states the compactness theorem looks very different from the style that we are working with. With the help of a corollary proved from Harrison's work by my supervisor, we have connected Harrison's work to our project by proving a version of the compactness theorem for $\mathcal{L}^1_\tau$-formulas, which is no more than a specialization to the standard version of compactness theorem to $\mathcal{L}^1_\tau$-formulas. We will use this version of compactness theorem for our work. The statement looks like:

Theorem 6.19 (`compactness_thm_L1tau`). *If α-is an infinite type, then for each set Δ of $\mathcal{L}^1_\tau$-formulas, if for every finite subset $\Delta_0 \subseteq \Delta$, there exists an α-model $\mathfrak{M}$ and a valuation σ such that every formula in Δ_0 is satisfied in $\mathfrak{M}$ under σ,* *then there exists an α-model $\mathfrak{M}$ and a valuation on $\mathfrak{M}$ such that all the formulas in Δ are satisfied.*

$$\begin{array}{l} \vdash \mathsf{INFINITE}\, \mathcal{U}(:\alpha) \ \wedge \ (\forall \phi.\ \phi \ \in \ \Delta \ \Rightarrow \ \mathcal{L}^1_\tau\ \phi) \ \wedge \\ \quad (\forall \Delta_0. \\ \qquad \mathsf{FINITE}\ \Delta_0 \ \wedge \ \Delta_0 \ \subseteq \ \Delta \ \Rightarrow \\ \qquad\quad \exists\, \mathfrak{M}\ \sigma.\ \mathsf{valuation}\ \mathfrak{M}\ \sigma \ \wedge \ \forall \phi.\ \phi \ \in \ \Delta_0 \ \Rightarrow \ \mathfrak{M}, \sigma \vDash \ \phi) \ \Rightarrow \\ \qquad \exists\, \mathfrak{M}\ \sigma.\ \mathsf{valuation}\ \mathfrak{M}\ \sigma \ \wedge \ \forall \phi.\ \phi \ \in \ \Delta \ \Rightarrow \ \mathfrak{M}, \sigma \vDash \ \phi \end{array}$$

The assumption on infiniteness of the type universe comes from similar reason as that of 2.2. Because of this assumption, every statement which requires compactness theorem will be required to include the assumption on the infiniteness of type universe.

As a consequence of the compactness theorem, we have:

Corollary 6.20 (`compactness_corollary_L1tau`). *Under the assumption that the type universe of α is infinite and Δ is a set of $\mathcal{L}^1_\tau$ formula. If for every α-model $\mathfrak{M}$ and valuation σ, once we have $\mathfrak{M}, \sigma \vDash \ \phi$ for every $\phi \ \in \ \Delta$, then $\mathfrak{M}, \sigma \vDash \ \delta$, then there exists a finite subset Δ_0 of Δ such that once every formula in Δ_0 is satisfied in an α-model $\mathfrak{M}$ under a valuation σ, then $\mathfrak{M}, \sigma \vDash \ \delta$.*

$$\begin{array}{l} \vdash \mathsf{INFINITE}\, \mathcal{U}(:\alpha) \ \wedge \ \mathcal{L}^1_\tau\ \delta \ \wedge \ (\forall \phi.\ \phi \ \in \ \Delta \ \Rightarrow \ \mathcal{L}^1_\tau\ \phi) \ \wedge \\ \quad (\forall\, \mathfrak{M}\ \sigma.\ \mathsf{valuation}\ \mathfrak{M}\ \sigma \ \Rightarrow \ (\forall \phi.\ \phi \ \in \ \Delta \ \Rightarrow \ \mathfrak{M}, \sigma \vDash \ \phi) \ \Rightarrow \ \mathfrak{M}, \sigma \vDash \ \delta) \ \Rightarrow \\ \quad \exists\, \Delta_0. \\ \qquad \mathsf{FINITE}\ \Delta_0 \ \wedge \ \Delta_0 \ \subseteq \ \Delta \ \wedge \\ \qquad \forall\, \mathfrak{M}\ \sigma.\ \mathsf{valuation}\ \mathfrak{M}\ \sigma \ \Rightarrow \ (\forall \phi.\ \phi \ \in \ \Delta_0 \ \Rightarrow \ \mathfrak{M}, \sigma \vDash \ \phi) \ \Rightarrow \ \mathfrak{M}, \sigma \vDash \ \delta \end{array}$$

