Some Intuitionistic Provability and Preservativity Logics [1]
(and their interrelations)

Chunlai Zhou [2]

September 30, 2003

[1]Advisors: Prof. dr. Dick de Jongh and Dr. Rosalie Iemhoff. First Version: July 30, 2003. Defended: August 19, 2003

[2]E-mail: czhou@indiana.edu

Contents

Acknowledgements

I have spent a fruitful and wonderful year in Amsterdam. Before I came here, I admired Amsterdam as my holy place in logic. Now when I am leaving, I admire and love Amsterdam more deeply because of the logics and the people that I have met in the year at ILLC.

First I want to express my greatest gratitude to my supervisor Prof. Dick de Jongh. He always tries to do everything to help me realize my Holland dream in logic not only intellectually but also socially. His teaching has greatly improved my understanding of logic especially of modal logic. Our beloved logical systems with diamonds and boxes are not artificial but just living organisms like us human beings. Absolutely his logic is a kind of magic to me. I am fascinated by the techniques in many of his theorems especially in his beautiful fixed point theorem. I clearly remember that one day I was so grasped by his proof of the fixed point theorem that I worked in front of one blackboard in the Euclides Building the whole day totally ignoring everything around me. After every weekly meeting with him, I felt intellectually stimulated by his ideas and had an irresistible impulse to implement them and even to generalize them. And then I kept toiling on them for the next whole week, which made me taste how great it would be to be a logician. Moreover, he has introduced me to many young Dutch logicians, and thinks carefully about the research environment for my future career.

I am more than grateful to Prof. van Benthem for his financial support of my study in Amsterdam through his famous Spinoza project "Logic in Action". Two years ago when I determined to do logic, I had no other expectations than going to the Netherlands to study from the masters in logic. His generous donation made my trip to Amsterdam. And I will never forget his email on the New Year:"Congratulations for your full score on the course Logic and Games,", which is the biggest evaluation that I have achieved so far. Nothing can be compared to such a kind encouragement from such a great logician to a young enterprizing logic student. I will cherish this forever!

Also I want to thank my co-supervisor Rosalie. She offered me such a wonderful thesis topic. And she gave me many precious suggestions to attack the problems in my thesis and to improve my results. Besides, she taught me many crystallized intuitions behind such a difficult subject as preservativity logic. In fact, in my thesis writing, I has always drawn inspirations from and tried to follow the examples that she has set for preservativity logic in her Ph. D thesis.

I have benefitted a lot from Yde's strict policy on presentations of proofs in modal logic. At the beginning of the modal logic course, he often wrote on the top of the first pages of my homework: "Chunlai: You MUST improve your presentation". Every time I got my homework back, I felt a little disappointed with the scores that he gave me because they were lower than I had expected. Sometimes, I complained that he saw only the trees instead of the woods. But, after I realized that a transparent presentation is one essential part of logic, I felt very lucky: what an invaluable training in logic I have received from Yde.

It was Joost who introduced me to provability logic. I have learned a lot from his devoted teaching on the course Provability and Interpretability Logic in Utrecht University. He is always able to make the most elusive stuff very easy to understand.

In addition, I would like to thank Spencer. We share not only the apartment 221 B[1] in Prinsegracht but also many interests in logic and mathematics. We live harmoniously the whole year. Whenever I have financial difficulty, he kindly helps me. Another thing that I will never forget is that, when he learned my admiration of Mac Lane, he took a lot of efforts to get his signature from Chicago for me.

Finally, I would like to mention many people that are very important to me in the past year. Two of my colleagues Wouter and Andreas are always available for technical helps. Nick and Clemens have answered me many questions in modal logic. My other Master of Logic classmates: Be, Dana, Tanja, Jill, Julia, Linh, Dirk, Oren, Guillaume, Giosue, Clive, Gilad, are also important all along my study here. My Amsterdam Chinese friends made my life quite enjoyable: Guo Meilin, Feng Jin, Yang Yongfang, Lin Guoxin, Chen Yebin, Zhang Mingzhi and Zhang Shibiao, but most of all Zhang Yuanpeng.

Thank you all!!

ILLC Amsterdam/August 18th

Chunlai ZHOU

[1] I was told that this number is very prestigious in some well-known western detective stories. It seems that everything in Amsterdam even my apartment number here is so nice to me.

Chapter 1

Introduction

1.1 Preservativity Logic and Provability Logic of HA

This thesis is about the preservativity logic and provability logic of Heyting arithmetic (or HA). It will be inspiring to us to look at its classical analogue GL first. The theorems of GL contains all tautologies of the classical propositional logic, contains all distribution axioms, i.e, all sentences of the form $\Box(A \to B) \to (\Box A \to \Box B)$, contains Löb's axiom, i.e. all sentences of the form $\Box(\Box A \to A) \to \Box A$, and is closed under modus ponens, substitution and necessitation. It is well-known that GL is modally complete with respect to all finite, transitive and conversely well-founded frames.

There are three important theorems in the provability logic of PA: Löb's theorem, Solovay's theorem and the fixed point theorem. The landmark paper by Solovay[1976] can be seen as the starting point of provability logic. But, before we can state these theorems, we need to define some basic notions in provability logic. As we know, one of Godël's discoveries is that we can reason about the meta-theories *in* their object theories. That is to say, we can arithmetize the informal mathematical theory or, as Boolos calls it, logicians's English of syntax of PA in itself. In particular, PA allows a formalization "being provable in PA". One can define $Proof(x, y)$ as the statement that x is the code of a finite sequence of proof in PA of a formula B whose code is y. $Prf(x)$ denotes $\exists y Proof(y, x)$. A *realization* is a function that assigns each propositional letter an arithmetical sentence such that

$p^* := *(p)$;
$\bot^* := 0 \equiv 1$;
For any $\circ \in \{\vee, \wedge, \to\}, (A \circ B)^* := A^* \circ B^*$;
$(\Box B)^* := Prf(\ulcorner B^* \urcorner)$.

The *provability logic* of PA can be defined as the logic of all the modal formulas G that are valid under any realization, or, $PA \vdash G^*$ for any realization $*$. In short, the provability logic is the logic about the proof predicate Prf. By Löb's theorem, it is easy to see that $GL \subseteq \{G | PA \vdash G^* \text{ for any realization}\}$. Moreover, the first Solovay arithmetical completeness theorem tells us that the converse also holds. In other words, if A^* is provable in PA for any realization $*$, then $GL \vdash A$.

Compared to the maturity and stability (and hence beauty) of axiomatization and arithmetical completeness of GL, the provability logic of HA, the intuitionistic counterpart of PA, is still a small child. Even today, we still don't know what is the provability logic of HA. But there are many insights and conjectures about this logic in Iemhoff [2001] (see also Visser[2002]). Just as we have done it for classical case, we will define a similar notion for HA. φ is said to *preserve* ψ with respect to HA, if, for all Σ_1 sentences θ, it holds that $HA \vdash (\theta \to \varphi)$ entails $HA \vdash (\theta \to \psi)$. We denote this as

$\varphi \rhd \psi$. Correspondingly if x and y are the codes of formulas A and B, respectively, $Pres(x, y)$ is the formalization of that 'A preserves B'. Similarly we define a realization as a function from propositional letters to arithmetic sentences such that

$p^* := *(p)$;
$\bot^* := 0 \equiv 1$;
For any $\circ \in \{\vee, \wedge, \rightarrow\}, (A \circ B)^* := A^* \circ B^*$;
$(A \rhd B)^* := Pres(\ulcorner A^* \urcorner, \ulcorner B^* \urcorner)$.

The *preservativity logic* of HA is the logic of all the $L_\rhd$ formulas that are valid in HA under any realization. As we will see, it includes the provability logic. Although we still don't know what is exactly the preservativity logic, Iemhoff[2001] has a well-justified conjecture: iPH (we will come up with it below) is the preservativity logic of HA. In other words, $HA \vdash \varphi^*$ for any realization $*$ iff $iPH \vdash \varphi$. One direction or the soundness of iPH has been proved in Visser[2002](see also Iemhoff [2001]). This logic covers all the preservation principles that are known so far:

The Principles of the preservativity logic of iPH (Iemhoff[2001]);
$\Box A \equiv_{def} \top \rhd A$.
Taut: all tautologies in iPC (intuitionistic propositional logic)[1]
$P1 : A \rhd B \wedge B \rhd C \rightarrow A \rhd C$
$P2 : C \rhd A \wedge C \rhd B \rightarrow C \rhd A \wedge B$
$Dp : A \rhd C \wedge B \rhd C \rightarrow A \vee B \rhd C$
$4p : A \rhd \Box A$
$Lp : (\Box A \rightarrow A) \rhd A$ (Löb's preservativity principle)
$Mp : A \rhd B \rightarrow (\Box C \rightarrow A) \rhd (\Box C \rightarrow B)$.
$Vp_n : \bigwedge_{i=1}^n (A_i \rightarrow B_i) \rightarrow A_{n+1} \vee A_{n+2}) \rhd (\bigwedge_{i=1}^n (A_i \rightarrow B_i)(A_1, \cdots, A_{n+2})$ (Visser's principles)
$Vp : Vp_1, Vp_2, \cdots$. (Visser's schema)

iPH is the logic of all of the above principles closed under modus ponens and the preservation rule: $A \rightarrow B / A \rhd B$.

Since we will not mention Visser's principles and Visser's scheme after this in the thesis, we choose not to give a detailed definition of these complicated rules. Clearly preservativity logic is a natural extension of the provability logic of HA. It can be shown (Iemhoff [2001]) that both $\Box(A \rightarrow B) \rightarrow A \rhd B$ and $A \rhd B \rightarrow (\Box A \rightarrow \Box B)$ are derivable from the above axioms. And the principles of HA have a more elegant presentation in the setting of preservativity logic. As we have met them only through preservativity logic, we don't know the principles of intuitionistic provability logic. As a matter of fact, the below-defined iH is strictly contained in *the* intuitionistic provability logic. But iH contains all the principles about the proof predicate that have been known so far. The principles for iH are as follows:

$K : \Box(A \rightarrow B) \rightarrow (\Box A \rightarrow \Box B)$
$4 : \Box A \rightarrow \Box\Box A$
$L : \Box(\Box A \rightarrow A) \rightarrow \Box A$
$Le : \Box(A \vee B) \rightarrow \Box(A \vee \Box B)$ (Leivant's axiom)
$Ma : \Box\neg\neg(\Box A \rightarrow \bigvee \Box B_i) \rightarrow \Box(\Box A \rightarrow \bigvee \Box B_i)$ (Formalized Markov Scheme)

The rules for iH are modus ponens and necessitation: $A / \Box A$.

As Iemhoff [2001] has shown, iH is not the $L_\Box$ fragment of iPH, not to mention the $L_\Box$ fragment of the expected preservativity logic.

[1] Consult Troelstra and van Dalen [1988].

1.2 Connection between Preservativity Logic and Interpretability Logic

In order to understand preservativity logic from a broader perspective[2], we briefly discuss[3] the connection between preservativity logic and interpretability logic and introduce a new principle
W: $A \rhd B \to (\Box B \to A) \rhd B$.

The connection is via a new notion: Π_1-conservativity. A theory T is Π_1-*conservative* over T' if T' proves all the Π_1 formulas that T proves. It is easy to see that, for classical theories such as PA, Π_1-conservativity is equivalent to Σ_1-preservativity in the sense that $\varphi \rhd_T \psi$ iff $\neg\varphi$ is Π_1 conservative over $\neg\psi$.

In interpretability logic, we use a binary connective $\rhd_i$. The arithmetical realization of $A \rhd_i B$ in a theory T will be that $T + A$ *interprets* $T + B$,or, alternatively, that T plus the realization of B is Π_1-conservative over T plus the interpretation of A. Thus, for PA, $\varphi \rhd_{PA} \psi$ iff PA plus $\neg\psi$ interprets PA plus $\neg\varphi$ (or $\neg\psi \rhd_i \neg\varphi$). It is well-known that the interpretability logic of PA is ILM. Define $\Diamond A := \neg\Box\neg A$.

Axioms of ILM:
$L : \Box(\Box A \to A) \to \Box A$
$J1 : \Box(A \to B) \to A \rhd_i B$
$J2 : A \rhd_i B \wedge B \rhd_i C \to A \rhd_i C$
$J3 : A \rhd_i C \wedge B \rhd_i C \to (A \vee B) \rhd_i C$
$J4 : A \rhd_i B \to (\Diamond A \to \Diamond B)$
$J5 : \Diamond A \rhd_i A$
$M : A \rhd_i B \to (A \wedge \Box C) \rhd_i (B \wedge \Box C)$
Rules of ILM:
Modus Ponens and necessitation.

Although for PA the notions of interpretability and preservativity are essentially equivalent, they are different for HA. As we have done for PA, we define a translation t from the $L_{\rhd_i}$ to $L_{\rhd}$: $(\neg A \rhd_i \neg B)^t := B \rhd A$. It is easy to check that all the translations of axioms of ILM are provable in iPH. But the converse does not hold. The only axiom in iPH that does not hold for classical theories is the principle Dp.

In this thesis we will introduce a new preservativity principle:
W: $A \rhd B \to (\Box B \to A) \rhd B$.
For classical theories, it is equivalent to the principle
$W_i : A \rhd_i B \to A \rhd_i B \wedge \Box\neg A$,
which is the reason why we call the new preservativity principle W after the well-known principle W_i in interpretability logic. To see this, we reason according to the fact that for classical thoeries, $\varphi \rhd_{PA} \psi$ iff $\neg\psi \rhd_i \neg\varphi$. The following principles (or schemas) are equivalent:
$A \rhd_i B \to A \rhd_i B \wedge \Box\neg A$
$\neg B \rhd \neg A \to \neg(B \wedge \Box\neg A) \rhd \neg A$
(Take $\neg B$ as A and $\neg A$ as B)
$A \rhd B \to (A \vee \neg\Box B) \rhd B$
$A \rhd B \to (\Box B \to A) \rhd B$.

[2] Moreover, we need the following translation between preservativity logic and interpretability logic in Chapter 4
[3] For detailed elaboration, see de Jongh and Japaridze [1998] and Iemhoff [2001].

1.3 Motivation and Overview

Our interests in this thesis are not about iPH or iH itself but about some natural sub-logical systems of iPH and iH. We attain some conservation results relating these preservativity and provability logics. Also we show the fixed point theorem for iL and iPL. There is an open question in Iemhoff [2001]: is there an elegant axiomatization of the $L_\Box$-fragment of iPH. So if the conjecture that iPH is the preservativity logic is true, then that axiomatization will be the intuitionistic provability logic, or the provability logic of HA. Although we will not answer this profound question in this thesis, the conservation results that we achieve here will contribute to our understanding of the close relation between the preservativity logic and the provability logic of HA[4]. In fact, those conservation results are closely related to some much more intuitive equivalence results.

Here we explain the guiding motivation behind the main equivalence results of this thesis. First we define two new rules:
$DR : \Box A \to \Box B / \Box(A \vee C) \to \Box(B \vee C)$
$MoR : \Box A \to \Box B / \Box(\Box C \to A) \to \Box(\Box C \to B)$.
We inherit a notation convention from Iemhoff [2001]. iP is the preservativity logic having as axioms all tautologies in iPC, the principles $P1, P2, Dp$, and as rules Modus Ponens and Preservation. iPX is the logic iP plus the principe X. If the principle is listed in the section 1.1, we will omit the subscript p. For example. we will write $iP4$ instead of $iP4_p$ for the iP plus the principle $4p$. Similarly we can write iX for the provability logic iK with the provability principle X.

If X is an axiom of the form $A \rhd B$, then it is likely that the $L_\Box$ fragment of iPX is the logic iK plus $\Box A \to \Box B$ (because $A \rhd B$ is in iPX) with the extra rule DR (because Dp is in iP and hence is in iPX), and the $L_\Box$ fragment of $iPXM$ is the logic iK plus $\Box A \to \Box B$ with the extra rules DR and MoR (because Mp is in $iPXM$). We don't know whether this holds in general. But, in this thesis, we will see that this is true for many preservativity logics. As an demonstration, we list the following equivalence results:

The $L_\Box$ fragment of iP is equivalent to the logic iK with the extra rule DR (Theorem 3.3.16 and Corollary 3.3.18).

The $L_\Box$ fragment of $iP4$ is equivalent to the logic $iK4$ with the extra rule DR (Theorem 3.1.11).

The $L_\Box$ fragment of iPL is equivalent to the logic iL with the extra rule DR (Theorem 3.2.3 and Corollary 3.2.11).

The $L_\Box$ fragment of iPM is equivalent to the logic iK with the extra rules DR and MoR(Theorems 3.3.21 and 3.3.22).

The $L_\Box$ fragment of iPW is equivalent to the logic iL with the extra rule DR (Theorem 3.5.7).

In the proofs of these results, especially in those of the admissibility of DR or MoR in the $L_\Box$ fragment of the preservativity logic in question, the rule $\Box A \to \Box B / A \rhd B$ will play a dominant role. Take iPL as an example. $iPL_\Box$ denotes the $L_\Box$ fragment of iPL. If we want to show that $iPL_\Box$ is the logic iL with the extra rule DR, we have to prove that DR is admissible in $iPL_\Box$. Instead of showing this directly, we will detour via the admissibility of the rule $\Box A \to \Box B / A \rhd B$ in iPL [5]. If we can show that iPL is closed under the rule $\Box A \to \Box B / A \rhd B$, then the admissibility of DR in $iPL_\Box$ will follow immediately:
$\vdash_{iPL_\Box} \Box A \to \Box B$
$\vdash_{iPL} \Box A \to \Box B$ (by the fact that $iPL_\Box \subseteq iPL$)
$\vdash_{iPL} A \rhd B$ (by the assumption that iPL is closed under the rule $\Box A \to \Box B / A \rhd B$)

[4] We call it analytical. Of course we admit that the sum of the parts are not the whole.

[5] There are some exceptions: when we show the admissibility of DR and MoR in iK, first we can show instead that iP is closed under $\Box A \to \Box B / \Box(A \to B)$, which is stronger than the rule $\Box A \to \Box B / A \rhd B$ (lemma 3.3.23).

$\vdash_{iPL} (A \vee C) \rhd (B \vee C)$ (by Dp)
$\vdash_{iPL_\Box} \Box(A \vee C) \to \Box(B \vee C)$ (by the lemma 2.3.1)
We will treat the admissibility of MoR in $L_\Box$ fragments in a similar way.

But, even with the admissibility of DR in $iPL_\Box$, we only show one direction: the logic iL with the extra rule DR is contained in $iPL_\Box$. In order to show the other direction, we need the conservation of iPL over $iLLe$ with respect to formulas in $L_\Box$. The conservation will give us a tamable normal modal logic $iLLe$ instead of the quite elusive $iPL_\Box$. So it suffices to show that Le is derivable in the logic iL with the extra rule DR, which is much easier to show than that $iPL_\Box$ is contained in the logic iL with the rule DR.

Next we will give an overview of this thesis. In Chapter 2, we will introduce the preliminaries and tools [6] that are needed in the later chapters. In Chapter 3, we present some conservations results and their corresponding equivalence results. In Chapter 4, we prove the fixed point theorems for iL, iPL and iPW.

In addition to the above equivalence results, the main conservation results that we have achieved so far are:

1. iK is the $L_\Box$ fragment of iP;
2. iLe is the $L_\Box$ fragment of $iP4$ (in Iemhoff [2001]);
3. $iLLe$ is the $L_\Box$ fragment of iPL;
4. iK is the $L_\Box$ fragment of iPM.
5. $iLLe$ is also the $L_\Box$ fragment of iPW (where iPW is the new preservativity system introduced in this thesis axiomatized by $P1, P2, Dp$ and the principle:
   $W : A \rhd B \to (\Box B \to A) \rhd B$.)

In the meantime, we will show the admissibility and non-admissibility of some important rules in provability logics:
$\Box A \to \Box B / A \rhd B, \Box A/A, \Box A \to \Box B / \boxdot A \to B$ and some others.

In the last chapter we will show the beautiful fixed point theorems for iL, iPL and iPW, and the interderivability between Beth definability and fixed point properties, which are independent of the other parts of this thesis.

Our presentation here is impartial: whenever a subject about preservativity logic is dealt with, the same subject about provability logic follows immediately.

[6] See Iemhoff [2001].

Chapter 2

Preliminaries and Tools

In this chapter we will develop the basic notions and propositions which we will need in the thesis.

2.1 Semantics for Preservativity Logic

It is common sense in modal logic that with the completeness and soundness at hand we can show not only the theoremhood but also the non-theoremhood of any given formula. In provability logic, however, the modal completeness is more versatile than that because it is the starting point of Solovay's proof of arithmetical completeness of GL. Assume that $GL \nvdash A$. Then by modal completeness, there is a finite, transitive and conversely well-founded model M and and a world r in this model such that $M, r \not\models A$. By applying a series of amazing techniques, we can arithmetize this situation and find a realization (*) such that $PA \nvdash A^*$. The strategy in classical logic has more or less given us a deep insight how to find the preservativity logic, which makes our modal analysis of preservativity logics more cherished. In the meantime, we will pay equal attention to the modal analysis of provability logics.

In the following we will give some basic definitions for the semantics of preservativity logic and intuitionistic provability logic.

Definition 2.1.1 *A frame* F is a triple $\langle W, R, \leq \rangle$, where W is a nonempty set of possible worlds, points or nodes, $\leq$ is a partial order and R is a binary relation satisfying
$\leq \circ R \subseteq R$.
A model M is a quadruple $\langle W, R, \leq, \models \rangle$ where $\langle W, R, \leq \rangle$ is a frame and $\models$ is a forcing relation between points in W and propositional letters which satisfies the following condition:
(persistence) If $x \models p$ and $x \leq y$, then $y \models p$.

$\lhd$

Next we model the whole language by extending the forcing relation $\models$ to relate points to complex formulae. As for the connectives in IPC, we interpret them in usual manner:
$M, w \models A \wedge B \equiv_{def} M, w \models A$ and $M, w \models B$;
$M, w \models A \vee B \equiv_{def} M, w \models A$ or $M, w \models B$;
$M, w \models A \rightarrow B \equiv_{def} \forall v \geq w (M, v \models A$ implies $M, v \models B)$;
$M, w \models \top$ for any w;
$M, w \not\models \bot$ for any w.
Define $\neg A$ as $A \rightarrow \bot$. From the above third and fifth clauses and the definition of $\neg A$ as $A \rightarrow \bot$, it is easy to deduce that

$M, w \models \neg A$ iff $\forall v \geq w(M, v \not\models A)$.

The most important and characteristic clause is the following one for $\rhd$ formulas:
$M, w \models A \rhd B \equiv_{def}$ for any v such that wRv, if $M, v \models A$, then $M, v \models B$.
It follows immediately that
$M, w \models \Box A$ iff for any v such that wRv, $M, v \models A$.
The reason why we have defined the clause for $A \rhd B$ in the above way is that we want it to be a natural extension of our traditional clause for $\Box$ formulas.

It is easy to check (see the following lemma 2.1.2) that
(persistence for all formulas) for any formula A in $L_{\rhd}$, if $M, w \models A$ and $w \leq v$, then $M, v \models A$.

As a matter of fact, given the persistence for propositional letters, the condition that $\leq \circ R \subseteq R$ is a necessary and sufficient condition to guarantee persistence for all formulas, which is different from the condition $\leq \circ R \subseteq R \circ \leq$ for intuitionistic modal logic.[1] We will discuss these in section 2.2.