Proof. Under the assumptions, suppose, in order to get a contradiction, that for every finite subset Δ_0 of Δ, there exists an α-model $\mathfrak{M}$ and a valuation σ where all the formulas in Δ_0 are satisfied by δ is not satisfied, then every finite subset of $\Delta_0 \ \cup \ \{ \ {}^{\mathsf{f}}\neg\ \delta \ \}$ is satisfied on some α-model $\mathfrak{M}$ under some valuation σ. As δ is an $\mathcal{L}^1_\tau$-formula, so does ${}^{\mathsf{f}}\neg\ \delta$. By Theorem 6.19, this implies the whole set $\Delta_0 \ \cup \ \{ \ {}^{\mathsf{f}}\neg\ \delta \ \}$ is satisfied on some α-model under some valuation, contradicting our assumption. □

Now we have all the ingredient for translating the hard direction of the standard proof of *Van Benthem Characterization Theorem* into HOL.

Theorem 6.21. [1, Theorem 2.68 (Van Benthem Characterization Theorem), hard direction] *For an infinite type α, if δ is a first-order formula which is invariant for bisimulation on* ***`num`*** $\rightarrow$ α $\rightarrow$ ***`bool`****-first-order models and the* *only free variable may appear in δ is x, then there exists a modal formula whose standard translation at x is equivalent to δ on α-first-order models.*

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

Proof. Under the given assumptions, consider the modal consequence of δ, which is the set of standard translations implied by δ on all first-order models with α-sets as their domains, defined in HOL as

$$MOC\ =\ \{\,\mathsf{ST}_x\ \phi \mid \phi \mid \forall\,\mathfrak{M}\ v.\ \mathsf{valuation}\ \mathfrak{M}\ v\ \Rightarrow\ \mathfrak{M}, v \vDash\ \delta\ \Rightarrow\ \mathfrak{M}, v \vDash\ \mathsf{ST}_x\ \phi\,\}$$

Our first claim is that it suffices to prove δ is implies by MOC. To see why it suffices, assume it is true, then by Corollary 6.20, there exists a finite subset of Σ_0 of MOC such that once all the formulas in Σ_0 are satisfied, then δ is satisfied. Also by definition of MOC, once δ is satisfied, every formula in Σ_0 is satisfied. Hence δ will be equivalent to the big conjunction of formulas in Σ_0, which is a standard translation.

Fix a model $\mathfrak{M}$ and suppose $\mathfrak{M}, \sigma \vDash\ \varphi$ for every $\varphi\ \in\ MOC$, we prove $\mathfrak{M}, \sigma\ \vDash\ \ \delta$. Consider of the set Σ of formulas of the form $\mathsf{ST}_x\ \phi$ such that $\mathfrak{M}, \sigma \vDash\ \mathsf{ST}_x\ \phi$. Pick a model $\mathfrak{N}$ and a valuation $\sigma_{\mathfrak{N}}$ satisfying each formula in $\Sigma\ \cup\ \{\,\delta\,\}$. Such a model does exist: Suppose, in order to get a contradiction, that such a model does not exist, then for every model, once all the formulas in Σ are satisfied, the formula δ will not be satisfied. Then by Corollary 6.20, there exists a finite subset of Σ implies ${}^{\mathsf{f}}\neg\ \delta$. Taking its contrapositive, then a implies the negation of the big conjunction ψ of finitely many elements in Σ. As a negated big conjunction of standard translations is again a standard translation, we have ${}^{\mathsf{f}}\neg\ \psi\ \in\ MOC$. Recall we have assumed $\mathfrak{M}, \sigma \vDash\ \varphi$ for every $\varphi\ \in\ MOC$, so $\mathfrak{M}, \sigma \vDash\ {}^{\mathsf{f}}\neg\ \psi$, but also $\mathfrak{M}, \sigma \vDash\ \psi$ by definition of Σ. This is a contradiction.

Now let w denote $\sigma\ x$ and v denote $\sigma_{\mathfrak{N}}\ x$, we claim that if we regard both $\mathfrak{M}$ and $\mathfrak{N}$ as modal models, then w and v are modal equivalent. To prove this, suppose $\mathsf{folm2mm}\ \mathfrak{M}, w\ \Vdash\ \ \phi$ for a modal formula ϕ, then $\mathsf{ST}_x\ \phi\ \ \in\ \ \Sigma$ by Proposition 4.3, Proposition 4.1 and the definition of Σ, hence $\mathfrak{N}, \sigma_{\mathfrak{N}} \vDash\ \mathsf{ST}_x\ \phi$. By these two propositions again, we can prove $\mathsf{folm2mm}\ \mathfrak{N}, v\ \Vdash\ \ \phi$. This proves $\forall\,\phi.\ \mathsf{folm2mm}\ \mathfrak{M}, w\ \Vdash\ \ \phi\ \Rightarrow\ \mathsf{folm2mm}\ \mathfrak{N}, v\ \Vdash\ \ \phi$. Conversely, if $\mathsf{folm2mm}\ \mathfrak{M}, w\ \nVdash\ \phi$, then $\mathsf{folm2mm}\ \mathfrak{M}, w\ \Vdash\ \ \neg\phi$ and we can deduce $\mathsf{folm2mm}\ \mathfrak{N}, v\ \nVdash\ \ \phi$ by a symmetric argument.

If modal equivalence implies bisimularity, then we are done: Suppose modal equivalence implies bisimularity, then as $w \in (\mathsf{folm2mm}\ \mathfrak{M})^W$ and $v \in (\mathsf{folm2mm}\ \mathfrak{N})^W$ are modal equivalent, there exists a bisimulation between them. As δ is invariant for bisimulation and is satisfied at v, then it is also satisfied at w.

Although it is not always the case that modal equivalence implies bisimularity, we can take a detour with the help of Theorem 6.18. By 6.18, we obtain an ultrafilter U on a set J such that for the ultraproduct models $\mathfrak{M}_* = \Pi_U\ (\lambda\, j.\ \mathsf{folm2mm}\ \mathfrak{M})$ and $\mathfrak{N}_* = \Pi_U\ (\lambda\, j.\ \mathsf{folm2mm}\ \mathfrak{N})$, the worlds $w_* = \{\, f \mid (\lambda\, j.\ w) \sim_U^{\mathsf{worlds}\ (\lambda\, j.\ \mathsf{folm2mm}\ \mathfrak{M})} f\,\}$ and $v_* = \{\, g \mid (\lambda\, j.\ v) \sim_U^{\mathsf{worlds}\ (\lambda\, j.\ \mathsf{folm2mm}\ \mathfrak{N})} g\,\}$ are bisimilar. As δ is invariant for bisimulation, δ holds at w_* in $\mathsf{mm2folm}\ \mathfrak{M}_*$ iff it holds at v_* in $\mathsf{mm2folm}\ \mathfrak{N}_*$. We are going to carry the δ from the model $\mathfrak{N}$ where it is satisfied at v, to the point v_* in $\mathsf{mm2folm}\ \mathfrak{N}_*$, then to the point w_* in $\mathsf{mm2folm}\ \mathfrak{M}_*$, and finally to w in $\mathfrak{M}$.

To carry δ around, it suffices to prove $\mathfrak{M}, \sigma \vDash\ \delta \iff \mathsf{mm2folm}\ \mathfrak{M}_*, (\lambda\, x.\ w_*) \vDash\ \delta$ and $\mathfrak{N}, \sigma_{\mathfrak{N}} \vDash\ \delta \iff \mathsf{mm2folm}\ \mathfrak{N}_*, (\lambda\, x.\ v_*) \vDash\ \delta$ under our assumptions by hand. These two equivalence are of the same pattern, hence we prove it as a lemma:

$$
\begin{aligned}
&\vdash \mathcal{L}^1_\tau\ \delta\ \land\ \mathsf{FV}\ \delta\ \subseteq\ \{\, x\,\}\ \land\ \mathsf{ultrafilter}\ U\ J\ \land\ \mathsf{valuation}\ \mathfrak{M}\ \sigma\ \Rightarrow\\
&\quad (\mathfrak{M}, \sigma \vDash\ \delta\ \iff\\
&\qquad \mathsf{mm2folm}\ (\Pi_U\ (\lambda\, j.\ \mathsf{folm2mm}\ \mathfrak{M})), (\lambda\, x.\\
&\qquad\quad \{\, f \mid (\lambda\, j.\ \sigma\ x) \sim_U^{\mathsf{worlds}\ (\lambda\, j.\ \mathsf{folm2mm}\ \mathfrak{M})} f\,\}\,) \vDash\\
&\qquad \delta)
\end{aligned}
$$

The lemma holds by Proposition 6.14, 6.15 and 4.1. Hence we are done.