Lemma 2.1.2 *(Persistence) In every model $M := \langle W, R, \leq, V\rangle$, for every $x, y \in W$ and for every formula A of $L_{\rhd}$,*
if $x \leq y$ and $x \models A$, then $y \models A$.

Proof. We prove this by induction on the complexity of A. It is easy to check that it is true for the base case, conjunction and disjunction. Now we just consider two non-trivial cases.

1. $A := B \rightarrow C$. Assume that $x \leq y \leq z$, $z \models B$ and $x \models A$. It suffices to show that $z \models C$. Since $x \leq y \leq z$, $x \leq z$. It follows that $z \models C$.
2. $A := B \rhd C$. Assume that $x \leq yRz$, $z \models B$ and $x \models B \rhd C$. It suffices to show that $z \models C$. Since $x \leq yRz$, xRz (by the definition 2.1.1). It follows immediately that $z \models C$.

QED

Lemma 2.1.3 *Let $\langle W, R, \leq\rangle$ be a frame such that W is a non-empty set, $\leq$ is a partial order and $(\leq \circ R) \not\subseteq R$. Then there is a formula A of $L_{\rhd}$ and a valuation V such that in $\langle W, R, \leq, V\rangle$ for some $x, y \in W$,*
$x \leq y$ and $x \models A$ but $y \not\models A$.

Proof. Since not $(\leq \circ R) \subseteq R$, there are two worlds $x, y \in W$ such that $x(\leq \circ R)y$ but not xRy. That is to say, there is a world $u \in W$ such that $x \leq uRy$. Take
$V(p) := \{z | y \leq z\}; V(q) := \{z | z \not\leq y\}; V(r) := \emptyset$ for any other propositional letter[2] .
It is easy to check that V is a valuation, i.e. it satisfies that
for any propositional letter t, if $x \leq y$ and $x \in V(t)$, then $y \in V(t)$.
On one hand, $u \not\models p \rhd q$. For $uRy, y \models p$, $y \not\models q$ and hence $u \not\models p \rhd q$.

On the other hand, $x \models p \rhd q$. To see this we shall show that if xRz and $z \models p$, then $z \models q$. Since xRz, $z \neq y$. Moreover, $y \leq z$, for $z \models p$. Therefore $y < z$. This implies that $z \not\leq y$ and hence $z \models q$.

So we get that $x \leq u$, $x \models p \rhd q$ but $u \not\models p \rhd q$.

QED

[1] So the Remark 3.4.1 in Iemhoff [2001] seems mistaken.
[2] We can easily find the equivalence between the presentation of valuation here and the forcing relation in definition 2.1.1

A is *valid in a model* M if for any $w \in W$ $M, w \models A$. A is *valid in a frame* F if A is valid on any model $M = \langle F, \models \rangle$ on the frame.

2.1.1 Canonicity

The usual method in classical modal logic for obtaining completeness is to construct the required counter-model by taking the maximal consistent sets as the nodes in the canonical model. But, as de Jongh-Veltman [1990] pointed out, it can't be applied in interpretability logic directly. It can't generally be applied here either. On the classical side, GL is a good example. Another illustrating example for this is our completeness proof for iL (we give the proof here as an illustration in Theorem 2.2.7 even though it can be found in Iemhoff [2001]pp. 69). We have to restrict maximal consistent sets to some finite set of formulae called an adequate set. Since adequate sets are required to deal with the truth definitions, they have to be closed under subformulas and contain $\top$ and $\bot$. For specific logics, they have to meet specific requirements in addition.

Definition 2.1.4 An *adequate set* X is a set of formulas that is closed under subformulas and contains $\top$ and $\bot$. A consistent subset Γ of X is called X*-saturated* if it satisfies the following two conditions:
For all $A \in X$, if $\Gamma \vdash A$, then $A \in \Gamma$.
For all $A \vee B \in X$, if $A \vee B \in \Gamma$, then $A \in \Gamma$ or $B \in \Gamma$.
If X is the set of all formulas, then an X-saturated set is just called saturated. But also when the context is clear, we just omit X from X-saturated.

$\lhd$

Some notational conventions:

1. For easy reading, we write $w \vdash s \rhd A$ for $w \vdash \bigwedge s \rhd A$ if s is a finite set of formulas.
2. If s is an infinite set, then $w \vdash s \rhd A$ iff there is a finite subset Δ of s such that $w \vdash \bigwedge \Delta \rhd A$.
3. For simplification, sometimes we write $iT \vdash$ instead of the standard $\vdash_{iT}$.
4. $\Gamma \vdash_{iT} A$ is short for the existence of a deduction from Γ to A without use of the rule of necessitation, in other words, A is derivable from the theorems in iT and the formulas in Γ by modus ponens.
5. $\bar{R}$ denotes $R \circ \leq$.

Definition 2.1.5 For any logic L and any adequate set X, the *canonical model* $M = \langle W, R, \leq, \models \rangle$ over X is defined as follows:

1. W consists of all X-saturated sets over X.
2. $wRv :=$ for any formula $A_1, \cdots, A_n, B$, if $w \vdash A_1, \cdots, A_n \rhd B$ and $A_1, \cdots, A_n \in v$, then $B \in v$[3];
3. $w \leq v := w \subseteq v$
4. $w \models p := p \in w$ for any proposition letter p.

$\lhd$

[3] There is a very good reason why we have not defined the accessibility relation as follows: $wRv :=$ for any formula A, B, if $A \rhd B \in w$ and $A \in v$, then $B \in v$. Take $iP4$ for example. With this definition, we can not even show that $iP4$ is complete with respect to the class of finite gathering frames via the standard method of producing a canonical model over X for some finite X.

In order to have an intuition about this definition (compared to what we have done in classical modal logic), we will demonstrate the completeness proof for iP. Before our proof of completeness, we need some general tools which will be very important and, in some sense, are indispensable to proofs of both the completeness for iP and many other completeness proofs in this thesis.

2.1.2 Extendible Properties

Definition 2.1.6 Let iT be a preservativity logic and X be an adequate set. A property $*(\bullet)$ on subsets of X is called an *extendible property* with respect to X if it satisfies the following conditions:

1. For all $A \in X$: if $*(x)$ and $x \vdash A$, then $*(x \cup \{A\})$;
2. For all $A \vee B \in X$: if $*(x \cup \{A \vee B\})$, then $*(x \cup \{A\})$ or $*(x \cup \{B\})$.

If in addition it satisfies
For all $A \in X$: if $*(x)$ and $y \vdash x \rhd A$, then $*(x \cup \{A\})$,
it is called an *iT-extensible y-successor property.*

$\lhd$

We can use the same construction as in Iemhoff [2001] to show that,

1. if $*$ is an iT-extendible property and $*(x)$, then there is an X-saturated set v such that $x \subseteq v$.
2. If if $*$ is an iT-extendible y-successor property and $*(x)$, then there is an X-saturated set v such that $x \subseteq v$ and yRv.

Lemma 2.1.7 [4] *For any logic iT containing iP, for any formula $C \in X$ and for any $w, v \in W$, the following property is an iT-extendible w-successor property:*
$*(x) : w \nvdash x \rhd C$.
If X is additionally closed under disjunction, then the following property is also an iT-extendible w-successor property:
$\star(x)$: *for all $D \in X$, if $w \vdash x \rhd D$, then $D \in v$.*

It is easy to see that if $*(x)$, then there is a saturated set v such that $wRv, x \subseteq v$ and $C \notin v$. And if $\star(x)$, then there is a saturated set u such that $wRu \subseteq v$ and $x \subseteq u$. Instead of showing the above lemma, we will prove a simpler and more-often-used lemma to get used to the above definition of X-saturated sets.

Lemma 2.1.8 *Let Γ be a subset of an adequate set X. If $\Gamma \nvdash A$, then there is a X saturated set w such that $\Gamma \subseteq w$ and $w \nvdash A$.*

Proof. We prove this by Zorn's lemma. Take $S := \{s \subseteq X | s \nvdash A \text{ and } \Gamma \subseteq s\}$. It is easy to see that S is non-empty because $\Gamma \in S$. Take any chain C in S: $C_0 \subseteq C_1 \subseteq \cdots \subseteq C_n \subseteq \cdots$.

Claim: $S_0 := \bigcup C_i$ is an upper bound for this chain. To see this, we only need to show that $S_0 \in S$.i.e. $\Gamma \subseteq S_0$, $S_0 \subseteq X$ and $S_0 \nvdash A$. Of course $\Gamma \subseteq S_0$ and $S_0 \subseteq X$ because $C_i \in X$ and $\Gamma \subseteq C_i$ for all C_i. Suppose that $S_0 \vdash A$, i.e. there are a finite number of formulas $E_1, E_2, \cdots, E_n$ in S_0 such that $E_1, E_2, \cdots, E_n \vdash A$. It is easy to see that there is a k such that all $E_1, E_2, \cdots, E_n \in C_k$. This implies that $C_k \vdash A$. So $C_k \notin S$, which contradicts our assumption that $C_k \in S$. We arrive at a contradiction. So $S_0 \nvdash A$. Conclude: every chain in S has an upper bound.

[4] lemma 3.4.3 in Iemhoff [2001]

By Zorn's lemma, there is a maximal set S^0 in S. It is easy to see that $\Gamma \subseteq S^0$ and $S^0 \nvdash A$. Claim S^0 is an X-saturated set.

1. Assume that $S^0 \vdash B$ and $B \in X$. Suppose that B is NOT in S^0. It suffices to show that $S^0 \cup \{B\} \nvdash A$ because it will contradict the fact that S^0 is a maximal set in S. Suppose that this is NOT the case. $S^0 \cup \{B\} \vdash A$. Then it follows from the assumption that $S^0 \vdash B$ that $S^0 \vdash A$, which contradicts that $S_0 \nvdash A$.

2. Assume that $S^0 \vdash C \vee D$ for some $C \vee D \in X$. Suppose that $C \notin S_0$ and $D \notin S_0$, which also implies that $S^0 \cup \{C\} \vdash A$ and $S^0 \cup \{D\} \vdash A$ because S^0 is a maxmal set in S . By IPC, we get that $S^0, C \vee D \vdash A$. Since $S^0 \vdash C \vee D$, it follows that $S^0 \vdash A$, which contradicts that $S^0 \nvdash A$. So $C \in S_0$ or $D \in S_0$.

QED

Corollary 2.1.9 *If $\nvdash_{iT} A$, then there is an X-saturated set w such that $A \notin w$.*

Proof. We can just take Γ to be $\{\top\}$. QED

2.1.3 Completeness of the Base of Preservativity Logic iP

iP is the logic containing all tautologies in IPC, $P1, P2$ and Dp with the two inference rules: modus ponens and the preservation rule. After the following completeness theorem[5], we will see why iP is called the base of preservativity logic.

Theorem 2.1.10 *$iP \vdash A$ iff A is valid on all frames iff A is valid on all finite frames.*

Proof. First the left-to-right direction. We only check that $(A \rhd C) \wedge (B \rhd C) \to (A \vee B) \rhd C$ is valid on all finite frames. Given any finite frame F, any model M on this frame and any world w in this model, it suffices to show that $M, w \models (A \rhd C) \wedge (B \rhd C) \to (A \vee B) \rhd C$. Take any point v above w and assume that $M, v \models (A \rhd C) \wedge (B \rhd C)$. That is to say, $M, v \models (A \rhd C)$ and $M, v \models (B \rhd C)$. Further we take any successor u of v such that $M, u \models A \vee B$. We can assume without loss of generality that $M, u \models A$. Then by the result that $M, v \models A \rhd C$, we get that $M, u \models C$. So $M, v \models (A \vee B) \rhd C$, which in turn implies that $M, w \models (A \rhd C) \wedge (B \rhd C) \to (A \vee B) \rhd C$.

Now the other and right-to-left direction. As usual, we prove by contraposition. Suppose that $iP \nvdash A$. It is easy to see that there is a finite adequate set X such that A is contained in an X-saturated set w (Corollary 2.1.9). We define the canonical X-model as prescribed above. It is easy to check that it is indeed a frame, i.e $\leq \circ R \subseteq R$. It remains to show the existence lemma and the truth lemma. Since the truth lemma for $\to$ formulas is just as in IPC, we only consider here the one for $\rhd$ formulas.

Lemma 2.1.11 *For any formula $A \rhd B \in X$ and $w \in W$, if $A \rhd B \notin w$, then there is a node v in the X-canonical model such that wRv, $A \in v$ but $B \notin v$.*

The proof of this is just an application of the lemma 2.1.7. Consider the property $*$:
$*(x) : w \nvdash x \rhd B$.
According to the above lemma 2.1.7, it is a w-successor extendible property. Since $A \rhd B \in X$ and $A \rhd B \notin w$, $*(\{A\})$. By the construction that we just mentioned before lemma 2.1.7, we can get that there is a node v such that wRv, $A \in v$ but $B \notin v$.

As a rule, the existence lemma is the crucial lemma in our proof of the truth lemma.

[5] As you will find, iP is one of the largest creditors in this thesis.

Lemma 2.1.12 *For any $B \in X$ and $w \in W$, $w \models B$ iff $B \in w$.*

PROVE this by induction. In fact the proof of the left-to-right direction is about the definition of canonical X-model. We only treated the right-to-left direction and two non-trivial cases in this direction: B is in the form of $C \to D$ and $C \rhd D$.

1. If $B := C \rhd D$ and $C \rhd D \notin w$, then by the above lemma there is a node v such that wRv, $C \in v$ but $D \notin v$, which means, by inductive hypothesis, that $v \models C$ but $v \not\models D$. So $w \not\models C \rhd D$.

2. Assume that $B := C \to D$ and $C \to D \notin w$. Suppose that $w, C \vdash D$. It follows that $w \vdash C \to D$, which implies that $C \to D \in w$ because w is a X-saturated set. We have arrived at a contradiction. So $w, C \nvdash D$. By the above lemma 2.1.7, there is an X-saturated set v such that $w \cup \{C\} \subseteq v$ and $D \notin v$. Of course $w \subseteq v$. By inductive hypothesis and the fact that $w \subseteq v$ implies $w \leq v$ in the canonical model, there is an X saturated set v such that $w \leq v, v \models C$ and $v \not\models D$. So $w \not\models C \to D$.

QED

As we can easily see, iP stands in the same position in preservativity logic as iK does in classical modal logic. In this sense, it is the base of preservativity logic.

2.2 Semantics for Intuitionistic Modal Logic and Brilliant Models

2.2.1 Semantics

It is easy to see that the semantics for intuitionistic modal logic should be one part of the above semantics for preservativity logic because we define $\Box A$ to be $\top \rhd A$. However, there are two essential differences: one is in the sufficient and necessary condition for persistence for all formulas; the other one is in canonical models. The first difference does not make a difference while the second one does.

As Bozic and Dosen [1984] pointed out, given the persistence for propositional letters, the condition that $\leq \circ R \subseteq R \circ \leq$ is a sufficient and necessary condition that guarantee the persistence for all formulas in $L_\Box$ (in fact $\leq \circ R \subseteq R$ implies $\leq \circ R \subseteq R \circ \leq$). So it is natural to define frames for intuitionistic modal logic as follows:

Definition 2.2.1 *A frame* F is a triple $< W, R, \leq >$ where W is a nonempty set of possible worlds, points or nodes, $\leq$ is a partial order and R is a binary relation satisfying
$\leq \circ R \subseteq R \circ \leq$.
A model is defined the same as above definition 2.1.1.

$\lhd$

We are not contented with the harmony here. We want uniformity. It is fine to work on the above defined frames. But it will be better if we can work on the same frames in intuitionistic modal logic as those that we have defined for preservativity logic, namely those satisfying $\leq \circ R \subseteq R$. For we don't need to bother with the difference when we deal with close relations between preservativity logics and provability logics such as the presented conservation results of $L_\rhd$ over $L_\Box$. In the following we will show that we can make a better choice.

Lemma 2.2.2 $x \models \Box A$ *iff* $\forall y(x(R\circ \leq)y \to y \models A)$.

Proof. Left to the reader. QED

Lemma 2.2.3 *In the above definition 2.2.1 of frames the condition* $(\leq \circ R) \subseteq (R\circ \leq)$ *can be replaced by* $(\leq \circ R\circ \leq) \subseteq (R\circ \leq)$.

Proof. Assume that $(\leq \circ R) \subseteq (R\circ \leq)$. Then $(\leq \circ R\circ \leq) \subseteq (R\circ \leq \circ \leq)$. By the transitivity of $\leq$, we get that $(\leq \circ R\circ \leq) \subseteq (R\circ \leq)$.

Assume that $(\leq \circ R\circ \leq) \subseteq (R\circ \leq)$. It is easy to see that $(\leq \circ R) \subseteq (\leq \circ R\circ \leq)$ by the reflexivity of $\leq$. It follows that $(\leq \circ R) \subseteq (R\circ \leq)$. QED

Given any above-defined model $M = \langle W, R, \leq, \models\rangle$, we define a new model $N := \langle W, R_\Box, \leq, \models\rangle$ where $R_\Box \equiv (R\circ \leq)$.

Lemma 2.2.4 $F_\Box = \langle W, R_\Box, \leq\rangle$ *satisfies* $(\leq \circ R_\Box) \subseteq R_\Box$ *and* $(R_\Box\circ \leq) \subseteq R_\Box$.

Proof. It immediately follows from the above lemma 2.2.3.

QED

One observation: the model N satisfies persistence for all formulas in $L_\Box$.

Lemma 2.2.5 *For any formula* A *in* $L_\Box$ *and any world* w *in* W, $M, w \models A$ *iff* $N, w \models A$.

Proof. We prove this by induction on the complexity of A. We will only show the case when $A \equiv \Box B$.

Assume that $M, w \models \Box B$ and $wR_\Box v$. It suffices to show that $N, v \models B$. Since $wR_\Box v$, there is a world $u \in W$ such that $wRu \leq v$. Then $M, u \models B$ for $M, w \models \Box B$. By inductive hypothesis, $N, u \models B$. It follows that $N, v \models B$ by the above observation. For the other direction, assume that $N, w \models \Box B$ and wRv. It suffices to show that $M, v \models B$. Since wRv, $wR_\Box v$. It follows that $N, v \models B$. By induction hypothesis, we get that $M, v \models B$.

QED

Given any model M on a frame satisfying $(\leq \circ R) \subseteq (R\circ \leq)$, we can always find another model N on a frame satisfying $(\leq \circ R) \subseteq R$ which validates the same formulas as the model M. Thus we have achieved uniformity and hence are assured to give a new and simpler definition of frames in intuitionistic modal logic:

Definition 2.2.6 *A frame* F is a triple $< W, R, \leq >$ where W is a nonempty set of possible worlds, points or nodes, $\leq$ is a partial order and R is a binary relation satisfying

$\leq \circ R \subseteq R$.

$\lhd$

Note that in the following we will work with this definition instead of the above old definition 2.2.1.

Now we define the canonical model for intuitionistic provability logic. For any logic T in $L_\Box$ and for any adequate set X, the T-canonical X-model is a quadrable $\langle W, \leq, R, \models\rangle$ defined as follows:

W consists of X-saturated sets

$w \leq v := w \subseteq v$

$wRv :=$ for any $\Box A \in w$, $A \in v$ (*).

$w \models p := p \in w$.

The clause for R is different from that in $L_\rhd$ for R. According to our previous definition for R in preservativity logic, the clause for R should have been

$wRv :=$ for all $A \in X$, if $w \vdash \Box A$, then $A \in v$ (**)

The difference between $\Box A \in w$ and $w \vdash \Box A$ is just of a technical nature (See the completeness proof of iL). Both clauses (*)and (**) make the canonical model defined in this section satisfy the following property:

(brilliancy)[6] $R\circ \leq \subseteq R$.

To see this, take any two nodes $w, v \in W$ and assume that $w(R\circ \leq)v$. Then there is a world x such that $wRx \leq v$. If $\Box A \in w$ (or $w \vdash \Box A$), then $A \in x$. It follows that $A \in v$. So wRv.

However, the canonical model for preservativity logic can't be brilliant. If the frame were brilliant, then the principle $A \rhd B \to \Box(A \to B)$ would be valid on this frame. But the principle is NOT arithmetically valid (and hence, by Iemhoff's soundness results, is certainly not modally valid). In contrast, all intuitionistic modal logics iT that we will consider below are complete with respect to some class of brilliant frames.

Similarly we can show that iK is complete w.r.t the class of finite brilliant frames. It is easy to see that all the notions and propositions above can be adapted into intuitionstic modal logic automatically. We will not go into details about that.

2.2.2 iL as an example

Instead iL stands at the center of the stage of this section[7]. We will give a detailed completeness proof of iL to get a sense of the above definitions. Consider the big difference between IPC and CPC, we might think that the completeness proof for iL will be very different from that for GL. As you will see in the following theorem, this is not the case. The proofs are almost the same. Besides, the proof of iL will be very instructive and exemplary when we try to show the completeness of iPL, iL's analogue in preservation setting.

Theorem 2.2.7 [8] *In* $L_\Box$,

1. *L corresponds to semi-transitivity plus conversely well-foundedness;*
2. *$iL \vdash A$ iff A is valid on all finite transitive and conversely well-founded frames*[9].

Proof. The PROOF of the first proposition is just a routine exercise in Correspondence Theory. We leave it to the reader and devote ourselves to the much more important completeness proof.

In fact the proof of the left-to-right direction of the second proposition is just an application of the correspondence result. Now we only show the other direction. We prove this part by contraposition.

Suppose that $iL \nvdash A$. According to the above lemma 2.1.8, there are an adequate set X and an X-saturated set w such that $A \notin w$. We define the canonical model M as prescribed above except R. It is defined the same as in the classical case (but there is an essential difference between the two proofs):

$wRv :=$ (i) if $\Box B \in w$, then both $\Box B$ and B are in v; (ii) there is a boxed formula $\Box E \in X$ such that $\Box E \in v$ but $\Box E \notin w$.

[6] Bozic and Dosen [1984] called it strongly condensed
[7] Behind the stage is the intimidating iPL
[8] Theorem 4.3.2 in Iemhoff [2001]
[9] Brilliancy of the canonical model for iL does play a role here: brilliancy plus semi-transitivity implies transitivity.

As usual, it remains to show the truth lemma:

For any $B \in X$ and for any $w \in W$, $B \in w$ iff $w \models B$.

We will prove this by induction on the complexity of B and treat the only non-trivial case: $B \equiv \Box C$. Assume that $w \not\models \Box C$. Then there is a node v such that wRv and $v \not\models C$. By inductive hypothesis, this also means that $C \notin v$. It follows that $\Box C \notin w$ because otherwise $C \in v$.

Next the other direction. Suppose that $\Box C \notin w$. Take $\Delta := \{\Box E, E | \Box E \in w\} \cup \{\Box C\}$. The mentioned difference comes here. If $\{\Box E, E | \Box E \in w\} \cup \{\Box C\} \vdash C$, then $\{\Box E, E | \Box E \in w\} \vdash \Box C \to C$. It follows that $\bigwedge_i \Box E_i \vdash \Box(\Box C \to C)$ where $\Box E_1, \cdots, \Box E_n$ is an enumeration of all the boxed formulas in w (Here we use the fact that the principle 4 is derivable from L as is shown in Theorem 2.4.2). So $\bigwedge_i \Box E_i \vdash \Box C$ (by Löb's principle). Since all the $\Box E_i$'s are in w and w is a saturated set, $\Box C \in w$, which contradicts our assumption that $\Box C \notin w$. Therefore $\{\Box E, E | \Box E \in w\} \cup \{\Box C\} \nvdash C$. According to the above lemma 2.1.8, there is an X-saturated set v such that $\{\Box E, E | \Box E \in w\} \cup \{\Box C\} \subseteq v$ and $C \notin v$. It is easy to see that wRv (by the definition of Δ and v) and $v \not\models C$ (by inductive hypothesis). That is to say, $w \not\models \Box C$.