□

Now we have formalized both directions of the Van Benthem Characterization theorem. A reader may expect we can put them together to get a double implication. However, as we have already mentioned, we cannot get an 'if an only if' result. To see the reason: given an $\mathcal{L}^1_\tau$-formula ϕ with no more then one free variable, by the result we have just proved, 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 $(\texttt{num} \rightarrow \alpha) \rightarrow \texttt{bool}$-world. But according to Proposition 4.5, we can only conclude ϕ is invariant for bisimulation for models of type α. If the type universe of $(\texttt{num} \rightarrow \alpha) \rightarrow \texttt{bool}$ is small enough to be embedded into α, then we will also done. However, the cardinality of the universe of $(\texttt{num} \rightarrow \alpha) \rightarrow \texttt{bool}$ is larger than that of α, and hence we cannot derive ϕ is invariant for bisimulation for models with $(\texttt{num} \rightarrow \alpha) \rightarrow \texttt{bool}$-worlds from the fact that ϕ is invariant for bisimulation for models with α-worlds.

We get into this situation because the statements we have proved for both directions are not precise translations of their set-theoretic statements. Consider the easy direction: its set-theoretic statement is that if ϕ is equivalent to a standard translation on models of every type, then it will be invariant for bisimulation on models of every type, whereas in our statement 'if ϕ is equivalent to a standard translation on models of type α, then it is invariant for models of type α'. Both the assumption and the conclusion are weakened. We cannot encode the original statement in HOL, since we cannot quantify over types and refer to all the types to state 'invariant for bisimulation for models of all types' and 'equivalent to a standard translation on models of all types', just as the problem we encountered when defining equivalence of modal formulas. If we could quantify over types (as we could in a theorem prover based on dependent type theory), then we could prove 'ϕ is invariant for bisimulation on models of every type if and only if ϕ is equivalent to a standard translation on models of every type' using the same proof we have written out. For the easy direction, the assumption is that ϕ is equivalent to a standard translation on models of every type, and we want to conclude that ϕ is invariant for bisimulation for models of type α. But under assumption, the formula ϕ is equivalent to a standard translation on models of type α where α is an arbitrary type, so we prove the result by Proposition 4.5. Conversely, for the other direction, the assumption is that 'ϕ is invariant for bisimulation on models of every type', and the goal is to prove ϕ is invariant on models of type α where α is an arbitrary type. By assumption, the formula ϕ is invariant for bisimulation on models of type $(\texttt{num} \rightarrow \alpha) \rightarrow \texttt{bool}$, and the result follows from Theorem 6.21.

6.2 Positive existential formulas and preservation under simulations

There exists a concept of 'half of a bisimulation', which is called *simulation.* In this section, we are interested in the $\mathcal{L}^1_\tau$-formulas which are *preserved under simulation.* We have a set-theoretic proof that these formulas can also be characterized using their syntax. This section aims to translate this characterization into simple type theory. For precisely the same reason as in the last section, after we translate the proof of implications in both directions, they cannot be unified into a double implication. Nevertheless, we will spell out our formalization of proofs for those two directions separately.

As we expect, the clauses defining simulation is 'half of' the clauses for defining a bisimulation:

Definition 6.8. [1, Definition 2.77 (Simulations)] *A simulation Z between two models $\mathfrak{M}_1$ and $\mathfrak{M}_2$ (notation: $\mathfrak{M}_1 \overset{Z}{\underline{\to}} \mathfrak{M}_2$) is a relation between their worlds such that for every $w_1 \in \mathfrak{M}_1^W$ and $w_2 \in \mathfrak{M}_2^W$, if Z relates w_1 and w_2, then we have:*

- *For each propositional letter which is satisfied at w_1, it is also satisfied at w_2.*
- *If there is a world v_1 in $\mathfrak{M}_1$ such that $\mathfrak{M}_1^R\ w_1\ v_1$, then there exists a world v_2 in $\mathfrak{M}_2$ such that $\mathfrak{M}_2^R\ w_2\ v_2$, and moreover, v_1 and v_2 are related by Z.*

$$\begin{aligned}
&\mathfrak{M}_1 \overset{Z}{\underline{\to}} \mathfrak{M}_2 \overset{\mathsf{def}}{=} \\
&\quad \forall\, w_1\ w_2. \\
&\qquad w_1 \in \mathfrak{M}_1^W \wedge w_2 \in \mathfrak{M}_2^W \wedge Z\ w_1\ w_2 \Rightarrow \\
&\qquad (\forall\, p.\ w_1 \in \mathfrak{M}_1^V\ p \Rightarrow w_2 \in \mathfrak{M}_2^V\ p) \wedge \\
&\qquad \forall\, v.\ v \in \mathfrak{M}_1^W \wedge \mathfrak{M}_1^R\ w_1\ v \Rightarrow \exists\, v'.\ v' \in \mathfrak{M}_2^W \wedge Z\ v\ v' \wedge \mathfrak{M}_2^R\ w_2\ v'
\end{aligned}$$

The concept which corresponds to 'invariant for bisimulation' is 'preserved under simulation'. In contrast to that of 'invariant for bisimulation', the concept 'preserved under simulation' is about modal formula.

Definition 6.9. [1, Definition 2.77 (Preserved Under Simulations)] *A modal formula ϕ is preserved under simulation if once we have $w_1 \in \mathfrak{M}_1^W$ and $w_2 \in \mathfrak{M}_2^W$ with a simulation relating w_1 to w_2, then if ϕ is satisfied at w_1, it is also satisfied* *at* w_2.

$$\begin{array}{l}\mathsf{preserved\_under\_sim}\ (:\alpha)\ (:\beta)\ \phi \overset{\mathsf{def}}{=} \\ \quad \forall \mathfrak{M}_1\ \mathfrak{M}_2\ Z\ w_1\ w_2. \\ \qquad w_1 \in \mathfrak{M}_1^W \wedge w_2 \in \mathfrak{M}_2^W \wedge \mathfrak{M}_1 \overset{Z}{\twoheadrightarrow} \mathfrak{M}_2 \wedge Z\ w_1\ w_2 \wedge \mathfrak{M}_1, w_1 \Vdash \phi \Rightarrow \\ \qquad \mathfrak{M}_2, w_2 \Vdash \phi\end{array}$$

The predicate preserved_under_sim takes type parameters by the same reason as discussed when we define invar4bisim.

The rest of the section aims to translate the proof that characterizes formulas preserved under bisimulation as *positively existential formulas.* A positive existential formula is a modal formula which does not contain 'negative' connectives. Such a formula is built up from $\top$, $\bot$ and propositional letters using only the connectives '$\wedge$','$\vee$' or '$\Diamond$':

Definition 6.10. [1, Page 111 (Positive Existential)] *The rules of positive existential formulas read:*

- *The formulas '$\bot$' and '$\top$' are positive existential.*
- *A propositional letter standing alone is positive existential.*
- *If ϕ_1 and ϕ_2 are both positive existential, then both their conjunction and their disjunction are positive existential.*
- *Adding a diamond before a positive existential formula gives a positive existential formula.*

$$\frac{}{\mathsf{PE}\ \bot} \qquad \frac{}{\mathsf{PE}\ \top} \qquad \frac{}{\mathsf{PE}\ (\mathsf{VAR}\ p)} \qquad \frac{\mathsf{PE}\ \phi_1 \quad \mathsf{PE}\ \phi_2}{\mathsf{PE}\ (\phi_1 \wedge \phi_2)} \qquad \frac{\mathsf{PE}\ \phi_1 \quad \mathsf{PE}\ \phi_2}{\mathsf{PE}\ (\phi_1 \vee \phi_2)} \qquad \frac{\mathsf{PE}\ \phi}{\mathsf{PE}\ (\Diamond\phi)}$$

By induction, every big conjunction or disjunction of positive existential formulas is again a positive existential formula. We can immediately prove by induction on positive existential formulas that every positive existential formula is preserved under simulation, but the converse only holds for 'good models'. In Chapter 5, we introduced the concept of M-saturated models, and we have already seen that they are 'good' models, which gives equivalence between modal equivalence and bisimulation. It turns out that M-saturated models do not only give rise to nice properties about bisimulations, but also work well for simulations.