QED

As you can tell, the above proof method for completeness is an alternative to that in the completeness for iP. In the proof of the completeness for iP, we define the accessibility relation on the canonical models as general as possible. But in the above proof for iL, we define a new and stronger accessibility relation, which also means that we have to show the truth lemma differently. We can't just borrow the proof of the truth lemma from the proof for iP. In this thesis, we will use these two methods differently in the completeness proofs for many other logics depending on which logic we are working on.

Compared to the easy-going proofs in iL and GL and their similarity, the completeness proof of iPL should look much more complicated and quite different.[10].

The most successful strategy in the world is always to aim at the best with a peaceful mind to prepare for the worst. As far as we know, the best possibility for the modal completeness of iPL does not exist. Now we have to accept the possibility that it is incomplete. To use the same pattern as before, we will give an example of incomplete logic in intuitionistic modal logic to get an insight about iPL.

Consider the principle $YS : \Box(\Box A \leftrightarrow A) \to \Box A$. Let us call H, iKH the system of modal logic K, iK plus the principle YS, respectively. A system L is of normal modal logic is called *complete* if every sentence that is valid in every frame appropriate to L is derivable in L. A system is called *incomplete* if it is not complete. The following lemma is adapted from Boolos [1993].

Lemma 2.2.8 *$\Box(\Box p \to p) \to \Box p$ and $\Box(\Box p \leftrightarrow p) \to \Box p$ are valid on the same frames.*

Proof. It is easy to see that $\Box(\Box p \leftrightarrow p) \to \Box p$ is valid in the frames where $\Box(\Box p \to p) \to \Box p$ is valid. Now we just show the other direction. We need to show that $\Box(\Box p \to p) \to \Box p$ is also valid on every frame on which $\Box(\Box p \leftrightarrow p) \to p$ is valid. Given a frame $\langle W, R, \leq \rangle$, assume that $\Box(\Box p \leftrightarrow p) \to \Box p$ is valid in this frame. Consider any $w', w, x \in W$ such that $w' \leq wRx$. Suppose that $M, w \models \Box(\Box p \to p)$. It suffices to show that $M, x \models p$.

Instead of showing $M, x \models p$ directly, we prove it by detour via a new model. Define $N := \langle W, R, \leq, U \rangle$ where U is defined as follows:

$t \in U(p)$ iff $M, t \models \Box^n p$ for all $n \geq 0$.

Next take $wRv \leq y$. One observation: $M, y \models \Box p \to p$. We will show[11] that

[10] See remarks in Page 68 of Iemhoff [2001]

[11] The proposition here transcends the barriers between intuitionistic modal logics and classical modal logics.

$N, y \models p$ iff $N, y \models \Box p$, and hence $v \models p \leftrightarrow \Box p$.

The following are equivalent:
$N, y \models p$
$M, y \models \Box^n p$ for all $n \geq 0$
$M, y \models \Box^n p$ for all $n \geq 1$ (the direction from this proposition to the above one is by the observation)
For each z such that yRz, $M, z \models \Box^n p$ for all $n \geq 0$.
For each z such that yRz, $N, z \models p$.
$N, y \models \Box p$.

So we have shown that $N, v \models \Box p \leftrightarrow p$. Since wRv, $N, w \models \Box(\Box p \leftrightarrow p)$. It follows from the assumption that $\Box(\Box p \leftrightarrow p) \rightarrow \Box p$ is valid in the frame $\langle W, R, \leq\rangle$ that $N, w \models \Box p$. Since wRx, $N, x \models p$, i.e. $M, x \models \Box^n p$ for all $n \geq 0$. If we take $n = 0$, then we get that $M, x \models p$.

QED

Theorem 2.2.9 *iKH is incomplete. $\Box p \rightarrow \Box\Box p$ is not a theorem of iKH.*

Proof. $\Box p \rightarrow \Box\Box p$ is not a theorem of iKH. For otherwise $\vdash_{iKH} \Box p \rightarrow \Box\Box p$ and hence $\vdash_H \Box p \rightarrow \Box\Box p$, which is impossible because it is well-known that $\Box p \rightarrow \Box\Box p$ is not a theorem of H[12]. On the other hand, $\Box p \rightarrow \Box\Box p$ is valid in all the frames in which YS is valid. Therefore, iKH is incomplete.

QED

How can we make this logic complete by adding some new principles? The cheapest way is to add 4. We will show that in the section 2.4. Moreover, we can also show that $(\Box A \rightarrow A) \rhd A$ and $(\Box A \leftrightarrow A) \rhd A$ are valid in the same frames[13].

2.3 Some Basic Propositions in Preservativity Logic

In the following we achieve some basic propositions in preservativity logics that will be very useful to other sections in this thesis. First we repeat the presentation of the logical system iPH:
Principles of the preservativity logic of HA (Iemhoff [2001]);
$\Box A \equiv_{def} \top \rhd A$.
Taut: all tautologies in iPC (intuitionistic propositional logic)
$P1 : A \rhd B \wedge B \rhd C \rightarrow A \rhd C$
$P2 : C \rhd A \wedge C \rhd B \rightarrow C \rhd A \wedge B$
$Dp : A \rhd C \wedge B \rhd C \rightarrow A \vee B \rhd C$
$4p : A \rhd \Box A$
$Lp : (\Box A \rightarrow A) \rhd A$ (Löb's preservativity principle)
$Mp : A \rhd B \rightarrow (\Box C \rightarrow A) \rhd (\Box C \rightarrow B)$.
$Vp_n : \bigwedge_{i=1}^n (A_i \rightarrow B_i) \rightarrow A_{n+1} \vee A_{n+2}) \rhd (\bigwedge_{i=1}^n (A_i \rightarrow B_i)(A_1, \cdots, A_{n+2})$ (Visser's principles)
$Vp : V_{p_1}, V_{p_2}, \cdots$. (Visser's schema)
Inference rules: Modus ponens and preservation rule.

Sometimes we will find that the following principle W (introduced in Chapter 1) is very handy:
$W : A \wedge \Box B \rhd B \rightarrow A \rhd B$.
Next we want to find the connection between the natural rule for preservativity logic: preservation rule and the more common used rule: necessitation rule. iP^- is the logic iP minus Dp.

[12] see theorem 2 in Page 152 of Boolos [1993].
[13] We confess that we still don't know whether the logic iP with the extra principle $(\Box A \leftrightarrow A) \rhd A$ is complete or not.

Lemma 2.3.1 $iP^- \vdash \Box(A \to B) \to (A \rhd B)$

Proof.[14] $\vdash_{iP^-} \Box(A \to B) \leftrightarrow T \rhd (A \to B)$
$\vdash_{iP^-} \Box(A \to B) \to A \rhd (A \to B)$ (because $\vdash_{iP^-} A \rhd \top$)
$\vdash_{iP^-} \Box(A \to B) \to A \rhd (A \wedge (A \to B)$
$\vdash_{iP^-} \Box(A \to B) \to A \rhd B$ (because $\vdash_{iP^-} (A \wedge (A \to B)) \to B$ and hence $\vdash_{iP^-} (A \wedge (A \to B)) \rhd B$)

QED

Theorem 2.3.2 *In any preservativity logic iT containing all theorems in iP^-, the preservation rule and the necessitation rule are equivalent*[15].

Proof. Assume that $A \to B / A \rhd B$ is admissible in iT and $iT \vdash A$. Then $iT \vdash \top \to A$ because $IPC \vdash A \to (\top \to A)$. By the preservation rule, we get that $iT \vdash \top \rhd A$, i.e $iT \vdash \Box A$.

Now for the other direction. Assume that the necessitation rule is admissible in iT and $iT \vdash A \to B$. By applying the necessitation rule, we get that $iT \vdash \Box(A \to B)$. It follows from the above lemma that $iT \vdash A \rhd B$.

QED

Thirdly we want to prove a substitution lemma which will be very important in our proof of the fixed point theorem for preservativity logic.

Lemma 2.3.3 $iP4 \vdash \boxdot(A \leftrightarrow B) \to (F_p(A) \leftrightarrow F_p(B))$.

Proof. PROOF by induction on the complexity of F.

1. Base case: $F := \bot, p, q$. Immediate. Here we can see why we use $\boxdot$ instead of $\Box$.

2. $F := C \to D$. IH: $iP4 \vdash \boxdot(A \leftrightarrow B) \to (C_p(A) \leftrightarrow C_p(B))$ (*) and $iP4 \vdash \boxdot(A \leftrightarrow B) \to (D_p(A) \leftrightarrow D_p(B))$ (**). You can find the argument in the following: $iP4 \vdash \boxdot(A \leftrightarrow B) \to ((C_p(A) \to D_p(A)) \to (C_p(B) \to D_p(A))$ (by (*)). And then $iP4 \vdash \boxdot(A \leftrightarrow B) \to ((C_p(A) \to D_p(A)) \to (C_p(B) \to D_p(B))$ (by (**)). Similarly we can show the other implication.

3. $F := C \vee D$. IH: $iP4 \vdash \boxdot(A \leftrightarrow B) \to (C_p(A) \leftrightarrow C_p(B))$ (*) and $iP4 \vdash \boxdot(A \leftrightarrow B) \to (D_p(A) \leftrightarrow D_p(B))$ (**). It follows immediately that $iP4 \vdash \boxdot(A \leftrightarrow B) \to (C_p(A) \to D_p(B) \vee C_p(B))$ and $iP4 \vdash \boxdot(A \leftrightarrow B) \to (D_p(A) \to C_p(B) \vee D_p(B))$, both of which together implies that $iP4 \vdash \boxdot(A \leftrightarrow B) \to (C_p(A) \vee D_p(A) \to C_p(B) \vee D_p(B))$ or $iP4 \vdash \boxdot(A \leftrightarrow B) \to ((C \vee D)_p(A) \to (C \vee D)_p(B))$. Similarly we can show the other direction.

4. We can show the case when $F := C \wedge D$ similarly.

5. $F := C \rhd D$. IH: $iP4 \vdash \boxdot(A \leftrightarrow B) \to (C_p(A) \leftrightarrow C_p(B))$ (*) and $iP4 \vdash \boxdot(A \leftrightarrow B) \to (D_p(A) \leftrightarrow D_p(B))$ (**). It is easy to see that $iP4 \vdash \boxdot(A \leftrightarrow B) \to \Box(C_p(A) \leftrightarrow C_p(B))$ and $iP4 \vdash \boxdot(A \leftrightarrow B) \to \Box(D_p(A) \leftrightarrow D_p(B))$. Further it follows form lemma 2.3.1 that $iP4 \vdash \boxdot(A \leftrightarrow B) \to (C_p(A) \lhd\rhd C_p(B))$ and $iP4 \vdash \boxdot(A \leftrightarrow B) \to (D_p(A) \lhd\rhd D_p(B))$. Then $iP4 \vdash \boxdot(A \leftrightarrow B) \to (C_p(A) \rhd D_p(A) \to C_p(B) \rhd D_p(B))$.i.e. $iP4 \vdash \boxdot(A \leftrightarrow B) \to ((C \rhd D)_p(A) \to (C \rhd D)_p(B))$. Similarly we can show the other direction: $iP4 \vdash \boxdot(A \leftrightarrow B) \to ((C \rhd D)_p(B) \to (C \rhd D)_p(A))$.

QED

[14] lemma 3.1.1 in Iemhoff [2001].
[15] This theorem is, in fact, implicit in Lemma 3.1.2 in Iemhoff [2001].

The following lemma [16] looks very trivial but it is very handy in many of our equivalence proofs in this thesis.

Lemma 2.3.4 $\vdash_{iP^-} [(A \to C) \wedge A] \rhd C$
$\vdash_{iP^-} A \rhd (C \to A \wedge C)$

Theorem 2.3.5 *The following two forms of the Mp principle are equivalent over iP^-:*

1. $A \rhd B \to (\Box C \to A) \rhd (\Box C \to B)$
2. $(A \wedge \Box C) \rhd B \to A \rhd (\Box C \to B)$.

Proof. First the direction from 2 to 1. Reason inside iP^-:
$[(\Box C \to A) \wedge \Box C] \rhd A$ (by Lemma 2.3.4)
$A \rhd B \to [(\Box C \to A) \wedge \Box C] \rhd B$ (by $P1$)
$[(\Box C \to A) \wedge \Box C] \rhd B \to (\Box C \to A) \rhd (\Box C \to B)$ (by 2)
$A \rhd B \to (\Box C \to A) \rhd (\Box C \to B)$

Next the other direction from 1 to 2. Reason inside iP^-:
$(A \wedge \Box C) \rhd B \to (\Box C \to A \wedge \Box C) \rhd (\Box C \to B)$ (by 1)
$A \rhd (\Box C \to A \wedge \Box C)$ (by Lemma 2.3.4)
$(A \wedge \Box C) \rhd B \to A \rhd (\Box C \to B)$.

QED

Another application of the above trivial-looking lemma is the following equivalence proof for the W principle.

Corollary 2.3.6 *The following two forms of W are equivalent over iP^-:*

1. $A \wedge \Box B \rhd B \to A \rhd B$
2. $A \rhd B \to (\Box B \to A) \rhd B$

Proof. First the direction from 1 to 2:
$((\Box B \to A) \wedge \Box B) \rhd A$ (by lemma 2.3.4)
$A \rhd B \to ((\Box B \to A) \wedge \Box B) \rhd B$
$A \rhd B \to (\Box B \to A) \rhd B$ (by 1)

Next the other and 2-to-1 direction.
$A \rhd (\Box B \to A \wedge \Box B)$ (by Lemma 2.3.4)
$A \wedge \Box B \rhd B \to (\Box B \to A \wedge \Box B) \rhd B$ (by 2)
$A \wedge \Box B \rhd B \to A \rhd B$

QED

$iPLM$ is axiomatized over iP by Mp and Lp. iPL is axiomatized over iP by Lp. After the above lemma 2.3.6 and theorem 2.3.7, we can prove a lemma which will be very useful in Section 3.5:

Lemma 2.3.7 *$iPLM \vdash W$ and $iPW \vdash Lp$.*

[16] In fact we have used it in the proof of lemma 2.3.1.

Proof. First we show that $iPLM \vdash W$
$iPLM \vdash (A \wedge \Box B) \rhd B \to A \rhd (\Box B \to B)$ (by Mp)
$iPLM \vdash (A \wedge \Box B) \rhd B \to A \rhd B$ (by Lp)

Next the second one: $iPW \vdash Lp$.
$iPW \vdash A \rhd B \to [(\Box B \to A) \wedge \Box B] \rhd B$ (by the trivial lemma)
$iPW \vdash A \rhd B \to (\Box B \to A) \rhd B$ (by W). After replacing A with B, we get that $iPW \vdash Lp$.
QED

In fact the above lemma tells us that iPW is contained in $iPLM$. As we will show below, this containment is strict.

Theorem 2.3.8 *The following two principles are equivalent in iP:*

1. $(\Box A \to A) \rhd A$;
2. $(A \rhd B) \to (A \to B)) \rhd (A \to B)$

Proof. If we take A to be $\top$, then we will get the direction from the 2 to 1 immediately. And the proof of the other direction follows from Lemma 2.3.1. QED

2.4 Some Basic Propositions in Intuitionistic Provability Logic

As we have mentioned above, it is still unknown what the provability logic of HA is. In the following we will again write the principles that are known so far:
$K : \Box(A \to B) \to (\Box A \to \Box B)$
$4 : \Box A \to \Box\Box A$
$L : \Box(\Box A \to A) \to \Box A$
$Le : \Box(A \vee B) \to \Box(A \vee \Box B)$ (Leivant's axiom)
$Ma : \Box\neg\neg(\Box A \to \bigvee \Box B_i) \to \Box(\Box A \to \bigvee \Box B_i)$ (Formalized Markov Scheme)
Inference rules for iH: modus ponens and necessitation.

We call iH the logic with the above principles and rules. Recall that it is NOT the provability logic of HA.

Similarly we can get a substitution lemma, which will be very important in showing the fixed point theorem for iL.

Lemma 2.4.1 $iK4 \vdash \boxdot(A \leftrightarrow B) \to (F_p(A) \leftrightarrow F_p(B))$

Proof. The proof is just a sub-proof of the above one for $iP4$. QED

Theorem 2.4.2 $iL \vdash \Box A \to \Box\Box A$ *for any formula A.*

Proof. The proof here is the same as that on the classical side. Reason inside iL:
$A \to ((\Box\Box A \wedge \Box A) \to (\Box A \wedge A)$
$A \to (\Box(\Box A \wedge A) \to (\Box A \wedge A))$
$\Box A \to \Box(\Box(\Box A \wedge A) \to (\Box A \wedge A))$
$\Box A \to \Box(\Box A \wedge A)$
$\Box(\Box A \wedge A) \to \Box\Box A$
$\Box A \to \Box\Box A.$
QED

The following proposition just as the above one shows that the principle 4 is a very basic principle in the setting of intuitionistic provability logic.

Theorem 2.4.3 $\vdash_{iLe} \Box A \to \Box\Box A$

Proof. $\vdash_{iLe} \Box A \to \Box(A \vee \Box A)$
$\vdash_{iLe} \Box A \to \Box(\Box A \vee \Box A)$
$\vdash_{iLe} \Box A \to \Box\Box A$ QED

Next we show a theorem that is common to the logics $K4, iK4$ and $iP4$ that capture the Löb principle in the same way: [17].

Theorem 2.4.4 $\vdash_{iK4} \Box(\Box(\Box A \to A) \to \Box A) \to (\Box(\Box A \to A) \to \Box A)$

Proof. Reason inside $iK4$:
$\Box(\Box(\Box A \to A) \to \Box A) \to (\Box(\Box(\Box A \to A)) \to \Box\Box A)$
$\Box(\Box(\Box A \to A) \to \Box A) \to (\Box(\Box A \to A) \to \Box\Box A)$
$\Box(\Box(\Box A \to A) \to \Box A) \to (\Box(\Box A \to A) \to \Box\Box A \wedge \Box(\Box A \to A))$
$\Box(\Box(\Box A \to A) \to \Box A) \to (\Box(\Box A \to A) \to \Box A)$
QED

The *Löb rule* is the modal-logical inference rule:
$\Box A \to A/A$
Let $i4LR$ be the system of modal logic whose axioms are those of $iK4$ and whose rules are modus ponens, necessitation and the Löb rule. It will make no difference whether you add the Löb principle or the Löb rule in to $iK4$ as the folloing theorem [18] shows.

Theorem 2.4.5 *$i4LR \vdash A$ iff $iL \vdash A$ for any formula A.*

Proof. The left-to-right direction is obvious. We just consider the other direction. According to the above lemma,
$\vdash_{i4LR} \Box(\Box(\Box A \to A) \to \Box A) \to (\Box(\Box A \to A) \to \Box A)$.
By applying LR once, we immediately get that
$\vdash_{i4LR} \Box(\Box A \to A) \to \Box A$.
QED

It is time to answer the question left in section 2.2: How to make iKH complete. Let us call $i4YR$ the logic $iK4$ with the extra rule YR: $A \leftrightarrow \Box A/A$. As we will show in the following lemma, it makes no difference adding LR or YR to the background system $iK4$.

Lemma 2.4.6 *$\vdash_{i4LR} A$ iff $\vdash_{i4YR} A$.*

Proof. We only need to show that LR is admissible in $i4YR$. Reason in $i4YR$:
1. $\Box A \to A$
2. $\Box\Box A \to \Box A$
3. $\Box A \to \Box\Box A$ (by 4)
4. $\Box A \leftrightarrow \Box\Box A$
5. $\Box A$ (by YR)
6. A (by modus ponens of 1 and 5)

QED

[17] See the folowing lemma 3.2.4.
[18] You can find a classical counterpart of the following theorem in Boolos [1993].

Since $i4LR$ is complete (by the above Theorem 2.2.7 and Theorem 2.4.5), $i4YR$ is complete. It is easy to see that $i4YR$ is the same as $i4YS$. For, on one hand, $i4YR \subseteq i4YS$; on the other hand, $i4YS \subseteq iL$. So, if we add 4 to iKH, then we will get a complete system, which is the same as iL.

Theorem 2.4.7 [19] *In $L_\Box$,*

1. 4 *corresponds to semi-transitivity;*
2. 4 *is canonical;*
3. $\vdash_{iK4} A$ *iff* A *is valid on all finite transitive frames.*

By the above theorem, we can immediately get that $iK4$ is NOT closed under the Löb rule. For otherwise the Löb principle would be valid in all finite transitive frames, which we can easily falsify by a counterexample.

[19] Proposition 4.2.1 in Iemhoff [2001].

Chapter 3

Conservation Results

We are ready to begin the main part of this thesis. In Introduction, we have given the motivation behind the main results in this thesis. Moreover, it will be inspiring to explain the importance of conservation results in those equivalence results that we presented in Chapter 1. Again, we will take iPL as an example. It seems that we can't show directly that the $L_\Box$-fragment of iPL is equivalent to the logic iL with the extra rule DR because we even don't know whether the logic iL with the extra rule DR is equivalent to a normal modal logic or not. We will divide the proof into two stages:

1. iPL is conservative over $iLLe$ with respect to formulas in $L_\Box$ (Theorem 3.2.3)

2. $iLLe$ is equivalent to the logic iL with the rule DR (Theorem 3.2.11) .

The proof of the first is based on a transformation procedure from $Le-$frames to gathering frames in Iemhoff [2001]. And the the proof of the second one is a combination of the first one and the admissibility of the rule $\Box A \to \Box B / A \rhd B$ in iPL. This will be a guiding proof pattern of the equivalence results in the following sections.

As you will see, the rule $\Box A \to \Box B / A \rhd B$ plays a dominant role in the following sections. This rule is discussed in Iemhoff [2001] Section 5.2 where a short proof sketch is given for the admissibility of the rule for iPH. We will give detailed proofs of the admissibility of this rule in many other logics, which gives an impression that we repeat the same proofs. This is not the case though. Every time we show the admissibility for a different logic, you will find some additional new ideas in the proof, which will make our modal analysis richer and more powerful. A good example is our proof of the admissibility in iPM.

We will divide the presentation of this chapter into several parts according to the results which we have previously mentioned.

3.1 iLe is Equivalent to the Logic $iK4$ with the Extra rule DR

First we will give two proofs of the admissibility of this rule in $iP4$. The reason why we give the first and more complicated proof is that we want to introduce a handy lemma. But the second one is simpler and is following a more-often-used pattern. A notational convention: $[z) := \{w|$ there is a sequence of $w_0 S_0 w_1 S_1 w_2 \cdots w_n = w$ for some worlds $w_0, w_1, \cdots, w_n$ in the model where $S_i \in \{R, \leq\}\}$. Put it in another way, $[z)$ stands for the sub-model generated by z.

Definition 3.1.1 A frame $F = \langle W, R, \leq, \models \rangle$ is *gathering* if, for any w, v and u in W, $wRvRu$, then $v \leq u$. $\lhd$

Lemma 3.1.2 [1]

1. *The principle* $4p$ *corresponds to gatheringness.*
2. *The logic* $iP4$ *is canonical.*
3. $\vdash_{iP4} A$ *iff* A *is valid on all finite gathering frames.*

Lemma 3.1.3 *(Generated Submodel Lemma for* $iP4$*) Let* M *be a model on a gathering frame and* x, y *be two worlds in this model such that* xRy*. If* $y \models A$*, then, for any* $z \in [y)$*,* $z \models \boxdot A$*.*

Proof. First one observation, for any $z \in [y)$, $y \leq z$. This follows immediately from the fact that M is on a gathering frame. So $z \models A$. Take any successor w of z, $w \models A$ because $w \in [y)$. So $z \models \Box A$ and hence $z \models \boxdot A$.