Proposition 6.22. [1, Exercise 2.7.1] *Suppose* $w_1 \in \mathfrak{M}_1^W$ *and* $w_2 \in \mathfrak{M}_2^W$ *and the models* $\mathfrak{M}_1$ *and* $\mathfrak{M}_2$ *are both M-saturated. If for every positive existential* *formula ϕ, the satisfaction of ϕ at w_1 implies the satisfaction of ϕ at w_2, then there exists a simulation relation between $\mathfrak{M}_1$ and $\mathfrak{M}_2$ which relates w_1 to w_2.*

$$\begin{aligned}&\vdash \mathsf{M\_sat}\,\mathfrak{M}_1 \ \wedge\ \mathsf{M\_sat}\,\mathfrak{M}_2 \ \wedge\ w_1 \ \in\ \mathfrak{M}_1^W \ \wedge\ w_2 \ \in\ \mathfrak{M}_2^W \ \wedge \\ &\quad (\forall\,\phi.\ \mathsf{PE}\,\phi \ \Rightarrow\ \mathfrak{M}_1, w_1 \Vdash\ \phi \ \Rightarrow\ \mathfrak{M}_2, w_2 \Vdash\ \phi) \ \Rightarrow \\ &\quad\ \exists\,Z.\ \mathfrak{M}_1 \xrightarrow{Z} \mathfrak{M}_2 \ \wedge\ Z\ w_1\ w_2\end{aligned}$$

Proof. Under the assumptions, the relation Z defined as $Z\ v_1\ v_2$ iff $\forall\,\phi.\ \mathsf{PE}\ \phi\ \wedge\ \mathfrak{M}_1, v_1 \Vdash\ \phi\ \Rightarrow\ \mathfrak{M}_2, v_2 \Vdash\ \phi$ is a simulation. Checking it is indeed a simulation is completely analogous to the proof of Proposition 5.1. □

As the last theorem that is proved in the project, we translate the proof of the theorem that says modal formulas which are preserved under simulations are exactly the ones which are equivalent to a positive existential formula into HOL. This proof will use a similar idea as the characterization theorem proved in the last section. But this time, we only need the modal version of compactness theorem and its corollary.

Theorem 6.23 (Compactness of modal logic)**.** *If α is an infinite type, then given a set Δ of **num**-modal formulas, if for every finite subset $\Delta_0\ \subseteq\ \Delta$, there exists a modal model $\mathfrak{M}$ with α-world set and a world $w\ \in\ \mathfrak{M}^W$ such that $\mathfrak{M}, w \Vdash\ \phi$ for every $\phi\ \in\ \Delta_0$, then there exists a model with α-world set and a world in $\mathfrak{M}$ which satisfies all the modal formulas in Δ.*

$$\begin{aligned}&\vdash \mathsf{INFINITE}\,\mathcal{U}(:\alpha)\ \wedge \\ &\quad (\forall\,\Delta_0. \\ &\qquad \mathsf{FINITE}\ \Delta_0\ \wedge\ \Delta_0\ \subseteq\ \Delta\ \Rightarrow \\ &\qquad\quad \exists\,\mathfrak{M}\ w.\ w\ \in\ \mathfrak{M}^W\ \wedge\ \forall\,\phi.\ \phi\ \in\ \Delta_0\ \Rightarrow\ \mathfrak{M}, w \Vdash\ \phi)\ \Rightarrow \\ &\quad\ \exists\,\mathfrak{M}\ w.\ w\ \in\ \mathfrak{M}^W\ \wedge\ \forall\,\phi.\ \phi\ \in\ \Delta\ \Rightarrow\ \mathfrak{M}, w \Vdash\ \phi\end{aligned}$$