QED

Lemma 3.1.4 $\vdash_{iP4} A \rhd B$ *iff* $\vdash_{iP4} (\Box A \to \Box B)$.

Proof. The direction from left to right is immediate. Now we only need to treat the other direction. We prove this by contraposition. Suppose that $iP4 \nvdash A \rhd B$. By the completeness of $iP4$, it follows that it is not true at a point w of some finite gathering model M. Then there is a point v such that wRv and $v \models A$ but $v \not\models B$. Now we define a new model from the original one and in fact a submodel of the old one. Take $W' := \{w\} \cup [v)$, $R' = R\restriction_{W'}$ and $\leq' = \leq\restriction_{W'}$. When the context is clear, we will not distinguish $R', \leq'$ from R and $\leq$, which is justified by the definition of M'. And $V'(p) = V(p)$ for any propositional variable p. And it is easy to see that $\leq' \circ R' \subseteq R'$. So M' is on a frame.

Fact 1: for any $x \in [v)$ and for any formula B in $L_{\rhd}$, $M', x \models B$ iff $M, x \models B$.

It is clear that $M', w \not\models \Box B$ because wRv and $M', v \not\models B$.

By the above lemma, we get that $M', w \models \Box A$ because $R[w] \subseteq [v)$ and for any $x \in [v), x \models A$. So $M', w \models \Box A$ but $M', w \not\models \Box B$, i.e. $M', w \not\models \Box A \to \Box B$. By the soundness of $iP4$, $\nvdash_{iP4} \Box A \to \Box B$.

Conclude: $\vdash_{iP4} A \rhd B$ if $\vdash_{iP4} (\Box A \to \Box B)$

QED

Proof. (Second and Simpler Proof) The direction from left to right is immediate. Now we only need to treat the other direction. We prove this by contraposition. Suppose that $\nvdash_{iP4} A \rhd B$. It follows that it is not true at a point w_0 of some finite gathering model M. Then there is a point v such that $w_0 Rv$ and $M, v \models A$ but $M, v \not\models B$. Now we define a new model from the original one. The new one is NOT necessarily a sub-model of the original one. Here w is a new world.Take $W' := \{w\} \cup [v)$, $R' = R\restriction_{[v)} \bigcup \{(w, v)\}$ and $\leq' = \leq\restriction_{[v)} \cup \{(w, w)\}$. And $V'(p) = V(p)$ for any propositional variable p.

Fact 1: for any $x \in [v)$ and for any formula B in $L_{\rhd}$, $M', x \models B \Leftrightarrow M, x \models B$.

It is clear that $M', w \not\models \Box B$ because wRv and $M', v \not\models B$. First we show that the new frame defined above is in fact a finite gathering one. Take $xR'yR'z$. We divide the proof into the following cases;

[1]See Proposition 4.2.1 in Iemhoff [2001].

1. $w \in \{x, y, z\}$. But, according to the above definition of R', $w = x$ and $v = y$. Also it is obvious that $w_0 RyRz$. Then it follows that $y \leq z$ because M is a gathering model. So $y \leq' z$ because $y, z \in [v)$.

2. $w \notin \{x, y, z\}$. It follows that $x, y, z \in [v)$. Therefore, $xRyRz$. By the same argument as above, $y \leq z$. Since $y, z \in [v)$, $y \leq' z$.

It is easy to see that $M', w \models \Box A$ because v is the only successor of w and $M', v \models A$. So $M', w \models \Box A$ but $M', w \not\models \Box B$, i.e. $M', w \not\models \Box A \to \Box B$. By the soundness of $iP4$, $\nvdash_{iP4} \Box A \to \Box B$.

QED

Theorem 3.1.5 [2]*In $L_\Box$,*
(i)On finite frames Le corresponds to the Le-property: $\forall wv(wRv \to \exists x(wRx \leq v \wedge \forall u(vRu \to x \leq u)))$.
(ii)$\vdash_{iLe} A$ iff A is valid on all finite brilliant Le-frames.
(iii)$\vdash_{iLLe} A$ iff A is valid on all finite transitive conversely well-founded brilliant Le-frames.

For the sake of completeness, we will repeat the conservation of $iP4$ over iLe in Iemhoff [2001] though we give a more transparent presentation. Besides we need again the following procedure (in Lemma 3.1.7) transforming Le-frames to gathering frames in the proof of lemmas 3.2.1 and 3.5.6.

Lemma 3.1.6 *Let $M := \langle W, R, \leq, V\rangle$ and $N := \langle W, R', \leq, V\rangle$ be two finite frames. If $R' \subseteq R \subseteq (R' \circ \leq)$, then $M, w \models B$ iff $N, w \models B$ for any formula B in $L_\Box$ and any world $w \in W$.* [3]

Proof. Prove by induction on B. We only show the case when $B := \Box C$. Assume that $M, w \models B$ (i.e. for any v, wRv , $M, v \models C$). Take any x such that $wR'x$. Of course wRx. By the above assumption, $M, x \models C$. By IH, $M', x \models C$. This implies that $M', w \models \Box C$.

Now we show the other direction. Assume that $M', w \models B$ (i.e. for any v, $wR'v$, $M', v \models C$). Take any x such that wRx. By the above assumption, $w(R' \circ \leq)x$. That is to say, there is a w' such that $wR'w'$ and $w' \leq v$. So $M', w' \models C$, which implies that $M', v \models C$. By IH, $M, v \models C$. This implies that $M, w \models \Box C$.

QED

Lemma 3.1.7 *(Transformation Lemma for Le-Frames) Let $M := \langle W, R, \leq, V\rangle$ be a finite Le brilliant frame. Then there is a finite gathering frame $N = \langle W, R', \leq, V\rangle$ such that $R' \subseteq R \subseteq (R' \circ \leq)$.*

Proof. Assume that $M := \langle W, R, \leq, V\rangle$ is a finite Le brilliant frame. Define:
$wR'v \equiv_{def} wRv$ and $\forall u(vRu \to v \leq u)$ and $N := \langle W, R', \leq, V\rangle$.
$S(x)$ denotes the property: $\forall u(xRu \to x \leq u)$. Assume that wRv. We need to find the x such that $wR'x \leq v$. That is to say, $wRx \leq v$ and $S(x)$. By the Le-property, there is a successor x_1 of w which is below v and all its own successors. If $x_1 = v$, then we have found such an x. If $x_1 \neq v$, then there is another successor x_2 of w which is below x_1 and all its own successors. If $x_2 = x_1$, then we have found such an x. If not, we will repeat the same argument as above. Then we will get a sequence $x_1 x_2 \cdots$. Since the frame is finite, there are two nodes $x_{n-1} = x_n$ for some n. So x_n is the x that we are looking for.

QED

Lemma 3.1.8 $\vdash_{iP4} \Box(A \vee B) \to \Box(A \vee \Box B)$[4]

[2]See Proposition 4.4.1 in Iemhoff [2001].
[3]This lemma is similar to a previous lemma 2.2.5.
[4]Page 53 of Iemhoff [2001].

Proof. Reason inside $iP4$:
$A \rhd A$, $B \rhd \Box B$
$A \rhd (A \vee \Box B)$
$B \rhd (A \vee \Box B)$
$(A \vee B) \rhd (A \vee \Box B)$ (by Dp)
$\Box(A \vee B) \to \Box(A \vee \Box B)$ (by lemma 2.3.1)

QED

Theorem 3.1.9 *$\vdash_{iLe} A$ iff A is valid on all finite gathering frames.*

Proof. The right-to-left direction follows from the fact that Le is derivable in $iP4$. Now we just need to show the other direction.

Suppose that $\nvdash_{iLe} A$. Then by the completeness we get that there are a finite brilliant Le model $M = \langle W, R, \leq, V\rangle$ and a world b in this model such that $M, b \not\models A$. By the above lemma 3.1.7 there is another new finite gathering model $N = \langle W, R', \leq, V\rangle$ such that $R' \subseteq R \subseteq (R' \circ \leq)$. By the lemma 3.1.6, it follows that $N, b \not\models A$.

QED

Corollary 3.1.10 *(Conservation) $\vdash_{iP4} A$ iff $\vdash_{iLe} A$ for all formulas A in $L_\Box$.*

Theorem 3.1.11 *iLe is equivalent to the logic $iK4$ with the extra rule DR (denoted by $iK4^*$).*

Proof. First show that iLe is contained in $iK4^*$. We only need to show that Le is derivable in the latter logic. Since $iK4^* \vdash \Box A \to \Box\Box A$, we get by DR that $\Box A \to \Box\Box A / \Box(A \vee B) \to \Box(\Box A \vee B)$. It follows immediately that $iK4^* \vdash \Box(A \vee B) \to \Box(\Box A \vee B)$ for any formula A and B in $L_\Box$, which is exactly Le.

Now we show the other direction. It is known that 4 is derivable in iLe [5]. It suffices to show that DR is an admissible rule in iLe. Assume that $iLe \vdash \Box A \to \Box B$. According to the fact that Le is derivable in $iP4$, $iP4 \vdash \Box A \to \Box B$. By Lemma 3.1.4, we get that $iP4 \vdash A \rhd B$. From the following deduction:
$iP4 \vdash A \rhd B \vee C$ (from the above)
$iP4 \vdash C \rhd B \vee C$
$iP4 \vdash (A \vee C) \rhd (B \vee C)$ (Dp and the above two),
we arrive at the result that $iP4 \vdash \Box(A \vee C) \to \Box(B \vee C)$. By the conservation theorem, $iLe \vdash \Box(A \vee C) \to \Box(B \vee C)$.

So DR is admissible in iLe. Conclude: iLe is equivalent to the logic $iK4$ with the extra rule DR.

QED

3.2 Conservation of iPL over $iLLe$

3.2.1 Conservation of iPL over $iLLe$

First we prove an analogue of Proposition 4.4.2 in Iemhoff [2001].

Lemma 3.2.1 *$iLLe \vdash A$ iff A is valid on all finite gathering conversely well-founded frames.*

[5] Theorem 2.4.3

Proof. First the easier and left-to-right direction. It suffices to show that both L and Le are valid on all finite gathering conversely well-founded frames. By Prop. 4.4.2 in Iemhoff[2001] or the above lemma 3.1.5, Le is valid on all finite gathering frames and hence on all finite gathering conversely well-founded frames. Since L corresponds to semi-transitivity plus conversely well-foundedness and gatheringness implies semi-transitivity, L is valid on all finite gathering conversely well-founded frames.

Now we show the more difficult direction. Suppose that $iLLe \nvdash A$ for any A in $L_\Box$. Since $iLLe \vdash A$ iff A is valid on all finite transitive conversely well-founded brilliant Le-frames, there are a model $M = \langle W, \leq, R, V \rangle$ which is finite transitive conversely well-founded brilliant Le-model and some point $b \in W$ such that $M, b \not\models A$.

We will use the same procedure as that in the above lemma 3.1.7[6]. Define $wR'v := wRv$ and $\forall u(vRu \to v \leq u)$. By the same argument as there, we can get that $R' \subseteq R \subseteq (R' \circ \leq)$. Define $M' = \langle W, R', \leq, V \rangle$. It is easy to see that M' is on a finite gathering frame because we impose this property through the definition of R'.

Claim : M' is conversely well-founded. Suppose NOT. Then there is a loop: $w_0R'w_1R' \cdots R'w_nR'w_0$. According to the definition of R', $w_0Rw_1R \cdots Rw_nRw_0$, which is impossible because R is conversely well-founded. So M' is on a finite gathering well-founded frame.

It follows immediately from the above lemma 3.1.6 that $M', b \not\models A$. Since M' is on a finite gathering conversely well-founded frame, A is NOT valid on all finite gathering conversely well-founded frames.

QED

Lemma 3.2.2 [7] *The principle Lp corresponds to gatheringness plus converse well-foundedness of the modal relation.*

Theorem 3.2.3 *(Conservation) $iLLe$ is the $L_\Box$-fragment of iPL.*

Proof. Suppose that $iLLe \nvdash A$. By the above lemma, A is NOT valid on all gathering conversely well founded frames. It follows that $iPL \nvdash A$ because B is valid on all gathering conversely well-founded frames if $iPL \vdash B$ (A result immediately following from Lemma 3.2.2).

On the other hand, it is easy to see that $iLLe$ is contained in iPL. For both L and Le are derivable in iPL. To put it more precisely, $iPL \vdash 4p$ and $iP4 \vdash Le$ (by the above lemma 3.1.8).

QED

3.2.2 iLLe is Equivalent to iL with the Extra Rule DR

Just like our perfect result about iLe and $iP4$, it is natural to ask the following question:

Is $iLLe$ equivalent to iL with the extra rule DR? (*)

It is easy to see that $iLLe$ is contained in the logic iL with the extra rule DR [8]. For the other direction, it seems that we need the completeness of iPL to show that $\Box A \to \Box B / A \rhd B$ is an admissible rule in iPL like what we have shown for $iP4$. If this were the case, the above aimed question would be reduced to the following *question*:

[6] In fact, Iemhoff's procedure is quite universal: It can be used to transform almost all Le-frames into gathering frames. The only thing that we need to check is whether the other frame properties are invariants under the transformations like the converse well-foundedness above.

[7] See Proposition 4.3.1 in Iemhoff [2001].

[8] see the above theorem 3.1.11.

is it true that $iPL \vdash \Box A \to \Box B$ iff $iPL \vdash A \rhd B$ (**).
Unfortunately, like what we have shown for $iP4$, this question would reduce to our most difficult question in this thesis:
Is iPL complete with respect to some class of frames? (***)
However, fortunately, we can work out the problem without the completeness of iPL. In the following we will show that $\Box A \to \Box B/A \rhd B$ is admissible in iPL by detouring via the admissibility of this rule in $iP4$ (Lemma 3.1.4).

In the following paragraphs we use two ways to show that
$iPL \vdash \Box A \to \Box B \Leftrightarrow iPL \vdash A \rhd B$.
The first is unique for iPL because of the lemma 3.2.4 below. The second one is quite general. In fact the general result can show besides that $iP4V_n$ and $iPLV_n$ are also closed under the rule: $\Box A \to \Box B/A \rhd B$.

Lemma 3.2.4 $iP4 \vdash \Box((\Box C \to C) \rhd C) \to (\Box C \to C) \rhd C$[9]

Proof.(Semantic Proof) Suppose that: $iP4 \nvdash \Box((\Box C \to C) \rhd C) \to (\Box C \to C) \rhd C$. Then by the completeness, we get that there is a finite gathering model M and a possible world $w_0 \in W$ such that
$w_0 \not\models \Box((\Box C \to C) \rhd C) \to (\Box C \to C) \rhd C$
This means that there is a world w_1 above w_0 such that
$w_1 \models \Box((\Box C \to C) \rhd C)$ but $w_1 \not\models (\Box C \to C) \rhd C$
Further it implies that there is a successor w_2 of w_1 such that
$w_2 \models \Box C \to C$ but $w_2 \not\models C$, which implies that $w_2 \not\models \Box C$.
Again it will implies that there is a successor w_3 of w_2 such that:
$w_3 \models \Box C \to C$ but $w_3 \not\models C$, which implies that $w_3 \not\models \Box C$.
Then we will get an infinite line (in fact a loop) such that for any w_j in the line:
$w_j \models \Box C \to C$ but $w_j \not\models C$, which implies that $w_j \not\models \Box C$
However, since $w_1 \models \Box((\Box C \to C) \rhd C)$, $w_2 \models (\Box C \to C) \rhd C$. Since $w_3 \models \Box C \to C$, $w_3 \models C$, which contradicts our above result that $w_3 \not\models C$.

Conclude: $iP4 \vdash \Box((\Box C \to C) \rhd C) \to (\Box C \to C) \rhd C)$

QED

Before we give a syntactic proof of the same lemma, we will show a lemma in iP at first.

Lemma 3.2.5 $iP \vdash \Box(A \rhd B) \to (\Box A \rhd \Box B)$

Proof. We can give a very simple semantic proof to the lemma which is quite similar to that in the above lemma. We choose to give a syntactic proof. The argument is as follows:
$iP \vdash A \rhd B \to (\Box A \to \Box B)$
$\Rightarrow iP \vdash \Box(A \rhd B) \to \Box((\Box A \to \Box B))$
$\Rightarrow iP \vdash \Box(A \rhd B) \to (\Box A \rhd \Box B)$

QED

Proof.(Syntactic Proof of Lemma 3.2.4) The argument is as follows:
$iP4 \vdash \Box((\Box C \to C) \rhd C) \to \Box(\Box C \to C) \rhd \Box C$ (by the above lemma)
$\Rightarrow iP4 \vdash \Box((\Box C \to C) \rhd C) \to (\Box C \to C) \rhd \Box C$ (by $4p$)
$\Rightarrow iP4 \vdash \Box((\Box C \to C) \rhd C) \to (\Box C \to C) \rhd (\Box C \wedge (\Box C \to C))$
$\Rightarrow iP4 \vdash \Box((\Box C \to C) \rhd C) \to (\Box C \to C) \rhd C$

QED

[9]We have achieved a similar lemma in $iK4$. See lemma 2.4.4.

Corollary 3.2.6 *The Löb rule* (LR) *is not admissible in* $iP4$*.*

Proof. Suppose that LR is admissible in $iP4$. Then by the above lemma, we would get that Lp is derivable in $iP4$, which is impossible. We can easily find a counterexample by giving a gathering model which is not conversely well-founded such that $4p$ is valid in this frame but Lp is not.

QED

Lemma 3.2.7 $iP4 \vdash \Box \boxdot L \leftrightarrow \boxdot L, \leftrightarrow \Box L$ *where* L *is* $(\Box C \to C) \rhd C$*.*

Proof. $iP4 \vdash \Box \boxdot L \to L \wedge \Box L$ (by lemma 3.2.4). That is to say,
$iP4 \vdash \Box \boxdot L \to \boxdot L$;
Now for the other direction.
$iP4 \vdash (L \wedge \Box L) \to (\Box L \wedge \Box\Box L)$ (by $4p$ or 4),i.e.$iP4 \vdash \Box \boxdot L \leftrightarrow \boxdot L$. Similarly we can show the second equivalence. QED

Lemma 3.2.8 *(Detour Lemma 1)* $iPL \vdash A$ *iff there exist* $C_1, C_2, \cdots, C_n$ *such that* $iP4 \vdash \Box((\Box C_1 \to C_1) \rhd C_1) \wedge \cdots \wedge \Box((\Box C_n \to C_n) \rhd C_n) \to A$*.*

Proof. Let C denote $((\Box C_1 \to C_1) \rhd C_1) \wedge \cdots \wedge (\Box C_n \to C_n) \rhd C_n))$. Then the above proposition can be put in the following simpler way:
$iPL \vdash A$ iff there exists C such that $iP4 \vdash \Box C \to A$.

The direction from right to left is obvious. Now we just show the other direction. Assume that $iPL \vdash A$. This also means that there is a finite sequence of proof[10]of A in iPL: $s_1 s_2 \cdots s_n = A$. Let $(\Box C_1 \to C_1) \rhd C_1), \cdots, ((\Box C_n \to C_n) \rhd C_n)$ be the instances of the Löb principle occurring in the sequence and C denote $(\Box C_1 \to C_1) \rhd C_1) \wedge \cdots \wedge (\Box C_n \to C_n) \rhd C_n))$. It suffices to show that there is a finite sequence of proof of $\Box C \to A$ in $iP4$. In fact this required sequence is based on the original one for A. Define $s'_i := \Box C \to s_i$.

Fact 1: $iP4 \vdash \Box C \to \boxdot C$. This follows from the lemma 3.2.4.

Claim 2: There is a finite sequence of proof of $\Box C \to A$ in $iP4$. Prove by cases: for any $s_i (i \leq n)$,

1. if s_i is an instance of axiom of iP, then s'_i is a theorem of iP and hence of $iP4$ because in fact $s_i \to (\Box C \to s_i)$ is a tautology;
2. if s_i is an instance of Lp, it is easy to see that s'_i is then a theorem of $iP4$ because of the following reasoning:

   $iP4 \vdash \Box C \to \boxdot C$

   $\Rightarrow iP4 \vdash \boxdot C \to s_i$ (tautology)

   $\Rightarrow iP4 \vdash \Box C \to s_i$ (from the above two)

   $\Rightarrow iP4 \vdash s'_i$
3. if there are s_j and $s_k (j, k < i)$ such that $s_j \equiv s_k \to s_i$, then $s'_j \equiv \Box C \to (s_k \to s_i) \equiv ((\Box C \to s_k) \to (\Box C \to s_i)) \equiv s'_k \to s'_i$;

[10]Here we will use an equivalent definition of proof in a logic T: We call a finite sequence s is a proof of A if, for any s_i, either s_i is a theorem of T, or there are s_j and $s_k(j, k < i)$ such that $s_j \equiv s_k \to s_i$ or there are C, D and $s_j (j < i)$ such that $s_j \equiv C \to D$ and $s_i \equiv C \rhd D$. We use this dense form in our above proof. Absolutely we can make the above proof more precise. But the above dense presentation is better because it makes us see the woods instead of the trees.

4. if there are B, D and $s_j (j < i)$ such that $s_j \equiv B \to D$ and $s_i \equiv B \rhd D$, then $s'_j \equiv \Box C \to (B \to D)$ and $s'_i \equiv \Box C \to (B \rhd D)$. Now we show that as a matter of fact we can get s'_i from s'_j in $iP4$. The following is the argument:

   $iP4 \vdash s'_j$

   $\Rightarrow iP4 \vdash \Box C \to (B \to D)$

   $\Rightarrow iP4 \vdash \Box\Box C \to \Box(B \to D)$ (Necessitation and distribution)

   $\Rightarrow iP4 \vdash \Box C \to \Box(B \to D)$ (by the fact that $iP4 \vdash \Box C \to \Box\Box C$)

   $\Rightarrow iP4 \vdash \Box C \to (B \rhd D)$ (by the fact that $iP4 \vdash \Box(B \to D) \to (B \rhd D)$)

It is easy to see that in $iP4$ there is a finite sequence $s'_{11} \cdots s'_{1i_1} (\equiv s'_1) s'_{21} \cdots s'_{2i_2} (\equiv s'_2) \cdots s'_{n1} \cdots s'_{ni_n} (\equiv s'_n \equiv (\Box C \to A))$ of proof of $\Box C \to A$. Conclude: there exist $C_1, C_2, \cdots, C_n$ such that $iP4 \vdash \Box((\Box C_1 \to C_1) \rhd C_1) \wedge \cdots \wedge \Box((\Box C_n \to C_n) \rhd C_n) \to A$.