Proof. By Proposition 4.3 and Theorem 6.19. □

Corollary 6.24 (`modal_compactness_corollary`)**.** *For α is an infinite type, given a modal formula δ and a set Δ of **num**-modal formulas, if for every modal model $\mathfrak{M}$ with α-world set, every world w which satisfies all the formulas in Δ will also satisfy δ, then there exists a finite subset $\Delta_0\ \subseteq\ \Delta$ such that for a world w in a model $\mathfrak{M}$ with α-world set, if every formula in Δ_0 is satisfied at w, then* *a is satisfied at w.*

$$\begin{array}{l}\vdash \mathsf{INFINITE}\, \mathcal{U}(:\alpha) \;\wedge \\ \quad (\forall\, \mathfrak{M}\; w.\; w \;\in\; \mathfrak{M}^W \;\Rightarrow\; (\forall\, \phi.\; \phi \;\in\; \varDelta \;\Rightarrow\; \mathfrak{M}, w \Vdash \phi) \;\Rightarrow\; \mathfrak{M}, w \Vdash \delta) \;\Rightarrow \\ \quad\; \exists\, \varDelta_0. \\ \qquad \mathsf{FINITE}\; \varDelta_0 \;\wedge\; \varDelta_0 \;\subseteq\; \varDelta \;\wedge \\ \qquad \forall\, \mathfrak{M}\; w.\; w \;\in\; \mathfrak{M}^W \;\Rightarrow\; (\forall\, \phi.\; \phi \;\in\; \varDelta_0 \;\Rightarrow\; \mathfrak{M}, w \Vdash \phi) \;\Rightarrow\; \mathfrak{M}, w \Vdash \delta\end{array}$$

Proof. Similar to the proof of Corollary 6.20. □

All the modal formulas appearing in the above theorems are required to be `num`-formulas. That is because we need to appeal to standard translation to prove them, and the standard translation is only defined on `num`-modal formulas. Also, we require the assumption on infiniteness of the type universe since we use first-order compactness theorems to prove the above two theorems.

For the same reason that we did it for Theorem 6.21, we only consider simulations between models of the same type here.

Theorem 6.25. [1, Theorem 2.78, hard direction] *Let β be an infinite type. For each* ***num****-modal formula ϕ, if ϕ is preserved under simulation on (****num****, (β → ****bool****) → ****bool****)-models, then there exists a positive existential* ***num****-modal formula which is equivalent to ϕ on (****num****, β)-models.*

$$\begin{array}{l}\vdash \mathsf{INFINITE}\, \mathcal{U}(:\beta) \;\wedge \\ \quad \mathsf{preserved\_under\_sim}\; (:(\beta \;\rightarrow\; \mathit{bool}) \;\rightarrow\; \mathit{bool})\; (:(\beta \;\rightarrow\; \mathit{bool}) \;\rightarrow\; \mathit{bool}) \\ \qquad (\phi : \mathit{num\; form}) \;\Rightarrow \\ \qquad \exists\, (\varphi : \mathit{num\; form}).\; \phi \equiv_{(:\beta)} \varphi \;\wedge\; \mathsf{PE}\; \varphi\end{array}$$

Proof. Suppose ϕ is preserved under simulation for models of (β → `bool`) → `bool`-worlds. Consider the set *PEC* of positive existential formulas φ such that for every (`num`, β)-model $\mathfrak{M}$ and every world $w \in \mathfrak{M}^W$, if all the formulas in *PEC* are satisfied at w, then φ is satisfied at w. In HOL, the set *PEC* is defined as:

$$PEC \;=\; \{\, \varphi \mid \mathsf{PE}\, \varphi \;\wedge\; \forall\, \mathfrak{M}\; w.\; w \;\in\; \mathfrak{M}^W \;\wedge\; \mathfrak{M}, w \Vdash \phi \;\Rightarrow\; \mathfrak{M}, w \Vdash \varphi \,\}$$

By Corollary 6.24, if we can prove for every (`num`, β)-model $\mathfrak{M}$ and $w \in \mathfrak{M}^W$, $\mathfrak{M}, w \Vdash \varphi$ for all $\varphi \in PEC$ implies $\mathfrak{M}, w \Vdash \phi$, then there exists a finite subset S of *PEC* that entails ϕ. This will prove ϕ is equivalent to the conjunction of all the formulas in S, which is again a positive existential formula.