QED

Lemma 3.2.9 *For ANY C* [11]*, if $iP4 \vdash \Box C \to (\Box A \to \Box B)$, then $iP4 \vdash \Box C \to (A \rhd B)$.*

Proof. Suppose that $iP4 \nvdash \Box C \to (A \rhd B)$. Then by the completeness of $iP4$, there is a finite gathering model M and a world w_0 in this model such that $M, w_0 \not\models \Box C \to (A \rhd B)$. There is a world w_1 above w_0 such that $w_0 \leq w_1$ and $w_1 \models \Box C$ but $w_1 \not\models A \rhd B$. It follows that there is a world w_2 such that $w_1 R w_2$ and $w_2 \models A$ but $w_2 \not\models B$. Of course $w_2 \models C$. Now we construct a new gathering model M':

1. $W' := \{w'_1\} \cup [w_2)$(including w_2) where w'_1 is a new world;
2. $R' := \{(w'_1, w_2)\} \cup R \upharpoonright_{[w_2)}$;
3. $\leq' := \{(w'_1, w'_1)\} \cup \leq \upharpoonright_{[w_2)}$
4. $v'(p) := v(p)$ for all propositional variables.

Fact 1: for any $x \in [w_2)$ and for any formula B, $M, x \models B$ iff $M', x \models B$.

Fact 2: $\leq' \circ R' \subseteq R'$.

It is easy to check that the new model is on a gathering frame and that $w'_1 \models \Box C, w'_1 \models \Box A$ but $w'_1 \not\models \Box B$, which also implies that $w'_1 \models \Box C$ but $w'_1 \not\models (\Box A \to \Box B)$. We are done.

QED

Theorem 3.2.10 *$iPL \vdash \Box A \to \Box B$ iff $iPL \vdash A \rhd B$*

Proof. The PROOF is just an immediate application of the above two Lemmas 3.2.8 and 3.2.9. Assume that $iPL \vdash \Box A \to \Box B$. Then, according the Lemma 3.2.8, there are some instances of L: $(\Box C_1 \to C_1) \rhd C_1, \cdots, (\Box C_n \to C_n) \rhd C_n$ (C denote the conjunction of them) such that $iP4 \vdash \Box C \to (\Box A \to \Box B)$. By Lemma 3.2.9, we get that $iP4 \vdash \Box C \to (A \rhd B)$. This implies, according to Lemma 3.2.8, that $iPL \vdash A \rhd B$.

QED

Corollary 3.2.11 *iLLe is equivalent to the logic iL with the extra rule DR.*

[11] Here C is quantified and has nothing to do with our specified C above and below.

Proof. By the above theorem, DR is admissible in $iLLe$ because it is the $L_\Box$ fragment of iPL (Theorem 3.2.3). QED

Remark: The above lemmas 3.2.7 and 3.2.8 are unique for iPL because Lemma 3.2.4 is unique for iPL while the following lemma is quite general for any finitely axiomatizable normal logics $iP4X$[12] containing $iP4$:

Lemma 3.2.12 *(Detour Lemma 2)$iP4X \vdash A$ iff there are some instances of X-principles: $X_1, \cdots, X_n$ such that $iP4 \vdash \boxdot(\bigwedge X_i) \to A$.*

Proof. The proof of this lemma is very similar to and in fact simpler than the above lemma 3.2.8. QED

As we can easily see, Lemma 3.2.12 is more general than lemma 3.2.8. But for iPL and maybe many others satisfying lemma 3.2.4, lemma 3.2.8 is stronger than lemma 3.2.12. The following lemma is also very general[13] .

Lemma 3.2.13 *For any formula A, B, C_i and $D_i (i \leq n)$, if $iP4 \vdash \bigwedge \boxdot(C_i \rhd D_i) \to (\Box A \to \Box B)$, then $iP4 \vdash \bigwedge \boxdot(C_i \rhd D_i) \to (A \rhd B)$.*

Proof. Here we will only treat the case when $n = 1$. In notation, we will omit the subscripts in the C_i's and D_i's. As usual, we will use a semantical proof here. Assume that $iP4 \nvdash \boxdot(C \rhd D) \to (A \rhd B)$. Then by completeness of $iP4$, there is a gathering model M and a world $w_0 \in W$ such that:
$w_0 \not\models \boxdot(C \rhd D) \to (A \rhd B)$
That is to say, there is a world w_1 above w_0 such that
$w_0 \leq w_1, w_1 \models \boxdot(C \rhd D)$ but $w_1 \not\models (A \rhd B)$.

It follows that there is a successor w_2 of w_1 such that $w_2 \models A$ but $w_2 \not\models B$. Now we construct a new model M' as follows:

1. $W' := \{w_1'\} \cup [w_2)$ where w_1' is a new world;
2. $R' := \{(w_1', w_2)\} \cup R \restriction_{[w_2)}$;
3. $\leq' := \{(w_1', w_1')\} \cup \leq \restriction_{[w_2)}$;
4. $v'(p) := v(p)$ for any propositional variable.

Observation 1: For any $x \in [w_2)$ and for any formula B, $M, x \models B$ iff $M', x \models B$.

Observation 2: $\langle W', R' \rangle$ is a frame which is gathering. We only need to check the cases involving the new world w_1'.

Observation 3: For any $x \in [w_2)$, $M', x \models \boxdot A$ because $M, x \models \boxdot A$, the proof of which is just an application of Observations 1 and 2.

Observation 4: $M', w_1' \models \Box A$, $M', w_1' \not\models \Box B$ and $M', w_1' \models \Box(C \rhd D)$. For w_2 is a successor (for $M', w_1' \not\models \Box B$) and the only successor (for $M', w_1' \models \Box A$ and $M', w_1' \models \Box(C \rhd D)$) of w_1', and

[12] Including $iP4$ because we can treat 4 as an axiom schema instead of a result following from Löb principle.

[13] It is less general than the above lemma 3.2.12 though because we require that X_i's are in the form of $C_i \rhd D_i$ for some C_i and D_i here. And if C in lemma 3.2.8 is in the form of preservation say L, then we can deduce the following lemma 3.2.13 from lemma 3.2.8.

$M', w_2 \models A$, $M', w_2 \models C \rhd D$ but $M', w_2 \not\models B$ (by observation 1).

After the four observations, it suffices to show that $M', w_1' \models C \rhd D$. Since w_1' has only one successor w_2, we only need to consider w_2. Assume that $M', w_2 \models C$. By observation 1, $M, w_2 \models C$. Now we argue in the original model M. As we saw a moment ago, $M, w_1 \models \boxdot(C \rhd D)$. It follows that $M, w_1 \models (C \rhd D)$. So $M, w_2 \models D$. For $w_1 R w_2$ and, as we shown, $M, w_2 \models C$. Applying Observation 1 again, we get that $M', w_2 \models D$. Conclude: $M', w_1' \models C \rhd D$. Moreover, $M', w_1' \models \boxdot(C \rhd D)$ (by Observation 4).

It follows immediately from the above results that $M', w_1' \not\models \boxdot(C \rhd D) \to (\Box A \to \Box B)$. According to the soundness of $iP4$, that is to say, $iP4 \nvdash \boxdot(C \rhd D) \to (\Box A \to \Box B)$.

QED

It is tempting to generalize the above Lemma 3.2.13 to the following:
for any formula A, B and $C_i (i \leq n)$, if $iP4 \vdash \bigwedge \boxdot C_i \to (\Box A \to \Box B)$, then $iP4 \vdash \bigwedge \boxdot C_i \to (A \rhd B)$. But this can't be true. It suffices to show that by giving a gathering countermodel where $\boxdot(\Box p \to \Box q) \to p \rhd q$ is not valid because it is obvious that $\vdash_{iP4} \boxdot(\Box p \to \Box q) \to (\Box p \to \Box q)$. The countermodel is as follows:

1. $W := \{w, v, u\}$;
2. $R := \{(w, v), (w, u)\}$;
3. $\leq := \{(w, w), (v, v), (u, u)\}$;
4. $V(p) := \{v\}; V(q) := \{u\}$.

It is easy to check that M is on a gathering frame, $w \models \boxdot(\Box p \to \Box q)$ and $w \not\models p \rhd q$ because $wRv, v \models p$ but $v \not\models q$.

That iPL is closed under $\Box A \to \Box B / A \rhd B$ will follow from the above two lemmas immediately.

Theorem 3.2.14 *iPL is closed under $\Box A \to \Box B / A \rhd B$.*

Proof. Assume that $iPL \vdash \Box A \to \Box B$. Then, according the second detour lemma, there are some instances of L: $(\Box C_1 \to C_1) \rhd C_1, \cdots, (\Box C_n \to C_n) \rhd C_n$ (C denotes the conjunction of them) such that $iP4 \vdash \boxdot C \to (\Box A \to \Box B)$. By lemma 3.2.13, we get that $iP4 \vdash \boxdot C \to (A \rhd B)$. This implies, according to lemma 3.2.8, that $iPL \vdash A \rhd B$.

QED

In fact we can get more: For any finitely axiomatizable logic $iP4X$ containing $iP4$ and X is is in the form of $C \rhd D$ for some C and D, if $iP4X \vdash \Box A \to \Box B$, then $iP4X \vdash A \rhd B$. For example, iPL, $iP4V_n$ and $iPLV_n$ are covered by this general result.

In fact, we can show the admissibility of DR in $iLLe$ without that of $\Box A \to \Box B / A \rhd B$ in iPL. The proof strategy here is similar to the above though. First we give a similar detor lemma:

Lemma 3.2.15 *For any formula A in $L_\Box$, $\vdash_{iLLe} A$ iff $\vdash_{iLe} \boxdot C \to A$ where C is the conjunction of some instances of Löb's principle.*

Proof. Here we only mention that, for any instance C of Löb's provability principle, $\vdash_{iLLe} \Box C \leftrightarrow \boxdot C$.

QED

Theorem 3.2.16 *DR is admissible in $iLLe$.*

Proof. $\vdash_{iLLe} \Box A \to \Box B$
$\Rightarrow \vdash_{iLe} \Box C \to (\Box A \to \Box B)$ (by the above detor lemma)
$\Rightarrow \vdash_{iLe} \Box(C \wedge A) \to \Box B$
$\Rightarrow \vdash_{iLe} \Box((C \wedge A) \vee D) \to \Box(B \vee D)$ (by the admissibility of DR in iLe)
$\Rightarrow \vdash_{iLe} \Box(C \vee D) \to (\Box(A \vee D) \to \Box(B \vee D))$
$\Rightarrow \vdash_{iLe} \Box C \to (\Box(A \vee D) \to \Box(B \vee D))$
$\Rightarrow \vdash_{iLLe} \Box(A \vee D) \to \Box(B \vee D)$ (by the above detor lemma).

QED

3.3 Conservation of iPM over iK

3.3.1 iPM is Complete with respect to a class of finite frames.

Lemma 3.3.1 *In any canonical model: $v\bar{R} \subseteq w\bar{R}$ iff $w_\Box \subseteq v_\Box$ where $x_\Box$ denotes the set of boxed formulas in x.*

Lemma 3.3.2 [14] *(i) The principle Mp corresponds to the Mp property:*
$\forall wvu(wR \leq u \to \exists x(wRx \wedge v \leq x \leq u \wedge x\bar{R} \subseteq u\bar{R})$.
(ii) The logic iPM is canonical.

In the following section we will show that iPM is complete with respect to the class of finite Mp frames. But before we begin to prove this statement, we need to define some important notions to the proof. As we know, the reason why the proof in Proposition 4.5.2 in Iemhoff [2001] did not go through for the finite case is that $\Box A \to B$ does not necessarily belong to the adequate set X if the adequate set is finite. If we had required that the adequate set meet the closure condition that $\Box A, B \in X$ implies $\Box A \to B \in X$, then X would explode[15] So, if we have a mechanism to control the tendency to explode, then we will have the hope that the proof in the thesis will go through. In the following, we will show that this is manageable.

Definition 3.3.3 Given any finite set X_0 that is closed under subformulas and contains $\top$ and $\bot$, we enumerate all the boxed formulas in $X_0 : \Box E_1, \Box E_2, \cdots, \Box E_n$, which are called *atomic boxed formulas*. A formula A is called an *M-boxed formula* if it is a conjunction of atomic boxed formulas in which any two atomic conjuncts are different. Let X_1 be the union of X_0 and the set of M-boxed formulas. It is easy to see that X_1 is closed under subformulas and contains $\top$ and $\bot$. X_1 can be divided into two sets: the set X_1^b of M-boxed formulas and the set X_1^n of formulas which are not M-boxed formulas. Define $X_2 := \{E \to B | E \in X_1^b, B \in X_1\}$ and $X := X_2 \cup X_1$. Obviously X is closed under subformulas and contains $\top$ and $\bot$. We call it *adequate.*

$\lhd$

It is easy to see that the set of M-boxed formulas is finite because it contains less than 2^n elements. If we assume the size of X_0 is m, then X_1 contains at most $2^n + m$ elements. Obviously the size of X_2 is less than $(2^n + m) \times 2^n$. So X is finite. We define the iPM-canonical X-model as usual.

One observation: If $\Box A \in X$, then $\Box A$ is an atomic boxed formula.

[14] proposition 4.5.2 in Iemhoff [2001]

[15] This is actually not really true. By Diego's theorem a finite set of propositional letters generates a finite number of equivalent formulas if one only uses $\to$. But this number is superexponential (in Hendriks [1996]) and we prefer to stay in smaller ranges.

Lemma 3.3.4 *The iPM-canonical X-model satisfies the Mp-property.*

Proof. Take $w, v, u \in X$. Assume that $wRv \subseteq u$. It is easy to see that the following property is a $*$-extendible w-successor property:

$*(x)$: for all $B \in X$, if $w \vdash x \rhd B$, then $B \in u$.

If we can show that $*(v \cup u_\Box)$, then there is a node $x \in X$ such that $wRx, v \subseteq x \subseteq u$ and $x\bar{R} \subseteq u\bar{R}$. So it remains to show that $*(v \cup u_\Box)$, i.e.

if $w \vdash v, u_\Box \rhd B$, then $B \in u$.

Assume that $w \vdash v, u_\Box \rhd B$. We may assume that $u_\Box$ consists of $\Box C_1, \cdots, \Box C_s$, all of which are atomic boxed formulas according to the above observation and are different from each other. We divide the proof into two cases on whether B is in X_1 or in X_2:

1. Let $B \in X_1$.

   $w \vdash (v, \Box C_1, \cdots, \Box C_s) \rhd B$

   $w \vdash v \rhd ((\Box C_1 \wedge \cdots \wedge \Box C_s) \rightarrow B)$ (by Mp)

   Since $(\Box C_1 \wedge \cdots \wedge \Box C_s) \in X_1^b$ and $B \in X_1$, $(\Box C_1 \wedge \cdots \wedge \Box C_s) \rightarrow B \in X_2$. Of course it is in X. By the definition of accessibility relation on the canonical model and the assumption that wRv, $(\Box C_1 \wedge \cdots \wedge \Box C_s) \rightarrow B \in v$. Of course,$(\Box C_1 \wedge \cdots \wedge \Box C_s) \rightarrow B \in u$. Since $\Box C_1, \cdots, \Box C_s$ are also in u, $u \vdash B$. $B \in u$ because u is saturated and $B \in X$.

2. Let $B \in X_2$ and $B \equiv C_b \rightarrow C_1$ where $C_b \in X_1^b$ and $C_1 \in X_1$. Define F to be the conjunction of all the atomic boxed formulas that occur in either $\Box C_1 \wedge \cdots \wedge \Box C_s$ or C_b in which any two conjuncts are different. It is easy to see that $F \in X_1^b$.

   $w \vdash (v, \Box C_1, \cdots, \Box C_s) \rhd B$

   $w \vdash v \rhd ((\Box C_1 \wedge \cdots \wedge \Box C_s) \rightarrow B)$ (by Mp)

   $w \vdash v \rhd ((\Box C_1 \wedge \cdots \wedge \Box C_s) \rightarrow (C_b \rightarrow C_1))$

   $w \vdash v \rhd (F \rightarrow C_1)$

   Since $F \in X_1^b$ and $C_1 \in X_1, F \rightarrow C_1 \in X_2$. Therefore $F \rightarrow C_1 \in X$. It follows that $F \rightarrow C_1 \in v$. Because $v \subseteq u$, it implies that $u \vdash (\Box C_1 \wedge \cdots \wedge C_s) \rightarrow (C_b \rightarrow C_1)$. Since $\Box C_1, \cdots, \Box C_s \in u$, $u \vdash C_b \rightarrow C_1$. Then we get that $C_b \rightarrow C_1 \in u$ because $C_b \rightarrow C_1 \in X_2$ and hence $C_b \rightarrow C_1 \in X$. So $B \in u$.

Conclude: $*(v \cup u_\Box)$.

QED

Theorem 3.3.5 *$\vdash_{iPM} A$ iff A is valid on all finite Mp frames.*

Proof. The left-to-right direction follows from the correspondence result. Now we just need to consider the other direction. Suppose that $\nvdash_{iPM} A$. First define Y_0 consisting of all sub-formulas of A, $\top$ and $\bot$. By using the same construction prescribed in the definition 3.3.3, we get a finite adequate set Y. It is easy to see that there is a saturated set w such that $A \notin w$ (by lemma 2.1.8). Then we construct the iPM-canonical Y model M as usual. As the above lemma shows, M satisfies the Mp property.

It remains to show the truth lemma. Since the accessibility relation on the canonical models is defined as usual, we can borrow the proof of truth lemma from the completeness proof of iP (Th. 2.1.10). It follows that $M, w \not\models A$. So A is not valid on all finite Mp frames.

QED

3.3.2 iK is the $L_\Box$-fragment of iPM

We can prove that iK is the $L_\Box$-fragment of iPM directly by the following lemma 3.3.20. But why should we kill the goose laying golden eggs? We choose to move into the destination slowly but surely.

It is not difficult to give an intuitive proof of the admissibility of rules in iPM. However it is not easy to give a precise proof of the admissibility. Probably the reason for that lies in the fact that the Mp property is an existence property. In order to manage it, we have to define some notions to help with our formalization of the proof[16].

Definition 3.3.6 A triple (w, v, u) in a frame is called a *problem* if it satisfies $wRv \leq u$. It is called an *unsolved problem* if it additionally satisfies:
there is no x such that $wRx, v \leq x \leq u$ and $x\bar{R} \subseteq u\bar{R}$.
Such an x is called a *solution* to the above problem. If such an x exists, then the problem is called *a solved problem.*

Let (w, v, u) and (w, v', u) be two problems. If $v \leq v'$, then we denote $(w, v, u) \preceq (w, v', u)$ and say (w, v, u) is *below* (w, v', u). If $v < v'$, then we denote $(w, v, u) \prec (w, v', u)$.

$\lhd$

Lemma 3.3.7 *If $x \leq y$, then $y\bar{R} \subseteq x\bar{R}$. Therefore, if $x \leq y$ and $x\bar{R} \subseteq y\bar{R}$, then $y\bar{R} = x\bar{R}$.*

Proof. This follows immediately from the fact that $(\leq \circ R) \subseteq R$.

QED

Lemma 3.3.8 *Let (w, v, u) and (w, v', u) be two problems. If $(w, v, u) \preceq (w, v', u)$ and x is a solution to the problem (w, v', u), then x is also a solution to the first problem (w, v, u).*

Proof. This follows immediately from the definition of solutions. QED

Corollary 3.3.9 *Let (w, v, u) and (w, v', u) be two problems. Suppose that $(w, v, u) \preceq (w, v', u)$. For any x, y, if both (x, v, y) and (x, v', y) are problems, then each solution to (x, v', y) is also a solution to (x, v, y).*

Definition 3.3.10 A problem (w, v, u) is called *a dispensable problem* if there is another different problem such that $(w, v, u) \prec (w, v', u)$. A problem is called *indispensable* if it is not dispensable.

$\lhd$

Definition 3.3.11 $< [x] := \{z | x < z\}$. A point w is called a *minimal point* in X if there is no point v such that $v < w$. $min[x]$ denotes the set of minimal points in the set $< [x]$.

$\lhd$

One observation: any problem (w, v, v) is a solved problem.

Lemma 3.3.12 *Let F be a frame. If there is no unsolved problem, then F satisfies the Mp property.*

Lemma 3.3.13 *$iPM \vdash A \rhd B$ iff $iPM \vdash (\Box A \to \Box B)$.*

[16] Also see the following lemma Theorem 3.4.12.

Proof. One direction is very clear. Now we just need to show the other direction. Assume that $iPM \nvdash A \rhd B$. Then, by the completeness of iPM (lemma 3.3.5), $A \rhd B$ is falsified at some point w_0 of some model M on some finite frame satisfying the Mp-property. There is a v_0 such that $w_0Rv_0, M, v_0 \models A$ but $M, v_0 \not\models B$. In the following we will construct a new Mp-model M' satisfying the requirement that $M' \not\models (\Box A \to \Box B)$.

One observation: if a problem (x, y, z) is inside $[v_0)$, i.e $x, y, z \in [v_0)$, then it is easy to see that there is a solution to this problem in $[v_0)$. This means that we don't need to consider the problems in $[v_0)$ because all of them are solved in $[v_0)$.

First we define $W_0 := [v_0) \cup \{w\}, R_0 := R\restriction_{[v_0)} \cup \{(w, v_0)\}; \leq_0 := \leq\restriction_{[v_0)}$ where w is a new world. Then enumerate all the elements in $min[v_0]$[17]: $u_0^1, u_0^2, \cdots, u_0^{j_0}$. For the unsolved problem (w, v_0, u_0^1), we add a new world x_1^1. After that we define a new frame $F_1^1 = \langle W_1^1, R_1^1, \leq_1^1\rangle$ where $W_1^1 := W_0 \cup \{x_1^1\}; R_1^1 := R_0 \cup \{(w, x_1^1)\} \cup \{(x_1^1, z) | (u_0^1, z) \in R_0\}; \leq_1^1 := \leq_0 \cup \{(v_0, x_1^1), (x_1^1, u_0^1)\} \cup \{(x_1^1, z) | (u_0^1, z) \in \leq_0\}$. It is easy to see that $x_1^1\bar{R} = u_0^1\bar{R}$ and that we can define a new valuation V_1^1 on the frame to guarantee persistence for all propositional letters. We can just define: $x_1^1 \in V_1^1(p)$ iff $v_0 \in V_1^1(p)$ for all propositional letters. It follows that $M_1^1, x_1^1 \models A$.

Now we give some observations here. Since we add a new point, we add new problems. It seems that we increase the number of the unsolved problems. But this is not the case. The new problems can be divided into following two cases:

1. Two solved problems (w, v_0, x_1^1) and (w, x_1^1, u_0^1). The solution to the first problem is x_1^1. The solution to the second one is also x_1^1. We no longer consider them because we only care about the unsolved and indispensable problems.
2. The problems (w, x_1^1, z) with $z \in< [u_0^1]$. It is easy to see that, for each problem (w, v_0, z) with $z \in< [u_0^1]$, there is a problem (w, x_1^1, z) with $z \in< [u_0^1]$ such that $(w, v_0, z) \preceq (w, x_1^1, z)$. In fact the number of the problems (w, x_1^1, z) with $z \in< [u_0^1]$ is the same as that of problems (w, v_0, z) with $z \in< [u_0^1]$. But, by a previous lemma 3.3.9, a solution to (w, x_1^1, z) is also a solution to (w, v_0, z). This implies that we only need to consider the problems (w, x_1^1, z) where $z \in< [u_0^1]$. This in turn implies that the number of the unsolved and indispensable problems, which we need to consider, decrease by 1 because we have solved the problem (w, v_0, u_0^1).

We can similarly solve problems (w, v_0, u_0^k) by adding a new world x_1^k such that wRx_1^k and $v \leq x_1^k \leq u_0^k$. Afterwards define a new frame according to the same procedure as that we have prescribed for x_1^1. We will repeat this argument until there is no unsolved and indispensable problem.