Therefore, our task is to prove the entailment from PEC to ϕ. Suppose $\mathfrak{M}, w \Vdash \varphi$ for all $\varphi \in PEC$, we prove $\mathfrak{M}, w \Vdash \phi$. Define $\Gamma = \{ \neg\psi \mid \mathsf{PE}\ \psi \wedge \mathfrak{M}, w \Vdash \neg\psi \}$. We claim that there exists a $(\mathtt{num}, \beta)$-model with a world that satisfies the set $\Gamma \cup \{ \phi \}$. By Theorem 6.23, it suffices to prove each finite subset of $\Gamma \cup \{ \phi \}$ is satisfied by some model. Suppose there exists a finite subset of $\Gamma \cup \{ \phi \}$ which can be satisfied by no model, then there exists $\neg\psi_0, \cdots, \neg\psi_n \in \Gamma$ such that for every $(\mathtt{num}, \beta)$-model $\mathfrak{N}$ and every world v of it, if $\mathfrak{N}, v \Vdash \phi$, then there exists some $0 \leq i \leq n$ such that $\mathfrak{N}, v \Vdash \psi_i$. As all these ψ's are positive existential, so does their big disjunction ψ, and hence $\psi \in PEC$. As $\mathfrak{M}$ entails PEC, we have $\mathfrak{M}, w \Vdash \psi$, and hence $\mathfrak{M}, w \Vdash \psi_i$ for some i by definition of the ψ. But on other hand, $\mathfrak{M}, w \Vdash \neg\psi_i$ for every ψ_i by definition of Γ. This is a contradiction.

Hence we obtain a model $\mathfrak{N}$ such that every element in $\Gamma \cup \{ \phi \}$ is satisfied at a point $v \in \mathfrak{N}^W$. For every positive existential formula ψ such that $\mathfrak{M}, w \nVdash \psi$, we have $\neg\psi \in \Gamma$, so $\mathfrak{N}, v \Vdash \neg\psi$. Hence for every positive existential ψ, if $\mathfrak{N}, v \Vdash \psi$, then $\mathfrak{M}, w \Vdash \psi$. Consider the ultrafilter extensions ${}^{ue}\mathfrak{M}$ and ${}^{ue}\mathfrak{N}$, we claim the worlds ${}^{ue}\mathfrak{N}, \pi_v^{\mathfrak{N}^W} \Vdash \psi$ is related to the world ${}^{ue}\mathfrak{M}, \pi_w^{\mathfrak{M}^W} \Vdash \psi$ by a simulation. By Proposition 6.22, it suffices to prove that every positive existential formula which is satisfied at ${}^{ue}\mathfrak{N}, \pi_v^{\mathfrak{N}^W} \Vdash \psi$ is also satisfied at ${}^{ue}\mathfrak{M}, \pi_w^{\mathfrak{M}^W} \Vdash \psi$. Consider a positive existential ψ, ${}^{ue}\mathfrak{N}, \pi_v^{\mathfrak{N}^W} \Vdash \psi$ implies $\mathfrak{N}, v \Vdash \psi$, by the discussion above, it implies $\mathfrak{M}, w \Vdash \phi$, and hence implies ${}^{ue}\mathfrak{M}, \pi_w^{\mathfrak{M}^W} \Vdash \psi$ by Proposition 5.7 again.

As $\mathfrak{N}, v \Vdash \phi$, Proposition 5.7 gives ${}^{ue}\mathfrak{N}, \pi_v^{\mathfrak{N}^W} \Vdash \phi$, as ϕ is preserved under simulation, we have ${}^{ue}\mathfrak{M}, \pi_w^{\mathfrak{M}^W} \Vdash \phi$. Again by Proposition 5.7 , it implies $\mathfrak{M}, w \Vdash \phi$. This completes the proof. □

Bibliography

[1] Patrick Blackburn, Maarten de Rijke, and Yde Venema. *Modal Logic*. Cambridge University Press, 2001.

[2] C. C. Chang and H. Jerome Keisler. *Model Theory*. North Holland, 1990.

[3] HOL Developers. *HOL Manual: Logic*. Available from `http://hol-theorem-prover.org`.

[4] John Harrison. Formalizing basic first order model theory. In *Theorem Proving in Higher Order Logics, 11th Internatinal Conference*, Lecture Notes in Computer Science, pages 153–170. Springer, 1998.

[5] Thomas Jech. *Set Theory*. Springer, 2006.

[6] Konrad Slind and Michael Norrish. A brief overview of HOL4. In *Theorem Proving in Higher Order Logics, 21st International Conference*, Lecture Notes in Computer Science, pages 28–32. Springer, 2008.