Why will the procedure terminate in a finite number of steps? The reason lies in two facts. The first one is that, since the frame is finite, the number of the unsolved and indispensable problems is finite. The second is that every time when we add a new world, we have solved at least one unsolved problem that we need to consider, which implies that the number of the unsolved and indispensable problems will decrease every time. Finally we will reach the stage where there is no unsolved problem. So, by the above lemma 3.3.12, the frame F' that we have achieved at that time satisfies the Mp property.

One observation: for any formula C and any world $y \in [v_0)$, $M, y \models C$ iff $M', y \models C$.

We define a valuation V' on the new frame F' to guarantee persistence for all propositional letters. Then it follows that, for all the new worlds y, $M', y \models A$. Since, except for v_0, all successors of w are

[17] The purpose of choosing minimal points is to just to make the construction more efficient.

new points, $M', w \models \Box A$. $M', w \not\models \Box B$ because $M, v_0 \not\models B$ (and hence $M', v_0 \not\models B$) and $wR'v_0$. So, $\Box A \to \Box B$ is not valid on all finite Mp frames. According to the soundness of iPM, $\nvdash_{iPM} \Box A \to \Box B$.

QED

Since the above method is quite general, we can extract a very powerful theorem from the proof, which turns out to be handy when we deal with the admissibility of many other rules in iPM (See the later theorems 3.4.8 and 3.4.12).

Theorem 3.3.14 *(Extension Theorem for iPM) Let $M = \langle W, R, \leq, V\rangle$ be a finite Mp model and $[v)$ be any generated submodel by v. Then there is a new finite Mp model $N = \langle W', R', \leq', V'\rangle$ such that*

1. *$[v) \subseteq W'$;*
2. *For any world $x \in [v)$ and any formula E, $M, x \models E$ iff $N, x \models E$;*
3. *There is a world $w \in W'$ such that*
   $wR'v$ and
   If $wR'y$ and $M, v \models A$, then $N, y \models A$.
4. *If $M, v \models A$ and $M, v \not\models B$, then there is a world $w' \in W'$ such that $w'R'v$ and $N, w' \not\models \Box A \to \Box B$.*

Theorem 3.3.15 *The logic iK with extra rules MoR and DR is contained in the $L_\Box$-fragment of iPM (call it $iPM_\Box$).*

Proof. It is trivial that iK is contained in $iPM_\Box$. By the same argument as that in Theorem 3.1.11, it is easy to see that DR is admissible in $iPM_\Box$. Now we just need to show that MoR is also admissible in $iPM_\Box$. The proof is as follows:

$iPM_\Box \vdash \Box A \to \Box B$
$\Rightarrow iPM \vdash \Box A \to \Box B$ (trivial)
$\Rightarrow iPM \vdash A \rhd B$ (by the above lemma)
$\Rightarrow iPM \vdash (\Box C \to A) \rhd (\Box C \to B)$ (Mp-principle)
$\Rightarrow iPM \vdash \Box(\Box C \to A) \to \Box(\Box C \to B)$ (by lemma 2.3.1)
$\Rightarrow iPM_\Box \vdash \Box(\Box C \to A) \to \Box(\Box C \to B)$ (from the definition of $iPM_\Box$).

QED

In order to show: the $L_\Box$ fragment of iPM is the logic iK with the extra rules MoR and DR, according to our usual analysis we only need to show the following conjecture:

$iK_{MD} \vdash A$ iff A is valid on all finite Mp frames.

PROOF of this should be divided into two directions. The left to right is evident. For the right to left, it is the main part of the proof. Suppose that the conjecture is true. Now we will show that the $L_\Box$ fragment of iPM is the logic iK with the extra rules MoR and DR. Assume $iK_{MD} \nvdash A$ for any formula A in $L_\Box$. Then it is falsified at some point of some finite M_p frame. It follows from the completeness of iPM that $iPM \nvdash A$. So, for any A in $L_\Box$, $iPM \vdash A$ implies $iK_{MD} \vdash A$. According to Theorem 3.3.15 that $iK_{MD} \vdash A$ implies $iPM \vdash A$, we “prove” the proposition:

the $L_\Box$ fragment of iPM is the logic iK with the extra rules MoR and DR.

But the problem is that the logic iK_{MD} is NOT a normal modal logic. We can not work with the usual correspondence theory and semantical proof method. Instead we will show this syntactically.

Theorem 3.3.16 *iP is conservative over iK with respect to formulas in $L_\Box$.*

Proof. We can come up with TWO proofs of this theorem: one is semantical one, the other syntactic. The semantical proof is just an application of the completeness and soundness of iP and iK. For the syntactic proof, you can find it in the proof of the following lemma 3.3.20.

QED

Lemma 3.3.17 *$iP \vdash \Box A \to \Box B$ iff $iP \vdash A \rhd B$.*

Proof. The right-to-left direction is trivial. Now we just show the other direction. Assume that $iP \nvdash A \rhd B$. By the completeness result about iP, there are a finite model M and a world w in the model such that $M, w \not\models A \rhd B$. This means that there is a R-successor v of w such that $M, v \models A$ but $M, w \not\models B$. Take $W_0 := \{w' | M, w' \models A \text{ and } wRw'\}$. It is easy to see that $W_0 \neq \emptyset$ because $v \in W_0$. Denote $\bigcup\{[w'] | w' \in W_0\}$ as W_1. Now we define a new model on the original one. Here w_0 is a new world. Define $W' := \bigcup\{[w'] | w' \in W_0\} \cup \{w_0\}$, $R' := \{(w_0, w') | w' \in W_0\} \bigcup R \restriction_{W_1}$, $\leq' := \leq \restriction_{W_1} \bigcup\{(w_0, w_0)\}$. And $V'(p) := V(p)$ for all propositional variable p.

Fact 1: For any $w' \in W_1$ and for any formula B, $M', w' \models B \Leftrightarrow M, w' \models B$. Since any R'-successor w' of w_0 is in W_0 and hence in W_1, $M', w_0 \models \Box A$.

Fact 2: $M', v \models A$ but $M', v \not\models B$. Therefore $M', w_0 \not\models \Box B$.

Fact 3: $\leq' \circ R' \subseteq R'$;

It follows immediately from the above two facts and another fact that $(w_0, w_0) \in \leq'$ that $M', w_0 \not\models \Box A \to \Box B$. By the soundness of iP, we get that $iP \nvdash \Box A \to \Box B$.

QED

Corollary 3.3.18 *iK is closed under DR.*

Proof. The proof pattern here is the same as before.

$iK \vdash \Box A \to \Box B$.
$\Rightarrow iP \vdash \Box A \to \Box B$ (iK is contained in iP)
$\Rightarrow iP \vdash A \rhd B$ (by above lemma)
$\Rightarrow iP \vdash (A \vee C) \rhd (B \vee C)$ (Dp)
$\Rightarrow iP \vdash \Box(A \vee C) \to \Box(B \vee C)$
$\Rightarrow iK \vdash \Box(A \vee C) \to \Box(B \vee C)$ (by Theorem 3.3.16)

QED

Definition 3.3.19 The translation $*$ [18] from formulas in $L_{\rhd}$ to those in $L_{\Box}$ is inductively defined as follows:

1. For p, $\top$and $\bot$, $p^* = p$, $\top^* = \top$ and $\bot^* = \bot$.
2. For $\circ \in \{\vee, \wedge, \to\}$, $(A \circ B)^* = A^* \circ B^*$.
3. $(\neg A)^* = \neg A^*$
4. $(A \rhd B)^* = \Box(A^* \to B^*)$.

$\lhd$

[18] As discerning eyes can recognize immediately, the translation here is similar to the well-known Godël-McKinsey-Tarski translation by which they showed that IPC can be embedded into $S4$. It looks great!

It is easy to see that after the above lemma we can reduce the original question to

Is the $L_\Box$ fragment of iPM the logic iK with extra rule and MoR (denote it iK_M)?

and that we only need to show the following proposition:

if $iPM \vdash A$, then $iK_M \vdash A$ for any formula A in $L_\Box$.

But we can do better than that:

Lemma 3.3.20 *If $iPM \vdash A$, then $iK \vdash A$ for any formula A in $L_\Box$* [19].

Proof. Assume that $iPM \vdash A$. Of course we can write A in $L_\rhd$ according to the definition $\Box A \equiv (\top \rhd A)$. It suffices to show that

$iPM \vdash A \Rightarrow iK \vdash A^*$. (*)

Please note the following two facts:

Fact 1: For any formula B in $L_\Box$, $B^* = B$. This also justifies our definition of the translation (*).

Fact 2: if we can show the above proposition, then we show that $iPM \vdash A \Rightarrow iK \vdash A$ for any formula A in $L_\Box$

We will prove (*) by induction on the finite sequence of the proof of A. Since $iPM \vdash A$, there is a finite sequence $s_1 s_2 \cdots s_n = A$ of formulas in $L_\rhd$ such that s_i is either

1. in the forms of K_p, P_1, P_2, D_p or M_p.
2. Or $s_i = A_1 \rhd A_2$ and $s_j = A_1 \to A_2$ for some $A_1, A_2, s_j \in L_\Box$ and $j < i$
3. Or for some $s_j, s_k (j, k < i)$, $s_k = s_j \to s_i$.

Claim: In fact $s_1^* s_2^* \cdots s_n^* = A^*$ of formulas in $L_\Box$ is a proof of A^*. We only need to check that any $s_i^* (i \leq n)$ is either

1. a theorem of iK (!!)
2. Or $s_i^* = \Box A_1^* \to A_2^*$ and $s_j^* = A_1^* \to A_2^*$ for some $A_1^*, A_2^*, s_j^* \in L_\Box$ and $j < i$ (Application of the Necessitation Rule)
3. Or for some $s_j^*, s_k^* (j, k < i)$, $s_k^* = s_j^* \to s_i^*$. (Application of Mp)

We prove by cases:

1. If s_i is an instance of K_p, P_1, P_2, D_p or M_p, then it is easy to see that s_i^* is a theorem of iK. For example $(Mp)^*$ is derivable in iK.
2. If $s_i = A_1 \rhd A_2$ and $s_j = A_1 \to A_2$ for some $A_1, A_2, s_j \in L_\Box$ and $j < i$, then $s_i^* = \Box(A_1^* \to A_2^*)$ and $s_j^* = A_1^* \to A_2^*$, which is what we need.
3. If for some $s_j, s_k (j, k < i)$, $s_k = s_j \to s_i$, then $s_k^* = s_j^* \to s_i^*$, which satisfies the condition for the third case.

QED

[19]The method here is quite general: if we can add a principle X to iP and X^* is a theorem in iK, we can always get that the $L_\Box$-fragment of iPX is iK. According to this result, IP is obviously conservative over iK with respect to formulas in $L_\Box$.

Theorem 3.3.21 *iK is the $L_\Box$-fragment of iPM.*

Proof. The proof is just a combination of lemma 3.3.20 and the fact that $iK \subseteq iP \subseteq iPM$.

QED

To our joy, we get a corollary from the above results:

Corollary 3.3.22 *MoR is admissible in iK*

Proof. MoR is admissible in the $L_\Box$ fragment of iPM (Theorem 3.3.15). And iK is the the $L_\Box$ fragment of iPM (Theorem 3.3.21).

QED

But if we want to just show DR and MR are admissible in iK, we can show these more quickly by using the following lemma:

Lemma 3.3.23 $\vdash_{iK} \Box A \to \Box B$ *iff* $\vdash_{iK} \Box(A \to B)$; $\vdash_{iP} \Box A \to \Box B$ *iff* $\vdash_{iP} \Box(A \to B)$

Proof. We can show these just by the completeness and soundness of iK and iP. QED

Probably it is helpful to note that this rule is stronger than the rule $\Box A \to \Box B / A \rhd B$. Let's first show the admissibility of DR in iK. Reason inside iK:

$\Box A \to \Box B$
$\Box(A \to B)$
$(A \to B) \wedge (A \vee C) \to (B \vee C)$
$\Box(A \vee C) \to \Box(B \vee C)$

Next we show that iK is also closed under MR. Reason inside iK:

$\Box A \to \Box B$
$\Box(A \to B)$
$(\Box C \to A) \wedge (A \to B) \to (\Box C \to B)$
$\Box(\Box C \to A) \to \Box(\Box C \to B)$.

3.4 Admissible Rules: $\Box A/A$, $\Box A \to \Box B/ \boxdot A \to B$ and others

In the following paragraphs, we will consider the reflection rule: $\Box A/A$.

Theorem 3.4.1 *IP is closed under the rule $\Box A/A$*

Proof. It is easy to show by using completeness and soundness of iP. QED

Theorem 3.4.2 *$iP4$ is closed under the rule $\Box A/A$*

Proof. Suppose that $iP4 \nvdash A$. Then by the completeness of $iP4$, there is a gathering model M and a world $w \in W$ such that $M, w \not\models A$. Now we make a new model M' based on the original one M:

1. $W' := \{w', v'\} \cup [w)$;
2. $R' := \{(w', v')\} \cup R \restriction_{[w)} \cup \{(v', y) | (w, x) \in R\}$;
3. $\leq' := \{(v', w), (w', w'), (v', v')\} \cup \leq \restriction_{[w)} \cup \{(v', x) | (w, x) \in \leq\}$;;

4. $v'(p) := v(p)$ for any propositional letter p.

Observation 1: (W', R') is a gathering frame;

Observation 2: For any $x \in [w)$ and for any formula B, $M', x \models B \Leftrightarrow M, x \models B$.

Since $M, w \not\models A$, $M', w \not\models A$. It follows from the fact that $v' \leq' w$ that $M', v' \not\models A$. This implies that $M', w' \not\models \Box A$ because $w'R'v'$.

QED

Lemma 3.4.3 *If $iP4 \vdash \Box(C \rhd D) \to \Box A$, then $iP4 \vdash \boxdot(C \rhd D) \to A$.*

Proof. We use the same strategy as before. Suppose that $iP4 \nvdash \boxdot(C \rhd D) \to A$. Then by the completeness of $iP4$, there is a gathering model M and a world $w_0 \in W$ such that $M, w_0 \not\models \boxdot(C \rhd D) \to A$. That is to say, there is w_1 such that
$w_0 \leq w_1$, $M, w_1 \models \boxdot(C \rhd D)$ but $M, w_1 \not\models A$.

We will build a new model M' as follows:

1. $W' := \{w', w_0'\} \cup [w_1)$;
2. $R' := \{(w', w_0')\} \cup R\restriction_{[w_1)} \cup \{(w_0', y) | (w_1, y) \in R\}$;
3. $\leq' := \{(w_0', w_1), (w', w'), (w_0', w_0')\} \cup \leq\restriction_{[w_1)} \cup \{(w_0', x) | (w_1, x) \in \leq\}$;
4. $v'(p) := v(p)$ for any propositional letter p.

Observation 1: $M', w_1 \models \boxdot(C \rhd D)$ and $M', w_1 \not\models A$.

Observation 2: (W', R') is gathering.

Observation 3: $M', w' \not\models \Box A$;

Claim 4: $M', w_0' \models \Box(C \rhd D)$. Take any x such that $w_0'R'x$. The it follows that w_1Rx. By observation 1, we get that $M, x \models \boxdot(C \rhd D)$ and hence $M, x \models C \rhd D$. This implies that $M', x \models C \rhd D$. So $M', w_0' \models \Box(C \rhd D)$.

Claim 5: $M', w_0' \models (C \rhd D)$. Take any x such that $w_0'R'x$ and $M', x \models C$. Then it follows that $x \in [w_1)$, w_1Rx and hence $M, x \models C$. Moreover, since $M, w_1 \models C \rhd D$, $M, x \models D$. Equivalently, $M', x \models D$. So, $M', w_0' \models (C \rhd D)$.

By the two claims:$M', w_0' \models \boxdot(C \rhd D)$. It follows that $M', w' \models \Box(C \rhd D)$. By the soundness of $iP4$, $iP4 \nvdash \Box(C \rhd D) \to \Box A$.

Similarly, we can get the following applicable general form:
If $iP4 \vdash \bigwedge_i \Box(C \rhd D) \to \Box A$, then $iP4 \vdash \bigwedge_i \boxdot(C \rhd D) \to A$.

QED

Corollary 3.4.4 *If $iP4 \vdash \boxdot((\Box C_i \to C_i) \rhd C_i) \to \Box A$, then $iP4 \vdash \boxdot((\Box C_i \to C_i) \rhd C_i) \to A$. Similarly, $iP4 \vdash \bigwedge_i \boxdot((\Box C_i \to C_i) \rhd C_i) \to \Box A$, then $iP4 \vdash \bigwedge_i \boxdot((\Box C_i \to C_i) \rhd C_i) \to A$*

Proof. The proof is just a combination of lemma 3.2.4 and the above lemma 3.4.3.

QED

Theorem 3.4.5 *iPL is closed under the rule $\Box A/A$.*

Proof. The proof is a just a combination of lemma 3.4.2 and the above corollary.

QED

Corollary 3.4.6 *Both of $iLLe, iLe$ are closed under the rule $\Box A/A$.*

Theorem 3.4.7 *$\Box A/A$ is admissible in both iL and $iK4$*

Proof. The proof is the same as the above semantical proof.

QED

Theorem 3.4.8 *iPM is closed under the reflection rule.*

Proof. Suppose that $\nvdash_{iPM} A$. The by the completeness of iPM, there are a finite Mp model M and a world v in this world such that $M, v \not\models A$. In other words, $M, v \models \top$ and $M, v \not\models A$. Then by Extension Theorem for iPM, there are a new Mp model $N = \langle W', R', \leq', V' \rangle$ and a world w in this new model such that $[v) \subseteq W', wR'v$ and $M, w \not\models \Box\top \to \Box A$, which is equivalent to that $M, w \not\models \Box A$. By the soundness of iPM, $\nvdash_{iPM} \Box A$.

QED

Corollary 3.4.9 *iK is closed under the reflection rule.*

Proof. It follows from the conservation of iPM over iK.

QED

Now we consider another rule: $\Box A \to \Box B / \boxdot A \to B$. Of course it is easy to see:

Theorem 3.4.10 *iP is closed under the rule $\Box A \to \Box B / \boxdot A \to B$.*

However, it is NOT admissible in $iP4$ or iPL. For $iP4$, it suffices to show that $iP4 \vdash \Box(A \vee B) \to \Box(A \vee \Box B)$ while $iP4 \nvdash \boxdot(A \vee B) \to (A \vee \Box B)$. To see this, we use the following model M:

1. $W := \{w, v, u\}$
2. $R := \{(v, u), (w, u)\}$;
3. $\leq := \{(w, w), (v, v), (u, u), (w, v)\}$;
4. $v(p) := \{u, \}, v(q) := \{v\}$.

It is easy to see the following list of items: $u \not\models q, (\Rightarrow) v \not\models \Box q, (\Rightarrow) v \not\models (p \vee \Box q); u \models p, (\Rightarrow) u \models p \vee q, (\Rightarrow) v \models \Box(p \vee q), v \models p \vee q, (\Rightarrow) v \models \boxdot(p \vee q)$. So $w \not\models \boxdot(p \vee q) \to (p \vee \Box q)$. Conclude: $iP4$ is NOT closed under the rule: $\Box A \to \Box B / \boxdot A \to B$.

We can use the same example to show that iPL is not closed under the rule $\Box A \to \Box B / \boxdot A \to B$ either. For Le is derivable in iPL, the above model M is gathering and conversely well-founded, and B is valid on all gathering conversely well-founded frames IF $iPL \vdash B$ for any formula B. The argument is also applicable for $iPLM$. So the rule is NOT admissible for $iPLM$ either.

Corollary 3.4.11 *Neither $iLLe$ nor iLe is closed under the rule $\Box A \to \Box B/ \boxdot A \to B$ while iK and iP are.*

There is another way to show the admissibility of $\Box A \to \Box B/ \boxdot A \to B$ in iP and iK. Here we illustrate the proof of the admissibility for iP. Reason inside iP:
$\Box A \to \Box B$
$\Box(A \to B)$ (by lemma 3.3.23)
$A \to B$ (by admissibility of reflection rule in iP (lemma 3.4.1))
$\boxdot A \to B$.

Theorem 3.4.12 *$\Box A \to \Box B/ \boxdot A \to B$ is admissible in iPM.*

Proof. The proof is another application of the Extension Theorem for iPM. Suppose that $\nvdash_{iPM}$ $\boxdot A \to B$. By the completeness of iPM, there are a Mp model M and a world v in this world such that $M, v \models A \wedge \Box A$ but $M, v \not\models B$. By the Extension Theorem for iPM, there are a new Mp model $N = \langle W', R', \leq', V' \rangle$ and a world $w \in W'$ such that $[v) \subseteq W', wR'v$ and $N, w \not\models \Box A \to \Box B$. By the soundness of iPM, $\nvdash_{iPM} \Box A \to \Box B$.

QED

Similarly to what we have shown for $iP4$ in Lemma 3.1.3, we will show a very handy lemma for $iK4$.

Lemma 3.4.13 *If M is on a finite semi-transitive brilliant frame and $M, v \models \boxdot A$, then, for any $x \in [v)$, $M, x \models \boxdot A$.*

The following theorem is an immediate application of the above lemma.

Theorem 3.4.14 *$\Box A \to \Box B/ \boxdot A \to B$ is admissible in $iK4$ and iL.*

It is natural to ask: what conditions should we impose on A and B to make the rule admissible in $iP4$ and iPL. The answer is that A is of the form of $C \rhd D$ for some C and D (Lemma 3.4.3 and the following Theorem 3.4.17).

Theorem 3.4.15 *If $iP4 \vdash \Box \bigwedge_i (C_i \rhd D_i) \to \Box B$, then $iPL \vdash \bigwedge_i \boxdot (C_i \rhd D_i) \to B$*

Proof. The strategy is similar to that in lemma 3.4.3. QED

Theorem 3.4.16 *If $iP4 \vdash \boxdot((\Box C_i \to C_i) \rhd C_i) \to (\Box(D_1 \rhd D_2) \to \Box A)$, then $iP4 \vdash \boxdot((\Box C_i \to C_i) \rhd C_i) \to (\boxdot(D_1 \rhd D_2) \to A)$; More generally, if $iP4 \vdash \boxdot \bigwedge_i ((\Box C_i \to C_i) \rhd C_i) \to (\Box(D_1 \rhd D_2) \to \Box A)$, then $iP4 \vdash \bigwedge_i \boxdot((\Box C_i \to C_i) \rhd C_i) \to (\boxdot(D_1 \rhd D_2) \to A)$.*

Proof. The proof of the special case to general case is semantical not syntactical. QED

Theorem 3.4.17 *iPL is closed under the rule: $\Box(C \rhd D) \to \Box B/ \boxdot (C \rhd D) \to B$.*

Proof. The proof is just an immediate application of the above theorem and lemma 3.2.12.

QED

For the disjunction property, we have the following theorem:

Theorem 3.4.18 *For any $L \in \{iP, iP4, iPL, iPM, iK, iLe, iLLe\}$, if $L \vdash A \vee B$, then $L \vdash A$ or $L \vdash B$.*

Proof. The regular method. Maybe we should mention iPM. Suppose that neither $iPM \nvdash A$ nor $iPM \nvdash B$. Then there are two Mp models M_1 and M_2 and two worlds $w_1 \in W_I$ and $w_2 \in W_2$ such that $M_1, w_1 \not\models A$ and $M_2, w_2 \not\models B$. The only thing that we want to say is that we add a reflexive new world w.

QED

We briefly mention the results about the rule $\Box A \to \Box B / \Box(A \to B)$.

Theorem 3.4.19 *None of $iP4, iK4, iPL, iLe, iLLe$ and iL is closed under $\Box A \to \Box B / \Box(A \to B)$ while iK, iP and iPM are.*

Proof. It is easy to see that the principle 4 is derivable in all of $iP4, iK4, iPL, iLe, iLLe$ and iL. Suppose that the rule is admissible in these logics. It follows that $\Box(A \to \Box A)$ is derivable in all of them, which looks very doubtful. By the fact that all of these logics are closed under the reflection rule, then $A \to \Box A$ is admissible in all of them. But it is very easy for us to give a counterexample for that. So, the rule $\Box A \to \Box B / \Box(A \to B)$ is admissible in none of the above mentioned logics.

As for the proof the admissibility of this rule in iPM, it is again an application of the Extension Theorem for iPM. We have proved the admissibility for the remaining two logics before.

QED

There are some other interesting propositions about $iP4$ and iPM:

Theorem 3.4.20 *If $iP4 \vdash (C \rhd D) \rhd B$, then $iP4 \vdash (C \rhd D) \to B$. As a corollary, if $iP4 \vdash \Box A \rhd B$, then $iP4 \vdash \Box A \to B$.*

Theorem 3.4.21 *iPM is also closed under the rule $\Box A \rhd B / \Box A \to B$. Moreover, if $iPM \vdash \Box A \rhd \Box B$, then $iPM \vdash A \rhd B$.*

Proof. We can prove the first part through the fact that iPM is closed under the reflection rule. Suppose that there are some A and B such that $iPM \vdash \Box A \rhd B$ but $iPM \nvdash \Box A \to B$. By Mp, we get that $iPM \vdash (\Box A \to \Box A) \rhd (\Box A \to B)$. Then $iPM \vdash \Box(\Box A \to B)$. So, by the Theorem 3.4.8, $iPM \vdash \Box A \to B$, which contradicts our assumption that $iPM \nvdash \Box A \to B$.

The proof of the second part is similar to that for the admissability of $\Box A \to \Box B / A \rhd B$.[20]

QED

3.5 W and iPW

3.5.1 Correspondence of W

As we have mentioned, the usual way for completeness like the proof method in the proof of completeness for iL breaks down for iPL[21]. It seems that Lp alone gives us too little for its completeness. With Mp, we can indeed make the usual method go through. But Mp as well as Lp will give us too much. After several rounds of selection, W looks like the best candidate to make the proof go through. In the following we will give a correspondence result of W and show that iPW is properly contained in $iPLM$.

[20] In fact we can deduce the admissability of $\Box A \to \Box B / A \rhd B$ in iPM from this theorem.
[21] See remarks on Page 68 of Iemhoff [2001].

Lemma 3.5.1 *Let F be a finite frame. $F \models (A \wedge \Box B) \rhd B \rightarrow A \rhd B$ iff F is gathering, conversely well-founded and satisfies the following property:*
$(*) : \forall wvu(wRvRu \rightarrow \exists x(wRx \wedge v < x \leq u))$ *(please note that $v < x$ not $v \leq x$).*

Proof. First the right-to-left direction. Prove by reduction to absurdity. Assume that F is gathering, conversely well-founded and satisfies the (*) property, and some instance $(A \wedge \Box B) \rhd B \rightarrow A \rhd B$ of the principle W is not valid in F. Then there is a model $M := (W, R, \leq, \models)$ on F and two worlds w', w in W such that $w' \leq w$ and $w \models (A \wedge \Box B) \rhd B$ but $w \not\models A \rhd B$. It follows that there is a world $v \in W$ such that wRv and $v \models A$ but $v \not\models B$. Therefore $v \not\models \Box B$ because otherwise $v \models B$. Then there is another world u such that $vRu, u \models A$ but $u \not\models B$.

By (*), there is a world x_1 such that $wRx_1, v < x_1 \leq u$. It is easy to see that $x_1 \models A$ but $x_1 \not\models B$. If $x_1 \models \Box B$, then $x_1 \models B$. Then we get a contradiction. We may assume that $x_1 \not\models \Box B$. Then there is a world u_1 such that $x_1 R u_1$ and $u_1 \not\models B$. Since F is gathering, $x_1 \leq u_1$ and hence $u_1 \models A$. By (*), there is a world $x_2 \cdots$. Now we encounter the same situation as above. Since $v < x_1 < x_2 \cdots$ and F is finite, we will finish this sequence after some finite number of circles. That is to say, there is a natural number n such that $wRx_n, x_{n-1} < x_n \leq u_{n-1}, x_n \models A, x_n \not\models B$ and, most importantly, $x_n \models \Box B$. It follows that $x_n \models B$. We arrive at a contradiction.

Now the other direction. Assume that the principle W is valid in F. It is easy to see that F has to be gathering and conversely-well-founded by the correspondence result of Lp and the fact that Lp is derivable in iPW (lemma 2.3.8). Suppose that F does NOT satisfy the (*) property, i.e.
$(**) : \exists wvu(wRvRu \wedge \forall x(wRx \wedge v < x \rightarrow x \not\leq u))$.

It suffices to define a model on the frame to falsify $(p \wedge \Box q) \rhd q \rightarrow p \rhd q$. Define $V(p) := \{z | v \leq z\}$ and $V(q) := \{z | z \not\leq u\}$. Obviously, $u \models p$, $u \not\models q, v \models p, v \not\models q$ and $v \not\models \Box q$. Of course $w \not\models p \rhd q$. Take any other successor x of w than v. If $x \models p$, then $v < x$ because $v \neq x$. By (**), we get that $x \not\leq u$, which implies that $x \models q$. In short, for any successor y of w such that $y \models p \wedge \Box q$, $y \models q$. So $w \models (p \wedge \Box q) \rhd q$. Recall that $w \not\models p \rhd q$. Finally we get that $w \not\models (p \wedge \Box q) \rhd q \rightarrow p \rhd q$.

QED

Since the $(*) : \forall wvu(wRvRu \rightarrow \exists x(wRx \wedge v < x \leq u))$ implies irreflexivity and gatheringness, the above correspondence result can be simplified to the following:

Theorem 3.5.2 *Let F be a finite frame. $F \models (A \wedge \Box B) \rhd B \rightarrow A \rhd B$ iff F satisfies the following property:*
$(*) : \forall wvu(wRvRu \rightarrow \exists x(wRx \wedge v < x \leq u))$.

Theorem 3.5.3 *iPW is strictly contained in $iPLM$.*

Proof. In lemma 2.3.7, we have shown that W is derivable in $iPLM$. It suffices to show that $\not\vdash_{iPW} Mp$. The counterexample is as follows:

1. $W := \{w, v, u, x, y\}$
2. $R := \{(w, v), (v, u), (w, x), (x, y), (w, y)\}$
3. $\leq := \{(v, u), (v, x), (x, y), (x, u)\}$
4. $v(p) := \{u, y\}; v(q) := \{y\}, v(r) := \emptyset$

We can see that Mp is falsified in the above model. On the other hand, W is valid on the above frame $< W, R, \leq >$ (it is easy to check that the above frame satisfies the above frame properties). This implies that $iPW \nvdash Mp$.

QED

3.5.2 Conservation of iPW over $iLLe$

Although iPW is a proper extension of iPL (it is easy to give an example to show that), their $L_\Box$ fragments are the same: $iLLe$ or the logic iL with the extra rule DR.

First we define a notion that will be indispensable in our inductive proof on any finite, transitive, gathering and conversely well-founded frame.

Definition 3.5.4 Let $< W, R >$ be a finite, transitive, gathering and conversely well-founded (hence irreflexive) frame. An end point w is a world such that there is no point w' such that wRw' or $w \leq w'$.[22] It is easy to see that for any $w \in W$, there is a finite sequence s of $w_n, w_{n-1}, \cdots, w_0$ such that $w = w_n S_n w_{n-1} \cdots S_1 w_0$ where s_i is either R or $\leq$ and w_0 is an end point. We define the *grade* $g_s(w)$ of w in this sequence inductively as follows:

1. $g_s(w_0) := 0$
2. If $g_s(w_{i-1}) = k$ and S_i is $\leq$, then $g_s(w_i) := k$; If $g_s(w_{i-1}) = k$ and S_i is R, then $g_s(w_i) := k + 1$.

Of course, $g_s(w_i) \leq n$ for any $i \leq n$.

For each $w \in W$, we define the *rank* $r(w)$ of w as the greatest such $g_s(w)$ (we omit the subscript $< W, R >$ here).

$\lhd$

Observation 1: If wRv, then $r(w) > r(v)$.

Observation 2: Let $< W, R >$ be a gathering frame and $w, v \in W$. If wRv and $v \models A$, then, for any $x \in [v)$, $x \models A \wedge \Box A$.

Theorem 3.5.5 *$iPW \vdash A$ only if A is valid on all finite gathering, transitive and conversely well-founded frames.*

Proof. It suffices to show that W is valid on all finite gathering, transitive and conversely-well-founded frame. Given a model M on such a frame $< W, R >$ and any $w', w, v \in W$ such that $w' \leq wRv$, assume that $M, w \models (A \wedge \Box B) \rhd B$ and $M, v \models A$. We need to show that $M, v \models B$. It suffices to show that $M, v \models \Box B$. Suppose that this is NOT the case: $M, v \not\models \Box B$.

Now we consider the v generated submodel M'. Obviously, $M', v \not\models \Box B, M', v \models A$ and M' is on a finite transitive,gathering and conversely well-founded frame. Then there is a world $v' \in W'$ with the least rank such that $M', v' \not\models \Box B$. This implies that, for any $v'' \in W'$ such that $v'Rv''$, $M', v'' \models \Box B$ and hence $M, v'' \models \Box B$. Such a v'' can always be found because at least every end point makes boxed formulas true there. It is easy to check that wRv'' by transitivity and that $M', v'' \models A$ (and hence $M, v'' \models A$) according to the fact that $M', v \models A$ and $v'' \in [v)$. Since $M, w \models (A \wedge \Box B) \rhd B$, $M, v'' \models B$ and hence $M', v'' \models B$. So $M', v' \models \Box B$. We have arrived at a contradiction. Conclude: $M, v \models \Box B$ and hence $(A \wedge \Box B) \rhd B$ is valid on any finite gathering, transitive and conversely-well-founded frame.

QED

[22]Please note that the definition here is a little different from our ordinary definition of end points.

But the converse of this lemma is NOT true. On one hand, it is easy to check that $(A \rhd B) \to \Box(A \rhd B)$ is valid on all transitive frames. On the other hand, it is well-known that this formula is NOT arithmetically valid in HA. Suppose that the converse were true. Then, according to the converse proposition, $iPW \vdash (A \rhd B) \to \Box(A \rhd B)$, which is impossible because W is a valid principle in HA while $(A \rhd B) \to \Box(A \rhd B)$ is NOT. So the converse can NOT be true.

Lemma 3.5.6 $\vdash_{iLLe} A$ *iff* A *is valid on all finite gathering* transitive *and conversely well-founded frames.*

Proof. In fact the lemma is not new, just an extension of a previous lemma 3.2.1. We only need to show that transitivity is preserved in the new model $N = \langle W, R', \leq, V \rangle$.

Assume that $w, v, u \in W$ and $wR'vR'u$. Then we need to show that $wR'u$. Since $wR'vR'u$, $wRvRu$ and hence wRu because R is transitive. It remains to show that, for any z such that $uR'z$, $u \leq z$. This immediately follows from the assumption that $vR'u$. So $wR'u$.

QED

Theorem 3.5.7 *iLLe or iL with the extra rule DR is the* $L_\Box$ *fragment of* iPW*.*

Proof. Since both L and Le are derivable in iPL and iPW is an proper extension of iPL, $iLLe$ is contained in the $L_\Box$-fragment of iPW. Now for the other direction. Suppose that $\nvdash_{iLLe} A$. By the completeness of $iLLe$ (Theorem 3.1.5), A is NOT valid on al finite transitive gathering and conversely well-founded frames. It follows from the above lemma 3.5.5 that $\nvdash_{iPW} A$.

Conclude: iPW is conservative over $iLLe$ with respect to formulas in $L_\Box$.

QED

Chapter 4

Fixed Point Theorems

As Gauss once put it, mathematics is the queen of the sciences and Arithmetic is the queen of mathematics. No sentence can be more appropriate than this one to describe the fixed point theorem in provability logic: fixed point is the diamond in the crown of provability logic. "The beautiful fixed point theorem (or de Jongh-Sambin Theorem) for GL is the most striking application of modal logic to the study of the concept of provability in formal systems."[1]. Here we will show the fixed point theorems for iL and iPL. It turns out that the proof of fixed point for iL is similar to that for GL; the proof of fixed point for iPL is similar to that for IL [2], the basic interpretability logic (de Jongh-Visser [1991]). In the last section, we will discuss the interderivability between fixed points and Beth definability (Definition 4.3.2) in the settings of both intuitionistic provability and preservativity logics.

A notational convention: AB is the result of substitution of B for p in the formula Ap. First we will lay (or borrow) a framework[3] for our following fixed point iPL. Let SR_0 be iL in the propositional modal logic extended with a binary operator $\bullet$ with the following extra principles:

$E1 :\vdash \Box(A \leftrightarrow B) \rightarrow (A \bullet C \leftrightarrow B \bullet C)$

$E2 :\vdash \Box(A \leftrightarrow B) \rightarrow (C \bullet A \leftrightarrow C \bullet B)$.

We can derive the following propositions immediately from SR_0

$S_1 :\vdash B \leftrightarrow C$ implies $\vdash AB \rightarrow AC$.

$S_2 :\vdash \boxdot(A \leftrightarrow B) \rightarrow (FA \leftrightarrow FB)$.

S_3 : Suppose that Ap is modalized in p, then $\vdash \Box(A \leftrightarrow B) \rightarrow (FA \leftrightarrow FB)$.

LR : Let B be a conjunction of formulas in the form of $\Box C$ or $\boxdot C$, then $\vdash B \rightarrow (\Box A \rightarrow A)$ implies $\vdash B \rightarrow A$. The proof of LR is just an application of Lob principle and the principle 4. We have shown the second substitution lemma (lemmas 2.3.3.and 2.4.1) for both iL and iPL. It is also easy to show the other two substitution lemmas by induction. Obviously, all the above four propositions also hold for iL.

Theorem 4.0.8 *(Uniqueness Theorem) Suppose that p occurs modalized in A, then* $\vdash_L (\boxdot(p \leftrightarrow Ap) \wedge$

[1] Page 104 of Boolos [1993]. It is very interesting to note the following relation chain from Sciences to the fixed point theorem:

Sciences⇒Mathematics⇒Arithmetic⇒Provability Logic⇒Fixed Point.

The first two arrows are based on Gauss's famous sentence. The last one is on Boolos's remarks. And the third one is on the very mathematical nature of Provability Logic.

[2] It is not iL.

[3] See Smorynski[1985] and de Jongh-Visser [1991]

$\boxdot(q \leftrightarrow Aq)) \rightarrow (p \leftrightarrow q)$ *where* $L \in \{SR_0, iL\}$.

Proof. Reason inside L:
$\Box(p \leftrightarrow q)$
$Ap \leftrightarrow Aq$ (by the substitution lemma $S3$)
$\boxdot(p \leftrightarrow Ap) \wedge \boxdot(q \leftrightarrow Aq) \rightarrow (\Box(p \leftrightarrow q) \rightarrow (p \leftrightarrow q))$.
$(\boxdot(p \leftrightarrow Ap) \wedge \boxdot(q \leftrightarrow Aq)) \rightarrow (p \leftrightarrow q)$ (by LR)
So if $\vdash_L C \leftrightarrow AC$ and $\vdash_L D \leftrightarrow AD$, then $\vdash_L \boxdot(C \leftrightarrow AC)$ and $\vdash_L \boxdot(D \leftrightarrow AD)$,and hence $\vdash_L C \leftrightarrow D$.
QED

In the inductive proof of the fixed point theorem for iPL, we will use the following theorem:[4]

Theorem 4.0.9 *Let U be any extension of SR_0 satisfying:*
FIX: Every formula Ap of the form $\Box Bp$ or $Bp \bullet Cp$ has a fixed point.
For every formula Ap with p modalized, there is a formula J such that p does not occur in J and $\vdash_U J \leftrightarrow AJ$.

Proof. The only essential thing that we should mention is that the proof of Theorem 2.4 in de Jongh-Visser [1991] did not use classical logic, which means that it is applicable here too. QED

4.1 Fixed Point Theorem for iL(iK with Löb's Principle)

Theorem 4.1.1 *(A Special Case)* $iL \vdash \Box A(\top) \leftrightarrow \Box A(\Box A(\top))$ *for all formulas A.*

Proof. (Syntactic Proof). $iL \vdash \Box A(\top) \rightarrow (\top \leftrightarrow \Box A(\top))$ (IPC)
$iL \vdash \Box A(\top) \rightarrow \boxdot(\top \leftrightarrow \Box A(\top))$ (4 is derivable in iL and the following rule is admissible in $iK4$: $\vdash \Box A \rightarrow B \Rightarrow \vdash \Box A \rightarrow \Box B$)
$iL \vdash \Box A(\top) \rightarrow (\Box A(\top) \leftrightarrow \Box A(\Box A(\top))$ (by substitution lemma shown above)
$iL \vdash \Box A(\top) \rightarrow \Box A(\Box A(\top))$ (IPC)

Now for the other direction. $iL \vdash \Box A(\top) \rightarrow \boxdot(\top \leftrightarrow \Box A(\top))$ (from above)
$iL \vdash \Box A(\top) \rightarrow (A(\top) \leftrightarrow A(\Box A(\top)))$. (substitution lemma)
$iL \vdash \Box A(\top) \rightarrow (A(\top) \leftarrow A(\Box A(\top)))$. (IPC)
$iL \vdash A(\Box A(\top)) \rightarrow (\Box A(\top) \rightarrow A(\top))$. (IPC)
$iL \vdash \Box A(\Box A(\top)) \rightarrow \Box(\Box A(\top) \rightarrow A(\top))$. (iK)
$iL \vdash \Box A(\Box A(\top)) \rightarrow \Box A(\top)$. (Löb's axiom).
QED

Corollary 4.1.2 *Let $A(p) := B[\Box C(p)]$. Then $iL \vdash AB(\top) \leftrightarrow A(AB(\top))$.*

Proof. By the above theorem, (Reason inside iL):
$\Box C(B(\top)) \leftrightarrow \Box CB(\Box CB(\top))$
$B(\Box C(B(\top))) \leftrightarrow B(\Box CB(\Box CB(\top)))$ (substitution S_1)
$A(B(\top)) \leftrightarrow A(A(B(\top)))$.
QED

In order to give a inductive proof of the following fixed point theorem for the general case we need a notion:

[4] See Theorem 2.4 in de Jongh-Visser [1991].

Definition 4.1.3 A formula φ is *k-decomposable* if for some sequence $B_1, \cdots, B_k$ of propositional letters, some formula $\xi(B_1, \cdots, B_k)$ not containing A but containing all $B_1, \cdots, B_k$ and some distinct formulas $\varphi_1(A), \cdots, \varphi_k(A)$, each containing A, $\varphi = \xi(\Box\varphi_1(A), \cdots, \Box\varphi_k(A))$. The smallest such a k is called the *decomposition degree* of φ.

$\lhd$

Theorem 4.1.4 *If in $\varphi(A, B_1, \cdots, B_n)$ the propositional letter A occurs exclusively under $\Box$, then there is a formula $\delta(B_1, \cdots, B_n)$ such that $iL \vdash \delta \leftrightarrow \varphi(\delta, B_1, \cdots, B_n)$.*[5]

Proof. There are many versions of proofs available in literature for this induction proof. For the best one, we refer to the inductive proof for GL in Page 69 of de Jongh and Veltman [1995]. For the sake of completeness we will repeat it here. Because the variables $B_1, \cdots, B_n$ plays no essential role in our following proof but only complicates the notations below, we will omit all of them.

One observation: Since φ is modalized in A, it is k-decomposable for some k. Prove by induction on the decomposition degree k of φ. It is easy to see that the proposition is true for the base case when $k = 0$. For φ is a fixed point of φ then.

Next we show the case when $k = n$. Assume that $\varphi(A)$ is of the form of $\xi(\Box\varphi_1(A), \Box\varphi_2(A), \cdots, \Box\varphi_n(A))$. Consider the formula $\varphi(A_1, A_2) = \xi(\Box\varphi_1(A_1), \Box\varphi_2(A_2), \cdots, \Box\varphi_n(A_2))$. Applying Corollary 4.1.2 with respect to A_1, we get that there is formula $\delta_1(A_2)$ of the form $\xi(\Box\varphi_1(\xi(\top, \Box\varphi_2(A_2), \cdots, \Box\varphi_n(A_2)), \Box\varphi_2(A_2)), \cdots, \Box\varphi_n(A$ such that

(*)$\vdash_{iL} \delta_1(A_2) \leftrightarrow \xi(\Box\varphi_1(\delta_1(A_2)), \Box\varphi_2(A_2), \cdots, \Box\varphi_n(A_2))$.

Since the formula $\xi(\Box\varphi_1(\xi(\top, \Box\varphi_2(A_2), \cdots, \Box\varphi_n(A_2)), \Box\varphi_2(A_2), \cdots, \Box\varphi_n(A_2))$ is modalized in A_2 and it is in fact of the form of $\xi'(\Box\varphi_2(A_2), \cdots, \Box\varphi_n(A_2))$ of decomposition degree $\leq n-1$ for some ξ', we can apply I.H. with respect to A_2. Then there is formula δ such that

$\vdash_{iL} \delta \leftrightarrow \delta_1(\delta)$.

If we substitute δ for A_2 in $*$, we will get that

(**)$\vdash_{iL} \delta_1(\delta) \leftrightarrow \xi(\Box\varphi_1(\delta_1(\delta)), \Box\varphi_2(\delta), \cdots, \Box\varphi_n(\delta))$.

Moreover, since $\vdash_{iL} \delta \leftrightarrow \delta_1(\delta)$, we are allowed to apply the first substitution lemma to (**) and get that

$\vdash_{iL} \delta \leftrightarrow \xi(\Box\varphi_1(\delta), \Box\varphi_2(\delta), \cdots, \Box\varphi_n(\delta))$,

which means that δ is a fixed point of $\varphi(A)$. Conclude:there is a formula δ such that $iL \vdash \delta \leftrightarrow \varphi(\delta)$.

QED

4.2 Fixed Point Theorem for iPL

The following proof is similar to the one for interpretability logic in de Jongh-Visser [1991]. To put it more precisely, the fixed point for the formula $A(p) \rhd B(p)$ in iPL is a kind of mirror image of that for the formula $A(p) \rhd_i B(p)$ in IL. To see this, we can just look at the following Theorem 4.2.3, which is in fact the crucial part of the following proof if we try to follow the steps in de Jongh-Visser [1991].

Define: $A \equiv B :\Leftrightarrow \vdash_{iPL} (A \rhd B) \wedge (B \rhd A)$.

Lemma 4.2.1 *$A \equiv A \wedge \Box A \equiv \Box A \rightarrow A$.*

[5]This is actually a form of Theorem 4.0.9.

Proof. As a matter of fact, we only need the first equivalence in our following proof. All the following reasoning is inside iPL. It is easy to see that $A \equiv A \wedge \Box A$.
$A \rhd \Box A$ and $A \rhd A$
$A \rhd \boxdot A$.
The other direction is clear. Next it is easy to see that
$A \equiv \Box A \to A$.

QED

Lemma 4.2.2 *If* $\vdash \Box B\top \to C$, *then* $\vdash B\top \wedge \Box B\top \leftrightarrow BC \wedge \Box BC$.

Proof. For the left-to-right direction we reason as follows:
$\vdash \Box B\top \to C$,
$\vdash \Box B\top \to (\top \leftrightarrow C)$,
$\vdash \Box\Box B\top \to \Box(\top \leftrightarrow C)$,
$\vdash \Box B\top \to \boxdot(\top \leftrightarrow C)$ (by the fact that 4 is derivable in iPL) ,
$\vdash \Box B\top \to (\boxdot B\top \leftrightarrow \boxdot BC)$ (substitution lemma 2.3.3)
$\vdash \boxdot B\top \to \boxdot BC$

Now for the other direction. Reason inside iPL:
$\Box B\top \to \boxdot(\top \leftrightarrow C)$ (*)
$\Box(\Box B\top \to \boxdot(\top \leftrightarrow C))$
$BC \wedge \Box BC \wedge \Box(\Box B\top \to \boxdot(\top \leftrightarrow C)) \to (BC \wedge \Box(\Box B\top \to B\top))$
$BC \wedge \Box BC \wedge \Box(\Box B\top \to \boxdot(\top \leftrightarrow C)) \to (BC \wedge \Box B\top))$ (by the Löb principle)
$BC \wedge \Box BC \to (BC \wedge \Box B\top))$
$BC \wedge \Box BC \to \boxdot(C \leftrightarrow \top)$ (by (*))
$(BC \wedge \Box BC) \to (B\top \wedge \Box B\top)$ (substitution lemma)

QED

Theorem 4.2.3 *If* $\vdash \Box B\top \to C$, *then* $\vdash B\top \equiv BC$.

Proof. It follows immediately from lemmas 4.2.1 and 4.2.2. QED

Corollary 4.2.4 $\vdash B\top \equiv B(A\Box B\top \rhd B\top)$

Proof. Since $\vdash A\Box B\top \rhd \top$, $\vdash \Box B\top \to (A\Box B\top \rhd B\top)$. It follows from the above theorem that $\vdash B\top \equiv B(A\Box B\top \rhd B\top)$. QED

Lemma 4.2.5 $\vdash \Box A\Box B\top \to (A\Box B\top \rhd B\top \leftrightarrow \Box B\top)$

Proof. For the left to right direction, the argument is just $\vdash (\top \rhd A\Box B\top) \wedge (A\Box B\top \rhd B\top) \to (\top \rhd B\top)$. For the right to left direction, it is just the following argument: $\vdash (\top \rhd B\top) \wedge (A\Box B\top \rhd \top) \to (A\Box B\top \rhd B\top)$ QED

Theorem 4.2.6 $\vdash \Box A\Box B\top \to \boxdot(A\Box B\top \rhd B\top \leftrightarrow \Box B\top)$

Theorem 4.2.7 $\vdash A\Box B\top \wedge \Box A\Box B\top \leftrightarrow A(A\Box B\top \rhd B\top) \wedge \Box A(A\Box B\top \rhd B\top)$

Proof. The left to right direction:
$\vdash \Box A\Box B\top \to \boxdot(A\Box B\top \rhd B\top \leftrightarrow \Box B\top)$ (from above theorem 4.2.6)
$\vdash \boxdot(A\Box B\top \rhd B\top \leftrightarrow \Box B\top) \to (A\Box B\top \wedge \Box A\Box B\top \leftrightarrow A(A\Box B\top \rhd B\top) \wedge \Box A(A\Box B\top \rhd B\top))$ (substitution lemma)
$\vdash A\Box B\top \wedge \Box A\Box B\top \to A(A\Box B\top \rhd B\top) \wedge \Box A(A\Box B\top \rhd B\top)$

Now the right to left direction: In fact it suffices to show:
$\vdash \Box A(A\Box B\top \rhd B\top) \to \Box A\Box B\top$ (*)
Because we can then combine (*) and lemma 4.2.3 to get the required result. The argument is similar to that for the right-to-left direction. We reason inside iPL as follows:
$\Box A\Box B\top \to \boxdot(A\Box B\top \rhd B\top \leftrightarrow \Box B\top)$ (theorem 4.2.6)
$A(A\Box B\top \rhd B\top) \to (\Box A\Box B\top \to A\Box B\top)$ (by substitution lemma)
$\Box A(A\Box B\top \rhd B\top) \to \Box(\Box A\Box B\top \to A\Box B\top)$
$\Box A(A\Box B\top \rhd B\top) \to \Box A\Box B\top$ (by Löb's principle)

QED

Lemma 4.2.8 $A\Box B\top \equiv A(A\Box B\top \rhd B\top)$.

Proof.$\vdash A\Box B\top \equiv A\Box B\top \wedge \Box A\Box B\top$ (by lemma 4.2.1). $\vdash A\Box B\top \wedge \Box A\Box B\top \equiv A(A\Box B\top \rhd B\top) \wedge \Box A(A\Box B\top \rhd B\top)$ (by theorem 4.2.7). And $\vdash A(A\Box B\top \rhd B\top) \equiv A(A\Box B\top \rhd B\top) \wedge \Box(A(A\Box B\top \rhd B\top))$

QED

Theorem 4.2.9 *(Fixed Point Theorem for $A(p) \rhd B(p)$)* $\vdash A\Box B\top \rhd B\top \leftrightarrow A(A\Box B\top \rhd B\top) \rhd B(A\Box B\top \rhd B\top)$[6].

Proof. This is just a combination of lemmas 4.2.4 and 4.2.8. QED

For boxed formulas, the proof of fixed point theorem is the same as that for iL.

Theorem 4.2.10 *(A Special Case for Boxed formulas in $L_{\rhd}$)* $\vdash_{iPL} \Box A(\top) \leftrightarrow \Box A(\Box A(\top))$ *for all formulas A in $L_{\rhd}$.*

So we have shown the first part of Explicit Definability Theorem.

Theorem 4.2.11 *(Explicit Definability Theorem: Part 1) Let Ap be either if the form $\Box Bp$ or $Bp \rhd Cp$, then there is a formula J such that $\vdash_{iPL} J \leftrightarrow AJ$.*

We can get a symmetric form of fixed point for formulas $Ap \rhd Bp$.

Theorem 4.2.12 $\vdash A\Box B\top \rhd B\top \leftrightarrow A\Box B\top \rhd B\Box B\top$

Proof. Since $\vdash_{iPL} \Box B\top \to \Box B\top$, $B\top \equiv B(\Box B\top)$.

QED

Theorem 4.2.13 *(Explicit Definability Theorem, Part 2) For every formula Ap with p modalized, there is formula J such that: p does not occur in J, and $\vdash_{iPL} J \leftrightarrow AJ$.*

[6] We can also get an interesting dual result to that $\vdash_{iPL} (A\top \rhd B\Box A\top) \leftrightarrow A(B\Box A\top \rhd A\top) \rhd B(B\Box A\top \rhd A\top)$. The proof of this proposition is similar to the above one.

Proof. First it is easy to check that
$\vdash_{iPL} \Box(A \leftrightarrow B) \rightarrow (A \rhd C \leftrightarrow B \rhd C)$
$\vdash_{iPL} \Box(A \leftrightarrow B) \rightarrow (C \rhd A \leftrightarrow C \rhd B)$
So iPL satisfies FIX in the theorem 4.0.9. From this theorem, our theorem here follows immediately.

QED

In iPW, we have a simpler form of fixed point for formulas $Ap \rhd Bp$.

Theorem 4.2.14 *In iPW, the fixed point is $A\top \rhd B\top$.*

Proof. We reason inside iPW as follows:
$\Box B\top \rightarrow (\Box B\top \leftrightarrow \top)$
$\Box B\top \rightarrow \boxdot(\Box B\top \leftrightarrow \top)$
$\Box B\top \rightarrow (A\top \leftrightarrow A\Box B\top)$
In other words, $\vdash A\Box B\top \wedge \Box B\top \leftrightarrow A\top \wedge \Box B\top$
In the following we will show that
$\vdash A\Box B\top \rhd B\top \leftrightarrow (A\Box B\top \wedge \Box B\top) \rhd B\top$ (by W)
$\vdash A\Box B\top \rhd B\top \leftrightarrow (A\top \wedge \Box B\top) \rhd B\top$ (by W)
$\vdash A\Box B\top \rhd B\top \leftrightarrow A\top \rhd B\top$ (by W)

QED

However, in iPL, we can't get such a simpler form. Consider the following formula $p \rhd q$. Suppose that the fixed point for formulas $Ap \rhd Bp$ were $A\top \rhd B\top$. Then $\Box q$ would be the fixed point of $p \rhd q$.i.e. $\vdash_{iPL} (\Box q \rhd q) \leftrightarrow \Box q$. It is easy to see that one direction is correct: $\vdash_{iPL} \Box q \rightarrow (\Box q \rhd q)$. But the other direction is NOT correct in iPL. Consider the following counterexample:

1. $W = \{w, v, u\}$
2. $R := \{(w, v), (v, u)\}$
3. $\leq := \{(w, w), (v, v), (u, u), (v, u)\}$
4. $V(q) := \emptyset$.

It is easy to see that the model $M := \langle W, R, \leq, V \rangle$ is on a finite gathering conversely well-founded frame and that $M, w \not\models (\Box q \rhd q) \rightarrow \Box q$. By the correspondence result of Lp, we get that $\nvdash_{iPL} (\Box q \rhd q) \rightarrow \Box q$.

Actually the fixed point theorem for IL and ILW (de Jongh and Visser [1991]) can be seen as a consequence of Theorem 4.2.13.

Corollary 4.2.15 *For every formula Ap with p modalized, there is formula J such that: p does not occur in J, and $\vdash_{IL} J \leftrightarrow AJ$.*

Proof. Just use the translation as discussed in Section 1.2 and note that the principle Dp has not been used in the above proof. Clearly ILW can be treated similarly.

QED

4.3 Beth Definability and Fixed Points

In the following, we will show for a general class of intuitionistic modal logics two theorems (Theorem 4.3.4 and Theorem 4.3.5)about the interderivability between Beth property (Definition 4.3.2) and fixed point property (Definition 4.3.3), which can be similarly proven to be true for the couterpart of intuitionistic preservativity logics.

Take any modal formula $A(\bar{p}, r)$ in $L_\Box$. When convenient, we omit $\bar{p}$. Define $A_1(r)$ as $A(r)$ with $\top$ substituting for all nonmodalized occurrences of r in $A(r)$. It is easy to see that $A_1(r)$ is modalized in r and, more importantly, $\vdash_{IPC} (r \leftrightarrow r') \to (A(r) \leftrightarrow A(r'))$. The following lemma is about the relation between $A(r)$ and $A_1(r)$.

Lemma 4.3.1 *For any intuitionistic modal logic $\mathcal{T}$,*
$\vdash_T r \wedge A(r) \leftrightarrow r \wedge A_1(r)$.

Proof. Reason inside $\mathcal{T}$:
$r \to (r \leftrightarrow \top)$
$(r \leftrightarrow \top) \to (A(r) \leftrightarrow A_1(r))$ (by the definition of $A_1(r)$)
$r \to (A(r) \leftrightarrow A_1(r))$
$r \wedge A(r) \leftrightarrow r \wedge A_1(r)$

QED

Definition 4.3.2 (Beth Definability Property) A logic $\mathcal{L}$ has *the Beth Property* iff for all formulas $A(\bar{p}, r)$ the following holds:
If $\vdash_L \boxdot A(\bar{p}, r) \wedge \boxdot A(\bar{p}, r') \to (r \leftrightarrow r')$, then there exists a formula $C(\bar{p})$ such that $\vdash_L \boxdot A(\bar{p}, r) \to (C(\bar{p}) \leftrightarrow r)$

$\lhd$

Definition 4.3.3 (Fixed Point Property) A logic $\mathcal{L}$ has the *fixed point property* iff for any formula $A(\bar{p}, r)$ which is modalized in r, there exists a formula $F(\bar{p})$ such that

- (existence) $\vdash_L F(\bar{p}) \leftrightarrow A(\bar{p}, F(\bar{p}))$
- (uniqueness) $\vdash_L \boxdot(r \leftrightarrow A(\bar{p}, r)) \wedge \boxdot(r' \leftrightarrow A(\bar{p}, r')) \to (r \leftrightarrow r')$.

$\lhd$

Theorem 4.3.4 *(From Beth Definability to Fixed Points) Let $\mathcal{L}$ be an intuitionistic modal logic in which*

1. $\vdash_L \Box A \to \Box\Box A$,
2. $\vdash_L \boxdot B \to (\Box A \to A)$ *implies* $\vdash_L \boxdot B \to A$,
3. *the Beth theorem holds.*

Then $\mathcal{L}$ has the fixed point property.

Proof. It is easy to check that the proof for the classical case in Hoogland [2001] is intuitionistically acceptable. QED

Theorem 4.3.5 *(From Fixed Points to Beth Definability) Let $\mathcal{L}$ be an intuitionistic modal logic in which*

1. *$\vdash_L \Box A \to \Box\Box A$,*
2. *$\vdash_L \boxdot B \to (\Box A \to A)$ implies $\vdash_L \boxdot B \to A$,*
3. *the fixed point theorem holds.*

Then $\mathcal{L}$ has the Beth property.

Proof. The proof here is similar to that in Hoogland [2001]. The only difference is that we will not use Maksimova's lemma (see in Hoogland [2001]) to reduce arbitrary formulas to ones that are"largely modalized" but apply the above lemma directly. Still the main proof procedure in Hoogland [2001] goes through here. As a matter of fact, Maksimova's lemma is not intuitionistically acceptable.

Assume that (*): $\vdash_L \boxdot A(r) \wedge \boxdot A(r') \to (r \leftrightarrow r')$. We need to show that there exists a formula C such that $\vdash_L \boxdot A(r) \to (r \leftrightarrow C)$. Take as $A_1(r)$ the one prescribed at the beginning of this section.

Since $A_1(r)$ is modalized in r, there exists a formula F_1 such that
(1) $\vdash_L F_1 \leftrightarrow A_1(F_1)$, and
(1') $\vdash_L \boxdot(A(r) \leftrightarrow r) \wedge \boxdot(A(r') \leftrightarrow r') \to (r \leftrightarrow r')$. Hence
(2) $\vdash_L \boxdot(r \leftrightarrow A(r)) \to (r \leftrightarrow F_1)$.
In the following we will show that the F_1 is the C that we are looking for. First we show the following claim:
(3) $\vdash_L \boxdot A(r) \to [\Box(A_1(r) \to r) \to (A_1(r) \to r)]$.

By the above lemma, we get that $\vdash_L A(r) \to (r \to A_1(r))$ and hence $\vdash_L \Box A(r) \to \Box(r \to A_1(r))$. Therefore
(4) $\vdash_L \boxdot A(r) \wedge \Box(A_1(r) \to r) \to \Box(r \leftrightarrow A_1(r))$.
Let C denote $\boxdot A(r) \wedge \Box(A_1(r) \to r)$. From (2) (the uniqueness of fixed point of A_1), we get that
(5) $\vdash_L \Box(r \leftrightarrow A_1(r)) \to \Box(r \leftrightarrow F_1)$.
Combining (4) and (5), we get that
(6) $\vdash_L C \to \Box(r \leftrightarrow F_1)$.
From (6) follow immediately the following two corollaries:
(7) $\vdash_L C \to \Box(A(F_1))$ (by the definition of C),
(8) $\vdash_L C \to (A_1(r) \to A_1(F_1))$ (by the fact that A_1 is modalized in r).

Since F_1 is a fixed point of A_1,
(9) $\vdash_L C \to [A_1(r) \to F_1]$ and hence
(10)$\vdash_L C \to [A_1(r) \to F_1 \wedge A_1(F_1)]$
According to the above lemma, it follows that
(11) $\vdash_L C \to [A_1(r) \to A(F_1)]$,
which together with (7) implies
(12) $\vdash_L C \to (A_1(r) \to \boxdot A(F_1))$.
According to the assumption (*), we get that
(13) $\vdash_L \boxdot A(r) \wedge \boxdot A(F_1) \to (r \leftrightarrow F_1)$.
Then it follows from the above (12) and (13) that
(14) $\vdash_L C \to (A_1(r) \to (r \leftrightarrow F_1))$.

Finally the claim (3) follows from (9) and (14). With the claim, we can show the main theorem easily. From the claim and the second condition on the logic $\mathcal{L}$,
(15) $\vdash_L \boxdot A(r) \to (A_1(r) \to r)$.
From the above lemma it is evident that $\vdash_L A(r) \to (r \to A_1(r))$ and hence

(16)$\vdash_L \boxdot A(r) \to \boxdot(r \leftrightarrow A_1(r))$.

So, with (2) and (16), we achieve our goal:
(17) $\vdash_L \boxdot A(r) \to (r \leftrightarrow F_1)$. That is to say, F_1 is the C that we are looking for.

QED

We have shown the fixed point theorem for iL and iPL. Since any extension $\mathcal{L}$ of iL or iPL will have the fixed point property, it should also have the Beth property according to the above theorem.

Corollary 4.3.6 *Let $\mathcal{T}$ be an extension of iL or iPL (of course in the appropriate language). Then $\mathcal{T}$ has the Beth property.*

Chapter 5

Conclusion

In this thesis, we have investigated some intuitionistic preservativity and provability logics and their interrelations through the conservation of these preservativity logics over their corresponding provability logics. In fact these conservation results are interweaved with the equivalence results that we have mentioned in the Introduction. First let us summarize what we have achieved in this thesis. The following results are the conservation results that are known so far:

1. iK is the $L_\Box$-fragment of iP (in Iemhoff [2001]);
2. iLe is the $L_\Box$-fragment of $iP4$ (in Iemhoff [2001]);
3. $iLLe$ is the $L_\Box$ fragment of iPL;
4. iK is the $L_\Box$ fragment of iPM;
5. $iLLe$ is also the $L_\Box$ fragment of iPW;

Correspondingly we have some equivalence results:

1. The $L_\Box$ fragment of iP is equivalent to the logic iK with the extra rule DR (Theorem 3.3.16 and Corollary 3.3.18).
2. The $L_\Box$ fragment of $iP4$ is equivalent to the logic $iK4$ with the extra rule DR (Theorem 3.1.11).
3. The $L_\Box$ fragment of iPL is equivalent to the logic iL with the extra rule DR (Theorem 3.2.3 and Corollary 3.2.11).
4. The $L_\Box$ fragment of iPM is equivalent to the logic iK with the extra rules DR and MoR(Theorems 3.3.21 and 3.3.22).
5. The $L_\Box$ fragment of iPW is equivalent to the logic iL with the extra rule DR (Theorem 3.5.7).

In the meanwhile we have shown the admissibility and non-admissibility of two important inference rules in provability logic: $\Box A/A$ and $\Box A \to \Box B/\boxdot A \to B$. Some other rules are also very interesting in preservativity logic on their own.

The fixed point theorem for iL and iPL is independent of the other parts of the thesis. Its significance in preservativity logic can be compared to its in interpretability logic (de Jongh-Visser [1991]).

Although we can show the admissibility of $\Box A \to \Box B / A \rhd B$ by detouring via that in $iP4$ without the completeness result of iPL, the completeness problem about iPL has its own importance[1]. The prominent open problem about modal completeness is as follows: is iPL complete or not, and, if yes, is it complete with respect to a class of finite frames. Our conjecture here is that it is incomplete. It seems that this problem is closely related to the completeness problem of iPW. To see this, we only need to look at our usual method showing the truth lemma for iPL. Up to now our understanding is that, if we want to show the truth lemma, the principle W is indispensable. Lp alone gives us too little for its completeness. We need the principle W to make the proof of the truth lemma go through, which makes us conjecture that iPL is probably incomplete. What is worse is that iPW is a proper extension of iPL. It is true that, if we can show the completeness (or incompleteness) of iPW, it will absolutely help us with the completeness problem with iPL. Unfortunately, the completeness problem of iPW is as stubborn as that for iPL.

How about conservation results and equivalence results if we put into iP two or more preservativity principles? For example, what is the $L_\Box$ fragment of $iPLM$? Is the $L_\Box$-fragment equivalent to iL with extra rules DR and MoR? We still don't know how to deal with this combined problem. The No. 1 difficulty is with the construction of the expected finite canonical model[2]. Because with the combined properties the canonical model has a tendency to explode. We have not found the mechanism to control that yet.

If we are able to answer the above combined conservation and equivalence problems, we have reached the level to aim at the ultimate question:

What is the $L_\Box$ fragment of iPH?

Moreover, if the conjecture in Iemhoff [2001] that iPH is *the* preservativity logic of HA holds, this logic will be *the* provability logic of HA.

[1] On classical side, we have mentioned the importance of the modal completeness of GL in provability logic for PA.

[2] The reason why we consider finite canonical models is that Lp is derivable in the logic in question.

Bibliography

[1] Johan van Benthem [1983]: Correspondence Theory, in Dov Gabbay and F. Günther (eds.) *Handbook of Philosohphical Logics*, Volume 2, Reidel Publishing Company.

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

[3] George Boolos [1993]: *the Logic of Provability*, Cambridge University Press.

[4] Milan Bozic and Kosta Dȯsen [1984], Models for normal intuitionistic modal logics, *Studia Logica*, XLIII, 3, pp. 217-245.

[5] A.Diego [1966],*Sur les Algebres de Hilbert*, Gauthier-Vilars, Paris.

[6] Alex Hendriks [1996], *Computations in Propositional Logic*, Ph.D thesis, ILLC-dissertation series, DS-1996-01, Universiteit van Amsterdam.

[7] Eva Hoogland [2001], *Definability and Interpolation*, Ph.D thesis, ILLC-dissertation series, DS-2001-05, Universiteit van Amsterdam.

[8] Rosalie Iemhoff [1998], A modal analysis of some principles of the provability logic of Heyting Arithmetic, in M, de Rijke, K. Segerberg, H. Wansing and M. Zakharyaschev (eds), *Advances in Modal Logic*, 2, CSLI Publications, Stanford.

[9] Rosalie Iemhoff [2001], *Provability Logic and Admissible Rules*, Ph.D thesis, ILLC-dissertation series, DS-2001-04, Universiteit van Amsterdam.

[10] Rosalie Iemhoff [2003], Preservativity Logic, *Mathematical Logic Quarterly*, 49(3), 2003, pp. 1-21.

[11] Dick de Jongh and Frank Veltman [1990], Provability Logics for relative interpretability, in P.Petkov (ed.), *Mathematical Logic, Proceedings of the Heything 1988 Summer School in Varna, Bulgaria*, Plenum Press, New York.

[12] Dick de Jongh and Albert Visser [1991], Explicit Fixed Point in Interpretability logic, *Studia Logica*,50,pp. 39-50

[13] Dick de Jongh and Frank Veltman [1995], *Intensional Logics*, Lecture Notes, Universiteit van Amsterdam.

[14] Dick de Jongh and Giorgi Japaridze [1998], The Logic of Provability, in Samuel R, Buss (ed.), *Handbook of Proof Theory*, Elsevier.

[15] Craig Smoryński [1985], *Self-reference and Modal Logic*, Springer-Verlag, Berlin.

[16] Robert Solovay [1976], Provability Interpretations of modal logic, *Israel Journal of Mathematics*, 25, pp, 287-304.

[17] Anne Troelstra and Dirk van Dalen [1988], *Constructivism in Mathematics*, Volume I, North Holland, Amsterdam.

[18] Albert Visser[2002], Substitutions of Σ_1^0 sentences, *Annals of Pure and Applied Logic* 114(1-3), pp.227-271 .